

Posicionamiento Web y dinámicas de información en motores de búsqueda

**Propuestas de análisis y estudio comparativo de visibilidad de
contenidos digitales en el caso de procesos electorales**

Carlos Gonzalo Penela

TESI DOCTORAL UPF / 2014

DIRECTOR DE LA TESI

Dr. Lluís Codina

DEPARTAMENT DE COMUNICACIÓ



A mis padres y hermano,
y al linaje de mi familia
hasta sus principios

Agradecimientos

En esta pequeña página de gloria y vida no sería justo dejar ahí la dedicatoria olvidando todas las familias que apuntan y se reúnen en mí, de ahí aprovechar la oportunidad para honrar a todas las generaciones sin las que hoy no podría estar aquí dedicándoles mi trabajo.

Gracias a todas las madres de Gonzalo, Penela, Rodríguez, Bolea, Álvarez, Colás, Salas, Pérez, Renales, Hernández, Sanz, Muñoz, García, Lorenzo, Fernández, Romero, Benito, Serrano, Azara, Palá, Gadea, González, Gómez, López, Velasco, Prieto, Martín, Martínez, Sancho, Tomey, Urgel, Hombradas, Arruego, Murillo, Alfranca, Gabarro, Lanud, Moliné, Valcarce, Sierra, Ximénez de Bagüés, Alberó, Giménez, Gabarró, Santín, Termenón, Vega, Vidal, Merayo, Gurdiel, Marcén, Sieso, Seral, Pin, Núñez y Gayoso, y a todas las que sin estar enumeradas, están.

Y mi más especial agradecimiento a todos aquellos maestros y guías que en el transcurso de los años me han ayudado en la forja y el temple.

ÍNDICE GENERAL

RESÚMENES	15
Resumen	17
Resum	18
Abstract	19

CAPÍTULO 0. Introducción	21
0.1. Objeto de estudio y objetivos	23
0.2. Contexto	30
0.3. Antecedentes y marco metodológico	37
0.4. Justificación de la investigación	42
0.5. Hipótesis de partida y preguntas de investigación	47
0.5.1. Preguntas de investigación	48
0.6. Tipo de investigación, metodología y fuentes de información	49
0.6.1. Tipo de investigación	49
0.6.2. Metodología	50
0.6.3. Fuentes de datos	53
0.6.4. Fuentes de información	53

CAPÍTULO 1. Estado de la cuestión	55
1.1. El Agenda Setting y la propaganda en la era de Internet	57
1.1.1. El modelo de la era Internet	66
1.1.2. La blogosfera	67
1.2. El modelo de Infoesfera	69
1.2.1. Origen y marco teórico del modelo de Infoesfera	70
1.2.2. Nuevo modelo de Infoesfera 2014	75
1.3. Google como modelo de Gatekeeper en la era Internet	78

1.3.1. Los riesgos de la Generación Google	80
1.3.2. La guerra de información	84
1.3.3. La guerra de información en la Blogosfera y el papel de los bloggers	90
1.4. Los buscadores y la lucha por el posicionamiento	97
1.4.1. Razones de la importancia del posicionamiento en SERP	100
1.4.2. Tipos de búsquedas	104
1.4.3. SERPs, Search Engine Results Pages	108
1.4.4. Los resultados orgánicos y los enlaces patrocinados.	110
1.4.5. División entre parte visible y área de desplazamiento	111
1.4.6. Visualización y Evaluación de la SERP	117
1.4.7. Relevancia percibida	118
1.4.8. CLICKS & visibilidad en SERP	122
1.4.9. Preferencias de clic: resultados orgánicos vs. enlaces patrocinados	125
1.4.10. Factores de ranking básicos	127
1.4.11. Posicionamiento natural vs planificado	130
1.5. Anatomía del buscador Google, 1998-2014	131
1.5.1. Componentes del buscador Google	133
Rastreador, robot, crawler o Spider	133
Sistema de recuperación de información	134
Sistema de indización	134
Sistema de análisis de los documentos	136
Interfaz de consulta	139
Algoritmo de cálculo de relevancia u ordenación.	140
1.5.2. Algoritmos del buscador Google	143
Google WebSearch>Universal Search & Blended Results	143
Google WebSearch>Google Caffeine	148
Google WebSearch>Google Hummingbird	151
Google News	157
Google BlogSearch	169
1.5.3. Listado completo de factores SEO	174

1.5.4. Web Spam	207
Aproximaciones a la detección de Spam	209
Content Farms, Spam y SEO Black Hat	209
Algoritmo Google Panda	213
Algoritmo Google Penguin	221

CAPÍTULO 2. Sistema de análisis	227
2.1. Antecedentes	229
2.2. Metodología	231
2.3. Técnicas e Instrumentos de Recolección de Datos	234
2.4. Técnicas de Procesamiento y Análisis de Datos	237
2.5. Selección de nichos temáticos: Uso de palabras clave	238
2.6. Acotación y Selección de las unidades de tiempo	239
2.7. Muestreo por conglomerados	239
2.8. Composición de los estratos y principales conglomerados de la muestra	240
2.9. Scraping de datos	249
2.9.1. Scraping de datos - Parlament de Catalunya 2010	250
2.9.2. Scraping de datos - Municipales de España 2011	251
2.9.3. Scraping de datos - Generales España 2011	252
2.9.4. Scraping de datos - Parlamento de Andalucía de 2012	253
2.9.5. Scraping de datos - Xunta de Galicia de 2012	254
2.9.6. Scraping de datos - Parlament de Catalunya 2012	255
2.10. Procedimiento para presentar e interpretar los datos	256

CAPÍTULO 3. Análisis de resultados	261
3.1. Introducción: la Infoesfera en Google, Yahoo y Bing	263
3.1.1. La composición por estratos en Google, Bing y Yahoo	265
3.1.1. Composición del Estrato Superior	267

3.1.2. Composición del Estrato Medio	273
3.1.3. Composición del Estrato Micro	275
3.1.4. Tendencias Google – Bing – Yahoo	278
3.2. Google Web Search	280
3.2.1. Generalitat Catalunya – 28 Noviembre 2010	282
3.2.2. Elecciones Municipales-22 Mayo 2011	288
3.2.3. Elecciones Generales- 20 Noviembre 2011	293
3.2.4. Junta-2012__ may-12 Andalucía - Mayo 2012	298
3.2.5. Xunta Galicia – 21 Octubre 2012	303
3.2.6. Generalitat Catalunya - 25 Noviembre 2012	308
3.2.7. Tendencias Google Web Search	313
3.3. Google News Search	318
3.3.1. Generalitat Catalunya – 28 Noviembre 2010	321
3.3.2. Elecciones Municipales-22 Mayo 2011	328
3.3.3. Elecciones Generales- 20 Noviembre 2011	335
3.3.4. Junta-2012__ may-12 Andalucía - Mayo 2012	342
3.3.5. Xunta Galicia – 21 Octubre 2012	349
3.3.6. Generalitat Catalunya - 25 Noviembre 2012	356
3.3.7. Tendencias Google News Search	363
3.4. Google Blog Search	366
3.4.1. Generalitat Catalunya – 28 Noviembre 2010	368
3.4.2. Elecciones Municipales-22 Mayo 2011	371
3.4.3. Elecciones Generales- 20 Noviembre 2011	373
3.4.4. Junta-2012__ may-12 Andalucía - Mayo 2012	377
3.4.5. Xunta Galicia – 21 Octubre 2012	380
3.4.6. Generalitat Catalunya - 25 Noviembre 2012	383
3.4.7. Tendencias Google Blog Search	386

CAPÍTULO 4. CONCLUSIONES	391
4.1 Hipótesis de partida	393
4.2 Preguntas de investigación	395
4.3 Consideraciones finales y futuros desarrollos	401
BIBLIOGRAFÍA	403
Anexo. Glosario de términos	417

ÍNDICE DE FIGURAS

Figura 1: Imagen 'Evolución de la Blogosfera'	68
Figura 2: Imagen 'Modelo Bow-Tie'	71
Figura 3: Imagen 'Modelo Corona'	71
Figura 4: Imagen 'Modelo de W.L. Bennet'	74
Figura 5: Tabla comparativa del modelo de comunicación Mass Media vs. Internet.	79
Figura 6: Imagen 'Modelo de Infoguerra según Schwartau'	86
Figura 7: Mapa 'Países en los que las redes sociales están prohibidas '	89
Figura 8: Imagen 'Cartel propagandístico de Radio libertad durante la guerra fría'	92
Figura 9: Imagen 'Titular del Wired refiriéndose a la nueva "Radio libertad" de las Redes sociales'	92
Figura 10: Imagen 'Noticia de Junio de 2010 sobre la detención del blogger iraní "Hoder"'	92
Figura 11: Gráfico circular 'State of the Blogosphere 2011'	94
Figura 12: Gráfico 'BLOG GROWTH. 2006'	96
Figura 13: Gráfico 'Evolución de interés de la búsqueda 'Facebook' 2006-2014'	97
Figura 14: Tabla 'Hierarchical classification of user intent by Web queries'	106
Figura 15: Imagen 'Captura de la parte visible de la SERP de la búsqueda 'Doctorado Comunicación Social''	110
Figura 16: Imagen 'Área visible y de desplazamiento de una SERP'	112
Figura 17: Imagen 'Google: principales componentes de las SERP'	114
Figura 18: Imagen 'Bing: principales componentes de las SERP'	114
Figura 19: Imagen 'Yahoo: principales componentes de las SERP'	115
Figura 20: Tabla 'Elementos que se utilizan para componer la SERP'	115
Figura 21: Tabla 'versión actualizada de la tabla de Höchstötter y Lewandowski (2009)'	116
Figura 22: Tabla 'Probabilidad de clic en SERP (SEO-SEM)'	122
Figura 23: Imagen 'Probabilidad de clic en SERP visible'	123
Figura 24: Gráfico 'Probabilidad de clic en Top10'	123
Figura 25: Diagrama 'The Anatomy of a Large-Scale Hypertextual Web Search Engine'	131
Figura 26: Diagrama 'Nuevos factores de ranking '	131

Figura 27: Imagen 'Evolución de los algoritmos de Google durante el periodo 2009-2013'	132
Figura 28: Imagen 'Modelo de índice invertido. '	135
Figura 29: Imagen 'Formulario de búsqueda simple de Google'	139
Figura 30: Imagen 'Formulario de búsqueda simple de Bing'	139
Figura 31: Imagen 'Formulario de búsqueda simple de Yahoo'	139
Figura 32: Imagen 'Búsqueda avanzada de Google'	140
Figura 33: Imagen 'Comparativa del modelo de búsqueda de Universal Search'	153
Figura 34: Tabla 'Tipos de contenidos son intercalados en SERPs'	147
Figura 35: Gráfico 'Comparativa de la estructura de índice de Google Caffeine Search'	148
Figura 36: Comparativa del tiempo de búsqueda de Google Caffeine Search	150
Figura 37: Diagrama 'Sistema de desambigüación búsqueda de Google Hummingbird Search'	154
Figura 38: Diagrama 'News Virtual link graph'	159
Figura 39: Imagen 'News Search SEO'	161
Figura 40: Diagrama 'Crawling e indexación'	161
Figura 41: Imagen 'FAQ Crawling'	161
Figura 42: Diagrama 'Google News -Agrupamiento & Clustering'	162
Figura 43: Diagrama 'Google News - Story Ranking'	162
Figura 44: Imagen 'Google News - Article Ranking'	162
Figura 45: Imagen 'Google News - Sitemaps'	165
Figura 46: Imagen 'Google News – Sitemaps (2)'	166
Figura 47: Imagen 'Google News - Multimedia content'	166
Figura 48: Imagen 'Google News - Images'	166
Figura 49: Imagen ' Comparativa de interfaces SERP'	167
Figura 50: Imagen 'Parte visible de resultados sobre una noticia del accidente aéreo en Mali de julio de 2014'	167
Figura 51: Diagrama 'Evolución de los medios, de la comunicación masiva a la personal'	229
Figura 52: Tabla 'Conglomerados homogéneos desarrollada en la fase metodológica'	235
Figura 53: Tabla 'Elecciones y calendario de scraping'	239

Figura 54: Imagen 'Composición de la Infoesfera'	240
Figura 55: Diagrama 'Modelo de coronas según la importancia y relación de los sitios del nicho temático principal objeto de estudio'	241
Figura 56: Imagen 'Conglomerados del Estrato Superior'	242
Figura 57: Imagen 'Conglomerados del Estrato Medio. Corona A'	245
Figura 58: Imagen 'Conglomerados del Estrato Medio. Coronas B y C'	245
Figura 59: Imagen 'Conglomerados del Estrato Micro. Corona B'	247
Figura 60: Imagen 'Conglomerados del Estrato Micro. Corona B y C'	248
Figura 61: Tabla de resultados de extracción de resultados de SERPs	249
Figura 62: Tabla de Scraping de datos - Parlament de Catalunya 2010	250
Figura 63: Tabla de Scraping de datos – Municipales 2011	251
Figura 64: Tabla de Scraping de datos – Generales 2011	252
Figura 65: Tabla de Scraping de datos – Andalucía 2012	253
Figura 66: Tabla de Scraping de datos – Galicia 2012	254
Figura 67: Tabla de Scraping de datos – Parlament de Catalunya 2012	261
Figura 68: Tabla 'Selección de búsquedas de la comparativa Google-Bing-Yahoo'	264
Figura 69: Gráficos circulares 'Comparativa Google-Bing-Yahoo – Elecciones Generales'	265
Figura 70: Gráficos circulares 'Comparativa Google-Bing-Yahoo – Elecciones Xunta de Galicia'	266
Figura 71: Gráficos de barras comparativas 'Estrato Superior - Sector conglomerados principales - Elecciones Generales'	268
Figura 72: Gráficos de barras comparativas 'Estrato Superior - Sector conglomerados principales - Elecciones Xunta de Galicia'	269
Figura 73: Gráficos de barras comparativas 'Estrato Superior - Sector conglomerados secundarios - Elecciones Generales'	270
Figura 74: Gráficos de barras comparativas 'Estrato Superior - Sector conglomerados secundarios - Elecciones Xunta de Galicia'	271
Figura 75: Gráficos de barras comparativas 'Estrato Medio - Elecciones Generales'	273
Figura 76: Gráficos de barras comparativas 'Estrato Medio - Elecciones Xunta de Galicia'	274
Figura 77: Gráficos de barras comparativas 'Estrato Micro - Elecciones Generales'	276
Figura 78: Gráficos de barras comparativas 'Estrato Micro - Elecciones Xunta de Galicia'	277
Figura 79: Tabla 'Web Search - Serie temporal de elecciones'	280

Figura 80: Gráfico de líneas 'Web Search - Estratos principales'	280
Figura 81: Gráfico de áreas 'Elecciones Generalitat 2010 - Estrato Superior - Conglomerados principales'	282
Figura 82: Gráfico de áreas 'Elecciones Generalitat 2010 - Estrato Superior - Conglomerados secundarios'	283
Figura 83: Gráfico de áreas 'Elecciones Generalitat 2010 - Estrato Medio'	284
Figura 84: Gráfico de áreas 'Elecciones Generalitat 2010 - Estrato Medio - Partidos'	285
Figura 85: Gráfico de áreas 'Elecciones Generalitat 2010 - Estrato Micro'	286
Figura 86: Gráfico de áreas 'Elecciones Municipales 2011 - Estrato Superior - Conglomerados principales'	288
Figura 87: Gráfico de áreas 'Elecciones Municipales 2011 - Estrato Superior - Conglomerados secundarios'	289
Figura 88: Gráfico de áreas 'Elecciones Municipales 2011 - Estrato Medio'	290
Figura 89: Gráfico de áreas 'Elecciones Municipales 2011 - Estrato Medio - Partidos'	291
Figura 90: Gráfico de áreas 'Elecciones Municipales 2011 - Estrato Micro'	292
Figura 91: Gráfico de áreas 'Elecciones Generales - Estrato Superior - Conglomerados principales'	293
Figura 92: Gráfico de áreas 'Elecciones Generales - Estrato Superior - Conglomerados secundarios'	294
Figura 93: Gráfico de áreas 'Elecciones Generales - Estrato Medio'	295
Figura 94: Gráfico de áreas 'Elecciones Generales - Estrato Medio - Partidos'	296
Figura 95: Gráfico de áreas 'Elecciones Generales - Estrato Micro'	297
Figura 96: Gráfico de áreas 'Elecciones Andalucía 2012 - Estrato Superior - Conglomerados principales'	298
Figura 97: Gráfico de áreas 'Elecciones Andalucía 2012 - Estrato Superior - Conglomerados secundarios'	299
Figura 98: Gráfico de áreas 'Elecciones Andalucía 2012 - Estrato Medio'	300
Figura 99: Gráfico de áreas 'Elecciones Andalucía 2012 - Estrato Medio - Partidos'	301
Figura 100: Gráfico de áreas 'Elecciones Andalucía 2012 - Estrato Micro'	302
Figura 101: Gráfico de áreas 'Elecciones Galicia 2012 - Estrato Superior - Conglomerados principales'	303
Figura 102: Gráfico de áreas 'Elecciones Galicia 2012 - Estrato Superior - Conglomerados secundarios'	304
Figura 103: Gráfico de áreas 'Elecciones Galicia 2012 - Estrato Medio'	305
Figura 104: Gráfico de áreas 'Elecciones Galicia 2012 - Estrato Medio - Partidos'	306

Figura 105: Gráfico de áreas 'Elecciones Galicia 2012 - Estrato Micro'	307
Figura 106: Gráfico de áreas 'Elecciones Generalitat 2012 - Estrato Superior - Conglomerados principales'	308
Figura 107: Gráfico de áreas 'Elecciones Generalitat 2012 - Estrato Superior - Conglomerados secundarios'	309
Figura 108: Gráfico de áreas 'Elecciones Generalitat 2012 - Estrato Medio'	310
Figura 109: Gráfico de áreas 'Elecciones Generalitat 2012 - Estrato Medio - Partidos'	311
Figura 110: Gráfico de áreas 'Elecciones Generalitat 2012 - Estrato Micro'	312
Figura 111: Tabla 'Google News Search - Serie temporal de elecciones'	319
Figura 112: Gráfico de líneas 'Google News - Generalitat 2010 - Estratos principales'	319
Figura 113: Gráfico de áreas 'Google News - Generalitat 2010 - Conglomerados'	321
Figura 114: Gráfico de áreas 'Google News - Generalitat 2010 - Prensa Estatal'	321
Figura 115: Gráfico de áreas 'Google News - Generalitat 2010 - Prensa Regional'	321
Figura 116: Gráfico de áreas 'Google News - Generalitat 2010 - Televisión'	322
Figura 117: Gráfico de áreas 'Google News - Generalitat 2010 - Cibermedios'	322
Figura 118: Gráfico circular 'Google News - Generalitat 2010 - Conglomerado'	324
Figura 119: Tabla 'Elecciones Generalitat 2010 - Conglomerados - Tabla de Serie temporal'	324
Figura 120: Gráfico de áreas 'Google News - Generalitat 2010 - Grupos mediáticos - Serie temporal'	325
Figura 121: Histograma 'Google News - Generalitat 2010 - Grupos mediáticos'	326
Figura 122: Tabla 'Elecciones Generalitat 2010 - Grupos mediáticos - Tabla de Serie temporal'	326
Figura 123: Tabla 'Elecciones Generalitat 2010 - Grupos mediáticos - Tabla de Serie temporal (2)'	327
Figura 124: Gráfico de áreas 'Google News - Municipales 2011 - Conglomerados'	328
Figura 125: Gráfico de áreas 'Google News - Municipales 2011 - Prensa Estatal'	328
Figura 126: Gráfico de áreas 'Google News - Municipales 2011 - Prensa Regional'	328
Figura 127: Gráfico de áreas 'Google News - Municipales 2011 - Televisión'	329
Figura 128: Gráfico de áreas 'Google News - Municipales 2011 - Cibermedios'	329
Figura 129: Gráfico circular 'Google News - Municipales 2011 - Conglomerado'	331
Figura 130: Tabla 'Elecciones Municipales 2011 - Conglomerado - Tabla de Serie temporal'	331

Figura 131: Gráfico de áreas 'Elecciones Municipales 2011 - Grupos mediáticos'	332
Figura 132: Histograma 'Google News - Municipales 2011 - Grupos mediáticos'	333
Figura 133: Tabla 'Elecciones Municipales 2011 - Conglomerado - Tabla de Serie temporal'	333
Figura 134: Tabla 'Elecciones Municipales 2011 - Conglomerados - Tabla de Serie temporal (2)'	334
Figura 135: Gráfico de áreas 'Google News - Generales 2011 - Conglomerados'	335
Figura 136: Gráfico de áreas 'Google News - Generales 2011 - Prensa Estatal'	335
Figura 137: Gráfico de áreas 'Google News - Generales 2011 - Prensa Regional'	335
Figura 138: Gráfico de áreas 'Google News - Generales 2011 - Televisión'	336
Figura 139: Gráfico de áreas 'Google News - Generales 2011 - Cibermedios'	336
Figura 140: Gráfico circular 'Google News - Generales 2011 - Conglomerado'	338
Figura 141: Tabla 'Elecciones Generales 2011 - Conglomerado - Tabla de Serie temporal'	338
Figura 142: Gráfico de áreas 'Elecciones Generales 2011 - Grupos mediáticos'	339
Figura 143: Histograma 'Google News - Generales 2011 - Grupos mediáticos'	340
Figura 144: Tabla 'Elecciones Generales 2011 - Conglomerado - Tabla de Serie temporal'	340
Figura 145: Tabla 'Elecciones Generales 2011 - Conglomerados - Tabla de Serie temporal (2)'	341
Figura 146: Gráfico de áreas 'Google News - Andalucía 2012 - Conglomerados'	342
Figura 147: Gráfico de áreas 'Google News - Andalucía 2012 - Prensa Estatal'	342
Figura 148: Gráfico de áreas 'Google News - Andalucía 2012 - Prensa Regional'	342
Figura 149: Gráfico de áreas 'Google News - Andalucía 2012 - Televisión'	343
Figura 150: Gráfico de áreas 'Google News - Andalucía 2012 - Cibermedios'	343
Figura 151: Gráfico circular 'Google News - Andalucía 2012 - Conglomerado'	345
Figura 152: Tabla 'Elecciones Andalucía 2012 - Conglomerado - Tabla de Serie temporal'	345
Figura 153: Gráfico de áreas 'Elecciones Andalucía 2012 - Grupos mediáticos'	346
Figura 154: Histograma 'Google News - Andalucía 2012 - Grupos mediáticos'	347
Figura 155: Tabla 'Elecciones Andalucía 2012 - Conglomerado - Tabla de Serie temporal'	347
Figura 156: Tabla 'Elecciones Andalucía 2012 - Conglomerados - Tabla de Serie temporal (2)'	348

Figura 157: Gráfico de áreas 'Google News - Xunta 2012 - Conglomerados'	349
Figura 158: Gráfico de áreas 'Google News - Xunta 2012 - Prensa Estatal'	349
Figura 159: Gráfico de áreas 'Google News - Xunta 2012 - Prensa Regional'	349
Figura 160: Gráfico de áreas 'Google News - Xunta 2012 - Televisión'	350
Figura 161: Gráfico de áreas 'Google News - Xunta 2012 - Cibermedios'	350
Figura 162: Gráfico circular 'Google News - Xunta 2012 - Conglomerado'	352
Figura 163: Tabla 'Elecciones Xunta 2012 - Conglomerado - Tabla de Serie temporal'	352
Figura 164: Gráfico de áreas 'Elecciones Xunta 2012 - Grupos mediáticos'	353
Figura 165: Histograma 'Google News - Xunta 2012 - Grupos mediáticos'	354
Figura 166: Tabla 'Elecciones Xunta 2012 - Conglomerado - Tabla de Serie temporal'	354
Figura 167: Tabla 'Elecciones Xunta 2012 - Conglomerados - Tabla de Serie temporal (2)'	355
Figura 168: Gráfico de áreas 'Google News - Generalitat 2012 - Conglomerados'	356
Figura 169: Gráfico de áreas 'Google News - Generalitat 2012 - Prensa Estatal'	356
Figura 170: Gráfico de áreas 'Google News - Generalitat 2012 - Prensa Regional'	356
Figura 171: Gráfico de áreas 'Google News - Generalitat 2012 - Televisión'	357
Figura 172: Gráfico de áreas 'Google News - Generalitat 2012 - Cibermedios'	357
Figura 173: Gráfico circular 'Google News - Generalitat 2012 - Conglomerado'	359
Figura 174: Tabla 'Elecciones Generalitat 2012 - Conglomerado - Tabla de Serie temporal'	359
Figura 175: Gráfico de áreas 'Elecciones Generalitat 2012 - Grupos mediáticos'	360
Figura 176: Histograma 'Google News - Generalitat 2012 - Grupos mediáticos'	361
Figura 177: Tabla 'Elecciones Generalitat 2012 - Conglomerado - Tabla de Serie temporal'	361
Figura 178: Tabla 'Elecciones Generalitat 2012 - Conglomerados - Tabla de Serie temporal (2)'	362
Figura 179: Tabla 'Blog Search - Serie temporal de elecciones'	366
Figura 180: Gráfico de líneas 'Blog Search - Estratos principales'	366
Figura 181: Gráfico de áreas 'Elecciones Generalitat 2010 - Estrato Superior'	368
Figura 182: Gráfico de áreas 'Elecciones Generalitat 2010 - Estrato Medio'	369

Figura 183: Gráfico de áreas 'Elecciones Generalitat 2010 - Estrato Micro'	370
Figura 184: Gráfico de áreas 'Elecciones municipales 2011 - Estrato Superior'	371
Figura 185: Gráfico de áreas 'Elecciones municipales 2011 - Estrato Medio'	372
Figura 186: Gráfico de áreas 'Elecciones municipales 2011 - Estrato Micro'	373
Figura 187: Gráfico de áreas 'Elecciones generales 2011 - Estrato Superior'	374
Figura 188: Gráfico de áreas 'Elecciones generales 2011 - Estrato Medio'	375
Figura 189: Gráfico de áreas 'Elecciones generales 2011 - Estrato Micro'	376
Figura 190: Gráfico de áreas 'Elecciones Junta de Andalucía 2012 - Estrato Superior'	377
Figura 191: Gráfico de áreas 'Elecciones Junta de Andalucía 2012 - Estrato Medio'	378
Figura 192: Gráfico de áreas 'Elecciones Junta de Andalucía 2012 - Estrato Micro'	379
Figura 193: Gráfico de áreas 'Elecciones Xunta de Galicia 2012 - Estrato Superior'	380
Figura 194: Gráfico de áreas 'Elecciones Xunta de Galicia 2012 - Estrato Medio'	381
Figura 195: Gráfico de áreas 'Elecciones Xunta de Galicia 2012 - Estrato Micro'	382
Figura 196: Gráfico de áreas 'Elecciones Generalitat 2012 - Estrato Superior'	383
Figura 197: Gráfico de áreas 'Elecciones Generalitat 2012 - Estrato Medio'	384
Figura 198: Gráfico de áreas 'Elecciones Generalitat 2012 - Estrato Micro'	385

RESÚMENES

Resumen

Posicionamiento Web y dinámicas de información en motores de búsqueda: propuestas de análisis y estudio comparativo de visibilidad de contenidos digitales en el caso de procesos electorales es un trabajo de investigación que tiene como objeto de estudio las características de las páginas web obtenidas a través de resultados de búsqueda (SERPs) durante los procesos electorales realizados en España entre los años 2010 y 2012.

Este trabajo parte del modelo teórico de Infoesfera expuesto por W.L. Bennet en 2003, armazón teórico que se ha desarrollado para construir un sistema de análisis capaz de englobar la totalidad de la web y analizar las proporciones con las que los buscadores muestran determinadas categorías de sitios web, en especial medios de comunicación, webs de la esfera pública/política y de la web social.

Los objetivos principales han sido: 1) llevar a cabo un estado de la cuestión sobre el funcionamiento de Google, el posicionamiento web y la interacción de los usuarios con sus páginas de resultados (llamadas SERPs); 2) desarrollar un sistema de análisis propio de tales páginas de resultados; 3) aplicar y presentar el modelo utilizando datos extraídos durante comicios electorales en España; 4) contrastar la viabilidad del modelo de la infoesfera al análisis de la web actual; 5) conocer la composición del ecosistema de la web en Cataluña y España para resultados obtenidos a través de buscadores.

La motivación que ha estado detrás de toda esta investigación ha sido comprobar si Internet, a través de los resultados de buscadores, ha diversificado y, por tanto, eventualmente, ha democratizado la visibilidad de los medios de comunicación digitales.

El resultado ha sido, además de una aportación metodológica en el campo del análisis web, a través de un nuevo modelo que permite visualizar las proporciones y evolución de los resultados de buscadores y medir el peso relativo de todos sus componentes, un estado de la cuestión sobre SEO y factores de posicionamiento, y una representación del estado de la web en los ámbitos indicados, vista a través de los resultados de los buscadores.

Resum

Posicionament Web i dinàmiques d'informació a motors de cerca: propostes d'anàlisi i estudi comparatiu de visibilitat de continguts digitals en el cas de processos electorals és un treball d'investigació que té com a objecte d'estudi les característiques de les pàgines web obtingudes a través de resultats de cerca (SERPs) durant els processos electorals duts a terme a Espanya entre els anys 2010 i 2012.

Aquest treball parteix del model teòric de Infoesfera exposat per W.L. Bennet l'any 2003, esquema teòric que s'ha desenvolupat per construir un sistema d'anàlisi capaç d'englobar la totalitat de la web i analitzar les proporcions amb què els cercadors mostren determinades categories de llocs web, especialment mitjans de comunicació, webs de l'esfera pública / política i de la web social.

Els objectius principals han estat: 1) dur a terme un estat de la qüestió sobre el funcionament de Google, el posicionament web i la interacció dels usuaris amb les pàgines de resultats (anomenades SERPs); 2) desenvolupar un sistema d'anàlisi propi d'aquestes pàgines de resultats; 3) aplicar i presentar el model utilitzant dades extretes durant comicis electorals a Espanya; 4) contrastar la viabilitat del model de Infoesfera a l'anàlisi de la web actual; 5) conèixer la composició de l'ecosistema de la web a Catalunya i Espanya per resultats obtinguts a través de cercadors.

La motivació que ha estat darrere de tota aquesta investigació ha estat comprovar si Internet, a través dels resultats de cercadors, ha diversificat i, per tant, eventualment, ha democratitzat la visibilitat dels mitjans de comunicació digitals.

El resultat ha estat, a més d'una aportació metodològica en el camp de l'anàlisi web, a través d'un nou model que permet visualitzar les proporcions i evolució dels resultats de cercadors i mesurar el pes relatiu de tots els seus components, un estat de la qüestió sobre SEO i factors de posicionament, i una representació de l'estat de la web en els àmbits indicats, vista a través dels resultats dels cercadors.

Abstract

Web Positioning and information dynamics in search engines: proposals for analysis and comparative study of the visibility of digital content in electoral processes.

This dissertation aims to study the characteristics of web pages retrieved through Search Engine Results Pages (SERPs) during elections held in Spain between 2010 and 2012.

The research was based on the Infosphere theoretical model put forward by W.L. Bennet in 2003, a theoretical framework that has been developed to build an analytical system that encompasses the entire Web and analyses the proportions with which search engines show particular categories of websites, especially media sites, websites in the public/political sphere and social networking sites.

The main objectives were: 1) to clarify the status of the issue regarding the operation of Google, web optimization and user interaction with search engine results pages (so-called SERPs); 2) to develop a system for analysing these SERPs; 3) to implement and present the model using data retrieved during elections in Spain; 4) to monitor the viability of the Infosphere model for the analysis of the web now; 5) to determine the composition of the web's ecosystem in Catalonia and Spain through SERPs.

The motivation behind all this research has been to determine whether the Internet, through SERPs, has diversified and thus has ultimately democratized the visibility of digital media.

The result has also been a methodological contribution in the field of web analytics, through a new model that displays the proportions and evolution of SERPs and measures the relative weight of all its components, a state of the art about SEO and ranking factors, and a representation of the state of the web in the above areas, seen through SERPs.

0. INTRODUCCIÓN

0.1. Objeto de estudio y objetivos

El objeto central de estudio de esta tesis son las páginas web que forman parte de los resultados de las búsquedas efectuadas a través de motores de búsqueda como Google.

Las páginas de resultados de los buscadores o SERP, por sus siglas en inglés, están compuestas de enlaces a páginas web. Como es sabido, la inmensa mayoría de los ciudadanos utilizan buscadores como Google de forma rutinaria para acceder a la información disponible en la web. La web que aparece a través de las SERP es la web, sin más, al menos para los millones de ciudadanos que, a diario, antes de entrar en un sitio lanzan una búsqueda a través de su buscador preferido, es decir, a través de Google en más del 90 por ciento de los casos para los ciudadanos de nuestro país y de la mayor parte de los países de la Unión Europea.

Las páginas de resultados de los motores de búsqueda (SERPs a partir de ahora) son uno de los objetos de estudio centrales de una de las disciplinas que, a su vez, forma parte del marco teórico de esta investigación, denominada «SEO» por las siglas de Search Engine Optimization.

El SEO, efectivamente, tiene como objeto de estudio las SERP porque en ellas queda de manifiesto la correlación (variable) que existe entre una pregunta determinada y lo que Google «considera» que corresponde a la mejor respuesta posible a esa pregunta. Es por ello que, solamente del estudio de las SERPs, puede llegar a saberse algo, con base sólida, de cómo atribuye Google el atributo de calidad a las páginas de la web que forman su índice. Si una página web es considerada por Google una página de referencia o simplemente la mejor respuesta para una palabra de búsqueda, esa página tendrá una alta posición en las SERP y, como consecuencia, se beneficiará del enorme caudal de tráfico que supone ocupar las primeras posiciones.

Las SERPs utilizadas en este estudio fueron extraídas de los buscadores durante los procesos electorales llevados a cabo en España durante los años 2010-12, siendo éste considerado como un escenario donde podrían llevarse a cabo acciones de guerra de información.

La evolución de las proporciones en las SERPs, dentro de un mismo proceso electoral o en comparación con otros, muestra las dinámicas de visibilidad durante las elecciones de sitios webs específicos, así como de diferentes grupos homogéneos de páginas web que denominaremos «conglomerados».

El análisis de proporciones y evolución de estas series temporales permite detectar cuáles han sido los conglomerados con más visibilidad en SERPs y los cambios que han sufrido a lo largo del periodo 2010-12 debido a la propia evolución de los algoritmos de Google.

El objeto de estudio está delimitado, además, en función de los siguientes cuatro parámetros:

a. Temático

El objeto de estudio a analizar son las SERPs extraídas del campo temático de la comunicación social y política a través de un conjunto de palabras clave relacionadas con los procesos electorales.

La palabra clave principal de todos los paquetes de búsquedas ha sido «elecciones», (junto a «*eleccions*» en los comicios a la Generalitat de Catalunya) que, combinada con un conjunto de palabras clave secundarias asociadas a cada proceso específico, han conformado el paquete de palabras clave seleccionadas para realizar la extracción de datos. Las palabras clave secundarias que han actuado a modo de atributo del término principal «elecciones», han sido los nombres de partidos políticos, nombres de candidatos y, en algunos casos, noticias específicas del periodo electoral (ie. «José Blanco» durante las elecciones generales). Las combinaciones resultantes son palabras clave que potencialmente podían ser utilizadas por los usuarios durante el periodo de captura.

El resultado de consultar en Google las palabras clave seleccionadas ha sido encontrar páginas web asociadas a estas búsquedas de tema político, de lo que se deriva que son los paquetes de palabra clave los que definen el tema de las SERPs.

b. Idiomático, Geográfico espacial

Los motores de búsqueda seleccionados donde se han realizado las consultas y de los que se han extraído las SERPs a través de un proceso de *scraping* han sido las opciones de búsqueda web (opción por defecto de los buscadores), la de idioma «español» y la de país «España».

En las elecciones a la Generalitat de Catalunya 2010 y 2012 se han incluido motores de búsqueda y palabras clave en idioma «catalán» dentro del paquete de consultas. En el caso de los resultados de Catalunya, los cálculos de visibilidad se han realizado agregando los resultados de ambos idiomas.

Durante el proceso de captura de datos también se extrajeron datos de las elecciones del País Vasco de octubre de 2012¹, si bien los resultados fueron desechados por el problema idiomático del euskera y por la dificultad en categorizar y clasificar webs escritas en este idioma.

c. Temporal

La temporalidad del estudio abarca los días de campaña de los seis procesos electorales objeto de estudio. Las fechas de estas elecciones fueron:

1. Elecciones al Parlament de Catalunya de 28-nov-10
2. Elecciones municipales de España de 22-may-11
3. Elecciones generales de España de 20-nov-11
4. Elecciones al Parlamento de Andalucía de 25-mar-12
5. Elecciones al Parlamento de Galicia de 21-oct-12
6. Elecciones al Parlament de Catalunya de 25-nov-12

El periodo temporal empezó con las elecciones al Parlament de Catalunya de Noviembre de 2010, finalizando con las elecciones al Parlament de Catalunya, que al ser adelantadas se celebraron dos años más tarde en noviembre de 2012.

¹Elecciones al Parlamento Vasco de 2012
<http://es.wikipedia.org/wiki/Elecciones_al_Parlamento_Vasco_de_2012>

Este periodo de dos años incluye, además, las elecciones municipales españolas de 2011, las elecciones Generales de España de 2011, las elecciones a la Junta de Andalucía de 2012 y las elecciones a la Xunta de Galicia de 2012.

Dada la gran cantidad de elecciones y datos, se ha decidido no incluir en el estudio las elecciones a la Junta General del Principado de Asturias de marzo 2012² ni las elecciones al Parlamento Vasco de octubre de 2012.

De cada una de las elecciones se capturaron un conjunto de días variable de los que finalmente se seleccionaron 14 días que abarcan el periodo de campaña. Se ha intentado seguir la norma de incluir en la muestra el día de las elecciones, el inmediatamente posterior y 12 días del periodo de campaña electoral. En el caso específico de las elecciones al *Parlament* de Catalunya de 2012 de 25 de noviembre de 2012, a causa de la pérdida de diversos días de datos del periodo electoral, se han seleccionado dos días inmediatamente posteriores a las elecciones en lugar de uno.

d. Unidades de observación

La unidad de observación básica es la URL³. Estas URLs pertenecen a páginas web que han aparecido en resultados de búsqueda durante alguno de los procesos electorales 2010-12, y que han sido registradas y categorizadas.

Las URLs están agrupadas en forma de SERPs, que son la unidad de captura básica del estudio. Las SERPs agrupan un número de diez resultados vinculados a una palabra clave.

Cada uno de estos resultados corresponde a URLs que apuntan a un contenido web de un formato específico (HTML, vídeo, PDF, PPT, TWEET, post) y que pertenece a un dominio, que a su vez ha sido clasificado en un conglomerado

² Elecciones a la Junta General del Principado de Asturias de 2012 <http://es.wikipedia.org/wiki/Elecciones_a_la_Junta_General_del_Principado_de_Asturias_de_2012>

³ Localizador uniforme de recursos <http://es.wikipedia.org/wiki/Localizador_uniforme_de_recursos>

temático dentro de un sector de alguno de los niveles de un estrato de la Infoesfera⁴. El formato que sigue la categorización es:

URL → Dominio → Conglomerado → Sector → Corona → Estrato de la Infoesfera

Las URLs aparecidas en una SERP se clasifican según indicadores de mención ordinales, es decir, según la posición en que han aparecido en resultados, pudiendo estar entre la primera y décima posición de la SERP.

La profundidad de recuperación de datos en cada iteración aplicada ha sido la de tres SERPs, es decir, se han capturado a diario los 30 primeros resultados (el Top30) de todas las palabras clave durante todo el periodo de observación, lo que permite estudiar diferentes niveles de resultados, ya sea el Top4, Top10, Top20 o Top30, si bien nos centramos en Top10.

Las iteraciones de una misma palabra clave están agrupadas en forma de series temporales de SERPs, ordenadas cronológicamente en torno al día de celebración de unas elecciones. Las series temporales son de un periodo no inferior a las dos semanas, si bien para el estudio final se han seleccionado únicamente 14 días de datos.

Las SERPs, agrupadas en forma de series temporales, permiten calcular la visibilidad de los tres principales estratos, así como de cualquier sitio web aparecido en el Top30 para cualquier palabra clave. Como se ha indicado, dentro de estas series, analizamos los indicadores de mención asociados a posiciones dentro del Top10.

⁴Daremos detalles a estos diferentes niveles de clasificación, pero la tendencia del análisis global se centrará en unidades superiores como son los estratos o los conglomerados.

A partir de aquí, un enunciado muy sintético de los objetivos de esta investigación han sido: **1)** llevar a cabo un estado de la cuestión sobre el funcionamiento de Google, el posicionamiento web y la interacción de los usuarios con sus páginas de resultados (llamadas SERPs); **2)** desarrollar un sistema de análisis propio de tales páginas de resultados; **3)** desarrollar, aplicar y presentar el modelo utilizando datos extraídos durante comicios electorales en España; **4)** contrastar la viabilidad del modelo de la infoesfera al análisis de la web actual y, finalmente, **5)** conocer la composición del ecosistema de la web en Cataluña y España para resultados obtenidos a través de los principales buscadores.

A continuación, desarrollamos un poco más estos objetivos poniéndolos en relación con algunas bases teóricas que conviene adelantar aquí.

A. Objetivo general

Aunque el primer objetivo, de tipo pragmático señalado ha sido el desarrollo de un estado de la cuestión, el objetivo más general que actuó como «disparador» principal de la tesis es el de desarrollar un modelo de Infoesfera enunciado por de J.Peretti y W.L.Bennet y pasar del planteamiento teórico que supone este modelo a una plasmación del modelo con datos reales, en nuestro caso en un contexto electoral.

Esta adaptación del modelo teórico ha sido desarrollado sobre datos de posicionamiento en buscadores capturados durante seis comicios electorales, y el principal objetivo ha sido desarrollar un estudio exploratorio de los datos, lo más completo y exhaustivo posible, aplicando el modelo de Infoesfera.

B. Objetivos de investigación específicos

Los objetivos específicos radican en el análisis de sectores, conglomerados y dominios concretos de la Infoesfera. Los datos de posicionamiento han sido analizados utilizando un esquema en tres estratos, que a su vez han sido desglosados en diferentes niveles y categorías que agrupan sitios web por grados de afinidad.

Los objetivos específicos de la investigación incluyen el análisis de varios de estos sectores y conglomerados temáticos en las diferentes opciones de búsqueda de Google (Websearch, Blogsearch y Google News) con el objetivo de conocer el peso

específico y relativo de sus actores, así como conocer la evolución de su visibilidad a lo largo de las campañas electorales de los seis comicios estudiados.

El **primer sector** objetivo de estudio pertenece al estrato convencional, compuesto por diferentes conglomerados que incluyen las versiones digitales de los diferentes medios de prensa escrita, radio y televisión, y todos los medios informativos digitales deportivos y de prensa extranjera.

El **segundo sector** objetivo es el de los partidos políticos, que está dividido en diferentes conglomerados que incluyen todos los partidos políticos españoles. El objetivo ha sido el análisis del comportamiento de las páginas web de partido y de las diversas plataformas digitales que fueron usadas en las elecciones por los partidos políticos, su visibilidad global durante los días de campaña y los contenidos que mejor posicionaron.

Otros objetivos relacionados de la investigación han sido estudiar el conglomerado de sitios web protesta, conocido como «web indignada» y los conglomerados correspondientes a las redes sociales.

En cualquier caso, antes de seguir presentando otros aspectos esenciales de esta investigación como la hipótesis inicial y las preguntas de investigación derivadas de ella, entendemos que será útil presentar algunas cuestiones del contexto teórico, así como sus antecedentes basados en un trabajo de investigación previo.

0.2. Contexto

En febrero de 2001, Luis Bassat escribió: «Con el nacimiento de internet, no solo ha nacido un nuevo canal de comunicación, ha nacido también un canal de investigación, un canal de venta y un canal de atención y de servicio al cliente. Con Internet ha nacido una nueva generación de consumidores; son aquellos críos que descubrieron la informática con el Spectrum y que ahora son consumados navegantes, que buscan, miran y compran utilizando su tarjeta de crédito. La televisión entró primero en la clase alta. Internet ha empezado por la gente joven».

Han pasado trece años desde entonces y los niños que crecieron con un Spectrum tienen en la actualidad una media de cinco dispositivos de tecnología digital para acceder a Internet y a infinidad de contenidos digitales. En 2013 las ventas mundiales de *tablets* superaron a las de PC⁵. En España⁶ representaron el 60% del mercado total en unidades y un 40% de la facturación con una venta total de 3,8 millones de tabletas frente a 1,7 millones de portátiles y 910.000 ordenadores de sobremesa.

Las cifras de *Radiografía del mercado móvil en España*⁷ presentadas en el Mobile World Centre de Barcelona en febrero de 2014 mostraron las cotas más altas de penetración del *smartphone* con un 80% del total en 2013, frente al 59% del año anterior. Los datos del informe mostraban que el uso del *smartphone* había alcanzado a todos los tramos de edad, con un 91% de penetración en la franja de edad de los 18 a los 25 años.

El uso del *smartphone* para el acceso a Internet es ya generalizado, con un 86% de usuarios que acceden diariamente a través de este dispositivo, casi igualando al acceso desde portátiles y ordenadores de sobremesa que alcanza el 88% de los usuarios. El uso del teléfono móvil para llamadas se limita al 16% del tiempo total de uso, mientras que el 84% restante del tiempo se invierte en el uso de las nuevas funcionalidades propias de los *smartphone* y que requieren de conexión a Internet. El

⁵ Las ventas de *tablets* superarán este año a las de PC por primera vez <<http://www.computing.es/informatica-profesional/tendencias/1069745001701/ventas-tablets-sobrepasaran-ano-pc-primera-vez.1.html#sthash.08PBzpVi.dpuf>>

⁶ La venta de *tablets* sube un 68% en España y salva el mercado del PC. | Cinco Días <http://cincodias.com/cincodias/2014/01/22/tecnologia/1390421558_103791.html>

⁷ Fuente: *Radiografía del mercado móvil en España* <http://www.iabspain.net/wp-content/uploads/downloads/2013/09/V_Estudio_Mobile_Marketing_version_corta.pdf>

50% de las visitas a Internet se realizan para acceder a funcionalidades de comunicación social tales como la mensajería instantánea o chats, redes sociales y correo electrónico; el 20% de entradas a Internet tienen un propósito lúdico asociado a contenidos como juegos, música o aplicaciones, y solo un 14% lo utiliza para navegar por Internet.

Como consecuencia de estas cifras de penetración de dispositivos móviles⁸ se podría decir que la generación más joven, aquellos que no aparecen en el estudio de IAB Spain por ser todavía menores de edad, ha nacido con un *smartphone* o una *tablet* debajo del brazo. En la actualidad es habitual ver a niños de corta edad usando los dispositivos de sus padres en restaurantes como un método para entretenerlos, y es raro no ver a adolescentes «enchufados» a sus móviles 4G camino del colegio, donde haciendo uso de las nuevas tecnologías ya no es extraña la distribución de deberes vía *whatsapp* ni los casos de hostigamiento o *bullying* digital.

Estos nuevos usuarios, los miembros más jóvenes de la sociedad del conocimiento que ha sido denominada la «Generación Google», son los que nutren de forma voluntaria a las redes sociales de miles de comentarios personales, cuelgan las fotografías de sus viajes más recientes y responden a encuestas online sobre sus gustos y hábitos de consumo. Además de toda esta información que la generación más joven sube desinteresadamente a las redes sociales existe gran cantidad de información que cede involuntariamente a través de la navegación y de su actividad dentro de cualquier sitio web y que se convierte en la materia prima de los profesionales de la Publicidad y del Marketing para programar campañas y acciones dirigidas a segmentos del mercado que correspondan con perfiles socio-demográficos o hábitos específicos.

Sin embargo, al igual que los hábitos de consumo online suponen grandes oportunidades para los profesionales del sector online, estos mismos hábitos pueden verse como debilidades, vulnerabilidades potenciales que podrían ser aprovechadas

⁸ Un problema derivado del acceso a la información derivada del incremento de dispositivos es la piratería. La oferta de *tablets* ha ido de la mano de las ventas de los *e-readers* que en la actualidad se ofertan en *outlets*. La masiva compra de estos dispositivos y la disponibilidad de cualquier título de novedad editorial de forma gratuita a través de buscadores como Google ha acabado con la prensa tradicional en papel y con parte del sector editorial. La piratería es uno de los problemas extendidos en la sociedad del conocimiento española, tal vez derivado de una educación deficiente respecto a la sociedad de la información.

no ya para seleccionar nichos de mercado, sino para seleccionar y ejecutar acciones abiertamente deshonestas de desinformación y/o de contrainformación.

Para entender mejor este posible escenario podemos retroceder hasta enero de 2008 cuando apareció un estudio que mostraba las características de esta nueva generación.

El estudio en cuestión fue publicado por el *University College London* y se titulaba *Information Behaviour of the Researcher of the Future* y analizaba el comportamiento de los jóvenes frente a los nuevos medios de comunicación, en especial Internet y los buscadores.

El estudio destacó que esta nueva generación de internautas, la llamada «Generación Google», posee una gran preferencia por la información visual en detrimento de la textual y tiene una gran dependencia y confianza en los buscadores, en especial Google, careciendo de sentido crítico y de habilidades analíticas para evaluar la veracidad de la información encontrada en los buscadores.

Los posibles problemas derivados de estos hábitos de consumo de información por parte de la generación Google han sido objeto de estudio durante los últimos años, saltando a la palestra mediática en junio de 2008 cuando la eurodiputada socialista Marianne Mikko presentó en el Parlamento Europeo el *Documento de trabajo sobre la concentración y el pluralismo de los medios de comunicación en la Unión Europea*, generando una gran polémica entre los autores de blogs (*bloggers*) de toda Europa. En dicho documento, la Comisión de Cultura y Educación reconoció que los *weblogs* son un medio de comunicación en expansión utilizado tanto por profesionales de la comunicación como por personas privadas.

La publicación de este documento desencadenó una gran controversia ya que para solucionar el problema de imparcialidad y fiabilidad que suponían los *weblogs* el documento proponía regular la Blogosfera, y más concretamente la identidad de los blogueros. En los días posteriores Marianne Mikko concedió diversas entrevistas en las que justificaba su iniciativa alegando que algunos blogueros podrían estar siguiendo las directrices de lo que ella llamó «agendas ocultas» que podrían llegar a ser una amenaza al pluralismo, debiendo realizarse un seguimiento de la identidad de los que escriben tras la Blogosfera. Sin duda esta preocupación puede parecer que muestra

tener tonos alarmistas; sin embargo, dejando el posible alarmismo detrás, se basa en circunstancias reales bien conocidas como la enorme importancia que la web social tiene en la sociedad actual.

Por otro lado, como dejó patente la preocupación mostrada por la UE, la generación Google posee unos rasgos comunes que refuerzan la teoría de que, aunque este riesgo nos parezca lejano, informaciones no contrastadas provenientes de la blogosfera podrían suponer un peligro y desestabilizar la sociedad de la información.

La razón es que estas características de los miembros de la generación Google, tan dependiente de los resultados de un solo buscador y de unas pocas redes sociales, son vistas como una debilidad que los convierte en vulnerables en un escenario de hipotética actividad planificada de desinformación («guerra de información»), de tal manera que este mismo comportamiento puede ser explotado por terceros convirtiendo a la generación Google en víctimas potenciales de diferentes tipos de ataques de desinformación.

Cierto es que todo lo anterior es una amenaza no concretada, similar a otras que abundan en nuestra sociedad, como la del cambio climático por mencionar una muy distinta, sin embargo no por falta de concreción parece que su estudio sea despreciable. Esta preocupación que, afortunadamente, no ha conocido ningún caso aún de amenaza realizada y sin embargo preocupa también a grandes ensayistas actuales como Nicholas Carr. Como tal amenaza no ha tenido (que sepamos nosotros) ninguna concreción, es obvio que esta investigación no se centra en ellas, sino como decimos ha actuado simplemente (o nada menos) como una de las motivaciones de esta investigación.

Para contextualizar esta problemática, en el resto de esta investigación nos basaremos en el modelo de tres niveles de «guerra de información» expuesto en el año 1995 por W. Schwartau en su libro titulado *Information Warfare: Chaos on the Electronic Superhighway*.

En su obra, el autor distingue tres niveles de guerra de información que se diferencian por los individuos a los que va dirigido y por el objetivo final del ataque. De esta manera encontraremos los niveles de 1) Guerra de Información Personal, dirigido contra individuos, 2) Guerra de Información Corporativa un segundo nivel dirigido

contra grupos de personas o corporaciones, y 3) Guerra de Información Global dirigida contra esferas de influencia política, contra el sistema económico o contra estados, con el objetivo último de desestabilizar la sociedad civil generando alarma y caos.

Dejando de lado el aparente dramatismo del término, la cuestión es que en el año 1995 cuando W. Schwartau elaboró su modelo de Guerra de Información, la Blogosfera todavía no existía. Para entender la preocupación de la Comisión de Cultura y Educación por la seguridad en la Blogosfera complementaremos el anterior modelo con uno más reciente, el modelo de esfera de la información o Infoesfera propuesto en 2003 por W. Lance Bennett, profesor de ciencia política de la Universidad de Washington.

Según el marco teórico propuesto por W. L. Bennett (*Political Communication, Citizenship and Democracy: From the Mass Society to Personal Information Networks*, 2004), la Infoesfera se componía de tres estratos de tamaño e importancia bien diferenciadas: 1) El *estrato superior* compuesto por los medios de comunicación de masas: la radio, la televisión y la prensa escrita; 2) El *estrato medio*, que incluye entre otros blogs destacados, revistas electrónicas y grupos de presión; y 3) El *estrato micro* compuesto por el correo electrónico, listas de distribución y blogs personales.

En este modelo el flujo dominante de información tiene su origen en el estrato superior o convencional compuesto por los medios de masas, ejerciendo su influencia sobre los otros dos estratos, el medio y el micro.

A diferencia del estrato convencional, los dos estratos inferiores no son influyentes. Sin embargo, existe la posibilidad de que una noticia proveniente de uno de ellos active toda la Infoesfera a través de una estrategia de comunicación exitosa alcanzando los medios de comunicación de masas del estrato convencional.

Dependiendo de cuál sea el objetivo final perseguido, nos encontraríamos ante dos tipos de actuaciones, unas propias del marketing o la publicidad y otras propias de la guerra de información. En el primer caso los encargados de ejecutar estas campañas de comunicación serían los profesionales de Social Media o de Marketing. Sin embargo, en el segundo caso, estaríamos frente a la actuación de «activistas» o «ciberterroristas» guiados por una posible «agenda oculta».

La posible combinación del modelo de W. Lance Bennett con el modelo de Guerra de Información de Schwartau muestra cómo es posible desestabilizar la sociedad civil y generar el caos si se alcanzan los medios de comunicación de masas a través de alguno de los dos estratos inferiores. Si bien todavía no se han dado casos de Guerra de Información Global, veremos casos que se ajustan al modelo de la Infoesfera y que supusieron una seria amenaza que podría haber afectado a la economía nacional generando una alarma social difícil de controlar y que podría haber culminado en una crisis económica global, como los casos del Bank of East Asia en septiembre de 2008 o en 2009 el caso del bloguero surcoreano Park Dae-sung, apodado «Minerva».

Como veremos, el modelo teórico de W. L. Bennett refuerza la preocupación por las vulnerabilidades de la generación Google, y ante el riesgo potencial que suponen se hace necesaria e imprescindible la alfabetización digital como una forma de prevención de una potencial amenaza al pluralismo derivada de las características tanto de los *weblogs* como de la estructura de la Infoesfera.

Las actuaciones propuestas por la Comisión de Cultura y Educación, la alfabetización de la generación Google y el control de la Blogosfera, son dos medidas de carácter preventivo que persiguen el objetivo de evitar que información difundida a través de Internet pueda convertirse en una amenaza potencial, no sólo para los miembros de la generación Google destinados a ser los futuros consumidores de la sociedad de la información, sino para evitar que, en un caso de guerra de información global, la alarma generada en los medios digitales se propague al mundo real en forma de manifestaciones populares, crisis económica o caos social.

Con esta clase de motivaciones iniciales, consideramos necesario analizar el impacto de blogs en los resultados de buscadores, desarrollando para esta investigación un modelo de análisis que permitiese observar el comportamiento de sus estratos. Por lo tanto, para analizar el comportamiento de los blogs en el ámbito de los resultados de buscadores se hizo necesario encontrar un modelo capaz de abarcar toda la realidad de la web contemporánea y capaz de ser aplicado, en principio, a cualquier sector o temática.

A modo de actualización, volveremos a definir los tres estratos de la esfera de la información añadiendo a cada uno de los estratos los diferentes tipos de páginas web aparecidas a raíz de la evolución de la web 2.0. Se ha intentado respetar al máximo la

clasificación original de sitios web establecida por Peretti y Bennett, en especial en lo relacionado a la autoría de los contenidos de tipo UGC (contenidos generados por los usuarios), principal bastión del estrato micro.

No obstante, como ya tendremos ocasión de presentar y discutir con detalle, lo que acabamos encontrando, como en los mejores relatos de viajes, es una realidad algo diferente. Más que amenazas por la existencia de una multitud de potenciales fuentes de influencia desconocida, lo que nuestra investigación longitudinal ha puesto de manifiesto es que, de existir alguna amenaza esta consiste en el papel de facto que Google ha adquirido como *gatekeeper* y la pre-eminencia que Google ha acabado otorgando a los grandes medios y corporaciones en detrimento de la visibilidad que podría haber dado a la web social. Por eso, avanzando resultados, hablamos en algunos momentos de una pérdida del potencial democratizador en cuanto a que el papel de Google ha consistido, a lo largo de los últimos años, en restar visibilidad a los medios que podemos llamar alternativos en favor de las grandes corporaciones tradicionales.

0.3. Antecedentes y marco metodológico

Durante el verano del año 2010 presenté un proyecto de tesis dirigido por el doctor Lluís Codina cuyo título provisional fue *Dinámicas de información y contrainformación política en medios digitales: propuestas de análisis y estudio comparativo de procesos electorales en Cataluña*.

El periodo de investigación se inició en noviembre de 2010 con las elecciones al Parlament de Catalunya, y se alargó hasta el mes de noviembre de 2012, cuando tras un adelanto electoral marcado por la posible convocatoria de un referéndum de autodeterminación catalán, se volvieron a celebrar las elecciones al Parlament de la Generalitat de Catalunya. Estos comicios electorales pusieron punto final al periodo objeto de estudio que abarcó un total de ocho comicios electorales celebrados en el estado español.

El proceso de *scraping* (extracción) de resultados se inició por tanto en noviembre del año 2010 con las elecciones al Parlamento de Cataluña del 20 de noviembre. Este primer ensayo permitió llevar a la práctica la metodología de extracción planificada y perfeccionar el proceso para los siguientes comicios. En el planteamiento original de la tesis se pretendía estudiar la visibilidad en buscadores de webs de partidos políticos y medios de comunicación tradicionales durante los días de campaña y pre-campaña de las elecciones a la Generalitat de noviembre de 2010 y las Municipales de mayo de 2011. Sin embargo, conforme llegamos a 2011, la selección original de procesos electorales se amplió a las Elecciones Generales de España del mes noviembre, que se saldó con la victoria del Partido Popular con mayoría absoluta.

El año 2012 comenzó con las elecciones a la Junta de Andalucía de mayo de 2012, seguidas por las elecciones al Parlamento de Galicia de octubre para finalizar con las elecciones al Parlamento de Cataluña de Noviembre 2012. Por tanto, la metodología ideada para las dos elecciones se aplicó a todos los procesos electorales que tuvieron lugar en España a lo largo de 2011 y 2012, añadiendo en la captura de datos de cada comicios mejoras metodológicas e instrumentales.

Además, también se capturaron datos de las elecciones a la Junta General del Principado de Asturias de 2012 y de las elecciones al Parlamento Vasco de 2012, pero

estos dos comicios electorales fueron finalmente rechazados de la muestra final por su dificultad, lingüística en el caso del País Vasco y por la duplicación de elecciones en el caso de Asturias⁹.

Con las elecciones al parlamento de Cataluña de noviembre de 2012, celebradas por segunda ocasión durante el periodo de análisis debido al adelanto electoral, se alcanzó una cifra superior al millón y medio de URLs y finalizó el proceso de captura de datos, dando inicio a la fase final de contrastación de hipótesis y de redacción de la tesis doctoral.

Para conseguir los objetivos señalados, ha sido necesario hacer un seguimiento de los contenidos volcados por medios de comunicación y partidos políticos durante los procesos electorales. Con tal fin, los sitios web han sido agrupados y clasificados según el modelo teórico de Infoesfera, capaz de medir la visibilidad tanto de grupos de webs como sitios individuales, y se ha llevado a cabo un análisis de su evolución a lo largo de los días de campaña de cada una de las elecciones objeto de estudio.

Entre los objetivos secundarios se encontraba el descubrir las características de la evolución de los resultados que hubiesen alcanzado grados significativos de visibilidad en el así llamado «top10» (reciben este nombre los primeros 10 resultados de una SERP). Con este objetivo se ha realizado un análisis basado en parámetros de visibilidad online que incluye los más de 33.000 sitios web identificados, haciendo especial hincapié en el análisis de grupos mediáticos durante la campaña, en sus ritmos de publicación, estancia de noticias en el Top10, temática y tendencia política.

La acumulación de datos de posicionamiento web de seis procesos electorales ha permitido analizar las muestras desde múltiples puntos de vista, permitiendo realizar tanto estudios transversales en los diferentes comicios de cualquier estrato de la esfera web, como estudios longitudinales centrados en estratos o comunidades online específicas de un proceso electoral en concreto.

El millón y medio de URLs capturadas a lo largo de dos años han permitido desarrollar una metodología de estudio y de análisis de contenidos que permite comparar los distintos procesos electorales al identificar los componentes principales de resultados

⁹ Las elecciones a la Junta General del Principado de Asturias se celebraron en dos ocasiones debido a la retirada de apoyo del Partido Popular al partido encabezado por Álvarez-Cascos.

de buscadores (SERPs, por las siglas en inglés de *Search Engine Results Pages*). El análisis de estos componentes y su evolución a raíz de los últimos algoritmos de Google es el resultado de investigación de esta tesis.

De entre las elecciones destacan las elecciones a la Generalitat, que son los comicios que dieron origen a este trabajo. La doble captura de un mismo paquete de palabras clave permitirá ver la evolución real del comportamiento de Google en esos dos años transcurridos entre 2010 y 2012.

Para poder estudiar los resultados de este periodo se ha desarrollado una metodología específica de análisis de la Web aplicando un modelo de clasificación a los resultados en buscadores, en especial el motor de búsqueda Google, basado en el modelo teórico de la Infoesfera.

La metodología desarrollada se basa en datos de posicionamiento web ¹⁰, entendiendo éste como el conjunto de procedimientos y técnicas que estudian las características que proporcionan a un sitio o una página web la máxima visibilidad en Internet. El conjunto de resultados estudiados en esta tesis pertenecen a aquellos sitios web que mejor han implementado las técnicas de optimización en buscadores, más conocido por sus siglas en inglés, SEO, de Search Engine Optimization, y que tras un procedimiento de optimización, han colocado un sitio o un conjunto de páginas web entre los diez primeros resultados proporcionados por un motor de búsqueda.

Los resultados aparecidos en estas posiciones de Top10 son calculados por los algoritmos de los buscadores, escogiendo entre todos los sitios web posibles de la Web aquellos considerados relevantes o pertinentes en relación con un evento central, concepto o tema planteado por los usuarios a través de palabras clave. Las páginas de resultados que dan respuesta a estas palabras clave, también conocidas por las siglas inglesas SERPs, de Search Engine Result Pages, están compuestas por enlaces a sitios web que albergan recursos digitales en diversos formatos y que a menudo están conectados por hipervínculos entrantes o salientes con otros sitios web.

¹⁰ Posicionamiento web: Conjunto de procedimientos y técnicas que estudian las características que proporcionan a un sitio o una página web la máxima visibilidad en Internet; conjunto de procedimiento que permiten colocar un sitio o una página web en un lugar óptimo entre los resultados proporcionados por un motor de búsqueda. Por extensión: optimizar una página web de cara a los resultados proporcionados por los motores de búsqueda. Codina (2004)

Las fronteras de la esfera web objeto de estudio están delimitados por una orientación temática adscrita a la comunicación política, y por un marco temporal múltiple, coincidente con los días de campaña electoral de los comicios estudiados. Sin embargo, un elemento importante en la conceptualización de la Infoesfera y de todos los sitios web aparecidos en resultados de SERPs es la naturaleza dinámica.

De esta forma, en el transcurso de la identificación y clasificación de sitios web pertenecientes a los límites de la esfera web política, en cada proceso electoral se han hallado nuevos sitios que han sido incluidos en la clasificación y que han incrementado la base de datos de sitios de referencia más allá de los 33.000 registros.

En este sentido, la definición de los límites y componentes de cualquier esfera de la web es un proceso recursivo en tanto que los sitios inicialmente incluidos, considerados como parte de la esfera en evaluación, se han visto alterados en cada nueva captura de datos a través del alta de nuevos sitios y la baja de sitios expirados. Por lo tanto, como una esfera web se analiza a lo largo del tiempo, sus límites cobran forma a través de las estrategias de identificación de los investigadores y de cambios en los propios sitios.

La base de datos resultante ha sido estructurada en forma de una ontología de categorías web que engloba todos los sitios aparecidos durante la captura de datos. Esta ontología ha requerido de un análisis previo que categorizase todos los sitios de la esfera de información española en diferentes niveles, categorías y subcategorías denominadas estratos, coronas, sectores y conglomerados. El resultado es una ontología web que permite categorizar cualquier resultado de SERPs de buscadores bajo un modelo general llamado Infoesfera, basado en el modelo teórico propuesto por W.L.Bennet y que hemos aplicado al análisis de resultados de ámbito político.

De forma paralela a la creación y categorización de la base de datos de sitios web aparecidos en SERPs, la metodología desarrollada para su análisis ha sido depurada a lo largo de todo el proceso de recolección de datos, permitiendo conocer a fondo, cómo y en qué proporción Google ha posicionado resultados web.

La parte principal de los resultados es un estudio global de visibilidad de sitios web agrupados en grupos homogéneos denominados conglomerados que engloban desde los grandes grupos mediáticos españoles hasta los blogs personales. Los cálculos

de visibilidad en Google realizados sobre la base de datos de resultados de elecciones han permitido medir cuál ha sido el posicionamiento orgánico de cualquier conglomerado o página web específica, permitiéndonos conocer y medir la visibilidad y su evolución durante el periodo 2010-2012.

Entre otras, las principales categorías que trataremos serán los sitios web de prensa y de partidos políticos, las blogosferas políticas de partido, los sitios web de las administraciones del estado y la visibilidad de los principales portales web y de diferentes grupos temáticos y colectivos. Haremos especial repaso a diferentes sectores de la Infoesfera, desde el papel de los grandes medios españoles y su evolución en las elecciones autonómicas españolas, hasta el detalle sobre la aparición y evolución a partir de 2011 de sitios web asociados a los movimientos 15M y «Democraciarealya».

El estudio final ha permitido además, conocer en profundidad cómo han afectado los algoritmos Panda y Penguin a los resultados web y si las páginas web han sido adaptadas a los nuevos factores *onpage* de un sitio optimizado para SEO. Aunque estos dos algoritmos de Google no se tuvieron en cuenta a la hora de plantear la investigación, su aparición durante el periodo de investigación ha supuesto modificaciones tanto en los resultados de Google como en las hipótesis de partida, y han ayudado a dar explicación a los resultados hallados.

Después de ampliar el ámbito de estudio con las elecciones de 2011 y 2012, y cerrarlo finalmente tras las elecciones a la Generalitat de 2012, el título final de la tesis es *Posicionamiento Web y Dinámicas de Información en motores de búsqueda: propuestas de análisis y estudio comparativo de visibilidad de contenidos digitales en el caso de procesos electorales*.

0.4. Justificación de la investigación

Como hemos avanzado, este proyecto nació como consecuencia de la revolución de la web 2.0 y la necesidad de medir su composición, para lo cual se ha aplicado el modelo teórico de Infoesfera ideado por Jonah Peretti en 2001 y desarrollado por W.L. Bennet en 2004.

El crecimiento de esta web 2.0 y la aparición de nuevas plataformas ha modificado los patrones de la comunicación tradicional, de tal manera que las teorías sobre comunicación vigentes hasta el momento deben ser revisadas y actualizadas de forma paralela a la evolución de las tecnologías sobre las que se sustenta la web. Esta nueva realidad hace necesaria la creación de nuevas teorías y estudios que hagan inteligibles las nuevas reglas de juego en comunicación digital en general, y en comunicación política en particular.

El valor potencial del estudio y la motivación principal para llevar a cabo esta investigación ha sido entender este nuevo entorno y de qué forma la web 2.0 y las redes sociales han modificado (si es que lo han hecho) las reglas de juego y han mutado la comunicación de masas tradicional de la segunda mitad del siglo XX hacia un nuevo sistema de comunicación personalizada. Este tránsito teórico del *onetomanyalmanytomany* debe ser medido y estudiado para su comprensión, de ahí el interés de este estudio.

A. Valor teórico

Entendemos que este proyecto desarrolla el modelo teórico de Infoesfera y abre nuevas vías para la investigación, no solo de campañas políticas en internet, sino que ayuda al desarrollo del conocimiento de la comunicación social en cualquier área temática, y establece nuevas bases para futuros estudios.

De forma más específica, uno de sus principales objetivos ha sido definir un nuevo modelo de análisis relativo a la comunicación en internet, estableciendo nuevas variables de seguimiento y estudiando la composición e interrelación de los diferentes estratos de la Infoesfera durante las elecciones. El proyecto pretende aportar nuevas

ideas para la investigación, y abre un abanico de recomendaciones y nuevas hipótesis para futuras investigaciones que destacaremos más adelante, tanto dentro del campo comunicación social como de la comunicación política.

El estudio de las campañas de comunicación y de las diferentes acciones de los diferentes partidos políticos ha servido además para desarrollar teorías acerca de cuales podrían ser las nuevas tendencias de comunicación política y electoral en las elecciones venideras.

B. Utilidad metodológica

El diseño y desarrollo de la metodología ha supuesto integrar toda una nueva tipología de sitios web que, como se destacará en el estado de la cuestión, no existía en el momento de teorizar el modelo dada su temprana aparición en el año 2001.

La metodología elaborada a raíz de esta tesis es extrapolable a cualquier paquete de palabras clave temáticas que correspondan a diferentes nichos de búsqueda. Estos nichos temáticos, delimitados por un conjunto limitado de palabras clave relacionadas temáticamente, mostrarán la misma estructura de estratos definidos en el modelo, solo que con una composición y proporciones diferentes. Las proporciones de estratos variarán dependiendo de las características del nicho temático. Por ejemplo, si nos centramos en el primer estrato del modelo compuesto por medios de prensa convencionales, cualquier nicho temático estará compuesto por distintas proporciones de webs de medios que serán propios de los resultados recuperados desde un país específico (es decir, es muy probable que cualquier nicho temático tenga noticias relacionadas provenientes de sitios web de medios, sin embargo, estos medios serán los propios del país desde donde se realice la consulta); en el caso del estrato medio y del estrato micro, las webs relevantes de un ámbito temático y los diferentes blogs relevantes que traten el tema, rara vez aparecerán para ningún otro nicho de búsqueda, ya que corresponderán a grupos propios de dicha temática. Sin embargo, a pesar de estas diferencias, encontraremos la misma estructura de estratos y una tipología similar de sitios web a la definida en el modelo en cualquier nicho de búsqueda.

Por otro lado, el desarrollo de este modelo llevado a cabo en esta tesis, ayudará a rediseñar y mejorar los métodos de análisis y de recolección de datos aplicados a proyectos similares, que más adelante se podrán utilizar para analizar en profundidad otros casos, no solo dentro del área de la comunicación electoral a nivel de política española o internacional, sino que permitirá analizar cualquier caso de comunicación social sea cual fuere su campo temático.

c. Implicaciones prácticas

Las implicaciones prácticas directas del análisis de los resultados web durante las campañas son las siguientes:

1. Se ha podido constatar el cambio de la composición de SERPs de Google durante el periodo 2010-12, tanto a nivel global de los estratos como a nivel concreto de los grupos más significativos.
2. El estudio de las dinámicas de comunicación a través de los diferentes estratos de la Infoesfera nos ha permitido detectar flujos característicos de una posible actividad de desinformación, identificados a través de los llamados «Key Performance Indicator» (KPI) de visibilidad y del control de sus valores medios.
3. El estudio ha permitido calibrar y comparar la actividad de los diferentes sitios web y grupos en cualquier parte de la Infoesfera, tanto en la esfera convencional, como en el estrato medio como en el estrato micro.

A medio plazo, los resultados de este estudio pueden ser una contribución a un amplio rango de aplicaciones que podrían ser usadas en campos muy diversos, desde la *Computer-Mediated Communication* hasta los estudios de Cultura Digital pasando por el SEO y otras disciplinas vinculadas con las Ciencias de la Web.

Desde un punto de vista más práctico, los conocimientos resultantes pueden contribuir a la mejora de la comunicación política en internet, al estudio de los llamados movimientos sociales 2.0, o incluso pueden complementarse con estudios de impactos de publicidad y marketing online, en especial con el llamado marketing de contenidos y *social media marketing*.

A largo plazo, conforme transcurra el tiempo y la importancia de internet aumente, los analistas predicen que los medios tradicionales serán reemplazados por una nueva generación de medios basados en el Web.

De cumplirse estos pronósticos, y en tal caso a sabiendas de que internet acabará siendo el medio de comunicación por antonomasia desbancando/absorbiendo a la televisión y a la prensa escrita, estudios como el aquí desarrollado serán indispensables para conocer y teorizar el futuro próximo de la «Sociedad de la Información».

Sin embargo, y por esta misma razón, la validez de los datos aquí mostrados conlleva una caducidad implícita a la composición de la búsqueda en el momento en que fueron obtenidos los datos. A pesar de ello, el modelo desarrollado en esta tesis podrá ser adaptado a la realidad de los buscadores mientras éstos sigan existiendo como tales, debiendo actualizar la base de datos de sitios y conglomerados, añadiendo los nuevos dominios y plataformas una vez vayan apareciendo en los resultados de buscadores.

Por otro lado, la investigación ha ayudado a definir y desarrollar un nuevo modelo de investigación y estudio de comunicación digital extrapolable a cualquier nicho temático. El conocimiento efectivo sobre la relación de los diferentes estratos de la Infoesfera en el ámbito temático político nos ha permitido crear un modelo que puede ser adaptado a cualquier otro ámbito temático en cualquier otro idioma.

Los conocimientos resultantes pueden extrapolarse fácilmente a otros marcos de estudio fuera de los ámbitos de la comunicación política, entre los que destacan los estudios de identidad digital, tanto a nivel de identidad corporativa como a nivel de identidad personal.

Finalmente, el estudio previo del estado de la cuestión llevado a cabo nos ha permitido aportar, entre otros resultados que entendemos valiosos para el sector de estudiosos y profesionales del SEO, la identificación de los más de 200 factores (plausibles) de posicionamiento que utiliza Google de acuerdo con la mayoría de los analistas. En nuestra aportación, además de identificar esos factores, los hemos sometido a un tratamiento sistematizado y los hemos caracterizado, uno a uno, de acuerdo con un grupo de atributos seleccionados por nosotros como resultado de nuestros análisis previos.

D. Relevancia social y política

Aunque ya hemos avanzado algunos argumentos en este sentido, ahora desde un punto de vista académico, entendemos que esta tesis puede ser relevante ya que se trata de un estudio pionero de análisis de contenidos de actualidad política en períodos electorales en internet. La aplicación del modelo de Infoesfera a los seis procesos de elecciones analizados permite observar la evolución de la comunicación política en el intervalo de tiempo de 2010-2012, así como el comportamientos de sus principales actores, pudiendo analizar la cantidad de contenidos aportados a la red de redes durante los días de campaña por los diferentes partidos políticos y por los principales medios convencionales.

Gracias a los resultados de esta tesis, tanto los estudiosos de la comunicación social en general como los profesionales de comunicación política o de la comunicación digital podrán acceder a un análisis pormenorizado del comportamiento de buscadores en esta nueva realidad comunicativa digital. Así, de igual forma que los medios convencionales podrán ver cuál ha sido su evolución en visibilidad durante el periodo objeto de estudio, los partidos políticos podrán aplicar y replicar el modelo de estudio para analizar la eficacia de sus campañas electorales y las de sus rivales en los próximos comicios electorales.

0.5. Hipótesis de partida y preguntas de investigación

El conjunto de antecedentes y el contexto mencionados en los apartados anteriores, nos llevó al planteamiento hipotético siguiente, en el que basamos el arranque de nuestra investigación:

El medio Internet, y más concretamente los resultados obtenidos a través de buscadores, podría suponer un aumento de la visibilidad para los medios digitales considerados alternativos, trayendo esto como consecuencia un aumento de la diversidad en las fuentes de información disponibles para los ciudadanos.

Ahora bien, dado el carácter altamente «conjetural» de la propuesta anterior, fue necesario derivar de ella unas preguntas de investigación que presentaran un aspecto mucho más operativo.

Entendimos que, si fuera posible diseñar una investigación con capacidad de responder a estas preguntas, entonces la hipótesis inicial podría verse como o bien como una propuesta descriptiva aceptable o bien como algo a desestimar.

Por el camino, dado el carácter (como se verá) mucho más «operacionalizable» de las preguntas de investigaciones teníamos la esperanza de obtener algunos resultados de valor, incluso si al final hubiera que descartar (o simplemente no fuera posible demostrar) la conjetura mencionada.

0.5.1. Preguntas de investigación

Partiendo de la conjetura inicial ya expuesta nos planteamos las siguientes preguntas de investigación:

1. ¿Es posible desarrollar el modelo de la Infoesfera enunciado por J. Peretti (2001) y W. L. Bennet (2004), pasando del planteamiento teórico a una plasmación del modelo con datos reales en un contexto electoral?
2. En tal caso, ¿sería posible aplicar y poner a prueba el modelo evolucionado de la Infoesfera a través de los resultados de los buscadores?
3. Una vez actualizado el modelo, ¿sería posible extrapolarlo a otros ámbitos temáticos?
4. ¿El modelo puede ayudar a identificar factores críticos de éxito en reputación e identidad digital?

0.6. Tipo de investigación, metodología y fuentes de información

0.6.1. Tipo de investigación

La investigación desarrollada en esta tesis es exploratoria descriptiva.

Se trata de una investigación exploratoria en tanto que desarrolla un modelo que fue teorizado por J. Peretti y W.L. Bennett entre los años 2001 y 2004 y que no se había aplicado desde entonces a un escenario real.

La investigación es descriptiva en tanto que, en base a resultados orgánicos de búsqueda extraídos mediante una selección de consultas electorales relevantes, busca especificar cuáles son los componentes principales, así como las propiedades, características y rasgos más importantes de los resultados observados.

Además, también se trata de una investigación descriptiva en tanto que, en base a una serie temporal de seis elecciones del periodo 2010-12, ha permitido observar los cambios que han producido los algoritmos Panda y Penguin en los resultados orgánicos de búsqueda de Google durante el periodo 2010-2012.

0.6.2. Metodología

Estrategia metodológica

La estrategia metodológica parte del marco teórico referencial del modelo de Infoesfera, expuesto en 2004 por los autores Jonah Peretti y W. L. Bennett. Este marco teórico se ha aplicado a diversos paquetes de datos que corresponden a un marco temático y temporal coincidentes con las fechas de seis procesos electorales llevados a cabo en España entre 2010 y 2012.

La puesta en práctica de dicho modelo ha requerido del desarrollo de una metodología que permitiese aplicar los conceptos teóricos a los datos recabados durante las elecciones. Dicho proceso se ha compuesto de las siguientes fases:

1. La creación de una metodología de extracción, recolección y tratamiento de los datos que permitiese recuperar la información de los resultados de buscadores para su posterior análisis.
2. Diseñar un método analítico que permitiese comparar las acciones de los diferentes agentes (sitios web o conglomerados) durante las elecciones.
3. Ejecutar el modelo de recolección de datos y sistematizarlo para poder aplicarlo a toda la serie temporal de elecciones.
4. Elaborar un modelo de interpretación para toda la información extraída.
5. Evaluar los datos extraídos y detectar casos de resultados anómalos en cualquiera de las opciones de búsqueda de los buscadores. Con tal objetivo, se analizan y comparan los resultados de visibilidad de diferentes conglomerados y/o páginas web específicas, en especial aquellos que puedan haber sido afectados de alguna manera por los agentes de la llamada *agenda-setting*, o por grupos con una agenda oculta como los mencionados por la eurodiputada socialista Marianne Mikko.

Enfoque metodológico

La complejidad del objeto de estudio nos ha obligado a que el diseño metodológico de la investigación combine métodos cuantitativos y cualitativos.

• Enfoque cuantitativo

El estudio cuantitativo se basa en los ratios de visibilidad observados, tanto de cada dominio individual, como agrupados en conglomerados y estratos. Las proporciones estándares y las variaciones de visibilidad de cada uno de los estratos en los diferentes comicios electorales varía para cada una de las opciones de búsqueda objeto de estudio. De igual forma veremos cuáles son los ciclos de vida de los contenidos que han alcanzado visibilidad.

1. Websearch. El estudio básico se basa en ratios de visibilidad y variaciones de visibilidad de los estratos, lo que incluye además, observaciones sobre diferentes sectores y conglomerados específicos, destacando en algunos casos los dominios más representativos o las URL cuya frecuencia en SERPS destaca por su permanencia y/o posicionamiento.

2. Blogosfera. De igual forma, los estudios sobre la Blogosfera se basan en un análisis de visibilidad de blogs individuales y de sus conglomerados. Como resultado, se han desarrollado *rankings* de conglomerados y de blogs destacados que permiten detectar variaciones anormales de visibilidad que pueden estar asociadas a intentos de manipulación de resultados en la Blogosfera. De forma paralela, se analiza el ciclo de vida de *posts* específicos.

3. Google News. Los datos extraídos del agregador de noticias Google News han sido sometidos a un análisis de frecuencia de aparición en SERPS. Los resultados muestran la evolución de los diferentes medios online a lo largo de los dos años de observación, la frecuencia de medios para diferentes grupos de búsquedas temáticas, la posición de las noticias en el ranking y su ciclo de vida y permanencia, así como la regularidad de publicación, entre otros.

Aunque en un principio se planteó hacer un estudio específico de las diferentes redes sociales, los diferentes portales de social media y sus dominios asociados se analizan

englobados en diversos conglomerados propios. Los datos referentes a Facebook, Twitter y otras redes sociales se limitarán a la exposición de ratios con los Tweets o páginas que tuvieron visibilidad dentro de cada una de las opciones de búsqueda, así como su ciclo de vida en SERPS.

- **Enfoque cualitativo**

El enfoque cuantitativo se complementa con el estudio cualitativo de las diferentes páginas web, que asociadas a conglomerados temáticos, ayudarán a matizar y responder a las preguntas de investigación acerca de la composición de los estratos de la Infoesfera.

1. Websearch. El estudio permite agrupar las URLS específicas en función de las características concretas de los sitios web donde han sido publicados. Las cualidades principales son su pertenencia a uno de los tres estratos, si bien se pueden clasificar utilizando una detallada categorización basada en múltiples factores tales como la nacionalidad del sitio, su cobertura, temática, colectivo, etc.

2. Blogosfera. Los estudios cualitativos de la Blogosfera se basan en el análisis de diversos factores característicos, en especial la orientación temática y si pertenecen o no a alguna red de blogs.

3. Google News. Los datos de Google News permiten conocer la adscripción a un grupo mediático u otro, el origen de las noticias, ya sean propias o de agencias, así como la tipología de los contenidos con mayor visibilidad (artículos informativos, de opinión, blogs del medio, etc).

Respecto a los conglomerados propios de las diferentes redes sociales, se muestran las cuentas de redes sociales con visibilidad clasificadas según su pertenencia o grado de afiliación con los conglomerados de partidos políticos (cuentas oficiales de partido, de políticos, simpatizantes u otros).

0.6.3. Fuentes de datos

Las fuentes primarias de información son las SERPs, es decir, la captura de datos de resultados en buscadores. Mediante su estudio y observación durante las campañas electorales se analizan las proporciones en resultados tanto de los medios de prensa digital como de los partidos políticos o de redes sociales específicas.

Los paquetes de búsquedas están compuestos por un conjunto de **palabras clave** que pertenecen a un **grupo temático** o **nicho de búsqueda** particular para cada uno de los procesos electorales. Estas palabras clave han conformado un paquete de búsquedas informativas que incluyen además de búsquedas genéricas, consultas sobre candidatos políticos, consultas centradas en el nombre de partidos políticos, consultas que incluyen términos geográficos como nombres de provincias, y búsquedas temáticas centradas en términos específicos como pueden ser «encuestas» o «resultados».

Estos paquetes de palabra clave temáticos, iterados cada día durante las campañas electorales, han permitido la extracción de diferentes **conjuntos de SERPs** de diez resultados que han sido agrupadas en **series temporales**. En cada una de las **elecciones** estudiadas, estas series temporales o cronológicas engloban 14 días de observaciones: el día de celebración de las elecciones, el día inmediatamente posterior y doce días del periodo de campaña electoral.

0.6.4. Fuentes de información

En lo referente a las fuentes de información que hemos utilizado para llevar a cabo el estado de la cuestión de la investigación, podemos dividirlas en tres grupos: las monografías, los artículos de publicaciones periódicas extraídos de bases de datos y las páginas web.

Internet ha sido la fuente madre de un alto porcentaje de las fuentes mencionadas en esta tesis de lo que se desprende que casi la totalidad de los materiales utilizados hayan sido en formato electrónico, incluso los libros.

Entre las monografías consultadas encontramos principalmente manuales de SEO y libros sobre comunicación política, temas que han sido reforzados con lecturas sobre las redes sociales que han surgido en su seno. De forma paralela al avance de la investigación, a raíz de la intersección de las dos principales temáticas de nuestra área de trabajo, se ha consultado bibliografía sobre historia de la propaganda y sobre temas de guerra de información reciente, en especial Wikileaks.

En las bibliotecas universitarias hemos consultado algún material en papel, si bien la mayor parte de las visitas han sido para acceder a revistas a través de bases de datos de acceso restringido. Éstas han servido para hacer una criba selectiva de autores y temas con el objetivo de acotar mejor las búsquedas afines a nuestro trabajo y garantizar la máxima calidad en los documentos consultados. Se ha trabajado con bases de datos como EBSCO y Scopus, pero también se han consultado bases de datos de noticias como la de la Agència Catalana de Notícies. Como no, también se ha utilizado Google Scholar para buscar referencias o contrastar las localizadas en otras bases de datos.

Por último, los sitios web han sido una fuente fundamental de información, tanto las redes sociales como las páginas personales de expertos en posicionamiento y comunicación política. A través de ellas hemos contactado con consultores políticos, responsables de comunicación de partidos y SEOs, profesionales a quienes se han realizado entrevistas para intercambiar información. Entre los SEOs destacan los responsables de grandes medios de prensa nacional y otros profesionales eminentes de sector que solicitaron permanecer en el más estricto anonimato al haber sido identificados como spammers de Google a través nuestras capturas de datos.

Capítulo 1. Estado de la cuestión

1.1. El Agenda Setting y la propaganda en la era de Internet

Durante este último cuarto de siglo las tecnologías de la información y la comunicación, el desarrollo de Internet y la generalización de su uso a través de los ordenadores personales y de los teléfonos móviles han desarrollado un nuevo tipo de comunicación mediada por ordenador que ha provocado una transformación en el sistema comunicativo que, a su vez, está teniendo efectos en el sistema político así como en los procesos de formación de la opinión pública¹¹.

A finales del siglo XX los medios de comunicación de masas eran la suma de los medios de prensa, radio y televisión. La radio y la televisión habían nacido en el periodo de entreguerras y tras la segunda guerra mundial los Estados vencedores aplicaron las técnicas de propaganda aprendidas durante la guerra a las nuevas tecnologías de comunicación de masas con fines ideológicos y políticos. Al inicio de la Guerra fría, desde 1945 a 1970, se vivió una etapa de expansión económica que repercutió en el desarrollo del sector informativo y junto al periodismo escrito se desarrolló el periodismo en la radio y la televisión.

En el contexto de la Guerra Fría, de acuerdo con algunas corrientes críticas con la función de los medios abanderadas por autores como Edward S. Herman y Noam Chomsky (1988)¹² o Guinsberg (2005)¹³, estos medios de comunicación de masas se convirtieron en instituciones ideológicas efectivas y poderosas, que llevan a cabo una función propagandística de apoyo al sistema. La propaganda, generada por las élites corporativas y gubernamentales, viajaba a través de los medios de masas e influenciaba la cosmovisión de los sujetos a través de signos y mensajes masivos que a su vez tenían incidencia en el ámbito social e ideológico. En ambos bandos los medios de comunicación de masas cumplieron una función social de servicio público, convirtiéndose en grandes máquinas de propaganda.

Después de la crisis del petróleo de los setenta el negocio informativo creció y las empresas de información aumentaron su poder a través de la concentración de los medios. La caída del muro de Berlín en 1989 aceleró el proceso y muchas emisoras de

¹¹ Blumler, J. G. (2001). *The third age of political communication*. Journal of public affairs, 1, (3), 201-209

¹²Edward S. Herman y Noam Chomsky (1988). "Manufacturing Consent: The Political Economy of the Mass Media"

¹³Guinsberg, Enrique (2005). *Control de los medios, control del hombre. Medios masivos y formación psicosocial*. Plaza y Valdés Editores.

radio y cadenas de televisión públicas pasaron a manos privadas concentrando todavía más el número de emisores y convirtiendo la información como un fenómeno supranacional con un claro predominio de las agencias y cadenas de televisión americanas.

Según las corrientes críticas mencionadas, la mayor concentración de los medios de comunicación de masas facilitó el control de los medios, estableciendo un sistema de propaganda cada vez más eficiente que permitía manipular a los públicos a través de la manipulación de los medios y la televisión. Según Guillermo López García (2006)¹⁴, las características del funcionamiento de este sistema es el que pretendía analizar la denominada Teoría de la **Agenda Setting**, según la cual, la prensa y los medios de comunicación no reflejan la realidad, pero sí la filtran y la moldean para sus públicos. Esta teoría se fundamentaba en dos parámetros, por un lado, la idea de que para configurar la realidad cognoscible el público se basa casi exclusivamente en fuentes de información proporcionadas por medios de comunicación de masas. Por otro lado, **el segundo parámetro** es que el sistema de medios de comunicación de masas se caracteriza por la existencia de unos pocos medios fuertemente jerarquizados y unidireccionales, que concentran audiencias masivas y transmiten unos contenidos **periodísticos estandarizados** tanto en el fondo, por los temas tratados a los que se da más relevancia, como en la forma, por la manera en la que los periodistas elaboran la información.

Guillermo López García (2006) muestra las características de estos contenidos periodísticos estandarizados en la denominada "mcdonalización de la sociedad", basada a su vez en las ideas expuestas por el sociólogo estadounidense George Ritzer (1996)¹⁵. Según esta corriente de pensamiento, a pesar del aparente aumento del pluralismo en los medios de comunicación, asistimos a una "mcdonalización" de las noticias mostradas por los medios convencionales. Las características que según Ritzer muestran los restaurantes McDonald's:

- En primer lugar, la **eficacia**. El sistema de los McDonald's garantiza los mejores medios disponibles, en apariencia al menos, para satisfacer nuestro apetito de la forma más rápida y eficaz posible.

¹⁴Guillermo López García. *Comunicación digital y líneas de fractura en el paradigma de la agenda setting*. 2006: 37-58

¹⁵Ritzer, G. (1996). *La Mcdonalización de la sociedad. Un análisis de la racionalización en la vida cotidiana*. Barcelona: Ariel.

– En segundo lugar, McDonald's se centra en ofrecernos un servicio en el que se nos da "**más por menos**", o lo máximo posible por la menor cantidad de dinero. Se trata, según Ritzer, de asimilar cantidad con calidad: mucha comida equivale a muy buena comida. Además, ahondando también en su eficacia, McDonald's garantiza que esta enorme cantidad de alimentos llega a nosotros en el menor tiempo posible, tanto en lo que se refiere a la recepción de los mismos como al consumo. Los restaurantes de comida rápida no lo son únicamente por la velocidad del servicio, sino también porque los productos adquiridos son de rápido consumo. McDonald's es un símbolo de la cultura de la **velocidad**.

– En tercer lugar, McDonald's nos ofrece **aquello que es previsible**, es decir, los alimentos consumidos son siempre los mismos, de sabor y cantidades idénticas, sin que importe lo más mínimo el lugar o el momento del consumo. El modelo McDonald's reduce (o elimina) la incertidumbre.

– Por último, McDonald's ejerce un **férreo control** sobre los individuos que forman parte de su mundo mcdonalizado, se trate de operarios o de clientes, y ese control se efectúa partiendo de la tecnología: la organización del espacio en función de la necesidad de que los clientes consuman lo antes posible y se marchen, la utilización de máquinas que garanticen que el servicio y el producto sean siempre los mismos, el adoctrinamiento de los trabajadores para que efectúen sus rutinas productivas siempre de la misma manera.

Según este reduccionismo de los autores anteriormente citados, estas pautas de funcionamiento de los McDonald's (**eficacia, cantidad y velocidad**) como sinónimos de calidad (ofrecimiento de productos previsibles y control) confunden la buena información con la abundancia de la misma. Este patrón de consumo y calidad se corresponden casi exactamente con unos criterios **de generación de informaciones** en los que prima la velocidad de transmisión de datos sobre su valor cualitativo, y donde las fuentes de las informaciones y agencias estandarizan cada vez en mayor medida el producto informativo.

En base a esta teoría de la "mcdonalización de la sociedad", el producto informativo derivado es una información escasa o nulamente elaborada por medios de

comunicación que no disponen ni de la infraestructura necesaria ni del incentivo económico para desarrollar informaciones propias y diferenciadas de la competencia, y que se dedican a replicar las noticias en base a criterios periodísticos de selección de los hechos relevantes que homogeneízan la actualidad para el gran público.

Según el primer nivel de la teoría de la Agenda setting¹⁶, este **enfoque estandarizado** que dan los medios de comunicación de masas sobre ciertos aspectos de la realidad **es poderoso** para indicar al público no qué pensar, sino sobre qué pensar. De esta manera, los medios de masas moldean la agenda informativa mostrando noticias que hacen pensar al público que son las noticias más importantes de lo que está realmente sucediendo.

Sin embargo, los dos parámetros en los que se fundamenta el Agenda Setting han sufrido una considerable variación en la última década, ya que el desarrollo de las **tecnologías digitales** ha provocado que el número de medios de comunicación a los que tienen acceso los ciudadanos se haya multiplicado de forma exponencial. Si a finales del siglo XX los tres medios de comunicación de masas eran la prensa, la radio y la televisión, a inicios del siglo XXI debimos añadir el **medio internet** y la suma de todos los medios de comunicación utilizados para canalizar la información hacia las grandes audiencias, es decir, todos los medios web que a día de hoy nos aseguran que los eventos se conozcan de forma inmediata en todo el planeta.

Frente al ecosistema mediático de finales del siglo XX caracterizado por un número menor de medios de comunicación, en gran parte fuertemente jerarquizados, y dirigidos a una audiencia de masas fundamentalmente pasiva, la nueva comunicación digital que ha supuesto Internet, ha posibilitado un aumento de los

¹⁶ “En el estudio retrospectivo que en 1993 presentaron McCombs y Shaw sobre la evolución de la teoría de la *agenda-setting* consideran que el desarrollo de la investigación ha atravesado cuatro fases. La primera es la confirmación del hallazgo de 1972. La segunda se centra en la observación de las posibles condiciones que amplían o limitan el planteamiento de la agenda de los medios de difusión, La tercera, por una parte, extiende la consideración del efecto más allá de la sola transferencia de temas, e indaga si los medios también consiguen influir en la consideración pública de aspectos y características de candidatos (la transferencia de la "agenda" de los temas se denomina entonces "primera dimensión" o "primer nivel" del efecto, y la transferencia de los aspectos "segunda dimensión" o "segundo nivel"); y, por otra parte, extiende la investigación a otros asuntos de interés personal, no limitándose a la política". Esteban LÓPEZ-ESCOBAR (1996) Agenda-setting: investigaciones sobre el primero y el segundo nivel. *Communication & Society* 9(1 y 2), 9-15. <http://www.unav.es/fcom/communication-society/es/articulo.php?art_id=286>

medios de comunicación disponibles para el público, tanto en el plano cuantitativo como al cualitativo¹⁷:

Cuantitativamente: gracias a la aparición de Internet, el nuevo soporte a través de redes digitales, junto a los medios de comunicación tradicionales encontramos un oferta mediática digital que aparentemente tiende al infinito y que en la actualidad incluye la blogosfera y las redes sociales.

Cualitativamente: si bien muchos medios de comunicación tradicionales tienden a seguir unas reglas de producción y presentación de noticias estandarizadas, al menos algunos de los medios de mayor calidad en cada país intenta adaptarse. Además, esta infinidad virtual de medios digitales cada vez más variadas, dispersas y especializadas, en ocasiones ajena a los flujos de información convencionales, complementa a los medios de comunicación de masas y ofrece al público un abanico de informaciones mucho más plural, aunque no siempre fiable.

Esta nueva realidad es la que supone una fractura con respecto a la supuesta Mcdonalización que apoyan algunos de los autores mencionados. Autores como Guillermo López abogaban en el 2006 por un incipiente **cambio de paradigma** en lo relativo a los efectos de los medios de comunicación sobre el público gracias a la aparición de Internet y las tecnologías digitales. Las audiencias masivas monopolio de los medios de masas convencionales se rompen con la llegada del entorno digital, ya que en lugar de una sola audiencia masiva que consume la misma información difundida por los grandes medios, aparecen múltiples audiencias, intereses y canales de información.

El punto importante para nuestra investigación es que, de acuerdo con la teoría de la Infoesfera, estos nuevos medios de comunicación a pesar de no ser tan influyentes en un principio, tienen una capacidad muy interesante, a saber, la capacidad de activar toda la esfera de información. Esta característica puede estar relacionada, a su vez, con la tradicional distinción entre "organismo de radiodifusión" y "público" que caracterizaba los medios de comunicación, pasando del modelo de difusión de la información del "uno a muchos" a un modelo del "muchos a muchos" en el que los usuarios pueden obtener información a través de infinitud de fuentes. A través de

¹⁷ Shaw, D. L.; Stevenson, R. L. y Hamm, B. J. (2001). "Agenda Setting Theory and Public Opinion Studies in a Post-Mass Media Age". *WAPOR Annual Conference*. Roma, septiembre de 2001.

medios como la blogosfera, las redes sociales e incluso las wikis en colaboración, los usuarios eligen sus propias fuentes para crear sus propios medios de comunicación, rompiéndose así el oligopolio imperante por los medios de comunicación tradicionales y dando pie a un nuevo escenario al aumentar el número de canales por los que la información llega al público.

Este nuevo escenario está suponiendo un **reto** para el antiguo modelo de difusión de noticias. Un modelo que según algunos analistas era en realidad un modelo de propaganda, como proponen Edward S. Herman y Noam Chomsky en 1988 en su libro "*Manufacturing Consent: The Political Economy of the Mass Media*" (*Fabricando el consenso: la economía política de los medios de comunicación*). Estos autores proponen un modelo que, según ellos, incorpora los mecanismos potenciales que han estado activando la corriente principal los medios de comunicación.

Según los autores citados, el **modelo de propaganda**¹⁸ se basaría en un marco coherente de **cinco filtros**¹⁹ que limitan el acceso a los medios de comunicación que seleccionan la información que llega a los medios convencionales:

1. **Primer filtro: magnitud, propiedad y orientación de los beneficios de los medios de comunicación.**
2. **Segundo filtro: beneplácito de la publicidad.**
3. **Tercer filtro: el suministro de noticias a los medios de comunicación.**
4. **Cuarto filtro: disensión (flak);**
5. **Quinto filtro: opiniones "anti-".**

Partiendo de la aceptación con reservas del modelo de Herman y Chomsky, el punto más destacado es el nuevo escenario de comunicación de Internet, caracterizado por la aparición de la blogosfera y las redes sociales que supondrían la superación de los cinco filtros y por lo tanto del antiguo sistema de comunicación que, según los autores citados, era una forma de propaganda:

Primer filtro: magnitud, propiedad y orientación de los beneficios de los medios de comunicación.

¹⁸Modelo de propaganda en Wikipedia <http://es.wikipedia.org/wiki/Modelo_de_propaganda>

¹⁹ Los cinco filtros son la propiedad, publicidad, el tamaño, y la orientación de beneficios de los medios de comunicación; abastecimiento, las opiniones anti- y el mecanismo de control ideológico <http://wiki.medialab-prado.es/images/0/0a/Textos_sobre_propaganda.pdf>.

Las corporaciones interesadas; buena parte de los medios de comunicación está en manos de grandes corporaciones, que pertenecen de hecho a las élites económicas. El primer filtro es la limitación de acceso a la propiedad de los medios. El coste de la maquinaria de cualquier rotativa de prensa supone una importante barrera de entrada por la alta inversión necesaria.

Sin embargo, el nuevo modelo de internet supone que cualquier persona puede lanzar su propio sitio web a cambio de lo que cuesta comprar un nombre de dominio y alquilar un hospedaje en un hosting web, un precio que en 2014 podía rondar los 40€ anuales. Además, no existen cargos extra si el sitio utiliza software de tecnología abierta (open source), como puede ser Wordpress o Drupal, sistemas que pueden crear un sitio web de alta calidad y sofisticación comparable al de cualquier versión online de un medio convencional. El **coste de entrada** es por tanto significativamente menor que el precio de entrada a los medios convencionales tradicionales como la prensa escrita o la televisión. Solo los sitios web más grandes tendrán gastos derivados del mayor tráfico en relación al mantenimiento de servidores, actualizaciones de software y creación de contenidos.

Segundo filtro: beneplácito de la publicidad.

Los medios dependen de la publicidad de las élites económicas para su existencia. Herman y Chomsky consideran la publicidad como el “segundo filtro”, argumentando que la elección de la publicidad influencia en la prosperidad de los medios y su supervivencia.

Los medios basados en anuncios reciben ingresos a través de la publicidad que le otorga una ventaja competitiva frente a rivales que no disponen de este tipo de financiación.

En el mundo online, aunque los propios anuncios son percibidos por los lectores como un fastidio y rara vez son percibidos como una ventaja cualitativa, los sitios web con mucha publicidad pueden atraer más ingresos que los sitios libres de anuncios. Ciertamente, algunos autores de blog (bloggers) tienen una fuerte dependencia de los **anuncios de la red de display de Google**, en los que la selección de anuncios para su inclusión en la página está basada en coincidencias de palabras clave.

A través de este mecanismo las empresas publicitadas buscan emplazar sus anuncios en sitios relevantes para sus productos. Pero gracias a la infinidad de sitios web, en la práctica siempre habrá un sitio web que contendrá contenido que encaje con alguna de las palabras clave de las empresas publicitadas, por lo que el sistema no impone demasiada presión en los sitios para que adapten sus contenidos a los deseos de las empresas publicitadas.

Tercer filtro: el suministro de noticias a los medios de comunicación.

Los medios deben producir un **flujo permanente de nuevas noticias**. Los principales proveedores de noticias están vinculados con el poder, son los departamentos de prensa de los gobiernos, de la oposición, de las grandes corporaciones y de los sindicatos, y no siempre la verdadera fuente de la noticia en sí, el evento a informar, salvo en el caso de los medios convencionales que practican el periodismo de investigación. Las fuentes gubernamentales y empresariales son consideradas como fuentes veraces debido a su estatus y prestigio²⁰, así como los principales partidos de la oposición y los sindicatos mayoritarios. Los periodistas tradicionales dependen en gran parte de estas fuentes oficiales, sin embargo, el nuevo modelo de internet ha facilitado la aparición del periodismo ciudadano, de tal forma que cualquier persona puede ser una fuente de información, alejándose de los límites del periodista convencional, aunque no faltan las voces críticas con el denominado "periodismo ciudadano" por tratarse de un oxímoron (aunque solamente sea por la falta de verificación típica del mismo).

Cuarto filtro: disensión (*flak*).

Herman y Chomsky definen *flak* como las "respuestas negativas a una declaración o un programa de los medios que puede tomar la forma de cartas, telégrafos, llamadas telefónicas, peticiones, demandas, discursos y proyectos de ley, así como otros medios de denuncia, la amenaza y la acción punitiva".

Los grupos de presión pueden organizar respuestas sistemáticas ante cualquier desviación de línea editorial. En el caso de los medios convencionales, la publicidad se centra principalmente en bienes de consumo que son fácilmente objetivo de

²⁰ Para mantener una imagen de objetividad y para protegerse de ataques a la parcialidad y de difamación los medios necesitan de material presuntamente «correcto».

boicot, por lo que los lobbies pueden corregir los sesgos mediáticos ante ataques a un determinado anunciante.

Estos sistemas de control no funcionan en la blogosfera, ya que en muchos casos los propios blogs son fuentes de disensión bajo el paraguas del anonimato de la red. Herman y Chomsky afirman que la popularidad de los sitios web depende precisamente de este grado de disensión, atrayendo a gran parte de su público por la agitación y el debate apasionado que generan.

Quinto filtro: opiniones "anti-".

Inicialmente, este filtro era el anticomunismo, por la rivalidad de la Guerra Fría. Las opiniones de algunas ideologías eran consideradas "anti-patrióticas". En EE.UU, con el fin de la guerra fría, el anticomunismo se desvaneció como ideología nacional y ha sido sustituida por un tipo de retórica de "lucha contra el terrorismo" o "anti-Islam" – o en sentido de anti-anti-americanismo.

Hoy en día se considera que el anti-comunismo era una ideología de filtrado que debería considerarse simplemente como un ejemplo de un filtro más amplio, el del "nacionalismo como un mecanismo de control ideológico".

En el caso de la blogosfera, el filtro del refuerzo del nacionalismo a través de la ideología carece de sentido por la propia característica internacional de la red en las que impera la libertad de expresión sin la sujeción de ningún tipo de control ideológico posible.

1.1.1. El modelo de la era Internet

En la actualidad la diversificación de los medios del modelo de la era Internet permite un acceso ubicuo a la producción y distribución de medios disponibles a través de blogs, videoblogs, transmisiones digitales, YouTube y redes sociales, considerados como medios de segundo nivel (*second-tier media forms*).

Las principales conclusiones que se desprenden de la superación de los cinco filtros de propaganda ponen de manifiesto que conceptualmente los blogs y las redes sociales pueden desafiar a los principales medios de comunicación en cuatro dimensiones: propiedad, financiación, fuentes y contenido.

- En términos de **propiedad**, los blogs son lo opuesto a las grandes estructuras de los medios de comunicación, ya que en la mayor parte de los casos pertenecen a particulares o a colectivos asociativos muy alejados de las necesidades de mercado de los grandes medios.
- Respecto a la **financiación**, si los medios de comunicación dependen de la publicidad para poder competir en el mercado, los blogs no dependen de las estrategias comerciales, salvo casos particulares. Además, en el desempeño informativo diario los principales medios de comunicación a menudo adquieren compromisos y participan en diversas interdependencias con las élites políticas y económicas. En el caso de la blogosfera, los bloggers son independientes y su financiación no está sujeta a los intereses de los lobbies.
- Como hemos visto, respecto a las **fuentes** los medios convencionales están muy determinados por las fuentes vinculadas con el poder (entendiendo que en una democracia, los partidos de la oposición y los sindicatos también son poder) y a intereses derivados, mientras que para un blogger cualquier cosa puede constituir una fuente, desde su rutina diaria hasta los grandes conflictos internacionales. Muchos bloggers se dedican a tareas descritas como de «gatewatching», es decir, observando las puertas de salida de los medios de noticias considerados de primer nivel o fuentes primarias, y localizando materiales relevantes a los intereses y preocupaciones de sus propias audiencias, introducen estos materiales en su propia cobertura de noticias.
- Con este amplio abanico de actuación, los blogs no se ven afectados en cuanto a sus **contenidos** por las élites políticas y económicas, sino que más bien

desafían a las ya existentes. En su trabajo, los bloggers reintroducen puntos de vista alternativos a los de los medios convencionales a través de la combinación de diferentes medios y la contrastación en la cobertura de noticias, gracias a lo cual destacan las diferencias en el énfasis o la interpretación de las noticias causada por la parcialidad política o por el trabajo de periodistas.

Sin embargo, este apartado no estaría completo sin mencionar los aspectos críticos de este nuevo modelo de sistema comunicativo y los peligros que comporta. Como veremos en los siguientes apartados, diversos analistas han señalado, entre otros, dos grandes problemas: 1) que el llamado periodismo ciudadano carece de verificación y puede ser alimentado por «agentes» partidistas con finalidades más cercanas a la desinformación que a la finalidad de informar; 2) nos enfrentamos a un gran problema derivado de la preeminencia que han obtenido los buscadores en el acceso a la información, de modo que se han constituido en los nuevos gatekeepers digitales, con el agravante de que en algunas sociedades, como la europea, prácticamente tenemos un solo buscador que ejerce de gatekeeper: Google. Por su importancia, a este apartado hemos dedicado una sección más adelante donde aparece tratado en detalle este problema derivado del nuevo modelo de Internet.

1.1.2. La blogosfera

Originalmente conocidos como weblogs, los blogs abarcan desde diarios y reflexiones de autopromoción a la cobertura y análisis en tiempo real de noticias altamente personalizadas.

En la actualidad los blogs forman parte integral de la cultura online y un porcentaje muy elevado de usuarios de Internet lee blogs²¹, ya sean blogs de noticias de medios tradicionales, blogs centrados en algún tema particular o blogs de entretenimiento, sin embargo el fenómeno blog que conocemos a día de hoy tiene una historia muy reciente cuyo «Bing Bang» encontramos en una fecha tan cercana en el tiempo como septiembre de 1999, fecha en la que PyraLabs lanzó Blogger, una aplicación que aunó el imparable crecimiento del fenómeno blog gracias a una tecnología que no requería de conocimientos de HTML.

²¹ Marketing Directo. El 52,2% de los internautas españoles leen blogs
<<http://www.marketingdirecto.com/actualidad/digital/el-522-de-los-internautas-espanoles-leen-blogs/>>

El crecimiento de los bloggers fue imparable y ya en 2002 se contabilizaron 970.000 blogs, pasando al millón cuatrocientos mil blogs activos un año más tarde. Los blogs poco tardaron en popularizarse y se calcula que a mediados de 2004 ya existían 3 millones de blogs, la misma cifra que en 2011 Technorati²² reportó como el número de nuevos blogs que se publicaban online cada mes, alcanzando una cifra global de 178 millones de blogs en diciembre de aquel mismo año.

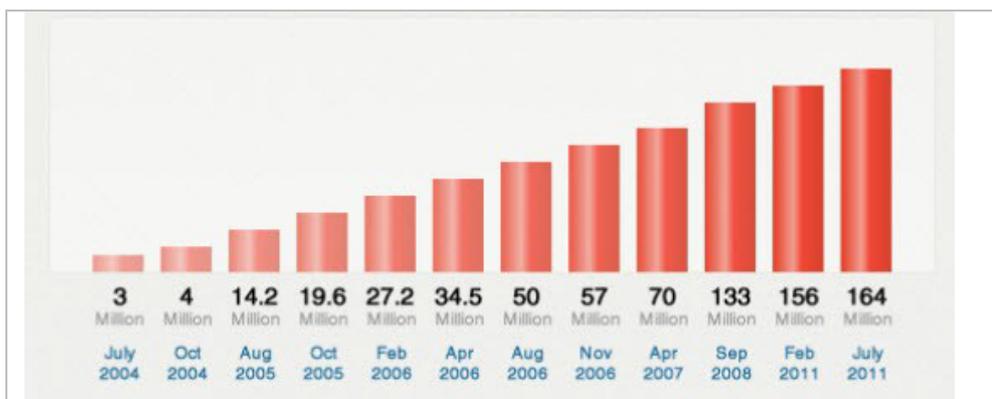


Figura 1. Evolución de la Blogosfera. Fuente: 7 Reasons to Rethink Your Blogging Strategy: New Research <<http://www.socialmediaexaminer.com/7-reasons-to-rethink-your-blogging-strategy-new-research/>>

²² State of the Blogosphere 2011 <<http://technorati.com/state-of-the-blogosphere-2011/>>

1.2. El modelo de Infoesfera

El 12 de marzo de 2014 se celebró el 25 aniversario de la creación de la World Wide Web por Sir Tim Berners-Lee en el CERN de Ginebra. La web había nacido en 1989 y en su primera década de vida pasó de ser un sistema de intercambio de información científica a ser un medio de comunicación y comercio de ámbito mundial.

La interconexión de redes y creación del lenguaje HTML a inicios de la década de los noventa del pasado siglo, permitieron el crecimiento masivo de Internet. Prueba de ello es que desde mediados de 1991 hasta la actualidad, la Web no ha cesado de desarrollarse y evolucionar con nuevas capacidades que completan y mejoran este sistema digital, hipertextual y en red.

El ejemplo más ilustrativo del crecimiento exponencial de la WWW es el del número de servidores web. En enero de 1993 había en marcha alrededor de 50 servidores web, creciendo hasta los 30 millones a finales de 2001 y alcanzando los 855 millones a fecha de mayo de 2014²³. Recientemente, tras tal incremento de servidores web, el antiguo protocolo de Internet, el sistema de direcciones IPv4, ha quedado agotado, mudando hacia el sistema actual IPv6.

La totalidad de la web es accesible vía Internet con un navegador web, sin embargo este sistema de distribución de documentos de hipertexto o hipermedios interconectados entre sí no puede navegarse a no ser que exista un mapa. A principios de Internet los directorios como Yahoo o Dmoz respondieron a esta necesidad, sin embargo, ese temprano crecimiento exponencial hizo imposible que los usuarios pudiesen absorber toda la información disponible, y supuso la aparición de distintos sistemas para localizar recursos relevantes sin perder demasiado tiempo, entre ellos los buscadores. Ante el imparable crecimiento de la web y la imposibilidad de catalogar todos los recursos que iban siendo publicados, los buscadores se hicieron indispensables.

El papel fundamental de los buscadores deriva de esta necesidad de navegar por la web. Para lograr ofrecer sus servicios, los buscadores copian la mayor parte de la web accesible a través de sus robots, y la muestran en forma de páginas de resultados. A

²³June 2014 Web Server Survey <<http://news.netcraft.com/archives/category/web-server-survey/>>

fecha de 21 de mayo de 2014, se estima que la web indexada contiene un mínimo de 2,68 billones de páginas web²⁴, de ahí que los buscadores se hayan convertido en herramientas indispensables para sacar el máximo partido a toda la información acumulada en la web durante el último cuarto de siglo.

1.2.1. Origen y marco teórico del modelo de Infoesfera

A lo largo de los últimos veinticinco años se han ideado diferentes modelos para visualizar la World Wide Web (WWW), la red informática mundial comúnmente conocida como la web. Uno de los primeros acercamientos a una categorización de la esfera web lo encontramos en el Modelo Bow-Tie presentado el año 2000 en el artículo *Graph structure in the web*²⁵.

Este primer modelo tuvo gran repercusión en el momento de su publicación, dando pie a consecuentes modelos basados en él, como el modelo corona publicado por

²⁴The size of the World Wide Web (The Internet) <<http://www.worldwidewebsite.com>>

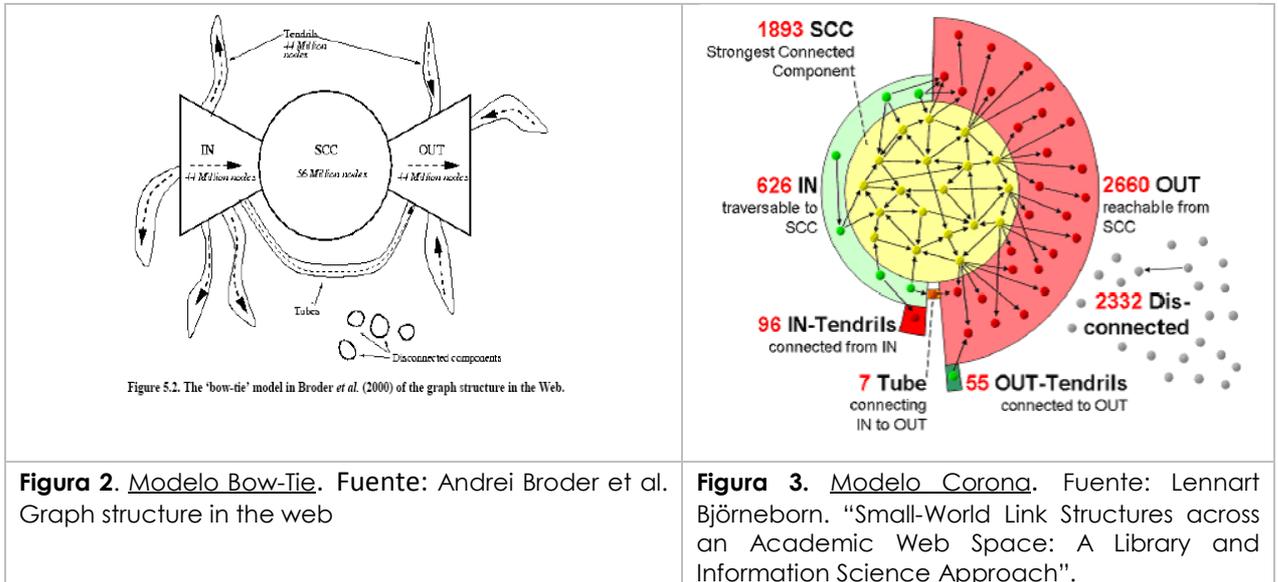
²⁵Andrei Broder et al. Graph structure in the web <<http://www9.org/w9cdrom/160/160.html>>

Este artículo fue escrito en colaboración por Andrei Broder y Farzin Maghoul del buscador AltaVista, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan y Andrew Tomkins de IBM Almaden Research Center, y por Raymie Stata y Janet Wiener de Compaq Systems Research Center. Andrei Broder et al. trabajaron en un proyecto que aunaba los esfuerzos del buscador AltaVista con lo mejor en empresas de computación de la época con el objetivo de medir el tamaño y forma de la web. En base a 200 millones de páginas se tuvo como resultado el modelo bow-tie, de la pajarita, que se componía de **cuatro grupos principales** de páginas web:

1. Un **núcleo fuertemente conectado (SCC)**, del inglés 'strongly connected component') de 56 millones de páginas.
Junto al SCC se detectaron dos grandes grupos, de cerca de 43 millones de páginas cada uno:
2. El primero consistía en todas las páginas que enlazaban con el núcleo pero que no tenían vínculos desde el núcleo de vuelta a ellos. Este es el grupo "**origen**" o "**In**".
3. La contrapartida era el conjunto de todas las páginas que recibían vínculos del núcleo, pero que no tenían enlaces de vuelta. Esta es la "**terminación**" o grupo de "**Out**", ya que contiene enlaces que conducen fuera del núcleo.
4. El cuarto y último grupo es el resto de **páginas inconexas**, que no enlazan con el núcleo ni están vinculados entre ellas.

Existen grupos más pequeños de las páginas web adicionales. Tanto el "In" como del "Out" surgen grupos en forma de "**zarcillos**" (del nombre original en inglés '**tendrill**') que surgen desde estos dos grupos de páginas y conducen hacia ellos. Estos consisten en páginas que enlazan con y desde los grupos "In" y "Out", pero que no son parte de cualquiera de los dos. En la versión más detallada del modelo, este esquema se puede reproducir hasta saciedad, y el modelo agrega zarcillos de los zarcillos, y así sucesivamente. Además, esta versión contiene otro importante grupo conocido como los "**tubos**". Este grupo está formado por las páginas accesibles desde "In" y que enlazan directamente con "Out", pero que no forman parte del núcleo de gran tamaño. Visualmente, forman rutas alternativas de "In" a "Out", como los tubos de flexión alrededor del elemento central fuertemente conectados.

Lennart Björneborn en el año 2004 en su "Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach"²⁶.



De forma paralela a este último artículo, en el año 2003 Jonah Peretti²⁷, emprendedor de Internet graduado en el MIT de Massachusetts, y fundador, entre otros, de sitios web como BuzzFeed²⁸ o el Huffington Post²⁹, definió en su artículo "Culture Jamming, Memes, Social Networks, and the Emerging Media Ecology The "Nike Sweatshop Email" as Object-To-Think-With"³⁰ un **modelo alternativo** que permitía clasificar el nuevo panorama mediático surgido a raíz de la aparición de internet, la web y la blogosfera.

²⁶ Lennart Björneborn. "Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach." <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.2353&rep=rep1&type=pdf>

El segundo modelo, derivado del anterior, es el del antes mencionado Lennart Björneborn, el llamado modelo '**Corona**'. Dentro de una metodología de cinco pasos, el Dr. Björneborn presentó en su tesis un nuevo modelo gráfico llamado "corona" que ilustra correspondencias de enlace entre componentes en el gráfico web de universidades del Reino Unido. El modelo identifica y caracteriza las propiedades de los vínculos dentro del espacio web académico del Reino Unido. Este análisis permitió establecer un modelo gráfico resultado de un "mapeo" de las estructuras de enlace entre los 7.669 sitios de universidades del Reino Unido y facilitó la selección de una muestra adecuada para su posterior análisis en una segunda fase de la metodología de investigación. Alrededor del 25% de los sitios investigados pertenecían al componente fuertemente conectado de la gráfica. En este modelo la característica de los SCC es que todos los pares de sitios están conectados entre sí a través de enlaces.

²⁷ Jonah Peretti en Wikipedia <http://en.wikipedia.org/wiki/Jonah_Peretti>

²⁸ BuzzFeed <<http://www.buzzfeed.com/>>

²⁹ The Huffington Post <<http://www.huffingtonpost.com/>>

³⁰ Jonah Peretti (2003). Culture Jamming, Memes, Social Networks, and the Emerging Media Ecology The "Nike Sweatshop Email" as Object-To-Think-With <<http://depts.washington.edu/ccce/polcommcampaigns/peretti.html>>

En este artículo, publicado en el sitio web de la universidad de Washington, Peretti narra un caso basado en lo que hoy se denominaría «crisis de reputación corporativa» contra Nike.

En 2001 el sitio web de Nike permitía a sus visitantes crear zapatillas personalizadas que llevaran una palabra o lema, un servicio que Nike anunció como un proyecto acerca de la libertad de elegir y la libertad de expresar quién eres. Ante este anuncio de Nike, el propio Peretti pidió un par de zapatillas con la palabra "sweatshop." (en inglés, "fábrica de explotación"), en referencia a las personas que en realidad elaboraban las zapatillas Nike en las fábricas de hacinamiento. Nike se negó a fabricar el pedido con aquel lema, lo que dio fruto a un intercambio de emails entre J.Peretti y Nike³¹ que fue reenviado a través de Internet, creando un enorme *buzz* de tal forma que la noticia de la disputa apareció en la televisión y en los principales periódicos y revistas de EE.UU.

En su ensayo Peretti afirma que sin las tecnologías actuales de comunicación, el caso de Nike y la "fábrica de explotación" nunca hubiera existido y que lo permitió el equivalente digital del boca a boca: el correo electrónico, los sitios web personales y los blogs.

Internet, a través de la web y las redes sociales disponibles en aquel 2001, proporcionó una red de distribución que consiguió que los memes (Peretti utiliza el término «meme» entendido como ideas auto-replicantes que se propagan por sí mismas) se extiendan de forma más rápida y poderosa. En palabras del propio Peretti: "*what began as a local act of protest grew into a quest to understand a global transformation*", a raíz de la cuál definió la ecología mediática por la que viajó su email sobre Nike.

Para el autor está claro que la eficiencia de las tecnologías de distribución es tan importante como el meme mismo, y analizando la propagación del correo electrónico de «Nike fábrica de explotación» visualizó esta ecología mediática contemporánea, compuesta por tres clases de tecnologías de distribución y estratos bien diferenciados y claramente identificables: la *micromedia*, *middle media*, y *mass media*.

Micromedia incluye todas las tecnologías de las comunicaciones personales, tanto el correo electrónico, teléfonos, sitios web personales, e incluso la comunicación verbal entre individuos. Aunque en 2001 la micromedia se utilizaba en la comunicación personal entre un pequeño número de personas,

³¹Lista de emails de Peretti vs Nike <<http://www.shey.net/niked.html>>

Peretti muestra el correo electrónico de «Nike fábrica de explotación» como un ejemplo de meme que se propaga a una audiencia mucho mayor a través de los canales de comunicación personales gracias a la recomendación personal de un conocido.

Peretti hace especial mención al software gratuito de blogger.com, a lo que él, en el año 2001, denominó "un emocionante movimiento de auto-publicación que se ha iniciado en la web". El blog, abreviatura de web log³², convierte a sus autores en curadores web personales mediante la compilación de enlaces comentados a sitios web que encuentran interesantes. Peretti indica que a estos sitios web personales tienden a agregarse pequeñas audiencias de amigos y personas de ideas afines, plataforma que fue utilizada para la propagación de su email.

Middle Media. Dentro de «Middle Media» Peretti incluye las tecnologías que ayudan a las comunidades a filtrar y agregar los contenidos creados por la micromedia. El autor incluye en este grupo millones de páginas web potencialmente interesantes que una vez identificadas y posteadas en redes de blogs ofrecen la posibilidad a una gran comunidad de personas de participar en la discusión acerca de los enlaces publicados. A ojos de Peretti, «Middle Media» es un foro público accesible a nivel mundial. Esto significa que este tipo de sitios intermedios tienen el potencial de transformar "*an obscure piece of net lore*" en una noticia con cobertura a nivel nacional. Esto es exactamente lo que sucedió en el caso del meme "Nike fábrica de explotación". Los medios de comunicación del «Middle Media» ayudaron a que el email se convirtiera en un tema de debate público mediante la transformación del buzz de la micromedia en un problema social de interés periodístico.

Medios de comunicación. Analizando lo ocurrido, Peretti llegó a la conclusión de que a pesar de los billones de dólares que cada año puede gastar una marca como Nike en mass media, los ciudadanos juegan a su favor con la posibilidad de que una campaña en Internet con éxito pueda dar lugar a una cobertura de medios de comunicación en forma de noticias de forma totalmente gratuita.

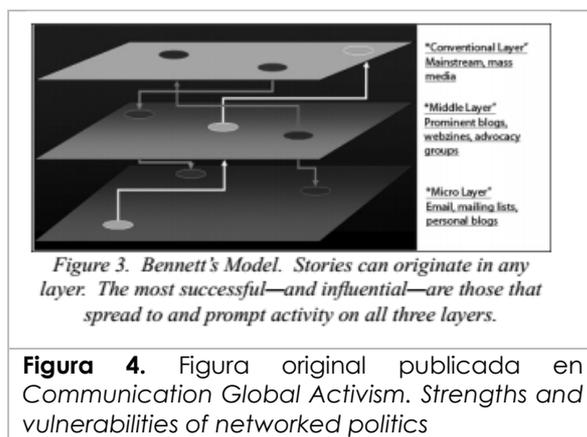
³² Nota: esta aclaración aparece en el artículo original escrito al poco tiempo del caso 'Nike, fábrica de explotación'.

Según el autor, esta es la razón por la que las campañas en Internet son tan atractivas, ya que un mero correo electrónico puede transformarse en noticias en directo en la televisión nacional e internacional como fue su caso.

Peretti halla la razón de la transmisión de un meme desde micromedia, pasando por sitios de «Middle Media», hasta de repente convertirse en una historia de los medios de comunicación de masas. La razón la encuentra en la psicología de los periodistas. Los reporteros son a menudo las personas que han dejado de lado profesiones más lucrativas porque quieren hacer un impacto social positivo informando al público. Sin embargo, muchos periodistas se encuentran cubriendo ruedas de prensa con guión o convirtiendo notas de prensa corporativas en noticias. Internet ofrece a estos periodistas descontentos la oportunidad de descubrir historias auténticas. En este caso los periodistas descubrieron el meme “Nike fábrica de explotación” a través de un correo electrónico o un blog, pero solo unos días más tarde, el meme “Nike fábrica de explotación” quedó cubierto/acallado por otra historia vendida por los medios de comunicación.

El modelo de J.Peretti fue sintetizado en el año 2003 por **W. L. Bennett**, quien recogió el artículo de Peretti en su comunicación *Communicating Global Activism: Strengths and vulnerabilities of networked politics*³³ donde exponía las ideas de Peretti en un modelo de análisis de la web esfera.

Un año más tarde W. L. Bennett propuso el marco teórico de este modelo en “*Political Communication, Citizenship and Democracy: From the Mass Society to Personal Information Networks*”, donde se muestra la composición en tres estratos de la Infoesfera, en los que cada uno tiene un tamaño e importancia bien diferenciados:



³³ Bennett, W.L. (2003). *Communicating Global Activism: Strengths and Vulnerabilities of Networked Politics*. <http://ccce.com.washington.edu/projects/assets/working_papers/communicatingglobalactivism.pdf>

1. El **estrato convencional** compuesto por medios de comunicación de masas: la radio, la televisión y la prensa escrita.
2. El **estrato medio** de blogs destacados, revistas electrónicas y grupos de presión.
3. El **estrato micro** formado por el correo electrónico, listas de distribución y blogs personales.

Según W. L. Bennett, a diferencia del estrato convencional, el estrato medio y el estrato micro no son directamente influyentes. El estrato convencional es el origen del flujo de información habitual. Esta corriente informativa va del estrato superior al medio y al micro-estrato, formada por las noticias surgidas de los medios de masas. Es decir, la relación de influencia iría desde arriba hacia abajo.

Sin embargo el modelo ofrece la posibilidad de excepciones, casos en los que una estrategia de comunicación viral consiga que una noticia de un estrato inferior alcance los medios de masas y active todos los estratos de la Infoesfera. Esta posibilidad teórica, evidenciada a raíz de uno de los primeros ataques reputacionales de la historia de internet ha sido una de las motivaciones de este trabajo de investigación, uno de cuyos objetivos era explorar la posibilidad real de este mecanismo de transmisión inverso de abajo hacia arriba, es decir, que actuaría desde los estratos micro y medio hacia el convencional.

1.2.2. Nuevo modelo de Infoesfera 2014

Para su ejemplificación, Bennet utilizó la figura que acabamos de ver en la que se representan los tres estratos de la Infoesfera y los diferentes flujos de información. Esta figura está entre los orígenes de esta tesis en la que presentamos un modelo de análisis centrado en las proporciones de los tres estratos de la Infoesfera.

Sin embargo, el modelo teorizado hace más de diez años ha quedado obsoleto dado que en el año 2003 la revolución que supuso la Web 2.0 no había sido ni siquiera planteada. Así pues, gran parte de los actores principales que componen la realidad web de 2014 no existían en el momento de su planteamiento inicial, requiriendo por tanto de una labor de actualización. Por este motivo, el modelo inicial, aunque nos parece útil como instrumento de partida entre otros, plantea varios problemas que hemos tenido que superar.

Como consecuencia de lo expuesto hasta ahora, la evolución de estos últimos doce años de la Infoesfera ha supuesto un cambio bastante radical en la composición de la Web que inspiró el modelo original.

El Estrato Superior, el correspondiente a los medios de prensa convencionales, ha sido el que ha crecido menos relativamente. En la actualidad, cualquier periódico, diario, revista que se publique en el mundo off-line tiene su representante en el mundo online a través de un sitio web. Como veremos, en el caso de España, la totalidad de sitios web de este primer estrato han sido localizados y categorizados dentro de su estrato correspondiente, si bien en los otros dos estratos el trabajo se ha centrado en los principales sitios de cada uno de ellos.

La evolución de los estratos Medio y Micro ha supuesto un gran cambio en su composición a raíz de la aparición de millones de nuevas páginas web. Focalizándonos en los últimos años de su evolución, en el año 2010, momento de inicio de esta tesis, la realidad de la web ya era social y había evolucionado mucho desde el planteamiento del modelo en 2001.

Durante el periodo 2010-2014, el desarrollo de la web móvil supuso un acelerón en la implantación de la web social, y en la actualidad los usuarios acceden en un porcentaje creciente a través de aplicaciones de sus Smartphone. Dentro de esta evolución hacia lo móvil, la web social ha seguido su crecimiento, y sus empresas más representativas han salido a la bolsa con diferentes niveles de éxito³⁴. LinkedIn salió al mercado en mayo de 2011³⁵. Un año más tarde, en mayo de 2012, lo hizo Facebook³⁶ con un precio de 45\$ por acción, que en la actualidad, dos años más tarde, se sitúan en torno a los 60\$. Twitter³⁷ ha sido la última en aparecer en el parqué, cosa que hizo en noviembre de 2013 con un precio de salida de 26\$ la acción. Apenas medio año más tarde las acciones de Twitter cotizan³⁸ a 44\$.

³⁴ La salida a bolsa de tres grandes empresas tecnológicas: Google, LinkedIn y Facebook <<http://www.genbeta.com/genbeta/la-salida-a-bolsa-de-tres-grandes-empresas-tecnologicas-google-linkedin-y-facebook>>

³⁵ LinkedIn sube más de un 130% en el mejor estreno en Wall Street desde 2000 <<http://www.elmundo.es/elmundo/2011/05/19/navegante/1305789520.html>>

³⁶ Las claves del batacazo de Facebook en Bolsa <http://cincodias.com/cincodias/2012/05/22/mercados/1337902006_850215.html>.

³⁷ Los motivos por los que Twitter ha sido un éxito en su salida a bolsa <<http://www.abc.es/economia/20131108/abci-motivos-twitter-sido-exito-201311081131.html>>

³⁸ Cotizaciones TWT <<https://es.finance.yahoo.com/echarts?s=TWTR#symbol=TWTR;range=6m>>

Solo estas noticias son un apunte de cómo ha cambiado Internet desde que se idease el modelo de Infoesfera. Ninguna de estas tres empresas existía entonces, sin embargo sí aparecen en los resultados de esta tesis.

Estas tres empresas son solo los ejemplos más destacados de cómo ha cambiado la web desde aquel lejano 2001 cuando el caso de «Nike fábrica de explotación» originó el nuevo modelo de análisis que pasamos a analizar en detalle. El desarrollo de la web móvil supone un cambio más, que recogeremos durante la exposición de los procesos de optimización SEO. Además de todos los sitios web derivados de la tecnología asociada a la web 2.0., al modelo deberíamos añadir nuevas plataformas de comunicación derivadas de la web móvil, como el caso de la mensajería instantánea si deseamos tener una visión completa y actualizada al contexto del año 2014. Sin embargo, como en todo trabajo de investigación, hemos tenido que centrarnos en el caso de Google como el principal gatekeeper virtual de Internet.

1.3. Google como modelo del Gatekeeper en la era Internet

Como hemos visto al inicio de este apartado, a pesar de la aparente ventaja competitiva de la blogosfera en las cuatro dimensiones señaladas de **propiedad, financiación, fuentes** y **contenido**, en la era Internet también existe la figura del gatekeeper encarnada en los buscadores, que a través de la dependencia de los usuarios de sus resultados actúan como una aparente evolución digital de los gatekeepers de los medios tradicionales del siglo pasado.

La teoría sobre los **gatekeeping** en las empresas periodísticas fue el preámbulo a la teoría de la agenda-setting (Becker, McCombs y McLeod, 1975, y Whitney y Becker, 1982). Esta teoría se refiere a cómo los contenidos de los sucesos del día y de las diferentes agendas políticas de los partidos son seleccionados por los “guardianes de las puertas” (gatekeepers): Editores jefes, responsables de los medios masivos (radio, prensa y televisión), periodistas y otros, basados en ciertos criterios previamente establecidos como: la línea editorial, las influencias de la organización y las influencias ideológicas entre otros³⁹.

En la actualidad los dos mecanismos principales de búsqueda de información en Internet son la “vertical” a través de portales y buscadores, y “horizontal” a través de redes de enlaces o hiperenlaces, o lo que es lo mismo: la recuperación de información mediante palabras clave y la navegación mediante links.

En el caso de los mecanismos de selección vertical, los buscadores guían al usuario presentando una selección ordenada por relevancia de los sitios web que tratan sobre el tema que el usuario ha definido mediante los criterios de búsqueda introducidos en el formulario de consulta. De esta manera se puede decir que los buscadores actúan como **gatekeepers** al ofrecer una selección de la realidad de acuerdo a unos criterios propios predefinidos mediante un algoritmo de cálculo de relevancia. En teoría un buscador ofrece todos los resultados de sitios web que responden al tema de la consulta (es fácil ver que Google, concretamente, informa de los miles, decenas de miles o millones de página que ha obtenido para una consulta). En la práctica el filtro (o la acción de gatekeeper) se produce porque los buscadores presentan los

³⁹ Luz Valdés Miranda “Análisis de Internet 2.0 y la influencia de la agenda-setting” <<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/19881/3/lgvaldesmTFM0113memoria.pdf>>

resultados ordenados por el mencionado cálculo de relevancia y los usuarios, por su parte, nunca pasan de la primera o segunda página de resultados.

Por lo tanto, es indiferente en realidad si una búsqueda arroja 10 resultados o 10 millones de resultados, la inmensa mayoría de los usuarios solamente revisarán los cinco o diez primeros. Es esta ordenación, en base a unos criterios prefijados, la que actúa como un gatekeeper, aunque sea un gatekeeper algorítmico y virtual. En esta tesis intentaremos mostrar cómo esta agenda se extiende hacia su vertiente digital debido al papel preponderante del buscador Google. Aunque éste no actúe como un gatekeeper de manera formal ni, en teoría, sus procedimientos sean manuales, sino algorítmicos, lo cierto es que, como resulta fácil de argumentar debido a su situación hegemónica, Google cumple de facto un papel similar.

De esta manera, podemos ver que la irrupción de la blogosfera a finales del s.XX, y más tarde el desarrollo de la web 2.0, aceleraron el cambio de paradigma de comunicación del *one to many* al *many to many*.

Era	Era Mass Media	Era internet
Periodo	S. XIX-XX	S. XXI
Dominio	Mass Media convencional	Multimedia y Blogosfera
Tipo de comunicación	one to many	Many to many
Modelo	Agenda Setting	Google Gatekeeper

Figura 5. Tabla comparativa del modelo de comunicación Mass Media vs. Internet. Elaboración propia

En la nueva era de Internet, la lucha por la visibilidad en los 10 resultados de búsqueda se convierte en indispensable para llegar a los públicos a través del tráfico web. En este nuevo escenario los criterios prefijados por los buscadores que actúan como un gatekeeper, y si bien Google no acalla las fuentes, es la no visibilidad, el no cumplimiento de estos criterios, lo que ejerce de filtro.

Como veremos a través del análisis de páginas de resultados, los mass media tradicionales se han adaptado a este nuevo escenario superando las limitaciones técnicas asociadas a los criterios de ordenación, ya que en muchos casos cuentan en sus equipos con profesionales SEO (de las siglas inglesas de *Search Engine Optimization*) que han permitido a los grandes medios alcanzar altas posiciones en el

ranking de buscadores. De igual forma, los cybermedios que más se ajusten a los requerimientos técnicos de buscadores como Google, alcanzarán visibilidad y tráfico, compitiendo así con los grandes medios convencionales.

1.3.1. Los riesgos de la Generación Google

En enero de 2008 el University College London había publicado el estudio *Information Behaviour of the Researcher of the Future* sobre el comportamiento de los jóvenes frente a los nuevos medios de comunicación, en especial Internet y los buscadores. Según el estudio, la nueva generación de internautas que en los próximos años serán los usuarios mayoritarios de Internet, la llamada «generación Google», posee unos rasgos comunes que refuerzan la teoría de que informaciones no contrastadas provenientes de la blogosfera y de redes sociales suponen riesgos.

Según el University College London, la generación Google además de tener una gran preferencia por la información visual en detrimento de la textual, tiene una gran dependencia y confianza en los buscadores, en especial Google, hecho que les lleva a aterrizar allá donde el buscador les lleva.

Además de estas características generales, el estudio apunta un rasgo más preocupante en lo referente al **consumo de información**, y es que esta generación, al parecer, carece de sentido crítico y de habilidades analíticas para evaluar la veracidad de la información encontrada en los buscadores.

La información que se desprende de este estudio se vio reforzada por artículos como el de Fabiola Cabra y Gloria Patricia Marciales titulado *Nativos digitales: ¿ocultamiento de factores generadores de fracaso escolar?*⁴⁰ Este artículo cuestiona los mitos contruidos en torno a los nativos digitales⁴¹ y a sus competencias informacionales, y afirma que son pocas las evidencias en la literatura sobre cambios cognitivos entre

⁴⁰ Fabiola Cabra y Gloria Patricia Marciales (2009) Nativos digitales: ¿ocultamiento de factores generadores de fracaso escolar? <<http://www.rieoei.org/rie50a06.pdf>>

⁴¹ Educación y nativos digitales: @blogoff desmontando mitos <<http://www.microsiervos.com/archivo/mundoreal/educacion-y-nativos-digitales-desmontando-mitos.html>>

generaciones, y que si bien algunos jóvenes son adeptos al uso de tecnologías en diferentes actividades de la vida cotidiana, una significativa proporción de ellos no tiene habilidades necesarias para el uso de las mismas.

Las autoras citan el trabajo de Combes (2007)⁴² como un estudio sobre la generación *net* que desmitifica al usuario intuitivo capaz de utilizar los recursos electrónicos para encontrar información. Otros estudios como el de Nicholas y otros (2003 y 2004)⁴³ demostraron que si bien los estudiantes manifestaron niveles altos de habilidad para el uso de la red (solamente entre el 10% y 12%), los porcentajes disminuyeron cuando se indagó sobre su habilidad para realizar tareas simples como recoger información (30,5%), organizarla (28,5%) y evaluarla (25%).

Las principales conclusiones de los investigadores en este campo indican que si bien los estudiantes sienten confianza en su habilidad para el uso de la tecnología y para encontrar información, sienten menos confianza para manipular y usar la información que encuentran, lo que justifica el denominado «pánico moral», fenómeno que se ha ido generando en la sociedad actual y que tiene lugar cuando un grupo social particular –en este caso el de los jóvenes de la generación *net* o generación Google– es retratado por los medios de comunicación como una amenaza potencial a los valores y normas sociales establecidos.

Fruto de este «pánico moral» y del posible uso fraudulento/malicioso de la Blogosfera, en junio de 2008 la eurodiputada socialista Marianne Mikko⁴⁴ presentó en el Parlamento Europeo el *Documento de trabajo sobre la concentración y el pluralismo de los medios de comunicación en la Unión Europea*⁴⁵.

El objetivo de este comunicado era fomentar la protección ante ataques de desinformación o amenazas de difusión de falsas noticias con la intención de alterar la realidad informativa en forma de propaganda difundida por redes sociales y blogs, a través de fuentes sin identificar dirigidas por lo que se denominó una «**agenda oculta**».

⁴² COMBES, Barbara (2007): «The Search for Information and the Net Generation», en *International Association of School Librarianship. Selected Papers from the 36th Annual Conference*.

⁴³ NICHOLAS, David y OTROS (2003): «Digital Information Consumers, Players and Purchasers: Information Seeking Behaviour in the New Digital Interactive Environment», en *Aslib Proceedings*, vol. 55, n.º 1/2.

⁴⁴ Parlamento Europeo. Página personal de Marianne Mikko

<http://www.europarl.europa.eu/meps/es/28421/MARIANNE_MIKKO_home.html>

⁴⁵ Documento: MOTION FOR A EUROPEAN PARLIAMENT RESOLUTION
<<http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A6-2008-0303&language=EN#title1>>

En la entrada 'AI' del documento se exponía lo siguiente:

Whereas weblogs represent an important new contribution to media pluralism and are an increasingly common medium for self-expression by media professionals as well as by private persons their proliferation implies a need to establish legal safeguards providing for the assignment of liability in the event of lawsuits, and establishing the right to reply”.

Como consecuencia de lo anterior, un poco más adelante, en el denominado ‘EXPLANATORY STATEMENT’, se exponía un sistema de salvaguarda:

“In this context the report points out that the undetermined and unindicated status of authors and publishers of weblogs causes uncertainties regarding impartiality, reliability, source protection, applicability of ethical codes and the assignment of liability in the event of lawsuits”.

“It recommends clarification of the legal status of different categories of weblog authors and publishers as well as disclosure of interests and voluntary labelling of weblogs.

En dicho documento por un lado se reconocía que los weblogs son un medio de comunicación en expansión utilizado tanto por profesionales de la comunicación como por personas privadas, pero por otro, alertaba que éstos podían ser una amenaza porque “el estatuto de los autores y de quienes los publican, incluido su estatuto jurídico, no se define ni se deja claro ante los lectores de los weblogs, lo que provoca incertidumbres respecto a la imparcialidad, fiabilidad, protección de las fuentes, aplicación de códigos éticos y atribución de responsabilidades en caso de acciones judiciales”.

La publicación de este documento por la Comisión de Cultura desencadenó una gran controversia entre autores de blogs de toda Europa, ya que para solucionar el problema de imparcialidad y fiabilidad que suponían los weblogs, el documento proponía regular la Blogosfera, y más concretamente la identidad de los bloggers.

Marianne Mikko justificó su iniciativa alegando que algunos bloggers estarían en una posición capaz de contaminar considerablemente el ciberespacio, bloggers que podrían estar siguiendo las directrices de lo que ella llamó «agendas ocultas». Las declaraciones de Marianne Mikko desataron reacciones adversas en la blogosfera ya que los bloggers veían en esta iniciativa una forma de controlar la libertad de los blogs. A pesar del alud de posts en contra de las pretensiones de la Comisión de Cultura y Educación de controlar la blogosfera, poco después el eurodiputado liberal alemán

Jorgo Chatzimarkakis⁴⁶ reafirmó las intenciones de la cámara europea y reforzó la argumentación de su colega destacando que los blogs, utilizados por grupos de presión, podrían llegar a ser una amenaza al pluralismo, debiendo realizarse un seguimiento de la identidad de los que escriben tras la Blogosfera.

Para prevenir la potencial amenaza al pluralismo que según tales teorías podrían suponer los weblogs, la Comisión de Cultura y Educación propuso la creación de un **dni para bloggers** para evitar el anonimato de la red⁴⁷ y conseguir que los blogs contasen con información sobre sus creadores que acreditase su relevancia, veracidad y responsabilidad social.

En el último punto de su documento Marianne Mikko "reconoce la necesidad de aumentar en la UE la alfabetización en materia de medios de comunicación, recomienda su inclusión entre las nueve competencias básicas y apoya el desarrollo de un programa de formación común europeo en esta materia".

Esta recomendación final de incrementar la educación sobre los nuevos medios de comunicación como una buena herramienta para salvaguardar el pluralismo, fue motivada por el estudio aparecido pocos meses antes sobre la generación Google que reforzaba la teoría de que los weblogs podían suponer una amenaza potencial.

Esta supuesta falta de conocimientos podría afectar también a la incapacidad para interpretar qué significa aparecer en las primeras páginas de los buscadores. Si los miembros de la generación Google carecen realmente de esos componentes críticos de los que alertaba el informe, entonces el problema consiste en que (como millones de ciudadanos de todas las edades) pueden creer que no solamente los primeros resultados son relevantes sino que son los mejores, cosa que obviamente un ciudadano informado y con la debida cultura digital sabe que son cosas muy distintas. Por ejemplo, si un ciudadano cree que los primeros resultados son siempre lo mejor que puede obtener, en caso de no obtener información adecuada, ya no buscará más. Un ciudadano que posea cultura digital crítica, en cambio, sabrá que existen otras fuentes (además de Google) si no encuentra lo que necesita o simplemente cambiará sus estrategias de búsqueda (aunque sea dentro del mismo Google).

⁴⁶Parlamento Europeo. Página personal de Jorgo CHATZIMARKAKIS

<http://www.europarl.europa.eu/meps/en/28243/JORGO_CHATZIMARKAKIS_home.html>

⁴⁷Europa regulará los blogs con un DNI para bloggers <<http://zonadigital.excite.es/europa-regulara-los-blogs-con-un-dni-para-bloggers-N1651.html>>

Todo esto hace que, según los razonamientos señalados, los miembros de la generación Google sean unos usuarios potencialmente vulnerables a diferentes tipos de ataques de información⁴⁸, siendo ésta una de las razones por las que la Comisión de Cultura y Educación aboga por aumentar la **alfabetización en materia de medios de comunicación** como una competencia básica de la UE, una educación destinada a corregir esta situación de debilidad de los que están destinados a ser los futuros protagonistas de la sociedad de la información.

1.3.2. La guerra de información

Aunque probablemente se pueda extrapolar el comportamiento de la Generación Google a personas de cualquier edad, sus hábitos de consumo de información, no solo la dependencia de los resultados de buscadores, sino también de un grupo muy reducido de redes sociales (un grupo compuesto por la "trilateral de facto" que forman Google+Twitter+Facebook") puede facilitar diferentes niveles y tipos de guerra de información, así como las denominadas «operaciones de información»⁴⁹. Para ver cuáles son las operaciones derivadas, debemos contextualizar los posibles escenarios de guerra de información en los que la generación Google puede suponer un riesgo.

El modelo de guerra de información de referencia, y que es el que también utilizaremos aquí, fue publicado en el año 1995 por W. Schwartz en el libro titulado *Information Warfare: Chaos on the Electronic Superhighway*. En su libro ofrece un modelo de tres niveles de guerra de información que se diferencian, no tanto por los métodos utilizados por los llamados «infowarriors» o «activistas», sino por los individuos a los que va dirigido el ataque y por el objetivo final.

⁴⁸Ingeniería social (seguridad informática)

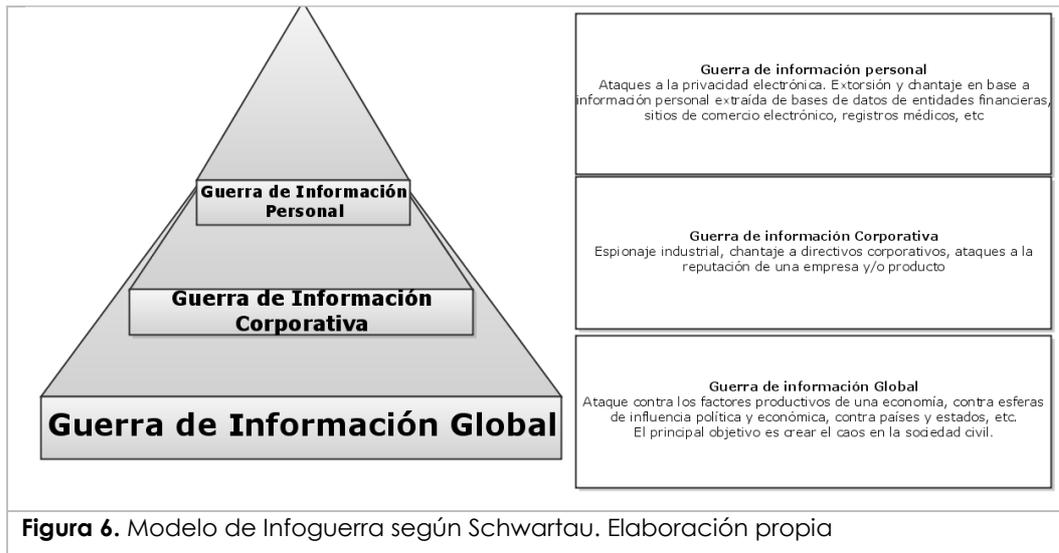
<[http://es.wikipedia.org/wiki/Ingenier%C3%ADa_social_\(seguridad_inform%C3%A1tica\)](http://es.wikipedia.org/wiki/Ingenier%C3%ADa_social_(seguridad_inform%C3%A1tica))>

⁴⁹La guerra de información es el subconjunto de operaciones de información realizadas durante el período de crisis o conflicto para alcanzar o promover objetivos específicos en contra de un adversario o adversarios específicos. Las operaciones de información incluyen una **variedad de disciplinas**: la seguridad operacional, operaciones psicológicas, la guerra electrónica, el engaño militar, la guerra de mando y control, operaciones de inteligencia, asuntos públicos, asuntos civiles, y ataques a la red física o informática

Según W. Schwartz, la guerra de información se libra contra individuos, contra las industrias, esferas políticas de influencia o contra las fuerzas económicas globales, e incluso contra países enteros. En esta guerra de información prima el uso de la tecnología con el objetivo de tratar de robar secretos, convirtiendo la información en un arma en contra de sus propietarios, un enemigo al que se trata de negar la posibilidad de utilizar tanto la tecnología como su información.

En este modelo encontramos los siguientes niveles:

1. El primero, conocido como **Guerra de Información Personal**, va dirigido contra individuos. En este nivel los métodos de ataque personal pueden ser la sustracción de la personalidad electrónica, el robo de información personal y el posterior chantaje con el objetivo de obtener un beneficio económico.
2. El segundo nivel va dirigido en contra de grupos de personas o corporaciones y se conoce como **Guerra de Información Corporativa**. Aquí el denominado infowarrior puede llevar a cabo acciones de espionaje industrial, utilizar métodos del primer nivel para desprestigiar a una persona influyente del grupo o bien puede usar métodos de desinformación y contrainformación con el objetivo de atacar la imagen y la reputación del grupo o corporación, hacer caer el valor de sus acciones o dañar la imagen de alguno de sus directivos, productos o servicios. En este segundo nivel incluimos también la confrontación de partidos y de sus políticos.
3. El último nivel teorizado sería el de la **Guerra de Información Global** e iría dirigido contra esferas de influencia política, contra los pilares del sistema económico y sus factores productivos, y en última instancia contra estados, con el objetivo último de desestabilizar la sociedad civil generando alarma y caos.



Ejemplos de algunos concretos de los diferentes niveles de ataque son los siguientes:

1. Guerra de Información Personal.

El nivel de guerra de información personal tiene muchas variantes que no pueden ser clasificadas y no son el objetivo de esta tesis. El caso de la publicación de contenido personal de ex-parejas es uno de los casos más comunes en este nivel. Sin embargo el campo es mucho más extenso y deberíamos introducirnos en el estudio de técnicas de ingeniería social (del inglés, *social engineering* como técnicas de engaño a través de la red -chat, email o hacking- para conseguir información comprometida de alguna persona).

2. Guerra de Información Corporativa.

Los ataques de guerra corporativa son más comunes de lo que podemos pensar a primera vista. Los ataques contra las **esferas de influencia política** son constantes y podemos constatarlas en la prensa diaria como reflejo de la pugna política entre partidos de cualquier estado democrático y sus medios afines.

Además del documento ya citado de Travers Scott (2004), el interés por este tipo de acciones de información contra la reputación política aparece en otros documentos especializados, como es el caso del libro de Jaime Durán Barba y Santiago Nieto titulado *El arte de ganar: Cómo usar el ataque en campañas electorales exitosas*⁵⁰.

⁵⁰ Jaime Durán Barba y Santiago Nieto. *El arte de ganar: Cómo usar el ataque en campañas electorales exitosas*. <http://books.google.es/books?id=1z82247MTfgC&hl=es&source=gbs_navlinks_s>

En la página 200 de dicho libro hay un apartado directamente relacionado con la comunicación política y uno de los aspectos de la guerra de información; este apartado está dedicado a los sicarios, a los consultores de comunicación política con una actitud ofensiva contra el adversario electoral.

Desde hace muchos años los académicos y los consultores han discutido acerca de los efectos de la campaña sucia en el resultado de las elecciones. En una primera etapa de nuestra profesión, algunos consultores se vendían casi como sicarios. Decían que eran 'consultores de ataque', se especializaban en difundir calumnias, hurtar en la vida privada de los adversarios, en inventar basura. La aparición de internet dio nuevos bríos a esa actividad, y algunos charlatanes han encontrado en ella un medio ideal para distribuir infundios y mentiras. Se pueden enviar mails, armar blogs, falsificar documentos, difundir todo tipo de pasquines con menos riesgo del que se corría en la época de la prensa escrita⁵¹. En ese entonces era más fácil ubicar a quién los escribía y obligarlo a denunciar a los actores intelectuales que le pagaban para hacerlo. Un viejo aforismo enunciaba que "la pared y la muralla son papel del canalla". Con internet tenemos una muralla cósmica que muchas campañas usan para insultar, agredir, hacer daño a los adversarios (...) Debemos distinguir entre distintos tipos de ataque que sólo se parecen, sin ser lo mismo. Joseph Napolitan desarrolló en sus textos algunas ideas sobre el tema. Una cosa es dar a conocer a los electores cierta información para demostrar que el adversario representa un peligro para la sociedad si gana las elecciones y otra es hacer campaña sucia. Lo primero es perfectamente legítimo y hasta una obligación moral.

3. Guerra de Información Global.

La guerra de información global incluiría aquellos casos en los que el ataque se efectúa contra el sistema económico o sus factores productivos, y en última instancia contra los estados. El objetivo último es desestabilizar la sociedad civil generando alarma y caos.

Los ataques al sistema económico pueden derivar en daños contra el sistema político de los estados, y durante los últimos años hemos asistido a este tipo de ataques en diversas ocasiones. Los casos más importantes de los últimos años han sido los del blogger Minerva de Corea del Sur, el caso del Hong Kong Bank, y más recientemente el caso de la guerra de información declarada entre EE.UU. y los Hackers chinos.

⁵¹ En este apartado, donde leemos 'consultores al ataque', también podríamos leer 'bloggers al ataque', dado que en gran parte de los casos la herramienta de descalificación online son blogs.

El blogger Minerva fue acusado de, a través de los contenidos vertidos en su blog, haber generado pérdidas al estado de Corea del Sur por valor de dos billones de dólares, y de ser él solo el causante de una pérdida del PIB del dos por ciento⁵². El caso del Hong Kong Bank⁵³, la combinación de cadenas de SMS⁵⁴ y listas de correo electrónico supuso la caída de las acciones del banco un 7% en las primeras cuatro horas de pánico hasta que el primer ministro apareció en televisión para desmentir los falsos rumores de falta de liquidez, frenando así un efecto dominó que podría haber supuesto la quiebra de todo el sistema financiero de Hong Kong con consecuencias catastróficas para la economía asiática y mundial.

El caso más cercano en el tiempo es el de la aparente guerra declarada entre EE.UU. y China a nivel electrónico y cibernético si atendemos a las noticias regulares de las que nos proveen los medios sobre los diversos episodios de las mismas. En particular, las noticias de supuestos hackers chinos⁵⁵ asaltando los ordenadores del Pentágono ponen de relieve la existencia de planes de ataque y defensa de los sistemas nacionales de inteligencia y el proceso de captación de información ante la posibilidad de una guerra de información global. En este ámbito las llamadas «Operaciones de Información» de ambas potencias se enfrentan en el campo de batalla de Internet para conseguir una ventaja competitiva en el caso de conflicto. Podemos imaginar cómo será la guerra de información Global en caso de ocurrir.

En lugar de la antigua Radio Libertad de la Guerra Fría, las directrices actuales del Pentágono son que en caso de guerra de información global la disponibilidad de

⁵² En 2009 el blogger surcoreano Park Dae-sung, apodado 'Minerva'. Minerva fue detenido y acusado de ser el causante de al menos 2.000 millones de dólares en pérdidas para la economía del país por difundir rumores perniciosos en su blog, siendo un claro ejemplo de información con origen en el micro-estrato. La popularidad de este blogger tuvo su eco en los estratos superiores generando con sus comentarios alarma social en su país y trascendiendo al resto del mundo con su detención <http://www.koreatimes.co.kr/www/news/biz/2009/01/123_37648.html>.

⁵³ Hong Kong Bank Mobbed By Savers <<http://news.sky.com/story/635509/hong-kong-bank-mobbed-by-savers>>

⁵⁴ Otro ejemplo de guerra de información lo encontramos en septiembre de 2008 en el caso del Bank of East Asia. A través de cadenas de SMS se difundieron rumores maliciosos sobre la estabilidad financiera del banco, generando así el pánico entre los ahorradores que lanzaron a las sucursales del banco en busca de su dinero. En unas pocas horas el pánico fue controlado, sin embargo las acciones de BEA cayeron aquel día un 11%, convirtiéndose en un claro ejemplo de los resultados que podía tener un ataque de guerra de información corporativa. En el caso que la situación de pánico se hubiese dilatado en el tiempo y que otros medios digitales del estrato medio y convencional hubiesen acreditado los rumores, el caso inicial de guerra corporativa podría haber afectado de la misma manera a otros bancos de la región en un efecto dominó, generando una alarma social difícil de controlar y culminando en una crisis económica global.

⁵⁵ 'Hackers' chinos han accedido al armamento secreto del Pentágono <<http://www.lavanguardia.com/internet/20130528/54374921591/hackers-chinos-han-accedido-al-armamento-secreto-del-pentagono.html>>

comunicación de redes sociales deben estar disponibles y funcionando en el momento del ataque. La conectividad total podría hacer que una noticia viral alcanzase a millones de usuarios en pocas horas o días, dependiendo entre otras cosas de la capacidad de reacción para desmentir la contrainformación.

Un ejemplo cercano lo encontramos en las revueltas de Irán de febrero de 2010. Durante esas semanas, apunta Xavier Peytibi⁵⁷, EE.UU. creó un software para evitar la censura en Internet por parte de Irán, incluso habló con responsables de Twitter para que no hiciera ninguno de sus típicos “parones por mantenimiento” durante las protestas. También dotó de dinero a agencias para que buscaran métodos para permitir el acceso a Internet desde Irán a través de fondos del departamento de Estado.

Como indica Noah Shachtman en Wired ⁵⁸, comparando esta situación con el anterior envío de mensajes a través de la radio, “la diferencia estriba en que **el contenido ya no es lo importante**; el énfasis está en **apoyar la infraestructura técnica** y dejar que la ciudadanía decida por sí misma qué decir. La propia comunicación puede hacer caer el despotismo”.



Figura 7. Turquía es un ejemplo más de un grupo de países en los que las redes sociales están prohibidas para evitar este tipo de conflicto de información. Fuente: Noticia del 15 de Marzo de 2007⁵⁶.

El día 21 de marzo de 2014 el gobierno turco de Recep Tayyip Erdogan bloqueó Twitter cosechando un amplio rechazo social y la condena internacional. A las pocas semanas el Tribunal Constitucional turco declaró ilegal la medida, lo que obligó a levantar el bloqueo⁵⁹. Según el veredicto del dos de abril, la prohibición violaba la libertad de expresión y los derechos individuales de los ciudadanos.

Por último, si bien no se ha dado ningún caso de Guerra de Información Global, la combinación del modelo de Infoesfera de Peretti/Bennett y del modelo de Guerra de

⁵⁶ Web censorship spreading globally <<http://www.ft.com/cms/s/2/1dbb5faa-d268-11db-a7c0-000b5df10621.html#axzz35UvzEG6c>>

⁵⁷El arte de la guerra (en las redes sociales) <<http://www.xavierpeytibi.com/2010/01/20/el-arte-de-la-guerra-en-las-redes-sociales/>>

⁵⁸ New U.S. Foreign Policy: Building Social Networks <<http://www.wired.com/2009/12/new-us-foreign-policy-building-social-networks/>>

⁵⁹El País. Turquía levanta el bloqueo de Twitter después de que el Constitucional lo declarara ilegal <http://internacional.elpais.com/internacional/2014/04/03/actualidad/1396534745_356724.html>

Información de Schwartau apunta a la Blogosfera como el **epicentro** de la Infoesfera desde donde en la peor de las hipótesis sería posible desestabilizar la sociedad civil si una o más noticias denominadas de «mass disruption» (trastorno masivo) logran alcanzar los medios de comunicación de masas si, como señalamos, los modelos anteriores resultan plausibles.

1.3.3. La guerra de información en la Blogosfera y el papel de los bloggers

En el 2004, cuando la blogosfera⁶⁰ había alcanzado una cifra cercana a los tres millones de blogs, el periodismo se interesó particularmente por este nuevo medio de publicación a causa de la cantidad de situaciones en las que los blogs fueron influenciando, suplantando, sobrepasando, e incluso socavando, a los medios de comunicación tradicionales.

El impacto de las noticias lanzadas desde blogs fue de particular interés por muchos públicos y se convirtió en el objeto de estudio no solo por parte de los medios de comunicación y de los propios blogs más leídos de la blogosfera, sino también por el mundo académico.

Este interés creciente apareció reflejado en el artículo *Tempests of the Blogosphere. Presidential Campaign Stories that Failed to Ignite Mainstream Media*, escrito por Travers Scott, profesor de la Universidad de Washington. En este artículo, uno de los inspiradores de esta tesis, el autor analizó la acción de los blogs durante la campaña presidencial norteamericana de 2004.

Travers Scott utilizó el modelo de los estratos de la Infoesfera elaborado por Bennett para analizar lo que él denominó “blogflops” (blogs fallidos/fracasos blog) e identificar historias y noticias que no llegaron a alcanzar a los medios convencionales. Según el autor, “las campañas de comunicación más exitosas incluyen métodos para dar acceso a una historia que active los tres estratos de la Infoesfera”. Travers explicó el proceso por el que un meme era capaz de llegar a todos los estratos de la Infoesfera con un coste mínimo. Según el análisis del autor en el contexto de las elecciones

⁶⁰Literalmente, la blogosfera son todos los blogs de la red, sin embargo, normalmente se utiliza para referirse al conjunto de los blogs más activos y más leídos, así como sus conversaciones con otros blogueros acerca de algún tema.

presidenciales de EE.UU. de 2004, "Buzz spread through conservative blogs and discussion forums such as *Power Line* and *Little Green Footballs*, sparking a flurry of research and, among liberal blogs, counter-research. The story quickly escalated to gossip site *The Drudge Report*, talk radio, cable news, and, finally mainstream broadcast and print".

Analizando los posts que no habían alcanzado el estrato convencional, Travers esbozó una teoría que asemejaba el papel de este tipo de información al de los globos sonda. Según el autor, cualquier noticia lanzada desde el Estrato Micro tiene la oportunidad de alcanzar altas posiciones de visibilidad si es citado por un medio convencional. En este caso, la noticia de un blog puede explotar en esta esfera convencional a través de la mención por un medio tradicional. Esta mención puede iniciar una segunda fase de visibilidad, ya que a través de la mención/réplica por un medio del estrato convencional puede alcanzar los estratos medio y micro.

Lo expuesto por Travers suscitó el interés por parte de las **agencias de inteligencia** de EE.UU. ya que esta información lanzada desde el micro-estrato podía estar vinculada a propaganda, contrainformación o desinformación, lo que nos lleva a otro tipo de blogger no recogido por la clasificación de Technoratti pero que sí ha sido estudiado y analizado por agencias de inteligencia de diversos países.

De esta manera, en el año 2006, asistimos a una evidencia de este creciente interés que despertó la Blogosfera con estudios como el de Kinniburgh (2006) de la Joint Special Operations University titulado *Blogs and Military Information Strategy. Information Operations Sphere*⁶¹. En este estudio los servicios de inteligencia de los Estados Unidos afirman que los blogs han de verse como riesgos potenciales de seguridad, pero también como valiosas fuentes de información para atacar a determinadas personas o promover mensajes específicos a través de equipos propios de bloggers o sobornando a terceros para que lo hagan.

⁶¹Kinniburgh (2006) *Blogs and Military Information Strategy. Information Operations Sphere*. Joint Special Operations University.

En el año 2008 apareció la publicación titulada “*Digital media and democracy: tactics in hard times*”⁶² editada por Megan Boler. Según la autora, a pesar de la desconfianza que generan los bloggers, la gente sigue utilizando los nuevos medios como una forma de aproximarse al debate público. Este acceso es facilitado por la tecnología de la Web 2.0. que llega incluso a retar a los medios tradicionales a raíz de la democratización del conocimiento y la multiplicación de fuentes de información provenientes del mundo digital. La propia autora refuerza este argumento con diferentes ejemplos que incluyen “conservatives offering training camps to bloggers; mainstream media adopting official bloggers to augment their behind-the-times print news; progressive bloggers redefining the dominant news agenda through investigation and muckraking (...)and viral video being used for savvy public relations and advertising”⁶³.

A pesar de la aparente neutralidad y beatitud de los bloggers descritos por Megan Boler, cuando en abril de 2009 apareció otra noticia titulada *America's Newest Profession: Bloggers for Hire*⁶⁴, el propio titular ofrecía un argumento más contra la peligrosidad del medio. Dicho artículo se hacía eco de la profesión más nueva en los Estados Unidos, la de blogger, una profesión que a pesar de ser muy reciente ya superaba en número a los programadores informáticos y que se acercaba cada vez más al número de abogados que ejercen en Estados Unidos.



Figura 8. Cartel propagandístico de Radio libertad durante la guerra fría.



Figura 9. Titular del Wired refiriéndose a la nueva “Radio libertad” de las Redes sociales



Figura 10. Noticia de Junio de 2010 sobre la detención del blogger iraní “Hoder”

⁶² Boler, M. (2008). *Digital media and democracy: tactics in hard times*. Massachusetts Institute of Technology.

⁶³Op cit. Pg 14.

⁶⁴America's Newest Profession: Bloggers for Hire<<http://online.wsj.com/news/articles/SB124026415808636575>>

Los llamados «bloggers en alquiler» reafirmaban de alguna forma la amenaza de que alguna persona, corporación o lobby podía contratar los servicios de escritores a sueldo que gracias al anonimato de la red podían ser utilizados para difundir información partidista. Los objetivos reales de estos bloggers podían estar relacionados con la agenda oculta mencionada por los eurodiputados y destacada por los servicios de inteligencia como un nuevo modo de ciberterrorismo en campañas de descrédito global en contra de personas, corporaciones, o incluso países.

Como indicaron los europarlamentarios, el uso ilícito con intenciones dañinas es una amenaza real debido a las características de navegación de los jóvenes de la generación Google. Para influenciar a esta generación Google, adquiere gran importancia la visibilidad que puedan alcanzar las diferentes páginas web, tanto en el Top10 como en el Top4, ya que la probabilidad de que el contenido sea clicado dependerá de su posición en la SERP (por las siglas inglesas de *Search Engine Results Pages*) de buscadores.

En este nuevo escenario de ciberterrorismo los actores son los bloggers, una gran cantidad de escritores mercenarios que escriben contenidos por encargo. Estos contenidos son el rey de la web 2.0 tienen como objetivo alcanzar a los usuarios de la Generación Google a través del medio de difusión más popular de la web, los buscadores, y en particular Google.

Como demostraremos a lo largo de este trabajo, la posibilidad de manipular resultados de búsqueda de los buscadores es algo real, factible, incluso sencillo, y recientemente diferentes grupos mediáticos han vuelto a destacar este peligro. En abril de 2013, un artículo de AIBRT titulado *Democracy at Risk: Manipulating Search Rankings Can Shift Voters' Preferences Substantially Without Their Awareness*⁶⁵ ponía de relieve esta amenaza que podía llegar a decidir los resultados electorales en una democracia como la estadounidense.

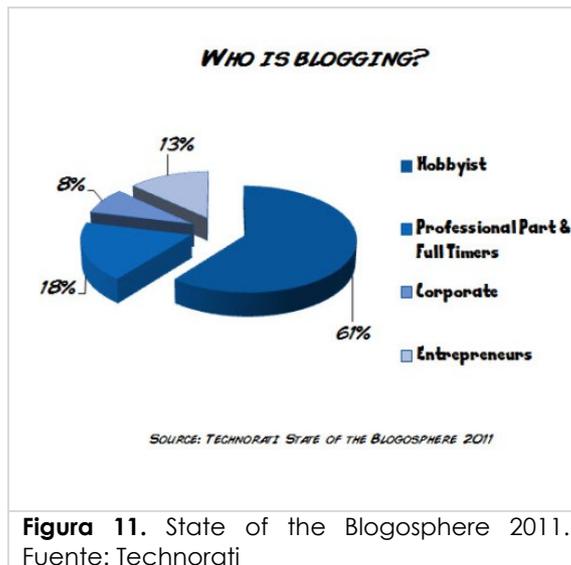
Así pues, la razón por la que blogs y bloggers se han convertido en el centro de atención tanto de agencias de inteligencia como del propio Parlamento Europeo, es por la posición central que ocupa la Blogosfera dentro de la Infoesfera y por la posibilidad de ser el origen de algún tipo de guerra de información.

⁶⁵Robert Epstein y Ronald E. Robertson. *Democracy at Risk: Manipulating Search Rankings Can Shift Voters' Preferences Substantially Without Their Awareness*. American Institute for Behavioral Research and Technology. 2013 <http://aibr.org/downloads/EPSTEIN_and_Robertson_2013-Democracy_at_Risk-APS-summary-5-13.pdf>.

En el estudio de Technorati mencionado anteriormente, se publicó los resultados de una encuesta realizada a bloggers para definir la diferente tipología de bloggers en base a su motivación y al grado de profesionalidad.

Los principales grupos o categorías de bloggers según la interpretación del resultado de la encuesta resultaron ser los siguientes:

- **Aficionados:** personas que bloguean por diversión y no tienen ningún ingreso. Representan alrededor del 60% de esta encuesta.
- **Profesionales** a tiempo parcial y bloggers a tiempo completo: son los bloggers independientes que, o bien usan su blog como una forma de complementar sus ingresos, o consideran que es su trabajo de tiempo completo. Ellos bloguean sobretudo



- acerca de temas personales y de tecnología. Estos bloggers representan el 18% de los encuestados.
- **Corporativos:** bloguean como parte de su trabajo a tiempo completo o bloguean a tiempo completo para una empresa o una organización para la que trabajan. Representan aproximadamente el 8% de este grupo.
- Los **emprendedores:** Un 13% de la blogosfera se caracteriza por ser emprendedores o empresarios que bloguean para su propia empresa u organización.

Esta clasificación global recoge los principales tipos de bloggers, sin embargo no es exhaustiva ya que obvia una motivación paralela derivada de la popularidad de los bloggers y que fue pronto objeto de estudio.

Por lo tanto, a la clasificación que acabamos de ver, entendemos que podemos añadir dos subcategorías que pueden ser clasificadas dentro de alguno de los cuatro tipos principales, pero que por sus características requieren de un estatus diferenciado. Estas dos subcategorías son la del **periodismo ciudadano** y la del **activista**.

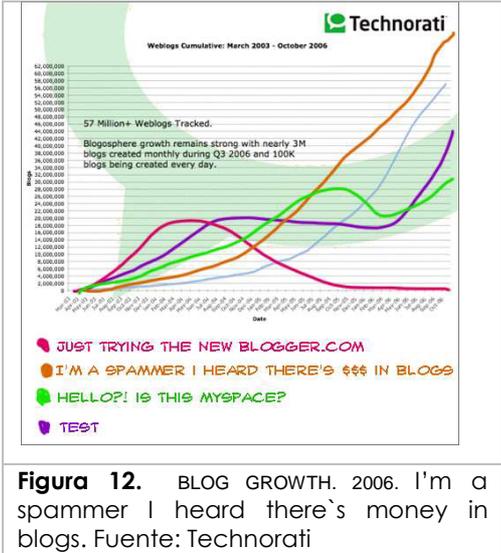
- El **periodismo ciudadano** ⁶⁶. Este tipo de periodismo independiente, también denominado periodismo 3.0., surgió gracias al uso generalizado de los blogs que se popularizó a finales de los 90 y principios de los 2000. De forma paralela, algunos medios digitales han fomentado la participación ciudadana a través de espacios alojados en sus páginas web para que den su visión de diversos acontecimientos. En ambos casos, en este tipo de periodismo los ciudadanos exponen su visión de diversos asuntos sociales, culturales, políticos, económicos, locales, deportivos, etc. Estos bloggers del periodismo alternativo pueden alcanzar relevancia dentro de grupos o comunidades específicas, desde comunidades geográficas pequeñas y medianas, hasta diversas comunidades de intereses. Este tipo de bloggers también podría ser clasificado en la categoría de «aficionados».
- El **activista**.⁶⁷ Este tipo de blogger puede trabajar de forma independiente pero suele estar asociado a una corporación o grupo de interés, poniéndose al servicio de intereses particulares. Los intereses defendidos por este tipo de blogger pueden ser lícitos o ilícitos y pueden estar vinculados a agendas específicas de lobbies, servicios de inteligencia o agencias gubernamentales y sus actuaciones estar relacionadas con la difusión de información, desinformación o contrainformación. Además de la manipulación de información, la especialización de algunos de estos activistas (también mencionados como *ciberwarriors*) incluyen la posibilidad de servicios relacionados con actuaciones de engaño (social engineering) o manipulación de servidores web.

⁶⁶El periodismo ciudadano o periodismo participativo o periodismo público o periodismo democrático (en inglés periodismo de guerrilla o periodismo de la calle) es un movimiento en el que son los propios ciudadanos quienes se convierten en informadores.

⁶⁷La razón de ser de esta tesis es la detección de casos de ciberguerra llevada a cabo por estos *ciberwarriors*.

Para entender la funcionalidad de este último tipo de blogger debemos remitirnos a los escenarios donde los activistas o *ciberwarriors* son actores necesarios e indispensables. Este tipo de escenarios son los de la guerra de información, en los que los bloggers pueden atender a intereses particulares y ser guiados por lo que se ha denominado una «agenda oculta». Este tipo de actuaciones partidistas pertenecen al ámbito de las «operaciones de información» y pueden suponer un peligro para

la sociedad. El marco de actuación de este tipo de operaciones es el de la guerra de información, como hemos visto, escenario definido y delimitado por Schartaw en 1995. No obstante, hemos de insistir en que se trata de un escenario hipotético previsto por modelos teóricos más o menos bien argumentados, pero no podemos decir que haya casos documentados en este sentido.



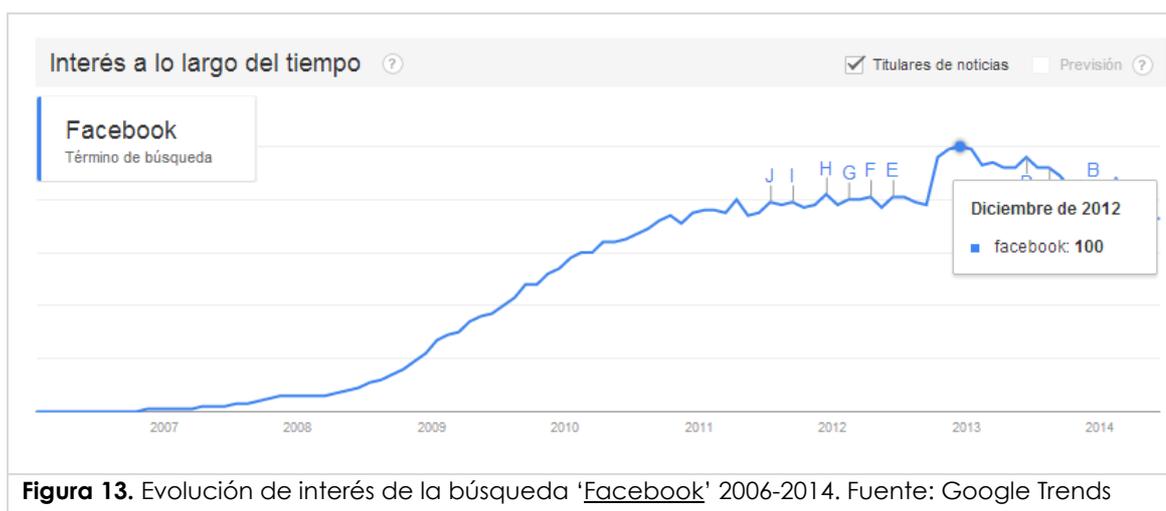
1.4. Los buscadores y la lucha por el posicionamiento

Si a mediados de los años 90 Internet fue considerado una novedosa vía alternativa para acceder a la información, dos décadas después la evolución de la tecnología ha cambiado la faz de Internet. La revolución de la web 2.0 permitió en unos pocos años la aparición y el crecimiento de las redes sociales, y en 2004 el nacimiento de Facebook.

Facebook no se abrió a todos los internautas hasta septiembre de 2006 y a lo largo de todo 2007 las redes sociales se convirtieron en la búsqueda estrella del año con unos 500 millones de visitas. El fenómeno fue global y afectó a prácticamente todos los países. Siete de las 10 palabras y términos de búsqueda que más crecieron en el buscador Google eran los nombres de redes sociales como Badoo y Facebook.

Por aquel entonces MySpace era la mayor de todas con cerca de 110 millones de usuarios, y Facebook alcanzaba ya los 55 millones de usuarios tras crecer un 100% ese mismo año.

Hasta **2007**, año de la eclosión de las redes sociales con Facebook a la cabeza, cerca del 70% de los usuarios de Internet usaban los buscadores como punto de entrada a la web. Ya entonces, los buscadores recibían cada día millones de búsquedas y presentan billones de resultados semanales en respuesta a esas búsquedas.



Dos años más tarde, a principios de 2009, aparecieron noticias en Estados Unidos en las que se podía leer que Internet había superado a la prensa escrita como principal fuente de noticias.

Cuando el 4 de febrero de 2014 Facebook cumplía su décimo aniversario con más de mil 200 millones de cuentas creadas, en apenas cinco años la evolución de la tecnología y la difusión de los teléfonos inteligentes han democratizado el acceso a Internet en una nueva realidad mediática que el buscador Google ha tenido que compartir con las redes sociales.

Una buena parte de esta realidad de 2014 la muestra el *Reuters Institute Digital News Report*⁶⁸, un estudio basado en una encuesta a más de 18.000 consumidores de noticias online en UK, EEUU, Alemania, Francia, Italia, España, Brasil, Japón, Dinamarca y Finlandia. En el informe podemos leer:

La forma más extendida de conocer noticias nuevas entre los usuarios españoles es acceder directamente al sitio web o la aplicación de un medio; el 46% de los encuestados se informa de esta manera [...]. Las **redes sociales** en conjunto son empleadas para informarse por un 38% de los usuarios: un 15% del total encuentra noticias a través de Twitter, y el 30% conoce informaciones nuevas a través de otras redes sociales, como Facebook, Google+, LinkedIn. Los **buscadores** son la tercera modalidad de acceso, con un 35% de personas que emplean internet para informarse y que descubren noticias a través de ellos. Eso sí: se emplean más para buscar el nombre de un sitio web concreto (23%), es decir, como mero acceso a sitios web decididos de antemano, que para buscar palabras clave de historias o temas específicos (20%).

De dicho informe se desprende que Facebook es la primera plataforma a la hora de seguir la actualidad informativa. En España el 50% de los usuarios accede habitualmente a diarios y noticias online a través de esta red social⁶⁹, mientras que los más de 400 millones de usuarios de WhatsApp a nivel mundial envían al día 50 mil millones de mensajes entre los que muchos enlazan con noticias online. La tercera red social es Twitter con un 15% de usuarios que utilizan esta red para acceder a las noticias.

La información que antes suplían de modo casi exclusivo los medios de prensa tiene ahora tres vías de acceso siendo según el informe de Reuters los **buscadores** la tercera modalidad de acceso, con un 35% de personas que los emplean para informarse tras el acceso directo (46%) y las redes sociales (38%). Además, los buscadores son la herramienta que mucha gente utiliza a diario no solo para acceder a las noticias

⁶⁸ Reuters Institute Digital News Report <<http://www.unav.edu/documents/3786985/0/Reuters-Institute-Digital-News-Report-2014-Espana.pdf>> o <<http://www.digitalnewsreport.org/>>

⁶⁹ Las redes sociales se consolidan como vía de acceso a las noticias <<http://www.puromarketing.com/16/22400/redes-sociales-consolidan-como-via-acceso-noticias.html>>

online, sino también para acceder a infinidad de sitios web, servicios y multitud de recursos y colecciones especializadas que ofrece la web a través de sus diferentes opciones de búsqueda.

Además de satisfacer necesidades informativas, los buscadores son herramientas de navegación que no sólo llevan a los usuarios a URLs específicas, sino que ofrecen acceso a colecciones de imágenes o música, e incluso a direcciones físicas a través de sus mapas, nos ayudan y guían en la búsqueda, incluso corrigiendo nuestras faltas ortográficas a través de thesaurus y diccionarios.

En el afán de devolver información relevante a los usuarios, los buscadores se han convertido en agregadores de servicios y colecciones de contenidos que, al ser ofrecidos en la parte alta de la página de resultados o SERP desplazan del área visible la página los diez resultados naturales u orgánicos seleccionados por los algoritmos de ordenación. En el campo del posicionamiento, denominado Top 10 a estos elementos (páginas web, documentos, etc.) que aparecen en la primera página de resultados a través de los algoritmos de ordenación en contraposición a los resultados que proceden de la publicidad (en el caso de Google, del programa AdWords) o de contenidos generados por los propios buscadores

Como demostraremos a lo largo del trabajo, en ocasiones solo dos resultados orgánicos (o naturales o gratuitos) del Top10 original aparecen en la zona visible de la interfaz, siendo desplazados no solo por la publicidad, sino también por elementos propios de los buscadores que convierten la zona alta de SERP en una zona muy competitiva.

Por otro lado, veremos que el conteo de los clics que recibe un resultado por parte de los usuarios, el denominado CTR (por la expresión click-through rate) es aplicado a la hora de reposicionar los resultados del Top10, permitiendo a sitios posicionados en la parte baja de la primera SERP, escalar posiciones hasta la parte visible de la interfaz gracias a los clics recibidos. Teniendo en cuenta este factor de interacción humana, aquellos sitios que consigan posiciones altas del ranking tendrán mayor probabilidad de verse beneficiadas del CTR en SERP, mientras que los sitios de la parte baja de la SERP, con una menor cantidad de clics, se verán perjudicados en tanto que podrán ser relegados más fácilmente del Top10 y pasar a aparecer en la segunda SERP con una probabilidad mucho menor de ser clicados.

Por tanto, podemos decir que sólo aquellos sitios que consigan aparecer en posiciones muy altas del ranking, además de conseguir tráfico, se verán beneficiadas del CTR en la SERP, mientras que el resto, a pesar de posicionarse en el Top10, conseguirán menos tráfico y al no ser tan clicados, podrán ser más fácilmente desbancados por otros sitios con contenidos nuevos u optimizados para aparecer en los resultados de esa misma SERP.

Vistos estos datos, se hace indispensable lograr altas cotas de visibilidad en la SERP a través del llamado posicionamiento web o SEO (por las siglas inglesas de *Search Engine Optimization*). El SEO es, a la vez, una dinámica rama profesional del mundo de la web y la publicación digital y como no podía ser de otro modo, un potente campo de estudios.

En particular, el SEO estudia las características de los motores de búsqueda y sus algoritmos de posicionamiento, y a partir de aquí, propone las mejores prácticas y estrategias para la optimización de una página web de cara a los resultados de los motores de búsqueda. En términos muy pragmáticos, optimizar una página en los buscadores implica conseguir que tal página aparezca en la primera página de resultados (SERP o SERPs, por las siglas en inglés de *Search Engine Results Pages*) de los buscadores para una o más palabras clave determinadas de antemano.

A continuación veremos cuáles son las principales razones de la importancia del SEO y la razón de este interés y preocupación por conseguir visibilidad en las SERPs de los buscadores.

1.4.1. Razones de la importancia del posicionamiento en SERP

La popularidad del uso de los buscadores y gran cantidad de tráfico de usuarios proporcionada a aquellos sitios web posicionados en la parte visible de la SERP, incrementó el interés y la preocupación de las empresas y organizaciones por conseguir visibilidad en las SERPs de los buscadores, y como consecuencia fomentó la aparición de especialistas en SEO. Este interés tiene su fundamento en las siguientes razones:

>Primacía del ranking. Aunque los buscadores pueden llegar a ofrecer millones de resultados para una búsqueda, en realidad esto carece de importancia. Literalmente es lo de menos que, para una palabra clave el buscador presente unas decenas de resultados o unos centenares de miles (o de millones). Los usuarios únicamente examinan los primeros resultados. Así que, para cualquier búsqueda, lo importante es aparecer en el llamado Top 10 o incluso en el Top 5. Para ordenar los resultados (el ranking) los buscadores utilizan algoritmos cuyos detalles conceptuales se conocen más o menos bien (no así los detalles operativos, que se mantienen en secreto). Por eso el SEO centra buena parte de sus esfuerzos en conocer, mediante ingeniería inversa al menos los aspectos conceptuales de los rankings de los buscadores.

>Visibilidad del negocio y Branding⁷⁰. Cuando los usuarios buscan productos y servicios, el posicionamiento SEO facilita que una página web aparezca en posiciones altas en los motores de búsqueda. Cuando un sitio web aparece en SERP para una o varias palabras clave, los usuarios pueden hacer clic en el enlace correspondiente de la SERP y acceder al sitio web en cuestión. La visibilidad en los resultados de búsqueda para diferentes términos con una vinculación semántica y/o temática (grupo semántico) directamente relacionados con una página web aumenta la posibilidad de que los usuarios hagan clic en los resultados, ya que la mayoría de usuarios de buscadores no se limitan solo a buscar una vez, sino que modifican los términos de búsqueda y renuevan los resultados varias veces perfeccionando así los términos de búsqueda.

Por otro lado, la aparición recurrente de una página web para un conjunto de búsquedas temáticas aporta reconocimiento de marca (*branding*) y un gran número de clientes potenciales que probablemente con el tiempo hagan clic y accedan al sitio web en cuestión que ha alcanzado visibilidad.

>Credibilidad a su negocio. Los usuarios de buscadores tienen en cuenta los resultados visibles ya que confían en el criterio de Google para mostrar resultados relevantes. De esta forma, aparecer en la primera página hace que los sitios sean más creíbles a ojos de los usuarios, cuyo clic será un voto de confianza hacia Google. En el caso que la página disponga de acceso al sistema de información denominado Google Analytics, el administrador de la web podrá corroborar el grado de satisfacción de los usuarios a través de la experiencia del propio usuario gracias a diferentes KPIs (del inglés Key

⁷⁰Why SEO is Important For Businesses <<http://firstviewonline.com/why-seo-is-important-for-businesses/>>

Performance Indicator) tales como el porcentaje de rebote de la página (usuarios que dejan la página sin hacer ningún clic en ella) o el tiempo medio de estancia en la página específica y en el sitio web.

>**Click-throught o CTR.** El principal indicador de los estudios de SEO está ligado al posicionamiento web y por tanto a la visibilidad de los sitios web en las SERPs de los buscadores. Lo cierto es que, en muchas ocasiones la fuente más importante de tráfico hacia un sitio web es Google a través de las búsquedas de los usuarios y los clics que hacen en las páginas de resultados. De esta interacción se deriva el click-throught, una métrica que calcula la cantidad de veces que un enlace ha sido clicado en relación a la cantidad de veces que ha sido visible⁷¹. Esta métrica es el principal indicador de rendimiento o KPI del tráfico generado a través de las SERPs de los buscadores y uno de los factores que más influyen en el posicionamiento. Por ello, los analistas estiman que también es uno de los que más utilizan los propios buscadores para la reconfiguración de sus SERP en base a la interacción con los usuarios⁷². Se considera que Google calcula el CTR global de la página así como el CTR para búsquedas específicas e introduce cambios en la posición del Top10 en base a la interacción de los usuarios.

>**Tráfico.** Como ya hemos señalado, el SEO facilita la creación de tráfico hacia el sitio web derivado de su visibilidad en la SERP de los buscadores. A mayor cantidad de tráfico mayor será la posibilidad de alcanzar algún objetivo de negocio. En el caso de la publicidad online, los ingresos generados dependerán en gran medida del tráfico generado, y a mayor cantidad de tráfico mayores ingresos publicitarios⁷³.

>**El retorno de la inversión.** Diversos estudios confirman que el ROI (por Return of Investment)⁷⁴ obtenido por una estrategia SEO es mejor que cualquier otra forma tradicional de publicidad offline u online. El motivo es que el SEO es una estrategia de marketing para generar tráfico de usuarios que están altamente incentivados a la

⁷¹Aunque el CTR es una popular métrica en los sistemas de enlaces patrocinados como Google Adwords, los buscadores calculan el CTR también de los resultados orgánicos.

⁷²Esta teoría no está documentada ni aceptada por parte de Google, sin embargo, buscadores como Ask.com tenían implementado este 'algoritmo de clics' desde que empezó a utilizar el motor Teoma en sus resultados.

⁷³El precio de la publicidad en los medios online depende de la cantidad de tráfico que puedan declarar. Agencias que establecen auditorías web como OJD y que constatan la cantidad de tráfico diaria de los medios online, sirve para incrementar el valor de la publicidad ofertada en el sitio del medio.

⁷⁴La fórmula del índice de retorno sobre la inversión es: $ROI = (\text{Utilidad neta o Ganancia} / \text{Inversión}) \times 100$

compra de los productos y servicios que ofrecen las empresas visibles en SERP para las búsquedas de su interés.

>Potencial de viralidad. El SEO es parte indispensable de cualquier proyecto de mercadotecnia viral ⁷⁵ para atraer masa crítica de usuarios hacia un punto determinado de una web, al igual que otros factores⁷⁶ como la originalidad del meme, la presencia en redes sociales o la facilidad de ser compartido. El SEO de una web bien posicionada es indispensable en mercadotecnia viral en cualquiera de sus fases, tanto en la visibilidad durante el lanzamiento como para controlar la resistencia de sus servidores durante los momentos de mayor tráfico⁷⁷.

Como analizaremos en la segunda parte de este capítulo, el SEO permite lograr cotas de visibilidad que pueden transformarse en tráfico en poco tiempo. El posicionamiento SEO incrementa la probabilidad de clic en una búsqueda informacional que puede convertirse en navegacional y/o transaccional si Google asocia el sitio como resultado más relevante para esta búsqueda específica. Grandes cantidades de tráfico en poco tiempo pueden provocar la viralidad de un contenido como veremos más adelante.

>Conocimiento de los usuarios. El software de minería de datos Google Analytics realiza un seguimiento de este tráfico. Los datos y las métricas son valiosos porque ofrecen una idea de cómo son los usuarios de la página web. Entre otros, nos muestra cómo buscan, cómo navegan, el lenguaje que utilizan, la tecnología que utilizan, la región en que viven, los días que son más activos, e incluso los momentos del día que concentran su navegación.

⁷⁵La mercadotecnia viral (omarketing viral) emplea técnicas de mercadotecnia para intentar explotar redes sociales y otros medios electrónicos tratando de producir incrementos exponenciales en "reconocimiento de marca" (brand awareness), mediante procesos de autorreplicación viral análogos a la expansión de un virus informático. Se suele basar en el boca a boca mediante medios electrónicos usa el efecto de "red social" creado por Internet y los modernos servicios de telefonía móvil para llegar a una gran cantidad de personas rápidamente.<http://es.wikipedia.org/wiki/Mercadotecnia_viral>

⁷⁶ Lista con 100 factores clave que encuentras en posts virales <<http://www.quondos.com/los-100-factores-para-crear-un-post-viral/>>

⁷⁷La opción 'Tiempo Real' de Google Analytics permite controlar en directo picos de tráfico que pueden desembocar en caídas de los servidores donde se aloja un sitio web.

1.4.2. Tipos de búsquedas

Los estudios han demostrado que la mayoría de los usuarios utilizan los buscadores para encontrar piezas específicas de información o para hacer preguntas (Rose y Levinson, 2004 Zhang, Jansen y Spink, 2009) a través de las páginas de resultados. Estas SERP son fruto de la interacción del usuario con el buscador. Una vez introduce una o varias palabras clave en el cajetín de búsqueda, obtiene una página compuesta por los resultados de búsqueda, normalmente diez enlaces que apuntan a uno o más sitios web.

A su vez, a través estudios de usuario divulgados por Google, se considera que las motivaciones de los usuarios para hacer consultas en buscadores se suelen dividir en tres grandes grupos: informacionales, transaccionales y navegacionales.

Las búsquedas o palabras clave **informacionales** son las consultas que buscan obtener informaciones sobre algún tema, y en concreto pueden buscar imágenes, canciones, vídeos o documentos. Las **transaccionales** son consultas que tienen una motivación comercial (típicamente adquirir un producto o servicio específico). Por último, las consultas **navegacionales** son aquellas en las cuales lo que los usuarios desean es poder ir a una URL específica.

En este sentido, el estudio de Jansen et al (2008) mostró que el 80% de las consultas de búsqueda web son predominantemente informativas, mientras que el restante 20% de las consultas se dividen entre las navegacionales y las transaccionales. Bernard et al (2008) ⁷⁸ utilizan esta categorización básica y la expanden a diferentes niveles basándose en diferentes artículos previos que habían intentado categorizar los diferentes tipos de palabras clave.

Según los autores, las características básicas de cada tipo de palabra clave, de acuerdo con la intencionalidad de la búsqueda, es la siguiente:

⁷⁸Bernard J., Booth D., Spink A. (2008). *Determining the informational, navigational, and transactional intent of Web queries.*

1. Consultas que indican intencionalidad navegacional:

- Las que contienen los nombres de compañía, empresa, organización o persona
- Las que contienen TLDs⁷⁹
- Con 'Web' como una de las palabras clave
- Con una longitud inferior a 3 términos
- Usuarios cuya necesidad informativa se ve cumplida visionando la primera SERP de los resultados de búsqueda⁸⁰.

2. Consultas con intencionalidad transaccional

- Las que contienen términos relacionados con películas, canciones, letras de canciones, recetas, imágenes, humor y pornografía
- Las que incluyen términos relacionados con la "obtención" de alguna cosa (por ejemplo, letras de canciones, recetas, etc.)
- Las que incluyen términos como 'descargar' (por ejemplo, software, películas, series, música, aplicaciones, etc.)
- Las que incluyen términos como 'audio', 'imágenes' o 'video'
- Con términos propios del sector del 'entretenimiento' (imágenes, juegos, etc.)
- Con términos relacionados con interacciones (por ejemplo, comprar, chat, etc.)
- Con términos directos como 'películas de [actor/director]', 'canciones de [grupo/cantante]', 'letras de [grupo/cantante]', 'imágenes de', 'multimedia' o palabras clave compuestas por extensiones de archivos o de compresión (jpeg, zip, etc.)

3. Consultas con intencionalidad informacional

- Incluyen signos o palabras de interrogación o pregunta (por ejemplo, 'maneras de "," cómo "," qué es ", etc.)
- Las que utilizan frases del lenguaje natural (en lugar de ecuaciones formales)
- Las que contienen términos que denotan aspectos de organización o presentación informativa (por ejemplo, lista, lista de reproducción, etc.)
- Búsquedas que consisten en un refinamiento de los términos de búsqueda de una consulta anterior

⁷⁹ En el artículo original los Top-Level Domains (TLD) aparecen mencionados como 'Internet domain suffixes'.

⁸⁰ A raíz de la implementación de datos provenientes de bases de datos tipo 'Knowledge Graph' los buscadores muestran en la SERP información gráfica que puede dar respuesta a la necesidad informacional sin necesidad de entrar en ninguna URL específica.

- Consultas efectuadas después de que el usuario haya visualizado múltiples instancias de una SERP
- Consultas con más de 2 términos
- Por exclusión: consultas que no cumplen los criterios navegacionales o transaccionales

A esta tipología básica de búsquedas Bernard et al (2008) aplican una taxonomía jerárquica de tres niveles basada en las características de cada tipo de consulta y que sintetiza los estudios previos.

El cuadro resultante de esta taxonomía es el siguiente:

Hierarchical classification of user intent as expressed by Web queries													
Level	User intent classification												
Level 01	Informational						Navigational			Transactional			
Level 02	Directed	Undirected	Find	List	Advice	Navigation to transactional	Navigation to informational	Online	Obtain	Download	Search engine results page	Interact	
Level 03	Closed	Open							Off-line	Free	Not free	Links	Other

Figura 14. Tabla 2 de Jansen, Booth y Spink (2008)

En los niveles jerárquicos de esta taxonomía encontramos las siguientes características:

- **Nivel 1:** en este nivel encontramos la división principal entre Informacionales, navegacionales o transaccionales.
- **Nivel 2:** las búsquedas informacionales se dividen entre **directas** cuando hay una pregunta específica, y **no directas** cuando son generalistas, o cuando el objetivo es encontrar un lugar del mundo físico, o conseguir listas de ítems o asesoramiento de algún tipo. Las búsquedas navegacionales se clasifican según si su objetivo es transaccional de forma directa o si el objetivo es obtener información previa a una posible transacción. Además, las búsquedas transaccionales pueden ser para 'obtener', para 'descargar' o 'Interactuar' o bien aquellas que tienen como objetivo final la propia SERP al dar solución a un problema de tipo informacional.
- **Nivel 3:** en este nivel, las navegacionales no tienen división. Por su parte, la división de Informacionales consiste en que puede ser abierta o cerrada según si buscan una respuesta concreta o amplia, y las transaccionales pueden ser del tipo off-line, gratis, no gratis, enlaces u otras.

En la tercera tabla de este mismo artículo se define cada uno de los diferentes tipos de búsqueda que hemos traducido y adaptado al castellano dado su interés para el

tema que discutimos en este capítulo. La clasificación de palabras clave es la siguiente:

El nivel uno

- (I) Informativa: consultas destinadas a obtener datos o información con el fin de hacer frente a una necesidad de información, el deseo, o la curiosidad
- (N) de navegación: consultas en busca de una URL específica
- (T) transaccional: consultas en busca de recursos que requieren un paso más para ser útil

El nivel dos

- (I, D) Dirigidas: pregunta específica
- (I, N-D) No dirigidas: "dime todo sobre un tema"
- (I, L) de lista: listado de elementos
- (I, B) Buscar: ubicar el lugar donde se pueden obtener algún servicio o producto del mundo real
- (I, A) Asesoramiento: asesoramiento, ideas, sugerencias, instrucciones
- (N, T) de navegación para una acción transaccional: la URL que el usuario desea es un sitio transaccional
- (N, I) Navegación por una motivación informativa: la URL que el usuario desea es un sitio informativo
- (T, O) Obtener: obtener un recurso u objeto específico
- (T, D) Descargar: encontrar un archivo para descargar
- (T, R) Página de resultados: obtener un recurso que se puede imprimir, guardar o leer desde la página de resultados del motor de búsqueda
- (T, I) Interacción: interactuar con el programa / recurso en otro sitio web

El nivel tres

- (I, D, C) Cerrada: se centra en un solo tema; respuesta unívoca
- (I, D, A) Abierta: se ocupa de dos o más temas
- (T, O, O) en línea u online: el recurso se puede obtener en línea
- (T, O, F) Fuera de línea: el recurso se obtuvo fuera de línea y puede requerir acciones adicionales por parte del usuario
- (T, D, G) gratuito: el archivo descargable es gratis
- (T, D, N) No es gratis: el archivo descargable no es necesariamente libre
- (T, R, E) Enlaces: los recursos aparecen en el título, resumen o el URL de uno o más de los resultados de la SERP
- (T, R, O) Otros: los recursos no aparecen en uno de los resultados, pero sí en otro lugar de la página de resultados del motor de búsqueda

1.4.3. SERPs, Search Engine Results Pages

En el proceso de búsqueda los usuarios obviamente piensan en conceptos, pero para expresar su necesidad de información no pueden usar conceptos de modo directo, ya que por definición es un inobservable. En su lugar utilizan términos concretos (para el ordenador, cadenas de caracteres). Estos términos se denominan palabras clave en la terminología especializada.

Una vez que una o más palabras clave son introducidas en el formulario de búsqueda, la respuesta toma la forma de lo que se denomina una SERP. Típicamente éstas contienen diez resultados (en rigor, diez URL más una breve información adicional denominado *snippet*). El resto de los resultados (si son más de diez) se organizan en sucesivas páginas de SERPs o instancias de la SERP⁸¹. Debido a esto a los resultados de la primera instancia de la SERP se les denomina el Top 10 puesto que son los primeros 10 resultados de una búsqueda (que puede haber arrojado millones de resultados teóricos). Por último, la SERP contiene al final una serie de enlaces, entre ellos los que permiten navegar por el resto de instancias de la SERP, o sea por el resto de resultados

⁸¹ Para hablar de las diferentes páginas de la página de resultados, establecemos que una SERP tiene instancias (páginas), hablando de la primera instancia de la SERP (10 primeros resultados), la segunda instancia (los resultados 11 a 20), etc. De tal manera que una SERP con 1000 resultados tiene 100 instancias (páginas).

proporcionados por la búsqueda más allá del resultado número diez (por tanto, el 11 y sucesivos).

Los buscadores no ordenan los resultados de modo "intelectual" o manual, cosa que haría inviable su modelo de negocio, sino mediante un algoritmo automático (otra cosa es que este algoritmo sea constantemente evaluado y refinado mediante juicios intelectuales de eficacia de sus resultados). Este algoritmo se conoce como cálculo de relevancia, puesto que intenta determinar, para cada consulta, cuáles son los resultados (las páginas o documentos) más relevantes. De ese modo, resulta indiferente que una búsqueda arroje unas pocas decenas de resultados o muchos millones de resultados. Se supone que, gracias al cálculo de relevancia el usuario solamente necesita revisar los N primeros (por ejemplo, los 10 primeros).

De aquí la importancia de aparecer en las primeras instancias de la SERP, ya que la probabilidad de que los usuarios hagan clic en un resultado depende de su visibilidad. De esta forma, la probabilidad de recibir un clic en un enlace posicionado en la primera SERP (Top10) es mucho mayor que la de un enlace posicionado más allá del Top10, y de igual forma, un enlace aparecido en las primeras 5 posiciones del Top 10 tiene una probabilidad mucho más alta de clic que cualquier otro resultado aparecido en la misma SERP.

Típicamente, las SERPs consisten en una combinación de **resultados gratuitos u orgánicos** (es decir, los resultados fruto de la indexación "natural" de la Web por el buscador) y resultados procedentes de la publicidad, compuesta por **enlaces patrocinados**, o sea enlaces por los que el propietario de la página hace un pago para poder aparecer cuando se buscan determinadas palabras clave. Además, puede existir una segunda división que depende del área visible de la SERP, dividiéndose en este caso entre **parte visible y área de desplazamiento**.



Figura 15. Captura de la parte visible de la SERP de la búsqueda 'Doctorado Comunicación Social'

1.4.4. Los resultados orgánicos y los enlaces patrocinados.

Como hemos señalado, los resultados devueltos por los buscadores a cualquier consulta pueden contener no solamente resultados orgánicos sino también un listado de enlaces patrocinados (Jansen y Resnick, 2005, Jansen y Molina, 2006).

Los resultados orgánicos consisten en los enlaces que el algoritmo del motor de búsqueda considera más relevante para la consulta, mientras que los enlaces patrocinados son enlaces que aparecen fruto de la (compra del derecho a) puja para aparecer en la SERP para una determinada palabra clave (Jansen y Resnick, 2005).

El núcleo de resultados de los motores de búsqueda proviene del denominado índice Web (sitios web en general), si bien gracias a la existencia de distintos índices especializados los motores de búsqueda ofrecen también diferentes opciones de búsqueda (como noticias, imágenes, videos, o documentos académicos), que provienen de índices especializados del buscador⁸². Estos índices, que en principio eran considerados secundarios, juegan un papel cada vez más importante para los usuarios.

⁸² Estos segmentos especiales son identificados en el índice de la web y agrupados por fuentes de medios comunes (with same kind of media source), como películas o fotografías. De esta manera algunas consultas de búsqueda pueden recuperar resultados de **varios índices**. Con esta segmentación, los motores de búsqueda ayudan a los usuarios al mostrar de forma automática los resultados de diferentes índices, y en el caso que un usuario esté interesado en una opción de búsqueda específica (imágenes, blogs, noticias) puede elegir directamente la interfaz de búsqueda correspondiente.

En la Figura 15, resultado de una captura de pantalla de la parte visible de la SERP para la búsqueda 'Doctorado Comunicación Social', se pueden apreciar las dos áreas correspondientes a los resultados orgánicos en un recuadro verde, y los resultados patrocinados en un recuadro rojo.

1.4.5. División entre parte visible y área de desplazamiento

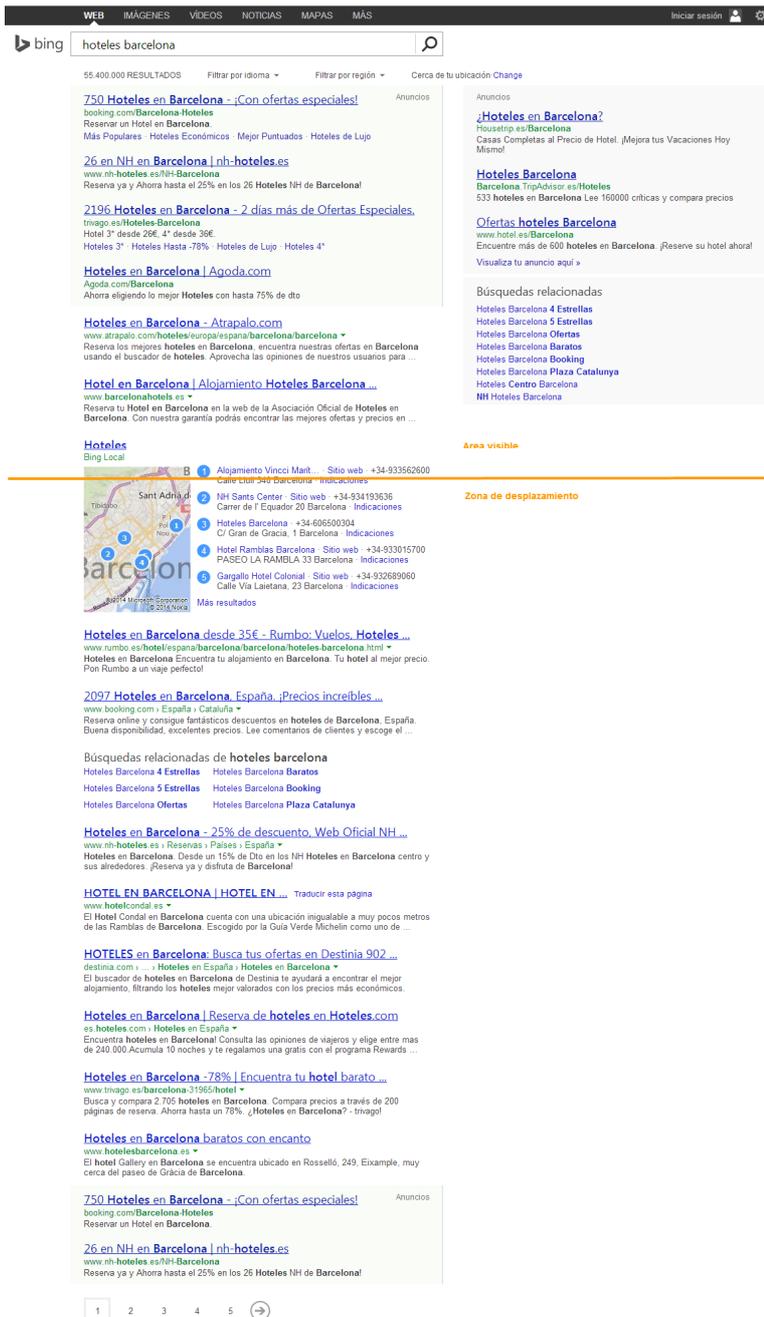
En una SERP se pueden definir diferentes áreas según diversos criterios. En este caso, nos referimos a las dos áreas principales formadas por la **zona visible**⁸³ y el **área de desplazamiento**, separadas por una línea imaginaria que variará en función del tamaño de la pantalla utilizada para la visualización. La zona visible es la que los usuarios pueden ver inmediatamente al abrir una página Web. El resto de información situada más abajo, la que no es visible sin "hacer scroll", es el área de desplazamiento.

A continuación ejemplificamos esta división con una captura de una SERP completa del buscador Bing de Microsoft.

⁸³ El número de resultados que se muestran en la primera pantalla depende de la pantalla y del tamaño de la ventana del navegador, respectivamente.

Página de resultados de Bing – Búsqueda ‘Hoteles barcelona’

La SERP está dividida por una línea imaginaria que delimita las dos principales secciones de la interfaz. Por encima de la línea naranja está la parte de la SERP que el usuario puede ver. Esta área enlaces patrocinados y resultados orgánicos, además de sugerencias de búsqueda y otros accesos directo dependiendo de la consulta.



Área visible

Incluye:

- Cajetín de búsqueda
- Resultados y filtros
- Enlaces patrocinados (parte superior y derecha)
- Enlaces orgánicos
- Búsquedas relacionadas (derecha)

Zona de desplazamiento

Incluye:

- Resultados primarios de búsqueda: Mapa y enlaces
- Resultados orgánicos

Figura 16. Área visible y de desplazamiento de una SERP. Elaboración propia.

De la figura anterior podemos observar que el espacio en la página de resultados es necesariamente limitado. En este caso, podemos ver cuatro anuncios en la parte superior y tres en la parte derecha, mientras que solo aparecen en la zona visible tres resultados orgánicos. En ocasiones, los resultados orgánicos pueden reducirse a dos o a uno solo si alguno de los resultados precedentes (orgánicos o patrocinados) tiene snippets (breves resúmenes de cada URL) enriquecidos, es decir, con información extra como valoraciones o fotografías. Ocurre lo mismo si el buscador en cuestión introduce resultados de otras opciones de búsqueda (p.e. artículos de la Wikipedia o noticias de actualidad) por encima de los resultados orgánicos.

En este sentido, Höchstötter y Lewandowski (2009)⁸⁴, muestran una representación esquemática de las interfaces de resultados de los cuatro principales motores de búsqueda (Google, Bing, Yahoo y Ask.com). Los autores se plantean la cuestión de qué es lo que los usuarios ven de verdad, teniendo en cuenta la mezcla de resultados en los motores de búsqueda y el limitado espacio de pantalla disponible para su visualización.

Los autores representaron en la Figura. 1 de su artículo los elementos principales que componen las SERPs de los buscadores. Los autores explican las principales áreas de las SERPs mismas con ejemplos de las versiones de EE.UU. de los motores de búsqueda. Ambos autores constatan que las páginas de resultados de diferentes países difieren en posicionamiento y que la cantidad de enlaces patrocinados no es la misma para todos los países.

La siguiente figura, de elaboración propia, es una adaptación de la figura mencionada y muestra los principales componentes de las SERPs de los tres buscadores objeto de estudio de esta tesis, **Google**, **Bing** y **Yahoo** a fecha de junio de 2014.

⁸⁴ Nadine Höchstötter y Dirk Lewandowski, *What Users See – Structures in Search Engine Results Pages* (Information Sciences 179, 2009), 1796–1812.

Google muestra primero los enlaces patrocinados (marcados por una etiqueta 'Ad' de color naranja). Las sugerencias de búsqueda (opcionales) aparecen cuando hay una falta de ortografía en la consulta de búsqueda, y preceden a los resultados orgánicos.

Se muestran búsquedas relacionadas tras los resultados orgánicos.

A la **derecha**, sin una separación por una línea vertical, aparecen un máximo de ocho enlaces patrocinados encabezados por una etiqueta 'Ad' de color naranja. En total hay un máximo de once resultados patrocinados por página.

Google.es

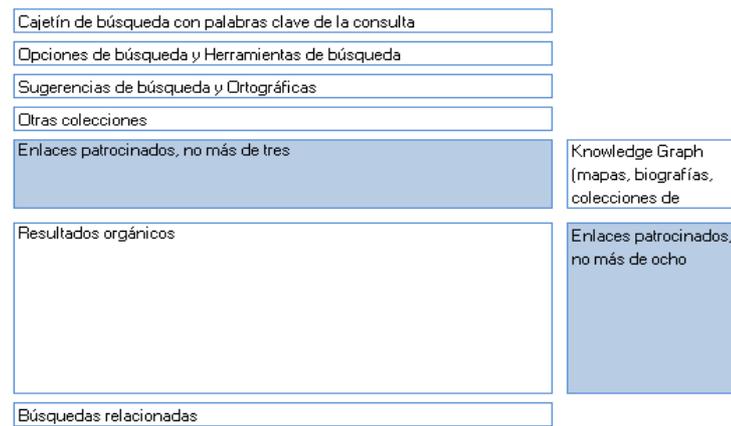


Figura 17. Google: principales componentes de las SERP.

Bing muestra primero enlaces de colecciones específicas (dependiendo del tipo de búsqueda) y sugerencias de corrección si hay faltas de ortografía en la consulta.

Bing.es

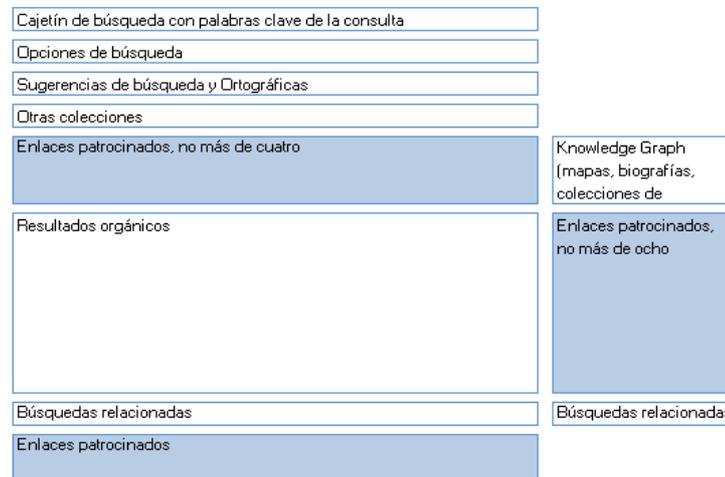


Figura 18. Bing: principales componentes de las SERP.

A continuación los **enlaces patrocinados** (marcados por una etiqueta 'Anuncios' y por una zona sombreada). Los resultados orgánicos ocupan la parte central y tras ellos otro grupo de enlaces patrocinados. En el lado derecho, muestra resultados primarios de búsqueda provenientes de su *Knowledge Graph*, enlaces patrocinados y sugerencias de búsqueda.

Yahoo tiene una pantalla de resultados similar a Bing, sin embargo las opciones de búsqueda aparecen en el lado izquierdo de los resultados. Tras un primer bloque de enlaces patrocinados aparecen los resultados orgánicos.

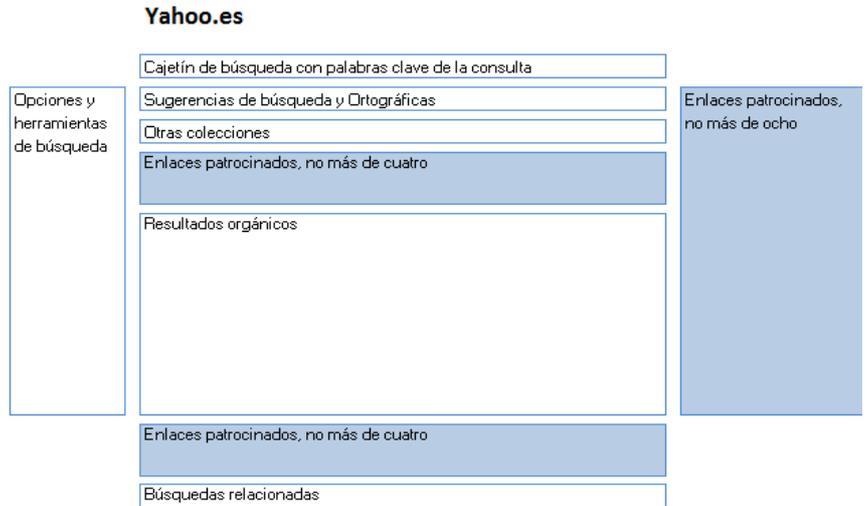


Figura 19. Yahoo: principales componentes de las SERP.

Del lado derecho y por debajo aparecen dos bloques más de enlaces patrocinados. A la derecha un máximo de ocho enlaces patrocinados.

En la tabla 2 del artículo los autores resumen los elementos que se utilizan para componer la SERP. Ambos autores destacan que la composición puede variar en función del tipo de búsqueda introducida.

Table 2: Elements on SERPs

Name	Description	Position
Organic	Results from Web crawl. "Objective hits" not influenced by direct payments.	Central on results page.
Sponsored	Paid results, separated from the organic results list.	Above or below organic results, on the right-hand side of the results list.
Shortcuts	Emphasised result pointing to results from a special collection.	Above organic results, within organic results list.
Primary search result	Extended result that points to different collections. It comes with an image and further information.	Above organic results, often within organic results.
Prefetch	Result from a preferred source, emphasised in the results set.	Above or within organic results.
Snippet	Regular organic result with result description extended by additional navigational links.	Within organic results list (usually first position only).
Child	Second result from the same server with a link to further results from same server.	Within organic results list; indented.

Figura 20. Elementos que se utilizan para componer la SERP.

A veces los resultados se presentan sin o con complementos provenientes de diferentes colecciones de datos como Knowledge Graph u otras opciones de búsqueda tales como imágenes, noticias, blogs o resultados académicos.

La **Figura 21** en una versión actualizada de la tabla de Höchstötter y Lewandowski (2009) en la que aparecen clasificados todos los elementos que pueden aparecer en una SERP en función del tipo de consulta realizado.

Nombre	Descripción	Posición
Orgánico	Resultados con origen en datos provenientes de procesos de crawling propios del buscador.	Cuerpo central de la SERP
Patrocinados	Resultados de pago, separados de los resultados orgánicos	Por encima o por debajo de los resultados orgánicos o en el lado derecho
Atajos	Resultado destacado apuntando a resultados de una colección especial	Por encima de los resultados orgánicos o en la parte superior derecha.
Resultados primarios de búsqueda	Resultados extendidos que apuntan a diferentes colecciones. Se muestra con una imagen e información complementaria	Por encima de los resultados orgánicos o en la parte superior derecha.
Prefetch (Pre-capturados)	Resultados de una fuente predefinida que destaca en el conjunto de resultados	Por encima o entre los resultados orgánicos.
Snippet extendido⁸⁵	Resultado orgánico con una descripción extendida que incluye diferentes enlaces navegacionales adicionales hacia el sitio	Dentro de los resultados orgánicos, normalmente solo para el resultado de primera posición
Hijo	Cuando el buscador encuentra más de un resultado en SERP de un mismo host o servidor, los agrupa de forma consecutiva	Dentro de los resultados orgánicos justo después del primer resultado posicionado del servidor

Los autores concluyen a partir de su análisis, que las SERPs están compuestas por muchos otros elementos además de los resultados orgánicos y enlaces patrocinados.

⁸⁵Los usuarios suelen ver los resultados que vienen con una serie de enlaces profundos del sitio. Esto es el snippet, descripción ampliada del resultado que muestra no sólo los elementos habituales (título, descripción corta, la URL y el tamaño de archivo), sino también los enlaces de navegación adicionales desde la página principal del sitio Web.

1.4.6. Visualización y Evaluación de la SERP

Los autores Chris Barry y Mark Lardner en su artículo *A Study of First Click Behaviour and User Interaction on the Google SERP*, destacan que una característica importante del comportamiento de los usuarios es el hecho que un porcentaje muy bajo visualiza la segunda página de resultados después de introducir una búsqueda en un buscador (Bar-Ilan, Levene, Mat-Hassan, 2006 Richardson, Dominowska and Ragno, 2007). Una vez la SERP aparece en pantalla existe una aparente impaciencia por los usuarios y es posible que refinen la búsqueda si no encuentran resultados relevantes en las posiciones más altas de la SERP. La velocidad para redefinir una búsqueda se incrementa cuando más insatisfechos se muestran con la SERP, y es cuando más tiempo invierten en la consulta (Browne et al, 2007).

En cambio, una vez satisfechos con la SERP, los usuarios adoptan dos tipos de aproximación a los resultados. Los autores Klockner, Wirschum y Jameson (2004) sostienen que los usuarios adoptan entonces uno de estos dos tipos de aproximaciones a los resultados: o bien un enfoque en **profundidad** (depth-first) o bien en **amplitud** (breadth-first):

- **Enfoque en profundidad (depth-first)**. Hace referencia a cuando un usuario evalúa cada una de las entradas de la lista empezando por la parte superior, y decide de forma sistemática si elegir este un enlace o pasar al siguiente de la lista.
- **Enfoque en amplitud (breadth-first)**. Esta aproximación se produce cuando el usuario revisa cierto número de entradas de la SERP y luego vuelve a visitar los enlaces más prometedores con el objetivo de abrirlos⁸⁶.

Por su parte, en un estudio de *eye-tracking* realizado por Klockner et al (2004), el 65% de los participantes utilizó el enfoque en profundidad mientras que el resto de participantes adoptó un enfoque en amplitud total o parcial.

⁸⁶ Las opiniones están divididas en cuanto a si un click en un enlace es representativo como un voto de relevancia. Joachims et al (2005) creen que un clic en un hipervínculo de un usuario en la SERP es signo aprobación de esa página. Por otro lado, otros autores creen que los clics en vínculos particulares no son una indicación de la importancia absoluta, pero sólo indica que los datos se hace clic se percibe como más relevante que otros enlaces que aparecieron en otras posiciones de la SERP arte, pero que no fueron clicados.

Joachims et al. (2005) destacan que debido a la confianza de los usuarios en los enlaces situados en los primeros lugares, éstos reciben muchos más clics que los siguientes enlaces, y también que los dos primeros reciben más atención que el resto de la SERP.

Así mismo, Barry y Lardner (2014) mostraron que los participantes parecían tener una confianza explícita en la mitad superior de la SERP. Los autores citan comentarios como: "... incluso si no estás seguro de lo que estás buscando, asumes que el primer resultado será relevante", "El primero que aparece es por lo general el más relevante " o " yo por lo general hago clic en algún enlace de la parte superior".

Keane, O'Brien y Smyth (2008) mostraron que los usuarios tienen tendencia a elegir los resultados en la parte superior en lugar de los resultados de la parte inferior de una lista, pero esta tendencia no es tan frecuente cuando se percibe la baja relevancia del resultado superior.

Jansen et al (2008) proporcionan algunas ideas útiles sobre la percepción del usuario sobre los enlaces patrocinados y enlaces orgánicos en consultas transaccionales (2007), a saber:

- En consultas de comercio electrónico los enlaces patrocinados son probablemente más relevantes que los vínculos orgánicos
- Para las consultas de comercio electrónico los enlaces patrocinados visualizados en la parte superior de la SERP son más relevantes que los enlaces patrocinados del lado derecho.
- El orden de los enlaces patrocinados no se correlacionan con su relevancia.

1.4.7. Relevancia percibida

Los autores Höchstötter y Lewandowski argumentan que para los buscadores cada vez es más importante producir conjuntos de resultados dentro del **área visible** ("*above the fold*") que sean suficientes para suplir las necesidades informativas de los usuarios, tales como imágenes, resultados de noticias y resultados de vídeo, entre otros.

Un elemento importante de este marco es la satisfacción de los usuarios con los motores de búsqueda, y en un alto grado, la satisfacción de un usuario puede provenir de la **composición de las SERPs**, es decir, del conjunto de todos los elementos presentados en la página, y de manera más concreta, en el área visible donde se encuentran los primeros resultados de búsqueda.

Debido a su modelo de negocio, el objetivo de los buscadores consiste en satisfacer las necesidades de información de los usuarios en base a conseguir mostrar contenidos de la máxima relevancia. Esto lleva a los motores de búsqueda a componer SERPs con diferentes fuentes de datos, lo que hace, a su vez, que las SERPs se conviertan en resultados agregados provenientes de diferentes índices.

En este sentido, Lewandowski (2008) enumera lo que constituiría los tres tipos de contenidos de calidad de acuerdo con sus investigaciones⁸⁷:

1. El grupo principal lo constituyen enlaces hacia aquellas que se consideran **páginas autoridad** en su temática. Pueden ser sitios oficiales que deben ser incluidos en el conjunto de resultados. Existen fuentes de autoridad para determinados temas. Por ejemplo se considerará una página autoridad el portal Softonic para cualquier búsqueda de descargas de software o el portal Medline (medicina en línea) cuando los usuarios busquen usando nombres de enfermedades o síntomas.

2. La segunda posibilidad es la inclusión de resultados directos procedentes de algunas **bases de datos**. Estos resultados son mostrados cuando un usuario introduce una consulta que produce un resultado en esa base de datos específica. Es el caso, por ejemplo, de la información proveniente de Google Maps o Google Places; cuya información proviene de usuarios que se han dado de alta en los servicios de Google. Además de esta información generada por los usuarios Google cuenta con su Knowledge Graph⁸⁸, una compleja base de datos para la gestión del conocimiento con la que Google mejora sus resultados mediante la vinculación de informaciones procedentes de diferentes recursos (p.e. la Wikipedia u otras bases de datos) y que puede combinar entre

⁸⁷ Dirk Lewandowski. *Search engine user behaviour: How can users be guided to quality content?*

⁸⁸ Knowledge Graph en Wikipedia <http://es.wikipedia.org/wiki/Knowledge_Graph>

sí para presentar una ficha unitaria compuesta por datos textuales e icónicos que permiten desambiguar las consultas y aporta informaciones adicionales al estilo de un sistema de respuestas inteligentes, lo cual evite en ocasiones que el usuario necesite hacer clic en los resultados de la búsqueda.

3. La tercera posibilidad es presentar enlaces hacia **bases de datos externas** en lugar de presentar directamente los resultados (como en el caso 2). Por encima de los resultados orgánicos puede aparecer una caja con información específica relacionada y con enlaces hacia esas fuentes de datos que no están conectadas con el buscador. Por ejemplo, estos enlaces externos aparecen cada vez que los usuarios buscan el nombre de una celebridad que puede apuntar a sitios especializados como 'the Internet Movie Database', 'the All Music Guide' o Wikipedia.

En el caso del primer grupo, el que denominamos **páginas autoridad**, existen diversos factores de relevancia que analizaremos en detalle más adelante, y que pueden incluir factores tan diversos como la fecha de los resultados o como la localización del usuario (que un buscador puede determinar por la procedencia del ordenador que hace la consulta).

En este sentido, la relevancia de los resultados para una misma búsqueda puede ser diferente según la provincia o el país desde el que se realiza la consulta. Así por ejemplo, la búsqueda «guerra de independencia» mostrará resultados completamente diferentes según el usuario se localice en México o en España. La razón es que los buscadores entienden que serán irrelevantes para un usuario mexicano los resultados asociados a la guerra de independencia de España contra el Imperio napoleónico, y para un español tendrá una relevancia menor los resultados de la «guerra de independencia» de las antiguas colonias españolas. Aunque esta clase de estimaciones, que forman parte del algoritmo de relevancia, suelen tener utilidad para muchos usuarios, no dejan de formar parte de lo que hemos denominado antes efecto gatekeeper de los buscadores.

De forma similar, el factor de edad de los resultados es de gran importancia en casos como las series de televisión. Los resultados antiguos se consideran resultados no relevantes, ya que puede haber miles de páginas hablando de esa serie pero solo unas pocas que muestren información de la última temporada.

Las diversas actualizaciones del algoritmo de Google han tenido como objetivo mejorar la relevancia de sus resultados teniendo en cuenta éstos y otros factores. En los siguientes capítulos de esta tesis veremos en detalle los últimos algoritmos de Google, los denominados Panda, Penguin y Hummingbird, y cómo han afectado a los resultados de SERP.

1.4.8. CLICKS & visibilidad en SERP

Como hemos expuesto, no todos los resultados de una SERP son visibles de forma directa ya que tanto la publicidad como los servicios de los buscadores que ya hemos mencionado desplazan los resultados orgánicos del área visible, llegando al extremo que en ocasiones solo dos resultados del Top10 algorítmico aparecen en la zona visible. Por lo tanto, sólo aquellos sitios que consigan tráfico en posiciones muy altas del ranking se verán beneficiadas de su visibilidad en SERP través de clics y visitas, mientras que el resto formarán parte del 'long tail' de resultados sobre los que (casi) nadie hace clic.

Los autores Chris Barry y Mark Lardner (2011) muestran cuál es el comportamiento de los usuarios de buscadores en lo relativo en su interacción con las SERPs y el primer clic que realizan. El artículo describe con precisión el proceso de búsqueda cuando el usuario tras a) plantear una o varias palabras clave en un buscador para satisfacer una necesidad de información; b) evalúa la lista de resultados devueltos en la SERP y c) escoge uno de los resultados como enlace a seguir (O'Brien, Keane and Smyth, 2006).

Nadine Höchstötter y Dirk Lewandowski (2009) afirman, que la presentación de los resultados en las SERPs está influida en gran medida por la selección previa de ciertos resultados por parte de los usuarios. De este modo, los resultados orgánicos en las primeras posiciones son los preferidos por los usuarios, pero alrededor de esta lista de resultados principales, se colocan además otros elementos adicionales, como publicidad, con el objetivo de beneficiarse del comportamiento típico de los usuarios en su selección de resultados.

Enlaces orgánicos		Enlaces patrocinados	
Posición 1	100%	Anuncio 1	50%
Posición 2	100%	Anuncio 2	40%
Posición 3	100%	Anuncio 3	30%
Posición 4	85%	Anuncio 4	20%
Posición 5	60%	Anuncio 5	10%
Posición 6	50%	Anuncio 6	10%
Posición 7	50%	Anuncio 7	10%
Posición 8	30%	Anuncio 8	10%
Posición 9	30%	Anuncio 9	10%
Posición 10	20%	Anuncio 10	10%

Figura 22. Probabilidad de clic en SERP (SEO-SEM). Fuente: Eyetools.

Estudios de eye-tracking realizados en 2005 por la empresa Eyetools⁸⁹ demostraron que la mayoría de los usuarios dirige su mirada a un triángulo imaginario denominado el triángulo de oro, que está formado a la izquierda

⁸⁹ Did-it, Enquiro, and Eyetools Uncover Google's Golden Triangle <<http://www.prweb.com/releases/2005/03/prweb213516.htm>>

de los resultados naturales. Por tanto, según su posición en SERP, la probabilidad de que un enlace sea visualizado y clicado es la que aparece en la figura 23.

La probabilidad de click (**CTR**) disminuye en los resultados por debajo de la línea de visibilidad (*below the fold*), es decir, en aquellos que sólo pueden verse cuando el usuario se desplaza por la lista de resultados. Este estudio es corroborado con otros como el informe de iProspect del año 2008 en el que se citaba que el 68% de los usuarios concentraban sus clics en la primera SERP.

En la figura 23 podemos ver cuál es la probabilidad de clic del área visible de una SERP. Como podemos observar, el área visible muestra una alta competencia y la probabilidad de clic es del 47% en la zona de enlaces patrocinados, el 30% corresponde a resultados primarios de búsqueda, mientras que solo el 23% es para los resultados orgánicos.

Diferentes estudios han analizado la probabilidad de clic de los resultados

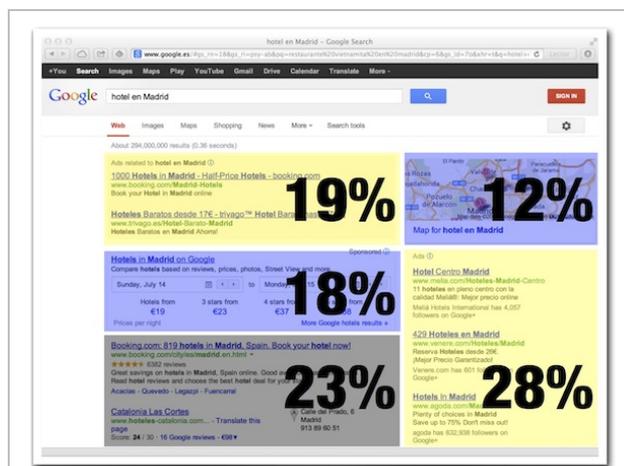


Figura 23. Probabilidad de clic en SERP visible. Fuente: <http://www.enriquedans.com/2013/07/la-preocupante-evolucion-de-google.html>

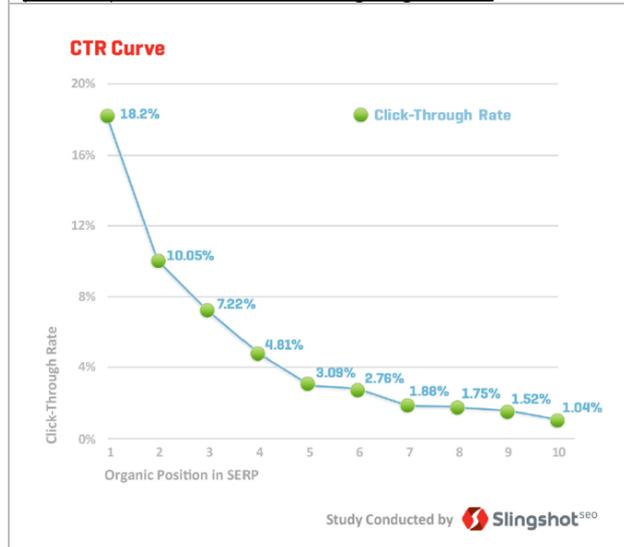


Figura 24. Probabilidad de clic en Top10. Fuente: <http://retelur.wordpress.com/2011/08/02/el-valor-de-aparecer-primero-en-los-resultados-organicos-de-busqueda/>

orgánicos de búsqueda teniendo en cuenta que el 47% del CTR es para los enlaces patrocinados. La agencia de posicionamiento web Slingshot SEO publicó un artículo en el que muestra (Figura 24) la probabilidad de recibir un clic de los resultados orgánicos de búsqueda⁹⁰.

⁹⁰La composición algorítmica del Top10 puede verse alterada por el factor de **click acumulado** o CTR, que se convierte en un input para reevaluaciones que el algoritmo puede hacer para recalculer la posición relativa de los enlaces (ver 200 factores de ranking) de tal manera que un sitio posicionado en una posición baja dentro del Top10 puede llegar a escalar posiciones hasta el Top1 gracias a la interacción de los usuarios.

Aunque las preferencias de clic pueden depender de la tipología de búsqueda, ambas figuras muestran cuál es la distribución efectiva de clic entre los resultados orgánicos y los enlaces patrocinados, lo que justifica la importancia del posicionamiento web en lo concerniente a la visibilidad y al tráfico conseguido a través de la SERP.

1.4.9. Preferencias de clic: resultados orgánicos vs. enlaces patrocinados

La mayoría de los motores de búsqueda tienen un interés personal en los enlaces patrocinados ya que son su principal fuente de ingresos⁹¹ y su modelo de negocio depende en gran medida de ellos. Por esta razón los resultados orgánicos son generalmente bordeados por anuncios contextuales, si bien la probabilidad de clic depende de diferentes factores que pasamos a mencionar.

Como se ha apuntado en el apartado anterior, las preferencias de clic dependen del tipo de búsqueda del usuario como han demostrado los trabajos de diversos investigadores. Así, Barry y Lardner (2014) muestran que en las **búsquedas informativas** el primer clic de todos los participantes fue para la primera posición, observación consistente con el estudio previo de Joachims et al (2005), quien atestiguaba que los primeros enlaces de SERP recibían mucho más clics que el siguiente enlace de la SERP, y que los usuarios prestaban mucha más atención al primer y segundo enlace de la SERP. Respecto a las **búsquedas transaccionales**, tanto Joachims et al (2005) como Barry y Lardner (2014) atestiguan que el 95% de todos los primeros clics ocurrió entre el primer enlace patrocinado y el tercer enlace orgánico, todos ellos dentro del área visible de la SERP.

Esta tendencia a las posiciones Top aparece en otros estudios (Richardson et al, 2007 Keane et al, 2008) y parece generalmente aceptado que los usuarios depositan más confianza en los dos primeros enlaces (Joachims et al 2005). Por tanto, los usuarios que utilizan una aproximación en **profundidad** harán su primer clic en la parte superior de la SERP, mientras que los sitios menos populares que no posicionan en lo alto de la SERP recibirán menos clics⁹².

Barry y Lardner (2014) incluyen comentarios de los entrevistados que muestran la tendencia a gravitar los sitios más altos en SERP *"I usually just go to the one on the top of the page...if it doesn't suit me I just go next one down"*, y la falta de paciencia en sus hábitos de búsqueda: *"If I don't get it, at a push, in the first five links, more often than not I just give up, I just go and type in something else"*. Este hecho reafirma la idea

⁹¹Los ingresos de Google en las relaciones con los inversores de Q1 de 2008 muestran que el porcentaje de los ingresos totales de los ingresos por publicidad es de 99%.

⁹²De este hecho se desprende que estos sitios deben mejorar su estrategia SEO para justificar y demostrar a Google a través de KPIs como el rebote o la estancia media en el sitio que merecen ser posicionados en una posición más alta.

comúnmente aceptada de que únicamente lleva un segundo hacer un juicio de valor/relevancia (Szetela, 2008).

La valoración de los resultados orgánicos frente a los enlaces patrocinados también ha sido objeto de estudio. El patrón general es que el primer enlace patrocinado recibe cuatro veces más primeros clicks que el enlace siguiente, y el primer enlace orgánico recibe cinco más primeros clicks que el siguiente enlace orgánico.

Cuando a los usuarios se les ha mostrado para examinar ambos listados, orgánicos y patrocinados, los enlaces orgánicos son percibidos como más relevantes (Jansen, Brown y Resnick, 2007), incluso si los enlaces patrocinados se presentan como igual de relevantes que los enlaces orgánicos (Jansen y Resnick, 2006). Laffey también muestra que en el caso específico del buscador Google, los usuarios son más propensos a hacer clic en enlaces orgánicos, en lugar de enlaces patrocinados (2007). Los usuarios también han manifestado reacciones emocionales negativas a los anuncios de búsqueda patrocinados (Marable, 2003).

Investigaciones como la de Richardson et al. (2007) o Barry y Charleton (2008) hacen hincapié en la importancia de aparecer en el ranking de anuncios patrocinados. Estos dos últimos autores abogan por que las estrategias de SEM (por Search Engine Marketing, es decir la búsqueda de posicionamiento basada en la adquisición de anuncios) son los métodos predominantes que las empresas utilizan para promover su visibilidad y que el objetivo principal de una estrategia de SEM es asegurar el sitio del aparezca cerca de la parte superior de la primera SERP, mientras que Barry y Lardner (2014) defienden que los hallazgos de sus estudios demuestran que las empresas están en desventaja si su enlace no aparece en el área visible de la SERP.

Barry y Lardner (2014) concluyen que los especialistas SEO deben asegurarse que los usuarios que clican en el enlace de una SERP son recibidos con una *landing page* que supla las necesidades del usuario y que se corresponda con la aparente relevancia del enlace de la SERP. Se considera un factor crítico en tanto que el usuario puede volver a la SERP y elegir otro enlace rival del Top10 (Todd, 2006, Szetela, 2008).

Respecto a los enlaces patrocinados, se aconseja a las empresas conseguir una posición alta en enlaces patrocinados como el objetivo principal de su estrategia de SEM. Las estrategias de SEM deben trabajar bajo la suposición de que los usuarios al evaluar la SERP utilizan el enfoque en profundidad, y las estrategias de SEM deberán

tomar todas las medidas necesarias (en la optimización del sitio y las campañas de búsqueda de pago) para asegurarse de que su sitio aparece y se mantiene en una posición de alta en la SERP con el fin de aumentar el tráfico a su sitio.

De estas observaciones se desprende que si se pretende alcanzar visibilidad y tráfico, tanto en resultados orgánicos como patrocinados, las estrategias de posicionamiento orgánico y de enlaces patrocinados deben estar focalizadas en colocar sus contenidos en las posiciones más altas de la primera SERP, de lo contrario, la visibilidad lograda no se transformará en tráfico efectivo.

1.4.10. Factores de ranking básicos

Partiendo del conjunto de conceptos vistos hasta ahora parece no sólo interesante, sino indispensable saber qué criterios utilizan los motores para calcular la relevancia de una página web, de ahí que para la comprensión del desarrollo expositivo de la tesis, se opta por aportar algunas nociones básicas en torno al posicionamiento web, nociones que irán complementándose a lo largo del resto de la tesis.

Codina (2004) recuerda que existe un cierto consenso entre los analistas sobre qué clase de criterios usan los motores de búsqueda para ordenar los resultados. En líneas generales, los principales factores de posicionamiento se pueden dividir en dos grupos, dependiendo si están en la misma página y, por tanto, bajo el pleno control del propietario del sitio (**on the page**) o externos a la página y, por tanto, relativamente alejados del control del propietario de la página (**off the page**). Dentro de cada uno de estos criterios generales encontramos diversos factores de posicionamiento.

A continuación veremos cuáles son los básicos y principales, aunque en los siguientes capítulos de esta tesis veremos que existen más de 200 factores de ranking a tener en cuenta para posicionar un sitio web.

Los criterios o factores **on the page**, intrínsecos a la página web, son⁹³:

- **Optimización de palabras clave.** La selección óptima de las palabras clave es la base de toda estrategia SEO, por lo que trataremos este factor en detalle en el siguiente apartado.
- **Optimización de los títulos.** Dado que el contenido que se encuentra situado en el elemento <title>título</title> de cada página es lo que los buscadores muestran en la lista de resultados, el objetivo de la optimización de este metadato es doble. Por un lado debe ser un reclamo para que los usuarios entren en la web, y por otro debe estar configurado de tal forma que los buscadores otorguen una buena posición a la web. Para alcanzar buenos rankings de relevancia se aconseja redactar títulos de entre 5 y 10 palabras en los que se mencione por lo menos una vez las palabras clave que optimizan la web.
- **Las meta etiquetas.** Junto con <title> hay otra serie de metadatos que se pueden incluir en la sección <head> de las páginas web y que tienen como objetivo ofrecer a los buscadores información complementaria acerca de la página. Los más importantes son los que forman parte de las etiquetas denominadas: META DESCRIPTION (describe el contenido de la página y sirve de descripción en los resultados de algunos buscadores), META KEYWORDS (actualmente casi todos los buscadores la ignoran debido a su manipulación para conseguir mayor relevancia), META LANGUAGE (indica el idioma de la página) y la etiqueta META ROBOTS (indica al buscador si se desea indexar la página y/o se desean seguir los links).
- **Palabras clave secundarias.** Las palabras clave secundarias son aquellas que complementan el sentido temático o semántico de una página, y se deben distribuir correctamente en el texto de la misma. Se recomienda una densidad (porcentaje de palabras clave sobre el total de palabras) de entre el 5% y el 8%. A su vez, según el principio del *keyword proximity* se recomienda situarlas lo más cerca posible del principio de la página. Para darles más importancia existen varias etiquetas: <Hn> (donde n es un número del 1 al 6), <i>, , y . Por su parte, la palabra clave principal se resalta con un <H1> y debe colocarse cerca del principio de la página.

⁹³Carlos Gonzalo (2004). *La selección de palabras clave para el posicionamiento en buscadores* <http://www.upf.edu/hipertextnet/numero-2/palabras_clave.html>

Los criterios o factores **off the page** son:

- **Los enlaces externos.** Dado que los motores de búsqueda usan medidas de popularidad de enlaces (cantidad de sitios "relacionados" o "autoritativos" que enlazan a una determinada web) para determinar los rangos de los resultados de las búsquedas, se recomienda la implementación de estrategias de para conseguir el máximo número de enlaces de entrada procedentes de sitios de buena reputación. Actualmente, la estrategia principal para esto pasa por la creación de contenidos de calidad.
- **El texto de los enlaces.** El texto de los enlaces que apuntan a una página es considerado por algunos como el principal recurso uno de la optimización. Las palabras clave en este caso se deben colocar en el texto que actúa como texto de anclaje de la etiqueta de un enlace. No obstante, los algoritmos actuales de los buscadores han relativizado el valor de los textos de los enlaces como veremos en siguientes apartados de esta tesis.
- **El PageRank (PR).** Es un valor entre 1 y 10 que depende de la cantidad y calidad de las webs que tengan enlaces hacia la web de referencia, así como de sus enlaces internos. El PR transmitido por las webs depende a su vez del PR propio y del número de enlaces salientes que tenga esa página [2]. La fórmula del PR es la siguiente: $PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$, donde **PR(A)** es el PageRank de la página de referencia, **d** es un factor de debilitación, por lo que **(1-d)** asegura que cualquier página, aunque no reciba ningún enlace, tendrá un PR mínimo de 0'15. **PR(Ti)/C(Ti)** es el PageRank (PR) de la página i-ésima que enlaza a la web de referencia, **(Ti)**, dividido por el número de enlaces **(C)** que también salen de esa página Ti, es decir, el PR que transmite. **i = 1, ..., n** ya que se suponen **n** páginas que enlacen a la de referencia.

En los siguientes apartados veremos que éstos son únicamente algunos de los factores básicos tenidos en cuenta por los buscadores, pero que la lista se extiende hasta los 200 factores. El peso de estos factores ha variado a lo largo del tiempo, y veremos cuáles son a día de hoy los factores más importantes a través del repaso de algoritmos de Google más recientes.

1.4.11. Posicionamiento natural vs planificado

En *Posicionamiento web: conceptos y ciclo de vida* (Codina, 2004)⁹⁴, se exponen las bases conceptuales del posicionamiento web como campo de estudio vinculado a la Biblioteconomía y la Documentación. El autor apunta los siguientes conceptos básicos en torno al posicionamiento:

1. **Posicionamiento natural:** el posicionamiento que consigue una página o un sitio de modo espontáneo, es decir, sin que sea consecuencia de una campaña consciente o planificada.
2. **Posicionamiento planificado:** el posicionamiento que consigue una página o un sitio web debido a una campaña consciente y planificada. El posicionamiento planificado puede ser ético o fraudulento:

A. Posicionamiento planificado ético: debe consistir en conseguir los mismos resultados que se obtendrían con un posicionamiento natural en una situación ideal en la cual los motores de búsqueda fueran exhaustivos y omniscientes y en la cual los responsables de sitios webs diseñaran espontáneamente sitios webs con visibilidad óptima para los buscadores.

B. Posicionamiento planificado fraudulento: el posicionamiento que consigue una página o un sitio como consecuencia de actividades deliberadas de engaño por parte del responsable de la página web actividades ILICITAS dirigidas conscientemente a forzar los resultados de los buscadores.

Dentro de este posicionamiento planificado fraudulento es donde localizamos las técnicas de posicionamiento también conocidas como spam o SEO *Black-hat*. Estas técnicas permiten incrementar artificialmente la relevancia de los sitios web en consultas para las que esa página no es relevante, consiguiendo posicionar algorítmicamente los contenidos en el Top10 a través de la manipulación de diferentes factores de ranking. En la lucha por el posicionamiento en SERP algunos SEO utilizan estas técnicas spam que incluyen diferentes acciones encaminadas a aumentar artificialmente el ranking, técnicas que han ido evolucionando a lo largo de los años y que han sido combatidas por numerosas actualizaciones de algoritmos anti-spam como Google Panda y Google Penguin que analizaremos un poco más adelante en los siguientes apartados de esta tesis.

⁹⁴ Codina, Lluís (2004). *Posicionamiento Web: Conceptos y Ciclo de Vida*. <http://www.upf.edu/hipertextnet/numero-2/posicion_web.html>

1.5. Anatomía del buscador Google, 1998-2014

Han pasado 16 años desde que en 1998 Sergey Brin y Lawrence Page publicasen su artículo *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Desde entonces Google ha evolucionado⁹⁵ mucho, tanto como para convertirse en uno de los actores más influyentes del ecosistema Internet y con un mayor potencial de riesgo global en caso de una utilización inadecuada, bien fuese por parte de la propia empresa Google o por terceros con capacidad de manipular los resultados del buscador.

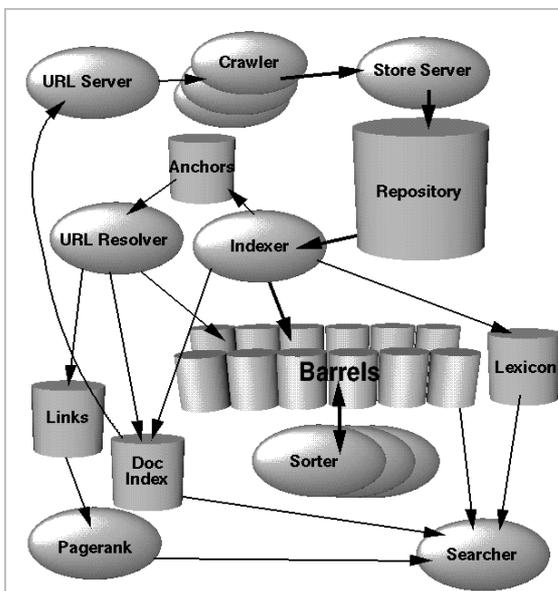


Figura 25. *The Anatomy of a Large-Scale Hypertextual Web Search Engine* Larry Page y Sergei Brin. En la URL original del artículo hoy encontramos un nuevo artículo con una versión actualizada del diagrama, una evolución del artículo original que no podemos representar aquí por su tamaño.



Figura 26. En comparación con la figura anterior mostramos un diagrama de MOZ que incluye los nuevos factores de ranking aparecidos hasta día de hoy. La etiqueta 'ALGO' de la figura se refiere al último algoritmo de Google, Hummingbird.

En concreto, en este apartado de la tesis analizaremos la posibilidad de manipular los resultados de búsqueda de Google a través de técnicas de posicionamiento natural, que cuando entran en prácticas claramente fraudulentas suelen ser denominadas *spam*. Se consideran técnicas fraudulentas o de *spam* cuando el objetivo de las mismas se centra directamente en influir los resultados del buscador en lugar de mejorar la calidad del sitio, la de sus contenidos o la experiencia del usuario. En estos

⁹⁵ The Anatomy of a Large-Scale Hypertextual Web Search Engine <<http://infolab.stanford.edu/~backrub/google.htm>>

casos, tales técnicas aprovechan las debilidades detectadas por expertos en los algoritmos de buscadores y son susceptibles de conseguir un posicionamiento que no se corresponde ni con la relevancia real de los contenidos ni con la credibilidad de la fuente.

Para mostrar la posibilidad de manipular los resultados de búsqueda analizaremos resultados reales de buscadores de seis elecciones, pero previamente debemos mostrar cómo funciona un buscador a día de hoy y cuáles son las técnicas y factores de posicionamiento web.

El conocimiento exhaustivo del buscador y los datos arrojados por tendencias y hábitos de los usuarios, permiten posicionar contenidos individuales o grupos de contenidos, llegando a la posibilidad de planificar de antemano estrategias de contenidos que permitan alcanzar altos grados de visibilidad ya sea con fines y/o medios lícitos o ilícitos.

Vamos a analizar la anatomía del buscador Google y desgranar sus funcionalidades, interfaces, y factores de ranking, desde lo más básico hasta lo más actual pasando por todos los algoritmos de Google desde el año 2009. Este repaso histórico tiene como objetivo finalizar la exposición con las técnicas de posicionamiento, algunas de ellas consideradas SPAM, y que permiten manipular los resultados de las SERPs y advertir con ello de que su potencial manipulación mediática es una realidad a la que se enfrenta la sociedad del conocimiento del siglo XXI.



Figura 27. Evolución de los algoritmos de Google durante el periodo 2009-2013
Fuente: <http://blog.odei.es/google-hummingbird-nuevo-algoritmo-motor-busqueda/>

>>Componentes del buscador Google

La principal función de un buscador es acceder a sitios web, localizar y descargar documentos de los que extrae el contenido textual y multimedia. El contenido de estos documentos descargados es analizado e indexado para construir los índices del motor que permiten la recuperación de los documentos a través de la formulación de consultas mediante palabras clave.

Posteriormente, para facilitar la ordenación de los resultados de una consulta, se realiza el análisis de enlaces de cada página, así como de diversos factores con el fin de otorgar a cada página alguna medida de relevancia que permita ordenarlos conforme a tales criterios.

Para realizar lo anteriormente expuesto, los motores de búsqueda se componen de las siguientes partes que veremos en detalle:

1. Rastreador, robot, crawler o Spider.
2. Sistema de recuperación de información:
 - Índices directos
 - Índices invertidos
3. Interfaz de consulta:
 - Formulario de consulta
 - Página de resultados o Search Engine Result Pages (SERP)
4. Algoritmo de relevancia u ordenación.

>>>Rastreador, robot, crawler o Spider

Un robot, *spider* (araña) o *crawler* (rastreador) es un programa que explora la Web de forma sistemática con dos **objetivos** principales:

- 1) Interactuar con los servidores de sitios web para descargar páginas web u otros documentos de acuerdo con el archivo robots.txt de cada sitio web.
- 2) Obtener nuevas direcciones (URL) para añadir a su lista de enlaces pendientes de revisar.

El rastreador, o sistema de exploración, parte de una lista de URL o direcciones de páginas web denominadas "simientes" (seed) proporcionada a su vez por un servidor de URL (URLServer). Los nuevos contenidos detectados, en su mayoría páginas web, pero también documentos ofimáticos, presentaciones, archivos de imágenes, vídeos, etc., son descargados en el almacén o repositorio de documentos del buscador. A partir de las páginas descargadas, el rastreador extrae nuevas URL que añade a su lista de URL pendientes de analizar.

>>>Sistema de recuperación de información

El sistema de recuperación de información se compone a su vez dos elementos:

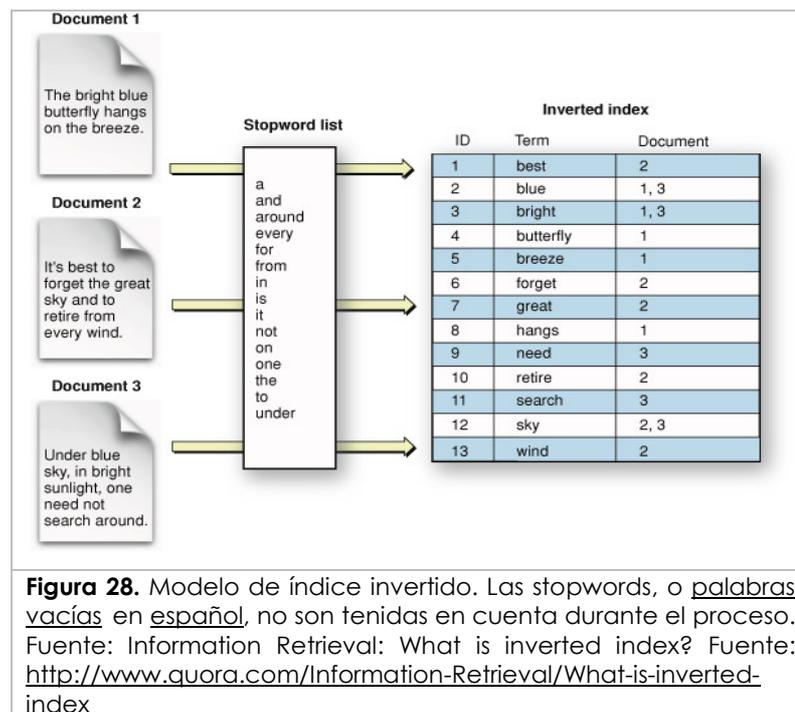
- **Sistema de indización:** su función es la del análisis de los documentos descargados de la web y de la creación de los índices que permitan luego realizar las consultas al motor de búsqueda (MDB).
- **Sistema de consulta:** es la interfaz visible del MDB con la que interactúan los usuarios

>>>Sistema de indización

Los MDB utilizan al menos dos clases de índices, denominados respectivamente índices directos e índices invertidos. Las consultas las resuelven los índices invertidos, mientras que los índices directos se utilizan como elementos de gestión y control internos.

- **Índices directos.** En un índice directo se compone de una lista de documentos (o de registros) en un orden cronológico o numérico.
- **Índices invertidos.** La estructura de un índice invertido consiste en una lista ordenada de términos o *tokens* correspondientes a formas individuales de palabra que aparecen en los distintos documentos, y junto a cada *token* una correspondencia a los documentos concretos en los que aparecen. Si no posee más información, se considera un índice booleano en tanto que sólo puede determinar si una palabra existe dentro de un documento en particular, sin almacenar información con respecto a la frecuencia y la posición de la palabra. Este índice determina que documentos coinciden con una consulta pero no clasifica los documentos coincidentes. En otros diseños de índices invertidos, como los que utilizan los sistemas de información documentales de Google, se incluye

información adicional, como la frecuencia de cada palabra en cada documento o las posiciones de una palabra en cada documento. La información de posición permite que el algoritmo de búsqueda identifique palabras por proximidad para apoyar la búsqueda de frases. La frecuencia se utiliza para calcular la pertinencia de los documentos para la consulta. La estructura típica de un índice invertido, por tanto es la siguiente:



El índice invertido puede crecer mediante técnicas reconstrucción o de fusión. La reconstrucción es similar a la fusión, pero elimina primero el contenido del índice invertido original. La fusión añade documentos recientemente indexados, por lo en general reside en la memoria virtual, mientras que el caché de índice se alberga en los discos duros de uno o más servidores. Esta arquitectura del índice diseñada para la indexación incremental, identifica el documento o los documentos que se agregaron o actualizaron y luego disecciona cada documento en tokens.

Después del análisis, la referencia de documento se añade a la lista de documentos correspondientes a cada término o token⁹⁶, donde se almacenan las palabras, la posición en el documento, el tamaño de fuente y mayúsculas. Estos documentos están distribuidos en diversos servidores (*barrels*).

⁹⁶En los grandes motores de búsquedas, el proceso de localización de cada token en el índice invertido se suele dividir en dos partes, el desarrollo de un *forward index* (índice hacia adelante) junto a un proceso que localiza el contenido de este segundo índice en el índice invertido.

>>>Sistema de análisis de los documentos

El proceso de análisis o *parsing* descompone los documentos en palabras o *tokens* para su inserción en ambos índices, tanto el invertido como el *forward index*. En procesos de indexación las palabras encontradas son llamadas tokens, por lo que en el ámbito del procesamiento del lenguaje natural, este análisis se conoce más comúnmente como "tokenización"⁹⁷.

A partir de 2006 el procesamiento del lenguaje natural ha sido objeto de continua investigación y mejora tecnológica. Los procesos de tokenization involucran múltiples tecnologías, que incluyen los siguientes procesos:

>>>>Tokenización. Durante la tokenización, el analizador identifica secuencias de caracteres que representan palabras y otros elementos, tales como puntuación, que se representan por códigos numéricos, algunos de los cuales son los caracteres de control no imprimibles. El analizador también puede identificar entidades tales como direcciones de correo electrónico, números de teléfono y direcciones URL. En la identificación de cada token, se pueden almacenar varias características, por ejemplo, el idioma o la codificación, la categoría léxica (parte de la oración, como "sustantivo" o "verbo"), la posición, el número de frase, la posición en la frase, la longitud y el número de línea.

>>>>Reconocimiento de idioma. En los buscadores que dan servicio en varios idiomas, un paso inicial común durante la tokenización es identificar el idioma del documento; muchos de los pasos posteriores son dependientes del idioma (como el etiquetado de lenguaje natural o el *stemming*). El reconocimiento de lenguaje es el proceso por el cual un software identifica de forma automática y categoriza el idioma de un documento⁹⁸.

>>>>Análisis de formato. Los motores de búsqueda tokenizan diversos formatos de documentos que contienen a su vez información de formato además de contenido textual. Por ejemplo, los documentos web contienen etiquetas del lenguaje HTML, que especifican información de formato, como si

⁹⁷ Este proceso también se conoce como desambiguación, etiquetado, segmentación de textos, análisis de contenido, análisis de textos, minería de texto, generación de concordancias, segmentación del discurso, léxico o análisis léxico. Los términos "indexación", "análisis" y "tokenización" se utilizan indistintamente en la jerga corporativa.

⁹⁸ Otros nombres para el reconocimiento del lenguaje incluyen clasificación de idioma, el análisis del lenguaje, la identificación del lenguaje, y el lenguaje de marcado.

una frase es un párrafo o un titular, énfasis en las negrita y tamaño de la fuente o el estilo. El análisis de formato es la identificación y el manejo de los contenidos de formato incrustados dentro de los documentos que definen cómo el documento en cuestión es interpretado por un programa de software o cómo se representa en una pantalla de ordenador.

El análisis de formato también se refiere a otros procesos como el análisis de la estructura, el análisis de formato, la extracción de etiquetas, el formato de extracción, la normalización de texto, la limpieza de texto, y la preparación de texto.

Los formatos de archivo más comunes que muchos motores de búsqueda incluyen son: HTML, archivos de texto ASCII (un documento de texto sin un formato específico para ser legible por ordenador), formato de documento Adobe (PDF), PostScript (PS), LaTeX, XML y derivados como RSS, SGML, formatos de metadatos Multimedia como ID3, Microsoft Word, Microsoft Excel, Microsoft PowerPoint o IBM Lotus Notes. Algunos motores de búsqueda soportan archivos en formato comprimido o cifrado como ZIP o RAR. Cuando se trabaja con un formato comprimido, el indexador descomprime primero el documento; de este paso pueden resultar uno o más archivos, cada uno de los cuales debe ser indexados por separado.

>>>>Reconocimiento de sección. Con anterioridad a la tokenización, algunos motores de búsqueda incorporan reconocimiento de sección, es decir, la identificación de las principales partes de un documento. No todos los documentos en un corpus se leen como un libro bien escrito, dividido en capítulos y páginas organizadas. Muchos de los documentos en la web, tales como boletines de noticias e informes corporativos, pueden incorporar contenido mal etiquetado y secciones laterales que no contienen el material primario de que trata el documento (por ejemplo, artículo con menús laterales y con enlaces a otras páginas web, o contenido formateado en las columnas). Sin el marcado de formato las diferentes áreas de la vista del contenido se pueden almacenar de forma secuencial, de tal forma que con el reconocimiento de secciones las palabras que aparecen de forma secuencial en el contenido de origen en bruto se indexan de forma secuencial a pesar de que estas frases y párrafos puedan aparecer en diferentes partes de la pantalla del ordenador.

>>>>Sistema de prioridades de HTML. Durante el proceso de indexación se utilizan también las etiquetas HTML para otorgar importancia o prioridad a determinadas partes del contenido basándose en el uso de etiquetas como ``, de forma que aunque un texto no aparezca al inicio del documento, puede ser reconocido como relevante.

>>>>Indexación de Meta-tags. Algunos documentos y las páginas web en especial, a menudo contienen meta información incrustada, como información sobre el autor, o palabras clave y descripciones que ayudan a un buscador a interpretar el contenido semántico del documento. También pueden incluir metadatos sobre aspectos como la lengua y el código de caracteres utilizado.

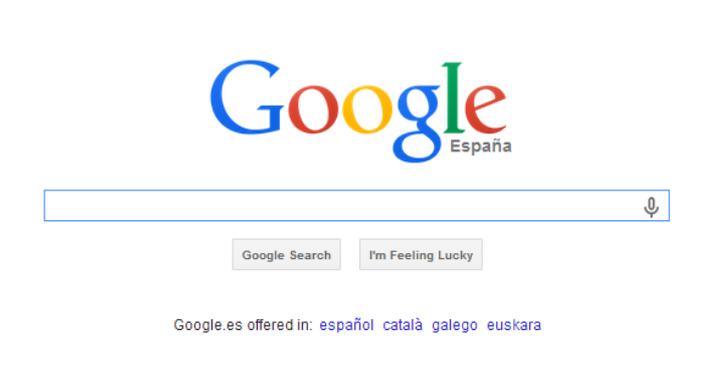
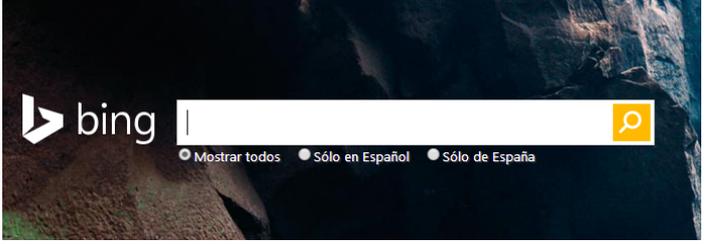
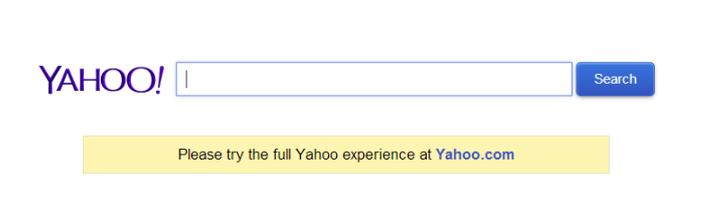
En el caso concreto de las páginas HTML, suele utilizar la etiqueta `<meta>` a tales fines. En los inicios de la tecnología de motores de búsqueda el documento completo no se indexaba y el índice invertido se componía sólo de estas palabras clave aparecidas en las meta etiquetas.

A medida que Internet creció a través de la década de 1990, muchas empresas establecieron sitios web corporativos. Entonces, las palabras clave utilizadas para describir las páginas web a través de la etiqueta `<meta>` cambiaron de descriptivas de modo objetivo a estar orientadas al marketing con el objetivo de impulsar las ventas mediante el posicionamiento de la página web en los resultados de búsqueda. La acumulación subjetiva de estas palabras clave fueron uno de los orígenes del spam, lo que llevó a muchos motores de búsqueda a adoptar tecnologías de indexación de texto completo en la década de 1990 por un lado y a dejar de considerar el contenido de algunas etiquetas `<meta>`, en particular, la etiqueta `<meta>` con la propiedad (*name*) `keyword`.

>>>Interfaz de consulta

Los motores de búsqueda se utilizan mediante **palabras clave** que representan la necesidad de información del usuario. Normalmente presentan dos opciones de búsqueda, simple y avanzada.

El formulario de búsqueda sirve para enviar la consulta al motor de búsqueda que compara los términos de la pregunta con el índice invertido y filtra de este modo las páginas web que contienen los términos de la pregunta. Después, interviene el algoritmo que calcula la relevancia de cada página recuperada para ordenar los resultados de la forma más clara y eficiente posible a través de la SERP del motor de búsqueda.

	<p>Figura 29. Formulario de búsqueda simple de Google. Muestra los botones "Google Search" y "I'm feeling Lucky". El cajetín incluye un micrófono, la opción de búsqueda por voz a partir del lanzamiento de Google Hummingbird</p>
	<p>Figura 30. Formulario de búsqueda simple de Bing. Muestra una lupa como botón de búsqueda y las opciones de búsqueda web, por idioma o nacional.</p>
	<p>Figura 31. Formulario de búsqueda simple de Yahoo. Muestra el botón de búsqueda y un enlace al portal nacional del buscador.</p>

Además de la opción de búsqueda simple encontramos la avanzada que permite enviar búsquedas parametrizadas y usar de forma explícita operadores booleanos para expresar de modo mucho más fiel la necesidad de información del usuario.

De este modo, los diferentes campos de la interfaz de búsqueda avanzada nos permiten realizar consultas simples, por frase, con palabras negativas, por idioma, región, fecha de publicación, dominio, nivel de lectura o formato. Cada uno de estos campos está asociado a un operador de búsqueda avanzada, palabras o símbolos que se pueden utilizar en las consultas desde el cajetín de búsqueda simple y que permiten encontrar lo que el usuario busca de una manera rápida y precisa.

Búsqueda avanzada

Buscar páginas con... Haz lo siguiente en el cuadro de búsqueda

todas estas palabras: Escribe las palabras importantes: terrier ratonero tricolor

esta palabra o frase exactas: Escribe las palabras exactas entre comillas: "terrier ratonero"

cualquiera de estas palabras: Escribe OR entre todas las palabras que quieras: minibus OR estándar

ninguna de estas palabras: Añade un signo menos delante de las palabras que no quieras que aparezcan: -roedor, -"Jack Russell"

números desde el: hasta Escribe dos puntos seguidos entre los números y añade una unidad de medida: 10..35 kg, 300..500 euros, 2010..2011

A continuación, limitar los resultados por...

idioma: cualquier idioma Busca páginas en el idioma que selecciones.

región: cualquier región Busca páginas publicadas en una región determinada.

última actualización: en cualquier momento Busca páginas actualizadas en el período de tiempo especificado.

sitio o dominio: Busca un sitio (como wikipedia.org) o limita los resultados a un dominio como, por ejemplo, .edu, .org o .gov.

los términos que aparecen: en cualquier lugar de la página Busca términos en toda la página, en el título de la página o en la dirección web, o enlaces a la página que estás buscando.

SafeSearch: Mostrar los resultados más relevantes Indica a SafeSearch si quieres que filtre contenido sexualmente explícito.

tipo de archivo: cualquier formato Busca páginas en el formato que prefieras.

derechos de uso: sin filtrar por licencia Busca páginas que puedas utilizar libremente.

Búsqueda avanzada

Figura 32. Búsqueda avanzada de Google

>>> Algoritmo de cálculo de relevancia u ordenación

Su función es poder presentar los resultados bajo algún criterio de relevancia, una opción imprescindible cuando pueden aparecer miles de resultados (o millones) como respuesta a una consulta. Para entender esta necesidad, que es a la vez, lo que otorga el mayor riesgo de manipulación por parte de un buscador, hay que pensar (siguiendo a Codina, 2004) en lo siguiente: en un contexto donde cualquier búsqueda arroja un número muy pequeño (o un único resultado) no haría falta ningún cálculo de relevancia. Una simple inspección de los cinco o seis resultados ofrecidos (por decir alguna cosa) daría al usuario la posibilidad de elegir el más relevante para él sin que nadie usurpara esa función. En cambio, si los resultados superan el centenar o más en

casi cada consulta, un sistema de información que ofreciera los resultados sin ordenar (o sin ordenar por una función de relevancia, por ejemplo, por orden alfabético o cronológico) sería totalmente inútil. Si la fórmula que indica el número de ítems que en promedio debería examinar un usuario es igual a $N/2$, siendo N el número de resultados obtenidos (por ejemplo, para encontrar un resultado relevante entre 10 mil, debería examinar 5 mil resultados). Para peor, si pudiera haber más de un resultado relevante, entonces la fórmula es igual a N (si obtiene 10 mil resultados, deberá examinar 10 resultados).

Es por ello que la SERP debe presentar la página de resultados con los documentos ordenados según su grado de relevancia. Esta ordenación, puesto que produce un ranking se denomina también así.

Como hemos avanzado en secciones anteriores, en el caso de las páginas web, los motores de búsqueda combinan dos tipos de criterios de ranking en función de si los factores a considerar están o no bajo el alcance del administrador del sitio (webmaster). Los criterios que puede modificar el webmaster son los denominados criterios o factores *on-page*, mientras que los criterios o factores que no puede (o apenas puede) manejar directamente son los criterios *off-page*⁹⁹.

Durante el periodo 2009-2013, Google llevó a cabo dos cambios importantes en sus algoritmos de ordenación de resultados, conocidos como algoritmo Google Caffeine (2009) y algoritmo Google Hummingbird (2013), con la consiguiente adaptación y evolución de los factores de ranking hasta alcanzar un número aproximado de 200 factores que analizaremos en el siguiente apartado, limitándonos ahora a esbozar los principales factores para poder proporcionar primero una visión global.

>>>>Factores internos u *on-page*.

Los criterios internos fundamentales se refieren a aspectos estadísticos o de frecuencia de ocurrencia de las palabras clave de la consulta. En principio, las páginas web con una mayor densidad de palabras clave, serán más relevantes. En líneas generales, los buscadores tienen en cuenta los siguientes factores relacionados con las palabras clave:

⁹⁹ Decimos que son los que no puede manejar directamente porque en realidad el SEO se basa en la capacidad de manipulación, aunque sea indirecta, de algunos de estos factores *off-page*. Por ejemplo, es evidente que si un webmaster establece acuerdos de enlace (link farms) con otros webmasters está consiguiendo una manipulación indirecta de su PageRank.

- La posición y frecuencia de palabras clave. Los buscadores tienen en cuenta la presencia de palabras clave en diferentes partes de la página, tales como los títulos, la URL o las meta etiquetas
- La presencia de palabras clave secundarias en la página
- La presencia de componentes multimedia
- La presencia de botones de redes sociales y el número de validaciones sociales por parte de los usuarios

>>>>Factores externos u *off-page*

Existen diversos factores externos entre los que históricamente destaca el análisis de los enlaces de entrada de la página considerada. Cada motor de búsqueda tiene sus propios conjuntos de criterios y sus propias reglas para asignar pesos a cada criterio, pero en general, el número de enlaces que recibe una página suele ser uno de los más importantes, al menos a igualdad de los otros factores. En origen el buscador Google tenía en cuenta dos factores relacionados con los enlaces:

- El texto de anclaje de los enlaces entrantes
- El denominado PageRank (PR), que se basa en:
 - la cantidad de enlaces entrantes
 - la calidad de los sitios que enlazan

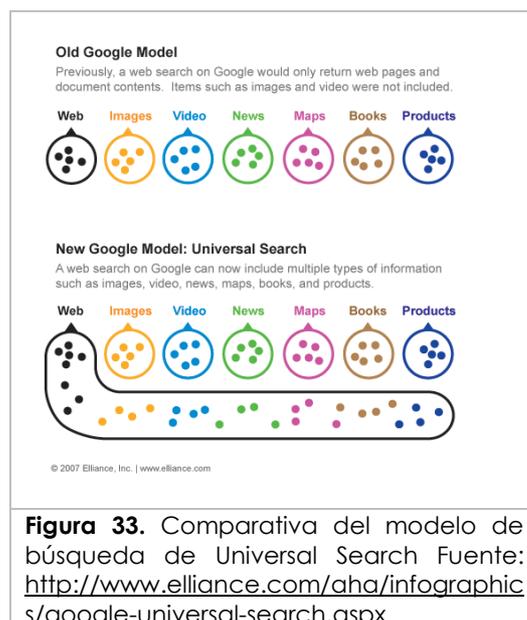
1.5.2. Algoritmos del buscador Google

>Google Search>Universal Search& Blended Results

En el pasado, al introducir un término de búsqueda en el formulario de Google se recibía como respuesta un listado de los sitios web que mejor encajaban con la consulta. En la actualidad, los resultados de Google incluyen diferentes contenidos tales como noticias, blogs, vídeos, imágenes y libros gracias a la denominada "búsqueda Universal" que Google integró a partir de 2007¹⁰⁰ en sus resultados y que combina resultados de diferentes bases de datos para componer las SERPs.

En este sentido, la patente de *Intercalado de Resultados de búsqueda* US 20080140647 A1¹⁰¹ está al parecer en el origen de la denominada *Google Universal Search*. La descripción de la patente tal como aparece en el registro original de la misma es la siguiente:

Métodos, sistemas y programas para el intercalado de resultados de búsqueda. El método incluye la presentación de varios resultados de búsqueda provenientes de un primer motor de búsqueda. Estos primeros resultados de búsqueda satisfacen una consulta de búsqueda dirigida al primer motor de búsqueda y se presentan en un orden. Un segundo resultado de búsqueda de un segundo motor de búsqueda se introduce en una posición entre dos resultados previamente adyacentes. El segundo resultado de búsqueda se recibe de un segundo motor de búsqueda en respuesta a la consulta de búsqueda¹⁰².



¹⁰⁰El 16 de mayo de 2007 apareció en el sitio web Searchengineland.com el artículo titulado Google Launches "Universal Search" & Blended Results <<http://searchengineland.com/google-20-google-universal-search-11232>>

¹⁰¹ Interleaving Search Results. Invented by David R. Bailey, Jonathan J. Effrat, and Amit Singhal. Assigned to Google. US Patent Application 20080140647. Published June 12, 2008. Filed: December 6, 2007. <<http://www.google.de/patents/US20080140647>>

¹⁰² "Methods, systems, and computer program products are provided for interleaving search results. A method includes presenting multiple first search results received from a first search engine. The first search results satisfy a search query directed to the first search engine and are presented in an order. A second

La patente es analizada en el artículo de Seobythesea.com¹⁰³ titulado 'HOW GOOGLE UNIVERSAL SEARCH AND BLENDED RESULTS MAY WORK', el cual seguimos para la exposición del funcionamiento del buscador. A continuación reproducimos parte de la presentación que realiza el citado artículo debido a su interés:

- 1) Una persona realiza una búsqueda, y el motor de búsqueda recibe la consulta, y, posiblemente metadatos, como el perfil de historial de búsquedas del usuario.
- 2) Los resultados de búsqueda se crean desde un motor genérico de búsqueda web.
- 3) Las puntuaciones de calidad de los resultados de la búsqueda se crean para cada uno de los resultados, en base a múltiples indicadores que pueden ser basados en:
 - I. Atributos de los recursos en cuestión
 - II. Características basados en el histórico de datos¹⁰⁴ que describen el acceso al recurso, el uso de los recursos¹⁰⁵, o ambos.

Algunos de los indicadores de puntuación están pre-calculados, mientras que el resto se crean dinámicamente en base al resultado de la búsqueda, la consulta del usuario y los metadatos del usuario como su perfil histórico de búsqueda.

- 4) Los diferentes indicadores tienen como resultado un nivel de calidad global para cada resultado de búsqueda individual.
- 5) Los resultados de la búsqueda web pueden ser vueltos a calcular en base a los indicadores de calidad de las búsquedas.
- 6) La consulta se envía a uno de los motores de búsqueda vertical, como Google News, donde se realiza la búsqueda y se calculan las puntuaciones de calidad de los resultados de búsqueda.
- 8) Al igual que los resultados de la Web, los resultados de la búsqueda de noticias se clasifican en base a las puntuaciones asociadas con las funciones/opciones de búsqueda.
- 9) Los resultados de la búsqueda web genérica y la de noticias se combinan para ser presentados en la SERP¹⁰⁶.

search result from a second search engine is inserted at a position between two otherwise adjacent first search results. The second search result is received from a second search engine in response to the search query“.

¹⁰³ HOW GOOGLE UNIVERSAL SEARCH AND BLENDED RESULTS MAY WORK <<http://www.seobythesea.com/2008/06/how-google-universal-search-and-blended-results-may-work/>>

¹⁰⁴ Las KPIs de awareness son tomadas de datos de Google Analytics>Adquisición>Palabras clave

¹⁰⁵ Las KPIs de Google Analytics>Visión general>Porcentaje de rebote | Porcentaje de salidas | Promedio de tiempo en la página

10) Al mezclar los resultados desde diferentes fuentes, el programa puede recalcularse las puntuaciones de calidad de los resultados de la búsqueda para cualquiera de los resultados de los diferentes motores de búsqueda.

11) El ranking de los resultados puede ser entonces recalculado en base a las nuevas puntuaciones de calidad de los resultados de la búsqueda con el objetivo de generar un único ranking. Existen diferentes formas que determinan cómo se listan los resultados. El mezclador de resultados puede no incluir resultados en base a las puntuaciones finales, pero puede agrupar estratégicamente los resultados que no pertenecen a resultados web de diferentes maneras:

- El mezclador de resultados puede insertar un resultado en cualquiera posición dentro de la lista de diez resultados web genéricos.
- El mezclador resultados puede determinar no introduzca ningún nuevo resultado entre los resultados genéricos al no alcanzar ninguno de los nuevos resultados un rango lo suficientemente alto.
- Un grupo de nuevos resultados de búsqueda web puede ser insertado en una posición fija entre los resultados genéricos, ya sea en la parte superior, inferior o central de una lista de resultados de búsqueda de páginas web genéricas.
- Puede haber algunas limitaciones en la inserción de nuevos recursos, como por ejemplo límites de posicionamiento más allá del tercer puesto del ranking, o distanciar en dos o más posiciones las nuevas noticias de las genéricas.

El ejemplo de cómo funciona este intercalado se centra en la búsqueda web y la búsqueda de noticias, pero las búsquedas adicionales como las de vídeo o imágenes también pueden aparecer en el resultado.

Por lo que vemos, parte de la información de *feedback* juega un papel determinante en el tipo de resultados que son mezclados en la SERP. Por ejemplo, existe un mecanismo de *feedback* que incluye datos de los clics de los usuarios para aprender características de las consultas y/o de su interacción con los resultados de búsqueda; en base a resultados de clics de alta calidad (páginas que reciben muchos clics) se comprueba que estén correlacionadas con los clics de otras consultas.

¹⁰⁶ Se nos dice que el propósito de mezclar conjuntamente los resultados es "aumentar la diversidad de los resultados de la búsqueda se presentan al usuario."

También existe cierta personalización de la información en base a las preferencias de los usuarios por algún tipo de resultado, tales como noticias, y esos resultados pueden tener una puntuación de calidad de los resultados de búsqueda más alta para ese usuario concreto.

Una información sobre la proporción en la que los diversos tipos de contenidos son intercalados en SERPs gracias a la *Universal Search* apareció publicada el 15 de octubre de 2013 en el portal SEO Searchengineland.com bajo el artículo '*Infographic: An Anatomy Of Google's Dynamic Search Results Page*'¹⁰⁷ basado a su vez en el artículo de Moz.com titulado *Mega-SERP: A Visual Guide to Google*¹⁰⁸. Este artículo muestra el porcentaje de veces en la que las diversas tipologías de contenidos aparecen en las SERPs de un test de diez mil palabras clave.

¹⁰⁷ Infographic: An Anatomy Of Google's Dynamic Search Results Page <<http://searchengineland.com/infographic-google-search-results-174190>>

¹⁰⁸ Mega-SERP: A Visual Guide to Google <<http://moz.com/blog/mega-serp-a-visual-guide-to-google>>

Características de SERP locales	
(S) Búsquedas relacionadas	ca. 99,00%
(C) Anuncios de AdWords (Top)	72,20%
(W) Anuncios de AdWords (derecha)	42,40%
(U) Panel de Local Knowledge	26,20%
(N) Resultados de Imagen	24,60% ¹⁰⁹
(L) Revisión de productos y marcas	24,00%
(K) Marcas de autoría	21,90%
(O) Resultados de Noticias	19,60%
(H) Enlaces de sitio (pack de 6)	19,40%
(I) Resultados de Vídeo	18,50% ¹¹⁰
(D) Resultados de compra (izquierda)	18,20%
(R) Los anuncios de AdWords (inferior)	16,50%
(Y) Google Map + Localizaciones	11,30%
(J) "Pack" de resultados locales	7,30%
(G) Panel Local Knowledge	6,30%
(Q) Artículos en profundidad	5,20%
(M) Resultados locales "Cercanos"	5,10%
(E) Answer Box	1,40%
(A) Carrusel de información local	1,00%
(T) Resultados comerciales (Derecha)	-
(B) Lista Carrusel	-
(V) Knowledge Graph (Marca)	-
(X) Caja de Desambigüación	-
(F) Mega-bloque de imagen	-
(P) Resultados Sociales	-

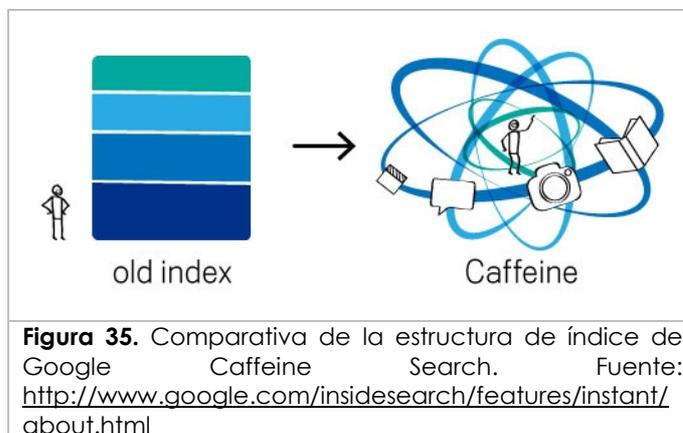
Figura 34. Tipos de contenidos son intercalados en SERPs. Elaboración propia. Fuente: Mega-SERP: A Visual Guide to Google

¹⁰⁹ Un estudio más reciente de 2014 titulado 'Study: Google Universal Results Show Up For 85% Of All Searches: Videos In 65% & Maps In 1%' sitúa la cifra de resultados de imagen en el 45% <<http://searchengineland.com/study-google-universal-results-show-85-searches-videos-65-maps-1-194477>>

¹¹⁰ El mismo estudio sitúa a los vídeos en el 65% de las búsquedas.

>Google WebSearch> Google Caffeine

En agosto de 2009¹¹¹ Google dio a conocer el adelanto de un cambio masivo de infraestructura, diseñado para acelerar el rastreo, expandir el índice, e integrar la indexación y clasificación casi en tiempo real a partir de servicios como Facebook y Twitter. La actualización de sus sistemas se extendió durante meses hasta el lanzamiento final que comenzó en EE.UU. a principios de 2010 y duró hasta el verano. La noticia oficial del lanzamiento de Google Caffeine¹¹² fue el 8 de junio de 2010, cuando Google anunció que el nuevo sistema de indexado había sido completado¹¹³ con unos resultados 50% más frescos debido a la continua actualización de su índice.



Google Caffeine supuso todo un reinicio del motor de búsqueda, fue una renovación y rediseño de la infraestructura de la indexación de Google, no un cambio en sus algoritmos de ranking. Hasta el momento los sistemas de rastreo y la indexación de Google trabajaron como procesos *batch* (no en tiempo real). Googlebot descubría nuevos conjuntos de páginas, los procesaba para extraer el contenido de los mismos y finalmente los añadirá al índice iniciando de nuevo el ciclo. Este sistema de ciclos de indexación fue continuo hasta la llegada de Caffeine, sin embargo todos los documentos de la nueva versión tenían que esperar hasta que se procesaba todo el lote de nuevas páginas para ser añadido a la nueva versión del índice. A partir de Caffeine, Google rastrea una página, la procesa a través de todo el canal de indexación y lo inyecta a la versión en vivo casi al instante.

¹¹¹ La medida fue interpretada en algunos sectores como una respuesta al reciente lanzamiento de Microsoft de una versión mejorada de su propio servicio de búsqueda, rebautizado como Bing.

¹¹² Google Caffeine en Wikipedia <http://en.wikipedia.org/wiki/Google_Search#Google_Caffeine>

¹¹³ Our new search index: Caffeine <<http://googlewebmastercentral.blogspot.com.es/2010/06/our-new-search-index-caffeine.html>>

Este rediseño tenía el objetivo de lograr una mayor velocidad, una mejor indexación de unos resultados de búsqueda más relevantes, como muestra el artículo *What is Google Caffeine* de About.com que destaca su triple mejora en velocidad, tamaño y relevancia¹¹⁴ según indicamos a continuación:

Velocidad. Este es el principal objetivo de Google Caffeine y los test realizados en la denominada "sandbox" (una sección de pruebas temporal) demostraron que Google había logrado cargar hasta dos veces más rápido que los resultados previos. Además Google Caffeine aceleró el tiempo que tardaba en encontrar una página en la web y agregarla a su índice.

Tamaño. A partir de 2009 la compañía ya no divulgó cifras del número de servidores y por tal motivo fue foco de especulación. El experto en energía Jonathan Koomey estimó que Google tenía 900.000 servidores basándose en una extrapolación de los datos que Google proporcionaba de su consumo total de energía¹¹⁵.

Relevancia. Además de la velocidad y el tamaño, la relevancia de los resultados de búsqueda de Google siguió marcando una diferencia respecto a sus competidores. Google había creado un algoritmo inteligente que encontraba resultados relevantes y mostraba las páginas más pertinentes para un amplio espectro de frases y palabras clave.

>Google WebSearch> Google Instant

El 8 de septiembre de 2010 llegó a EE.UU. la actualización Google Instant¹¹⁶, una de las grandes innovaciones del momento, aunque como muchas de éstas tienen la virtud de que los usuarios las utilizan sin apenas ser conscientes de ellas. Esta mejora de búsqueda muestra los resultados a medida que escribe y es capaz de ofrecer a los usuarios ideas de lo que podían buscar a la vez que teclean su búsqueda. Antes de Google Instant, el usuario medio tardaba 9 segundos en introducir un término de búsqueda, si bien Google poseía registros de hasta 30-90 segundos para escribir una búsqueda.

¹¹⁴What is Google Caffeine <<http://webtrends.about.com/od/webportals/a/what-is-google-caffeine.htm>>

¹¹⁵ En esas mismas fechas en términos de tamaño en bruto, Bing de Microsoft parecía tener el índice más grande <<http://www.datacenterknowledge.com/archives/2009/05/14/whos-got-the-most-web-servers/>>.

¹¹⁶ Google Instant en Wikipedia <http://en.wikipedia.org/wiki/Google_Search#Instant_Search>

En el momento del lanzamiento, Google estimó que Instant podía ahorrar a los usuarios 2 a 5 segundos en cada búsqueda, en conjunto unos 11 millones de segundos por hora. Junto al lanzamiento de Google Instant, Google deshabilitó la capacidad de optar para ver más de 10 resultados de la búsqueda por página¹¹⁷. Los beneficios que Google alegaba fueron:

Búsquedas más rápidas: al predecir tu búsqueda y mostrar resultados antes de que termine de escribir, Google Instant puede ahorrar 2-5 segundos por búsqueda.

Las predicciones más inteligentes: aún cuando el usuario no supiese exactamente lo que estaba buscando, las predicciones ayudaban a guiar su búsqueda. La predicción se muestra en texto gris directamente en el cuadro de búsqueda, por lo que puede dejar de escribir tan pronto como el usuario ve lo que necesita. .



Figura 36. Comparativa del tiempo de búsqueda de Google Caffeine Search. Fuente: <http://www.google.com/insidesearch/features/instant/about.html>

¹¹⁷ Búsqueda instantánea se puede desactivar a través del menú "Preferencias" de Google, pero las sugerencias de búsqueda al estilo de autocompletar no se puede desactivar, por la intención.

>Google WebSearch>Hummingbird

Después de Caffeine, Google Hummingbird ¹¹⁸ (Colibrí), es la actualización más reciente del algoritmo de Google, siendo su principal objetivo proporcionar una mejor experiencia de búsqueda a través de resultados más relevantes y significativos que se adapten al contexto del usuario.

En esta ocasión el nuevo algoritmo de Google no se trata de añadir un nuevo factor de ponderación de relevancia ya existente, sino una **actualización completa del algoritmo** de ordenación de resultados (ranking), cuyo precedente más cercano tuvo lugar en agosto de 2009 con el ya mencionado lanzamiento de Google Caffeine. Según sus creadores, su nombre se debe a su precisión y rapidez en la comprensión de búsquedas más complejas y largas, y supone dar un paso más hacia las búsquedas semánticas.

De acuerdo con la empresa, los **objetivos** operativos del algoritmo Hummingbird son:

- Comprender de una manera más precisa lo que el usuario está buscando.
- Dar respuestas a consultas de voz efectuadas desde dispositivos móviles y que han sido enunciadas con lenguaje natural.

>>Historia reciente de Google Hummingbird

Coincidiendo con su 15 aniversario, el pasado 28 de septiembre de 2013 Google hizo la presentación oficial de su nuevo algoritmo Hummingbird (Colibrí), que ya estaba en funcionamiento desde finales del mes de agosto, y cuyas primeras alteraciones en los rankings de las SERPs ya habían sido detectadas por los webmasters desde el día 4 de septiembre (cuando Google realiza cambios importantes en sus algoritmos, muchos webmasters lo detectan antes del anuncio oficial porque observan cambios de posicionamiento en las páginas que monitorizan).

¹¹⁸A pesar de que los datos de elecciones de esta tesis no muestran resultados de la última versión del algoritmo de Google, es necesario referirse a este nuevo algoritmo porque a partir del momento de su lanzamiento pasa a ser el principal algoritmo de ordenación de Google. Todos los ejemplos y casos vistos en esta tesis pertenecen a resultados de 2010.12, años en los que los resultados tenían como base Caffeine y sus updates antispam Panda y Penguin. A partir de ahora, todos los estudios incluirán resultados ordenados por este nuevo algoritmo, de ahí que se haga indispensable plantear cuál es la nueva realidad de ordenación en factores, y cómo puede afectar eso a la tipología de sitios web elegida.

En paralelo al lanzamiento de Hummingbird, Google actualizó el *Knowledge Graph*¹¹⁹ lanzado en mayo de 2012, una base de datos para la gestión del conocimiento con la que Google mejora los resultados, proporciona respuestas directas mediante información de redes semánticas y permite desambiguar las consultas.

En esta primera actualización de Colibrí, Google estimó que había afectado al 90% de las búsquedas, si bien desde Google también se aseguraba que no había habido cambios radicales en las posiciones de ranking ya que no se trataba de un filtro anti-spam. Afirmaciones aparentemente contradictorias como éstas, han generado miles de comentarios en la blogosfera durante los últimos meses, y desde portales tan relevantes como Moz se ha puesto en entredicho la gran cantidad de rumores y poca información que circula en la web acerca de Colibrí. La cuestión es que, tras la fecha de lanzamiento no existe información por parte de Google de actualizaciones (*updates*) oficiales de Google Hummingbird, aunque los analistas creen haber detectado uno oficioso que ha pasado a denominarse Hummingbird #2 y que está fechado en el 14 de Octubre de 2013. Lo cierto es que Google no ha confirmado una actualización, pero diversas herramientas mostraron cambios en las SERPs el día señalado. Una de las empresas de análisis más importantes, Moz, señaló los cambios detectados en un post que obtuvo un gran reconocimiento en el sector de los estudiosos del tema¹²⁰.

>>Cómo funciona Google Colibrí

En la actualidad el usuario medio realiza búsquedas por palabras clave, obviando artículos y preposiciones, las llamadas *stopwords* (palabras que en teoría carecen de significado y/o que no poseen capacidad de discriminación por aparecer en casi la totalidad de los documentos) que el propio buscador ignora en la mayor parte de los casos. Sin embargo, el auge de los dispositivos móviles y la llegada de una nueva generación de usuarios jóvenes no entrenados para buscar por palabras clave, ha hecho que Hummingbird se adapte a las búsquedas por voz, búsquedas compuestas por cadenas de palabras de lenguaje natural más largas que, obviamente contienen esa clase de partículas (artículos, preposiciones, conjunciones, etc.).

¹¹⁹ Knowledge Graph <http://es.wikipedia.org/wiki/Gr%C3%A1fico_de_conocimiento>

¹²⁰ Was There a November 14th Google Update? <<http://moz.com/blog/was-there-a-november-14th-google-update>>

Como hemos señalado, Hummingbird supone un paso más hacia la web semántica, una transición hacia el uso de lenguaje conversacional en las consultas de búsqueda que empezamos a ver en el Google I/O (la conferencia de desarrolladores oficial de Google) de ese año¹²¹. El nuevo algoritmo ha mejorado la precisión de su sistema detector de voz online y es capaz de entender qué estamos buscando al analizar todas las palabras de la cadena de búsqueda en su conjunto y su relación entre ellas.

El objetivo de Google Hummingbird no es tanto ofrecer resultados para palabras clave, sino comprender de una manera más precisa lo que el usuario está buscando y dar respuestas a las consultas de búsquedas que han sido enunciadas a través de lenguaje natural, incluyendo términos muy coloquiales. Por ejemplo, el tipo de consulta en el que potencialmente funcionaría mejor podría ser algo como: "¿Cuál es el mejor lugar para tomar un bocata de calamares en Madrid?". Para dar una respuesta, ahora Google puede utilizar reglas para sustituir los términos de la consulta por sinónimos, y en combinación con el análisis de otras palabras de la propia consulta, entender el contexto de una palabra y encontrar un término potencial con el que reformular (o sustituir) la búsqueda original y proporcionar mejores resultados.

Hasta la llegada de Hummingbird, Google mostraba resultados de búsqueda con las palabras clave que encontraba en la consulta: "mejor sitio", "bocata de calamares", "Madrid". Ahora con Colibrí no solamente va a tener en cuenta las palabras clave de la consulta, sino que entrarán en juego otros factores para desambiguar el significado de las palabras por el contexto, e identificará los sinónimos más adecuados dentro de un amplio abanico de sinónimos potenciales. De esta manera será capaz de entender que siendo "bocata de calamares" un alimento, el sinónimo más adecuado de "tomar" es "comer", y probablemente el usuario encontrará más relevante una respuesta que utilice el término "bar" o "restaurante" en lugar de "sitio", así como "bocadillo" lo es de "bocata".

¹²¹ Google's Impressive "Conversational Search" Goes Live On Chrome <<http://searchengineland.com/googles-impressive-conversational-search-goes-live-on-chrome-160445>>

aquellos sitios web que hayan sido adaptados a la navegación en dispositivos móviles serán destacados como resultados relevantes para ser visualizados en un smartphone o tablet.

>>Adaptarse a Hummingbird

La web móvil ha cambiado la manera en que los usuarios realizan sus búsquedas, y al igual que Hummingbird se adapta a los nuevos hábitos de búsqueda por voz, las empresas de comunicación que deseen mantener un buen posicionamiento de sus contenidos deberán adaptarse al nuevo algoritmo. Con este objetivo, la comunidad SEO apuesta por adaptarse a una afluencia creciente de consultas en lenguaje natural. Algunas recomendaciones derivadas del análisis precedente del nuevo algoritmo podrían ser las siguientes:

- Los usuarios de voz no utilizan palabras clave al hablar, sino que realizan búsquedas utilizando un lenguaje natural y conversacional. Los contenidos optimizados para Hummingbird, en lugar de concentrarse en palabras clave (entendidas como meras cadenas de caracteres), deberán crearse en torno a los temas e ideas de las consultas de conversación. Con este objetivo se recomienda generar contenido de calidad enfocado más a **términos** y **significados**, y no tanto a palabras clave concretas. Usar un lenguaje natural que combine **sinónimos** y términos relacionados con la palabra clave para la que nos queremos posicionar, y optimizar el sitio web utilizando palabras clave más conversacionales, con **variaciones** diferentes que sean más susceptibles de ser utilizadas en el **lenguaje natural**.
- Google busca respuestas claras a preguntas formuladas en lenguaje natural. Es recomendable adaptar el contenido a un formato **pregunta-respuesta** que ofrezca respuestas claras a consultas de voz.
- Utilizar los **Rich Snippets** (o fragmentos enriquecidos) agregando información útil que, mostrada en las SERPs de los buscadores, dote a los contenidos de un valor semántico y responda directamente a las preguntas de los usuarios (ver [microdatos](#) y [schema.org](#)).

- Las **señales sociales** tendrán un mayor impacto en los resultados de Google. Además de potenciar las señales sociales de Twitter, Facebook y LinkedIn, es recomendable ofrecer a Google datos estructurados a través de Google+ y Google Places.
- Incrementar la **autoridad** del sitio web tema específico. Google dará más visibilidad a aquellos sitios considerados fuentes relevantes de información para un campo temático específico.
- Borrar **enlaces de tipo Black Hat** (enlaces que no cumplen las directrices de calidad de Google) que puedan penalizarnos por provenir de sitios spam, o desautorizarlos a través de las herramientas de Google.
- Adecuar los sitios a una navegación creciente de **dispositivos móviles** tratando los contenidos para ser fácilmente visualizados desde móviles y tablets utilizando sistemas como los denominados *responsive design*.
- Adaptar la redacción de contenidos teniendo en cuenta la mayor importancia otorgada a la **geolocalización**, a la **búsqueda local** y a los resultados personalizados.

>Google News Search

En 2005 Gianna M. Del Corso, Antonio Gullí y Francesco Romani de la Universidad de Pisa en su artículo "*Ranking a Stream of News*" describen las cinco propiedades deseables en el ranking de noticias.

- 1. Ranking de noticias y de fuentes de noticias.** Los algoritmos deben calcular dos rankings separados e independientes, uno de artículos de noticias y otro de fuentes de noticias por cada uno de los temas más leídos.
- 2. Clúster de noticias importantes.** Una noticia importante n es probablemente replicada por muchas fuentes en forma completa o parcial. Por ejemplo, la medida de importancia de un artículo n originado por una agencia de prensa se expresa por el número de medios online que replican n o partes del texto de n en forma de extracto. Los buscadores tienen en cuenta el tamaño del clúster formado alrededor de n , siendo éste tamaño una medida de su importancia.
- 3. Refuerzo mutuo entre Artículos de noticias y fuentes de noticias.** Los algoritmos asignan una importancia determinada a diferentes fuentes de información de acuerdo con la importancia de los artículos que producen, de tal manera que una noticia proveniente de "El Mundo" tendrá más autoridad que un artículo similar procedente de "ACME prensa", ya que "El Mundo" es conocida como una fuentes de artículos.
- 4. La conciencia del tiempo.** La importancia de una noticia cambia a lo largo del tiempo. Se trata con un flujo de información en donde una noticia reciente debe ser considerada más importante que una vieja y a medida que el flujo de noticias avanza, su importancia decrecerá en el ranking.
- 5. El procesamiento online/en línea.** El tratamiento de las noticias requiere que el algoritmo de clasificación procese en línea toda la información asociada a un hecho noticioso desde el inicio de su observación, pudiendo en algún momento ser complejo técnicamente dependiendo de la cantidad media de artículos.

A esta clasificación, en 2014 los autores Bhavana Mahour y Akhilesh Tiwari en su artículo *A Ranking Algorithm for News Data Streams* añaden dos propiedades nuevas, 'Autoridad de la fuente' y 'Refuerzo mutuo de fuentes imitadoras' y sintetizan el modelo con un grafo que muestra la relación entre las noticias.

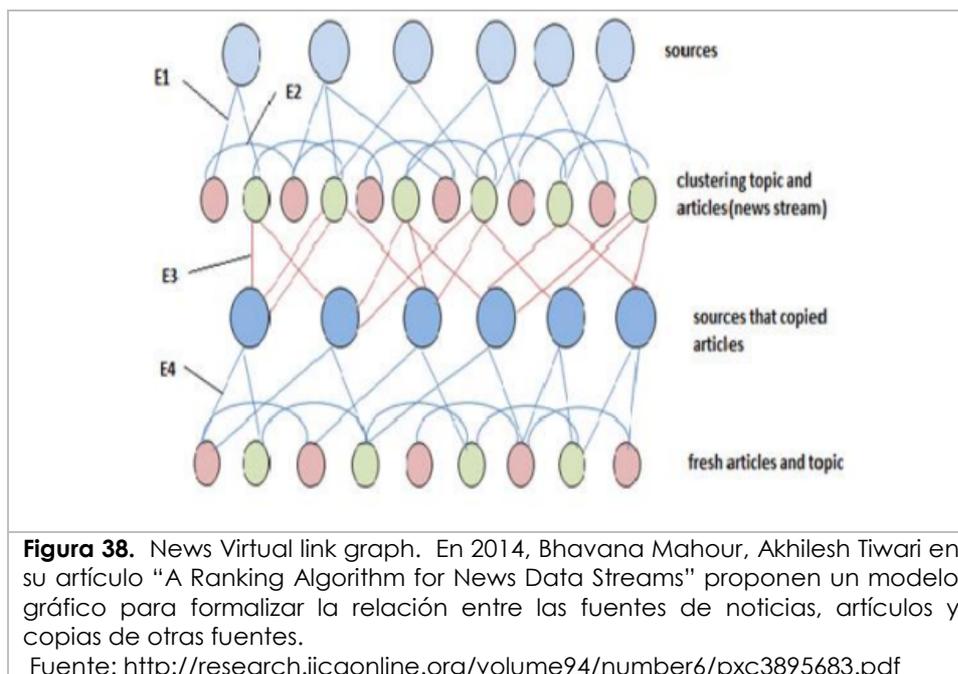
6. Autoridad de la Fuente. Los algoritmos atribuyen diferentes niveles de autoridad a los sitios web de noticias. De igual forma que en el caso de artículos de prensa casi idénticos un sitio de noticias tendrá preponderancia sobre otras fuentes menores, ocurre lo mismo en determinados temas de noticias asociados a medios que se superponen a otros medios.

7. Refuerzo mutuo de Fuentes imitadoras. Se debe tener en cuenta si las copias de las fuentes son totales o parciales. Esta métrica es usada para calcular un ranking de la fuente origen, aumentando su visibilidad en función del número de copias y del hecho que otros medios están copiando de ella; esto reduce la credibilidad de las fuentes secundarias reduciendo su ranking en función de la cantidad de texto copiado.

El hecho que existan pocos enlaces HTML explícitos entre las páginas web de noticias, imposibilita el uso de un grafo de enlaces como PageRank, por lo que Bhavana Mahour y Akhilesh Tiwari añaden a su modelo enlaces virtuales entre fuentes, temas, artículos y artículos copiados.

El modelo está representado en forma de grafo de enlaces no dirigidos $G = (V, E)$, donde $V = S \cup C \cup N \cup N$, siendo S el conjunto de la fuente, N el conjunto de artículos; C el conjunto de temas y N los artículos que han sido copiados.

El set virtual E contiene cuatro particiones separadas, $E1$ - $E4$. $E1$ representan el proceso de creación de noticias. $E2$ representan la similitud entre tema y artículos del flujo de noticias; cuando los artículos a_i pertenecen al tema t_j hay una arista no dirigida entre a_i y t_j . $E3$ es la relación entre el tema y la fuente en la que un artículo a_j es copiado por una fuente s_j creando un nodo que indica una relación de propiedad que reducirá en el ranking el artículo de la fuente s_j . $E4$ son los nuevos artículos publicados por s_j que están copiando otros artículos de otras fuentes.



>>Algoritmo Google News Search

La búsqueda de noticias de Google es una funcionalidad que permite buscar contenidos publicado en sitios de medios. La página de inicio es <http://news.google.es> y a través de sus resultados podemos acceder a una SERP con las noticias que han sido publicadas en los últimos días sobre los temas consultados a través de las palabras clave. Al igual que en la búsqueda web y de blogs, el sistema de consulta permite aplicar diversos filtros, como idioma, tipos de noticias o rangos de fecha. Igualmente, es posible ordenar los resultados por relevancia o por la antigüedad de la noticia.

El funcionamiento de Google News se aparta de modo notable del de su "hermano mayor" dedicado a la búsqueda de páginas web. En primer lugar, la incorporación de nuevos sitios al índice de Google News no es completamente automática. En su lugar, tal incorporación deben solicitarla los sitios aspirantes a formar parte del índice cumplimentando un formulario relativamente complejo (comparado con la facilidad para sugerir nuevos sitios web, que únicamente requiere indicar la URL correspondiente). Una vez cumplimentado, la solicitud debe superar una inspección manual antes de que el sitio sea admitido y por tanto sus noticias pasen a formar parte del índice de Google News.

Por un lado, el mero análisis del formulario permite comprobar que Google no acepta en su índice sitios web que no dispongan de una mínima estructura redaccional, así como del análisis de sus criterios de calidad, se deduce que no aceptan sitios que no produzcan noticias siguiendo criterios periodísticos, entre otros, los de producir información propia y la de citar fuentes.

Además, Google News actualiza su índice con nuevas noticias de todos los sitios de medios utilizando un sistema de sitemaps específicos de noticias que deben llegarle desde las redacciones digitales de los medios. De este modo, el sistema está permanentemente **actualizado**. En cada página de resultados el usuario encuentra enlaces que le ofrecen entre los 10 y los 100 mejores resultados de noticias para sus palabras de consulta.

Respecto a su sistema de cálculo de relevancia, algunas bases del funcionamiento de Google News Search aparecen en la patente *Systems and methods for improving the ranking of news articles*¹²², US 20050060312 A1 con fecha de publicación de 2005. El resumen de la patente indica literalmente lo siguiente:

Un sistema clasifica los resultados. El sistema puede recibir una lista de enlaces. El sistema puede identificar una fuente con la que cada uno de los enlaces está asociado y clasificar la lista de enlaces basados, al menos en parte, en la calidad de las fuentes identificadas.

Para facilitar la comprensión del funcionamiento de Google News, la empresa publicó en septiembre de 2009 un documento audiovisual, concretamente el video titulado *News Search SEO. Background and FAQs for publishers interested in Google News* donde se facilita información sobre cómo optimizar noticias para Google News.

La empresa afirma que en Google News ha realizado una apuesta por la pluralidad de fuentes. Las fuentes son fundamentalmente medios, pero también se incorporan una "amplia gama de tipos de sitios y publicaciones diferentes con una enorme variedad de perspectivas y opiniones"¹²³, entre los que se incluyen blogs, publicaciones especializadas, portales, ayuntamientos., es decir, toda fuente que pueda ser

¹²² Systems and methods for improving the ranking of news articles. US 20050060312 A1 <<http://www.google.com/patents/US20050060312>>

¹²³ Sitios incluidos en Google Noticias <<https://support.google.com/news/publisher/answer/40249?hl=es&topic=2511925&ctx=topic>>

relevante en cualquier categoría temática que analizaremos en el apartado de resultados.

En este sentido, según la descripción que podemos ver en el vídeo, el robot de Google accede a los sitios web de noticias y las indexa basándose en los sitemaps que publica el sitio y, si es el caso, en la información contenida en el fichero robots.txt que permite gestionar el acceso a las diferentes partes del sitio tanto de Googlebot como de Googlebot-News.

Google News refresca sus datos periódicamente, borrando las entradas más viejas de cierto periodo de días y sustituyéndolas por las nuevas noticias indexadas.

Los editores informan a Google a través de un formato de sitemap específico para noticias que el propio Google pone a su disposición. Se trata del "News sitemaps" al que ya nos hemos referido. Cabe señalar que los sitemap, en general, son ficheros en formato XML que describen la estructura de contenidos de un sitio. El News sitemap es un formato específicamente adaptado a informar sobre contenidos de noticias.

En base a este documento, que incluye la URL, el título del artículo y palabras clave, el buscador es capaz de indexar las nuevas noticias y mostrarlas en resultados en un tiempo inferior a los 20 minutos.



Figura 39. News Search SEO Fuente: http://www.youtube.com/watch?feature=player_embedded&v=hg8xgoULIIE



Figura 40. Crawling e indexación

FAQ: Crawling Updated Content

Can Googlebot visit our URLs more than once?

- Find new content often within a matter of minutes
 - Discovery crawl
 - News Sitemaps
- Re-crawl news articles to check for updates
 - Varied re-crawl rate
 - Often check for changes in article content within the first 12 hours after discovery

Figura 41. FAQ Crawling

En este proceso, el primer paso es **geolocalizar** el medio a través de la IP del servidor que publica el contenido, seguido por la clasificación en categorías con el objetivo de intentar mostrar al lector qué noticias son más importantes para cada tema y, en teoría, quién está informando mejor sobre el tema de la consulta.

En el proceso de **agrupación de noticias**, una vez ponderadas, se asignan automáticamente a un mismo tema (utilizando técnicas estadísticas de clustering) las que traten de un mismo hecho noticioso, elaborando luego un ranking dentro del mismo y destacando la mejor versión.

En esta fase entran en juego dos algoritmos, el '**Story Ranking**' y el '**Article Ranking**'. El 'Story Ranking' determina la formación de los clústers en base a la cantidad artículos escritos y a una puntuación de las noticias de última hora, mientras que el 'Article Ranking' sirve para ordenar las noticias dentro del clúster y por tanto, cuáles se sitúan en la parte superior.



Figura 42. Google News -Agrupamiento & Clustering

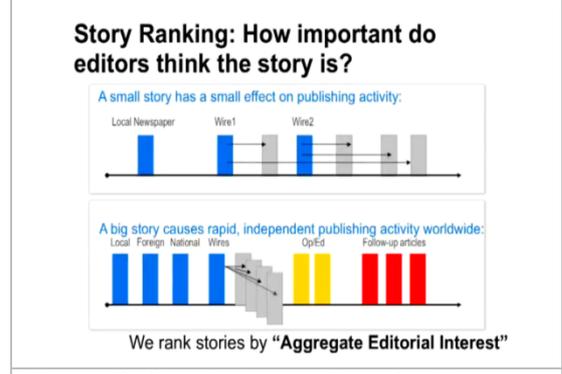


Figura 43. Google News - Story Ranking



Figura 44. Google News - Article Ranking

>Factores posicionamiento en el 'Article Rank'

Para que un artículo aparezca en las primeras posiciones dentro del clúster de un tema, y siempre de acuerdo con el análisis de su documentación oficial, Google tiene en cuenta entre otros los siguientes factores de relevancia:

- **Histórico de artículos del sitio.** Número de artículos producidos por el medio en un primer periodo determinado de tiempo
- La **longitud media** de un artículo producido por la fuente de noticias
- La **cobertura** que el medio produce en un **segundo período** de tiempo (actualizaciones, comentarios, artículos relacionados)
- La cantidad de **tráfico** de red a la fuente de noticias

- **Opiniones** de usuarios del medio
- Las estadísticas de **circulación** del medio (citado desde otros sitios de medios, portales o blogs)
- El número de **periodistas** asociado al medio y de oficinas del medio
- La **amplitud de la cobertura** del medio
- El **tráfico** originado hacia el medio **desde diferentes países**
- El **estilo de escritura** utilizado por la fuente de noticias

>Indicadores positivos de la calidad

Además, de los factores de posicionamiento de noticias, a partir de una consulta, que acabamos de examinar, la patente que dio lugar a Google News, identificaba los siguientes indicadores positivos de calidad, que por tal motivo pueden considerarse como una propuesta de “buenas prácticas” en este ámbito¹²⁴:

- **Utilizar palabras clave en los títulos de artículos, títulos de página, entradas y meta-datos.** Tras una investigación básica por parte de los editores, periodistas o documentalistas del medio sobre los términos clave relacionados con el contenido, las palabras clave principales deben aparecer en el título). Se recomienda que, siempre que no contradiga criterios de necesidad periodística, y para aumentar al máximo el peso de las palabras clave, los titulares deben ser breves (10 palabras como máximo). La URL también debe ser representativa e incluir las palabras clave. Respecto al cuerpo del texto, se sugiere el uso de sinónimos y palabras relacionadas. Se recomienda evitar los saltos de texto y espacios en blanco entre párrafos. Es preferible incluir bloques de texto compactos y los saltos o enlaces externos, es mejor incluirlos al final del texto.
- **Optimizar el dominio principal.** Es recomendable utilizar las tácticas de SEO estándar para aumentar el número de enlaces entrantes al dominio principal, creando una arquitectura de enlaces optimizada que reparta el Pagerank interno y se focalice en los apartados a optimizar.
- **Optimizar el contenido para su propagación en redes sociales.** La velocidad de las acciones sociales es importante. Es necesario mejorar la retroalimentación

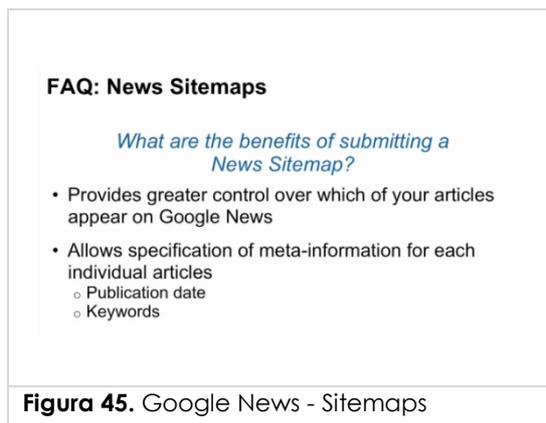
¹²⁴ Google News: Key Ranking Factors [Study]. Rob D. Young, September 26, 2011 <<http://searchenginewatch.com/article/2112126/Google-News-Key-Ranking-Factors-Study>>

social tanto en el dominio principal como en todas sus páginas, situando herramientas de intercambio social a través de botones visibles en espacios destacados del contenido. Las validaciones sociales de los lectores y el CTR de la SERP modificará el comportamiento del algoritmo a través del algoritmo de clicks que permitirá resituar una noticia en el Top10.

- **Ser la primera historia.** Lanzar la historia ayuda a que el contenido tenga éxito, tanto porque aporta al contenido tráfico sin competencia por un breve espacio de tiempo, como porque mejora el posicionamiento del sitio a largo plazo. Los artículos más recientes tienen preferencia sobre los antiguos, por lo que será necesario refrescar la noticia.
- **Trabajar las fuentes.** Es importante citar la fuente original. Las citas de fuentes reputadas son vistas como un elemento importante en la clasificación de un determinado artículo, lo cual a su vez es un excelente ejercicio periodístico. Es importante dar a conocer el contenido citado a través del llamado a veces "periodismo de enlaces".
- **Proporcionar contenido único.** Es recomendable mantener la cantidad de material citado al mínimo, proporcionando en su lugar una visión lo más fresca posible con recursos propios de la comunicación y el periodismo original sobre el tema.
- **Optimizar el título y entradilla para mejorar el CTR y bajar el porcentaje de abandonos.** Es indispensable crear titulares pertinentes, entradillas atractivas y contenido interesante. Google examina el CTR de la página y del artículo para determinar el valor que los contenidos aportan a los usuarios. Además, el rebote hacia la página de Google News después de visibilizar las noticias, es probablemente un indicador crítico para determinar la tasa de éxito.
- **Contenido de calidad.** El contenido de calidad mejorará el éxito del sitio. La calidad podrá ser contrastada a través del seguimiento del comportamiento de los usuarios (por ejemplo, el intercambio social y CTR) y la obtención de enlaces y citas de otros artículos.
- **Diseños responsive.** Los diseños de las páginas deben ser *responsive* para que permitan el acceso a los contenidos desde distintos tipos de dispositivos, incluyendo PC, tablets y smartphones.
- **Prácticas de optimización estándar.** También se aplican factores clásicos de SEO como el tiempo de carga de página, optimización de imagen, la relación-

entre anuncios y contenidos, y la longitud de contenido, que por regla general debe tener entre 200 y 800 palabras.

- **Utilizar Google News Sitemaps.** Además del título y la URL, el formato de sitemap permite a los editores especificar detalles adicionales del artículo (palabras clave, género, ubicación geográfica, el nombre de la publicación, etc), proporcionando metadatos relevantes para la clasificación de los contenidos que permitirán una mejor visualización en los resultados y aumentarán la velocidad a la que se indexa el contenido. Además de enviar periódicamente los sitemap con los nuevos artículos publicados, es recomendable dar de alta un sitemap que contenga enlaces actualizados a las diferentes secciones.



Por otro lado, existen indicadores que, aunque relevantes en el SEO tradicional, carecen de valor en Google News. Por ejemplo, el PageRank de una URL no es relevante a la hora de obtener una posición destacada en Google News. Más importante que el PR son factores como la relevancia local o la temporalidad. Tampoco es necesario añadir el nombre de una localidad al título ya que esto no ayuda a destacar el contenido local. Google extrae la información de la ubicación geográfica y la localización directamente "leyendo" el contenido del artículo o a través de su IP.

>Indicadores positivos de la calidad en contenidos multimedia e imágenes

Google no puede incluir archivos de audio ni contenidos multimedia, pero puede mostrar vídeos de Youtube si rastrea palabras clave de una consulta en el texto complementario de las páginas en las que se muestra los vídeos.

Para optimizar el **contenido multimedia** es necesario acompañar los vídeos y fotos de una descripción en texto (atributos ALT) o una transcripción del contenido, en el caso de los vídeos. Para los vídeos, se recomienda **crear un canal propio en YouTube** y darlo de alta en Google News.

Para las imágenes es recomendable usar el **formato jpeg**, un formato mínimo de 60x60 píxeles, que no sean clicables y ubicar las imágenes en el dominio principal del site.

Podemos encontrar el resto de la información en los tutoriales oficiales en el sitio de soporte de Google <<https://support.google.com/news/publisher/?hl=es#topic=4359865>>

FAQ: News Sitemaps

What are the benefits of submitting a News Sitemap?

- Provides greater control over which of your articles appear on Google News
- Allows specification of meta-information for each individual articles
 - Publication date
 - Keywords

Figura 46. Google News – Sitemaps (2)

FAQ: Multimedia Content

How do I optimize my multimedia content?

Video

- Create YouTube channel and submit it to us
- Textual descriptions and transcripts helpful



Figura 47. Google News - Multimedia content

Images

- Large sizes with good aspect ratios
- Descriptive captions and alt text
- Near title
- Inline, non-clickable
- Prefer JPEG



Figura 48. Google News - Images

>Google News en Universal Search

Los resultados de noticias pueden aparecer en los resultados principales del buscador a través de Universal Search. En este caso, la inclusión de noticias en la SERP general sigue algunas pautas diferenciadoras de cuando una SERP solamente contiene resultados de sitios web (y no de noticias). Para el análisis que sigue, nos basamos principalmente el estudio *10 Keys to Ranking on Google & Bing During Breaking News*

Events¹²⁵ publicado en <http://searchenginewatch.com/> en marzo de 2012, así como en el post de Searchmetrics.com titulado “6 tips to make things work with Google News”¹²⁶. En este sentido se han detectado las siguientes pautas/patrones:

Publicidad. Los estudios citados destacan que los anuncios patrocinados desaparecen de los resultados tanto en Google como en Bing. Los autores especulan que los anunciantes no quieren ser asociados con las últimas noticias y da el ejemplo: “si usted fuera una compañía aérea, ¿desearía que su anuncio apareciese junto a las noticias sobre un accidente aéreo?” En nuestra opinión otro motivo adicional puede estar relacionado con la acusación recurrente que recibe Google de lucrarse con contenidos ajenos.

Resultados en la parte visible. Otro cambio notable es el desplazamiento de los resultados orgánicos, que pierden visibilidad frente a las noticias de última hora. En el caso de que la búsqueda tenga una noticia asociada, sólo los tres primeros resultados orgánicos aparecen en la zona visible.

Por último, los **resultados multimedia** con imágenes y/o videos se presentan de manera más prominente en los resultados para sucesos de mayor relevancia.

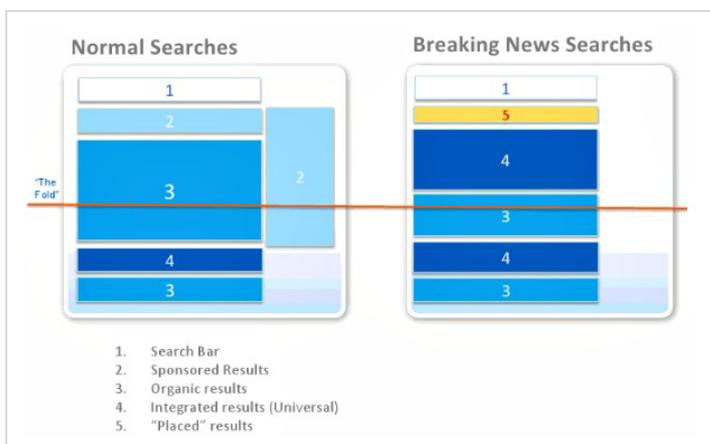


Figura 49. Comparativa de interfaces SERP. Fuente: <http://searchenginewatch.com/article/2159184/10-Keys-to-Ranking-on-Google-Bing-During-Breaking-News-Events-Study>



Figura 50. Parte visible de resultados sobre una noticia del accidente aéreo en Mali de julio de 2014. Elaboración propia

¹²⁵ 10 Keys to Ranking on Google & Bing During Breaking News Events' <<http://searchenginewatch.com/article/2159184/10-Keys-to-Ranking-on-Google-Bing-During-Breaking-News-Events-Study>>

¹²⁶ 6 tips to make things work with Google News. <<http://blog.searchmetrics.com/us/2011/07/20/6-tips-to-make-things-work-with-google-news/>>

Además de la inclusión de un documento sitemap.xml y de aplicar las mejores prácticas SEO al sitio y sus contenidos, los analistas mencionados citan diversos consejos relevantes para la conducir tráfico orgánico hacia un medio digital de noticias a través de las noticias/evento de última hora.

- **Investigación de palabras clave:** utilizar herramientas de monitorización en tiempo real para aprender cuáles son los temas de moda de las noticias/eventos y cómo resuena en los medios.
- **Estrategia de palabras clave:** Desarrollar una estrategia de palabras clave basada en aprender cómo los usuarios buscan estos eventos.
- **Usar los medios sociales para atraer tráfico:** El uso de los medios sociales durante el estallido de la noticia es esencial.
- **Validaciones/acciones en Redes sociales:** El uso de las redes sociales y facilitar que la gente interactúe con el contenido, aumentará sus posibilidades de obtener una mejor clasificación en las SERPs.
- **Google+:** Google da más peso a los contenidos con una gran cantidad de +1's.
- **Formato:** Los motores de búsqueda posicionan los resultados multiformato en la parte superior de la SERP.

>Google BlogSearch

La búsqueda de blogs de Google es una funcionalidad que permite buscar contenidos publicados, teóricamente, sólo en blogs. La página de inicio directa es <http://www.google.com/blogsearch> y a través de sus resultados podemos acceder a las SERPs donde se agrupan los blogs que han escrito en las últimas 24 horas sobre los temas mostrados en las palabras clave. El menú de «Herramientas de búsqueda» permite filtrar por idioma, entradas o blogs, rangos de fecha y ordenar por importancia o por la antigüedad del post.

Google Blog Search actualiza su índice con nuevas entradas de todos los blogs que incluyen protocolos para syndicar los contenidos, como RSS o Atom. El sistema está constantemente actualizado. En cada página de resultados el usuario encuentra varios enlaces que le ofrecen entre los 10 y los 100 mejores resultados de fuentes RSS y Atom. Para que un blog aparezca en el grupo de un tema, y de acuerdo con los analistas, Google tiene en cuenta los siguientes factores: Pagerank, periodicidad con la que publica nuevos contenidos, enlaces entrantes, comentarios que recibe, el número de suscritos vía RSS y la antigüedad del blog. El funcionamiento detallado en el cual se basa Google Blog Search aparece en dos patentes de 2005. Las patentes en cuestión son¹²⁷:

Indexing and retrieval of blogs¹²⁸. Filed Sept. 13 2005 – Awarded July 27th 2010. Esta patente muestra un sistema que puede recibir un feed asociado a un blog. El sistema puede extraer información del feed del blog y crear un documento híbrido¹²⁹ basado en la información extraída. El sistema puede reutilizar el documento híbrido para determinar la relevancia del blog para una búsqueda.

¹²⁷Los internautas saben al instante de qué se habla en Internet. A través de un sindicador de noticias, los internautas reciben las últimas actualizaciones de los blogs siempre que se publiquen nuevas entradas o comentarios que coincidan con su consulta. Incluye los operadores de búsqueda avanzada de Google lo que permite al usuario buscar por título, autor, fecha o dirección url del blog; también le habilita para conocer, por ejemplo, cuántos blogs enlazan con el suyo propio.

¹²⁸Indexing and retrieval of blogs <<http://www.google.de/patents/US7765209>>

¹²⁹ Como se explica en la definición de World Wide Web de Wikipedia, el usuario visualiza sitios web compuestos de páginas web que pueden contener texto, imágenes, vídeos u otros contenidos multimedia, y navega a través de esas páginas usando hiperenlaces. Los nuevos algoritmos de Google identifican estos componentes de las páginas y los transforman en factores de ranking a la hora de posicionarse en SERPs.

Ranking of blog documents¹³⁰. Filed Sept. 13 2005 – Awarded March 15 2007. La patente muestra el sistema por el que un motor de búsqueda de blogs recibe una búsqueda y determina las puntuaciones de un grupo de documentos blog en respuesta a la búsqueda, donde las puntuaciones se basan en la pertinencia de los documentos blog para la consulta y en la calidad del conjunto de documentos del blog. El motor de búsqueda de blogs también puede proporcionar información sobre el grupo de documentos blog sobre la base de puntuaciones determinadas.

Los especialistas en SEO Bill Slawski¹³¹ y Alister Cameron¹³² que gozan de una excelente reputación por la calidad de sus trabajos, analizaron las patentes y publicaron un análisis que, por su interés para esta tesis recogemos a continuación¹³³ centrado en los elementos más relevantes a la hora de posicionar un blog, el ranking de resultados y los indicadores de calidad tenidos en cuenta por Google. De acuerdo con los analistas mencionados, las áreas que Google examina a la hora de considerar los blogs son las siguientes:

- La existencia de *Rich Site Summary* (RSS) o Atom
- El contenido real del mensaje
- El título del blog
- El título del post
- El Autor o perfil del autor
- Calidad de enlaces del Blog roll
- Fecha de mensajes o actualizaciones (relevancia temporal)
- Información geográfica relativa al autor

El cumplimiento de estos elementos es preferencial. Como parte del proceso de indexación real, los robots escrutan no sólo la relevancia de los mensajes de un post, sino también del blog en su totalidad: "A method may include receiving a feed; fetching a blog and one or more posts associated with the feed; extracting information from the feed, the blog, and one or more posts"¹³⁴.

¹³⁰ Google. Ranking blog documents <<http://www.google.de/patents/US8244720>>

¹³¹ Bill Slawski. Seo by the Sea <<http://www.seobythesea.com/about-seo-by-the-sea/>>, o <<http://searchengineland.com/author/bill-slawski>>

¹³² Web personal de Alister Cameron <<http://www.alistercameron.com/>>

¹³³ The Fire horse trail. Google Blog Search under the hood <<http://www.huomah.com/Search-Engines/Algorithm-Matters/Google-Blog-Search-under-the-hood.html>>

¹³⁴ Indexing and retrieval of blogs <<http://www.google.com/patents/US7765209>>

La misma patente también discute acerca de fuentes secundarias de relevancia:

“Determine a relevance of a blog or a blog post to the search query based on information extracted from the blog or blog post and **information** extracted from at least one other source, and provide information relating to the blog or the blog post when the blog or the blog post is determined to be relevant to the search query.”

En cuanto a la evaluación interna, Google la realiza en base a la información de los feeds y de los documentos híbridos generados. La primera línea del snippet de un post y los primeros párrafos son decisivos a la hora de calcular la relevancia.

Otro aspecto que juega a favor son las «named entities» o atribuciones de autoría que enlazan el post con la biografía del autor. Esto se ve apoyado por el funcionamiento de Google News, haciendo de este un aspecto importante para cualquier blog.

En conjunto, para determinar el ranking de resultados el algoritmo tiene en cuenta la página y el blog en su conjunto además de los siguientes factores:

- **Número de ocurrencias del término de búsqueda** (densidad de palabras clave).
- **Ubicación de los términos en el documento** (título, URL, contenido, etc.)
- **Características del sitio:** tecnología, *theme*, tamaño de fuente, color, etc.
- **Puntuaciones adicionales de búsquedas multitérmino:** posicionamiento del blog para paquetes de palabra clave de términos relacionados.
- **Enlaces apuntando hacia el post:** el total de enlaces de otros sitios web que apuntan a un post específico.
- **Enlaces apuntando hacia en blog:** la suma global de todos los enlaces que apuntan a cualquiera de las páginas de un blog.
- **Localización:** para búsquedas geo-localizadas la ubicación declarada por el blogger también servirá para localizar resultados geo-localizados.

Además, las patentes identifican los siguientes **indicadores positivos de calidad** independientes de la palabra clave o términos de búsqueda:

- **Popularidad del blog:** la popularidad se basa en el número de suscriptores, tanto al boletín de correo electrónico como al feed RSS. El factor tiene en cuenta que, cuantas más personas se suscriban, más popular se considera un

blog. Google obtiene los datos de las suscripciones a través de Google Reader, el lector de feeds RSS de Google, así como a través de Feedburner, uno de los servicios de RSS más populares de la web comprada por Google.

- **Popularidad implícita del blog:** este factor se basa en el *Click Through Ratio* (CTR) de las SERPs. A cuanta más gente haga clic en un blog a través de resultados de búsqueda, mayor es su popularidad implícita.
- **La inclusión del blog en el blogroll:** los blogs aparecidos en un blogroll se estiman como de alta calidad y refuerzan la relevancia y confianza del blog de destino. Es decir, cada vínculo del blogroll a otro blog contará como un voto para ese blog que incrementará su PageRank.
- **Existencia del blog en blogroll de alta calidad**
- **Etiquetado del blog:** sitios como Delicious.com o Twitter.com permiten a los usuarios etiquetar y categorizar blogs. El etiquetado de los documentos por parte de los usuarios, una vez indexado por Google, puede ayudar a dar una indicación de palabras clave secundaria del blog.
- **Las referencias al blog mediante otras fuentes:** las referencias desde otros documentos web pueden ser un indicador de calidad de los blogs. Por ejemplo, contenidos de emails o chats que incluyan la URL de un blog a los que Google tiene acceso a través de su servicio de correo electrónico Gmail.
- **Un pagerank del blog:** los blogs, al igual que el resto de páginas web, son clasificados en una escala de 0 a 10 en función de los enlaces que apuntan al blog.

Además, los buscadores tienen en cuenta **indicadores negativos**, es decir, un conjunto de factores asociados a la lucha contra el spam que pueden afectar negativamente el ranking.

- **Frecuencia** de nuevos posts: la frecuencia de la publicación se puede utilizar en combinación con los datos históricos que Google conoce acerca de los splogs con el fin de luchar contra el spam. La patente de búsqueda de blogs se refiere a patrones de frecuencia de actualizaciones que tienen una alta correlación con blogs spam (por ejemplo, los spammers suben muchos posts en un corto período de tiempo o en intervalos específicos); cuando se detectan estos patrones, el blog ve su ranking afectado negativamente.

- El **contenido** y **tamaño** de los mensajes: existen factores relacionados con el tamaño de los mensajes y el contenido, como pueden ser determinadas palabras clave o contenidos duplicados. El relleno de palabras clave, en especial las relacionadas con otras industrias spam como el juego o los productos farmacéuticos genéricos, son un buen indicador de spam. El **contenido duplicado** no original es también un indicador para Google que un sitio es spam. La patente también sugiere que la disparidad de contenidos entre los contenidos aparecidos en el RSS del blog y los contenidos que en realidad aparecen (el feed RSS muestra un contenido con el fin de posicionar mientras que los visitantes visualizan un contenido diferente) es un indicador de sitios spam. Este tipo de contenidos falsos, copiados o auto-generados son reportados automáticamente como una señal potencial de spam.
- La **distribución** de enlaces del blog: un alto porcentaje de enlaces salientes en un post apuntando a un sitio específico puede ser indicador de spam.
- El factor **publicidad**: la presencia de un gran número de **anuncios** en el blog puede ser una indicación negativa de la calidad del documento blog.

>>Google Blog Search en Universal Search

Los resultados de blogs pueden aparecer en los resultados principales del buscador a través de Universal Search. En este sentido Google sigue las siguientes pautas:

- Universal Search sólo muestra 3 resultados de Google blog search en la primera de las SERPs.
- Los resultados de Google blog search normalmente aparecen en la parte inferior de la primera página, aunque en ocasiones pueden aparecer en las primeras posiciones si Google los considera relevantes.
- La selección de los 3 resultados blogs se va alternando de un total de 15 blogs. A medida que transcurre el tiempo, los posts más antiguos van desapareciendo.
- La autoridad del blog es decisiva para aparecer en la búsqueda universal.
- A nivel de SEO, el título de la página deberá contener la palabra clave exacta de la búsqueda, mejor cuanto más al principio de la frase esté.

1.5.3. Listado completo de factores SEO

En lo que sigue, intentaremos establecer cuáles son, según los mejores análisis y fuentes disponibles, los famosos 200 factores que, en su documentación oficial, Google afirma tener en cuenta para determinar la calidad de un sitio en relación a una necesidad de información determinada. Cabe señalar también que Google menciona explícitamente el uso de 200 factores, pero por supuesto, no entre en ningún momento a detallarlos.

No obstante, como decía Codina (2004), desde tiempo existe un cierto consenso entre los analistas sobre los criterios de los buscadores para ordenar los resultados, debido a la monitorización constante que hacen los estudiosos sobre el tema y al uso de técnicas de ingeniería inversa para poner a prueba los elementos de este consenso.

Los factores de ranking fundamentales son los que ya hemos citados en el apartado 'Factores de ranking básicos', sin embargo, a lo largo los años los buscadores han ido añadiendo nuevos elementos de ponderación adicionales hasta llegar a una cifra aproximada de 200 factores, y que son recogidos por algunas webs especializadas. En este apartado se han actualizado, comentado y complementado los contenidos aparecidos en los titulado *Google's 200 Ranking Factors: The Complete List*¹³⁵, contrastándolos previamente con otros análisis y sitios de referencia especializados.

A continuación, ofrecemos una versión de elaboración propia de cada factor, para lo cual hemos seguido además de la fuente indicada una amplia variedad de estudios y análisis disponibles en la web (ver bibliografía), los estudios que hemos llevado a cabo del estado de la cuestión y, por último, nuestra experiencia empírica no solo en los resultados de nuestra investigación sino también en proyectos de posicionamiento. Para dar mayor sentido a la lista de factores, los hemos sistematizado mediante una ficha de diseño propio con los siguientes campos:

- o **Número de factor:** es un número correlativo no significativo. Únicamente a efectos de identificación.
- o **Nombre:** un título o denominación del factor
- o **Tipo de factor:** tipología según el ámbito en el que incide el factor, ya sea de dominio, página, enlaces, etc.

¹³⁵Google's 200 Ranking Factors: The Complete List <<http://backlinko.com/google-ranking-factors>>

- **Ámbito:** tipo de factor según la posibilidad de ser modificado por el propio webmaster (on page) o si depende de terceros (off page)
- **Influencia:** indica si un factor es positivo o negativo para el posicionamiento
- **Importancia:** indica el impacto que el factor tiene en el posicionamiento.
- **Evidencia:** indica la cantidad de evidencia empírica y/o teórica que existe sobre el uso real que Google hace este factor. Por ejemplo, una evidencia Alta indica que, o bien Google ha reconocido formalmente la existencia del factor o bien existe una evidencia clara de su existencia.
- **Descripción:** Una explicación del factor, acompañada en su caso, de ejemplos.

A partir de la revisión de estos factores, es evidente que las empresas de cualquier tipo, y por supuesto las empresas de comunicación deberían centrarse sobre todo en los factores que:

- Tienen una Importancia entre Media y Alta
- Tienen una Evidencia entre Media y Alta

A continuación, presentamos la lista de los 204 factores sistematizados mediante los campos indicados:

1. Antigüedad del dominio				
Factor	Ámbito	Influencia	Importancia	Evidencia
de dominio	on page	Positiva	Media	Alta
De acuerdo con la mayor parte de los analistas, la antigüedad del dominio es un factor tenido en cuenta por Google como parte de la reputación de un sitio. No obstante, según el propio Matt Cutts (portavoz de Google en temas de spam) no es un factor significativo: "The difference between a domain that's six months old verses one year old is really not that big at all". Nótese que el ejemplo que cita no es muy significativo.				

2. Palabra clave en el nombre de dominio				
Factor	Ámbito	Influencia	Importancia	Evidencia
de dominio	on page	Positiva	Alta	Alta
Sigue siendo uno de los factores más relevantes a la hora de posicionar una palabra clave específica. Significa que un dominio con una URL que contenga las palabras "museos" y "barcelona", como en www.museosbarcelona.com , será muy relevante para esas búsquedas				

3. Palabra clave como primer elemento del dominio				
Factor	Ámbito	Influencia	Importancia	Evidencia
de dominio	on page	Positiva	Media	Media
Según datos de MOZ de 2011, tienen mayor preferencia a la hora de posicionar las páginas que comienzan con la palabra clave en el dominio ante otras que incluyen la palabra clave en el medio o al final.				

4. Duración del tiempo de registro del dominio				
Factor	Ámbito	Influencia	Importancia	Evidencia
de dominio	on page	Positiva	Media	Alta
Google aporta una mayor relevancia / confianza a dominios registrados durante más de un año. La razón es que las páginas que se utilizan para hacer spam y otro tipo de prácticas fraudulentas rara vez se registran durante más de un año, por tanto el tiempo de registro es una métrica que tenemos en cuenta. Esto reforzaría, contra lo declarado por el portavoz de Google, el criterio número 1.				

5. Palabra clave en subdominio				
Factor	Ámbito	Influencia	Importancia	Evidencia
de dominio	on page	Positiva	Media	Alta
Los subdominios de una página que empiecen con la palabra clave tendrán muchas más opciones de posicionar dicha palabra en los buscadores.				

6. Historial del dominio				
Factor	Ámbito	Influencia	Importancia	Evidencia
de dominio	on page	Negativa	Alta	Alta
Un dominio comprado en un parking de dominios puede haber sido utilizado anteriormente con fines fraudulentos. En tal caso, el factor de historial del dominio será en negativo. Pero se puede notificar a Google que el dominio ha cambiado y pedir que desautorice links anteriores mediante una herramienta denominada disallow.				

7. Exact Match Domain

Factor	Ámbito	Influencia	Importancia	Evidencia
de dominio	on page	Negativa	Media	Alta

Son nombres de dominio que incluyen palabras claves que se corresponden con términos de búsqueda exactas muy habituales. Este tipo de nombres de dominio ayudaban mucho en el posicionamiento SEO de una palabra clave concreta en el pasado. Sin embargo, a partir de septiembre de 2012 se implementó en el algoritmo Panda la actualización Exact-Match-Domain para reducir expresamente la visibilidad de este tipo de dominios.

8. Whois público

Factor	Ámbito	Influencia	Importancia	Evidencia
de dominio	on page	Positiva	Media	Alta

Los datos básicos del propietario de un dominio pueden ser públicos o privados. Google prima que Whois sea público y que cualquier persona pueda conocer esos datos personales del propietario de la página (básicamente nombre y datos de contacto).

9. Whois de persona que ha realizado spam

Factor	Ámbito	Influencia	Importancia	Evidencia
de dominio	on page	Negativa	Media	Alta

Los propietarios de sitios web que hayan sido penalizados por spam y cuyos datos se puedan rastrear por Whois hace que los futuros los sitios que registre a su nombre estarán serán revisados, pero no considerados spam por defecto.

10. La extensión del dominio (tld)

Factor	Ámbito	Influencia	Importancia	Evidencia
de dominio	on page	Positiva	Media	Alta

La extensión cuenta. La extensión de dominio ayuda a posicionar para el país de la extensión, siendo el .com el dominio universal.

11. Palabra clave en la etiqueta title

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

La palabra clave en el título de la página está considerado como uno de los factores de más relevancia on page. Es fundamental que las páginas estén tituladas con la palabra clave principal para la que pretende optimizar el sitio.

12. Etiqueta título que empieza por una palabra clave

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

Una URL posicionará mejor si el título empieza por una palabra clave que coincide con un término de búsqueda. Para una búsqueda como "SEO", posicionarán mejor las URLs tituladas "SEO, Search Engine Optimization" a que "Search Engine Optimization, SEO".

13. Palabra clave dentro de la etiqueta "description"

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

La palabra clave dentro de la descripción no es relevante en términos de posicionamiento Seo On Page. Su misión es aparecer en los snippets de las SERPs. Su optimización es sumamente importante, sobre todo a través de lo que se conoce como llamadas a la acción.

14. Palabras clave que aparece en el H1

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

Este es uno de los factores más relevantes junto con el de la etiqueta 'Title'. Google reconoce el peso de los contenidos principales a través de la etiqueta 'Title' y de las palabras marcadas con H1.

15. Palabra clave utilizada con mayor frecuencia

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

La importancia de las palabras clave dentro de un texto viene determinada por el número de veces que aparece. Repetir una palabra clave a lo largo de un texto es una señal más de notoriedad que Google tiene en cuenta. La densidad recomendada para las palabras clave principales es del 5%. Densidades superiores al 7% pueden ser consideradas spam.

16. La longitud de los textos

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

A igualdad de condiciones, un texto con más palabras clave tiene preferencia a la hora de posicionar. Los contenidos con una extensión mayor de palabras son percibidos como de mayor calidad.

17. Densidad de palabra clave

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

La densidad de la palabra clave es un factores secundario de posicionamiento, si bien en el pasado era un parámetro con más relevancia y que aun tiene algo de valor.

18. Indexación semántica latente (ISL)

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Baja

En especial a raíz de la actualización del algoritmo Hummingbird, el contenido semántico uno de los puntos más valorados por Google a la hora de poder deducir el significado de un post y sus palabras clave principales. La presencia o ausencia de ISL actúa probablemente como una señal de relevancia.

19. Palabras clave semánticas ISL en título y descripción

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Baja

A raíz del algoritmo Hummingbird, el uso de palabras claves o palabras clave contextuales y sinónimas entre sí en campos como el título y la meta-descripción ayudan a Google a entender la relevancia de una palabra clave y a discernir entre dos sinónimos.

20. Velocidad de carga de la página

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

Tanto Google como Bing utilizan la velocidad de carga como un factor de ranking. A menor tiempo de carga mejor repercute en nuestros contenidos respecto al posicionamiento y viceversa. Los motores de búsqueda pueden estimar la velocidad del sitio con bastante precisión basado en el código de una página y tamaño de archivo.

21. Contenido duplicado en página

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Negativa	Alta	Alta

Contenido duplicado en el mismo sitio (incluso ligeramente modificada) puede influir negativamente en la visibilidad de un sitio. El objetivo principal del algoritmo Panda es la eliminación en SERPs de sitios de contenido duplicado. Es un factor negativo tenido en cuenta por los motores de búsqueda a la hora de posicionar un post.

22. Rel = canonical

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

El elemento <link> con el atributo rel = canonical en un sitio web indica a los motores de búsqueda cual es la versión preferida para ser indexada entre varias páginas duplicadas en tu sitio web. Es una buena forma de prevenir el contenido duplicado.

23. Velocidad de carga de la página vía Chrome

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Media

Google utiliza datos de usuario a través de Chrome para obtener una estimación del tiempo de carga de las páginas. Es recomendable reducir los tiempos de carga. Se puede comprobar mediante herramientas específicas como GTMetrix, Pingdom o Page Speed de la propia Google.

24. Optimización de imágenes

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

Las imágenes optimizadas de un sitio web son fundamentales como factor de posicionamiento. En ocasiones una gran parte del tráfico que recibe un sitio puede ser debido al buen posicionamiento de una imagen. Es necesario escribir un texto alt y un título asociado a las imágenes. De forma colateral, incrementa la definición del nicho de búsqueda con el que se puede asociar el contenido general del artículo.

25. Actualizaciones regular y constante de contenido				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta
Desde Google Caffeine uno de los factores para determinar la importancia de una página es la actualización regular y constante de contenidos. Google favorece la información actual y reciente, en especial la creación de contenidos que den respuesta nuevas búsquedas asociadas a tendencias.				

26. Magnitud de la actualización del contenido				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media
La cantidad de texto que se añada a los contenidos ya indexados se tiene en cuenta en términos de posicionamiento, repercutiendo en un posicionamiento normalmente mejorado.				

27. Actualización histórica de una página específica				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta
Google tiene en cuenta la frescura de una página en función al número de actualizaciones que reciba y cuantas más mejor. Añadir secciones o apéndices explicativos mejora el posicionamiento SEO.				

28. Palabras clave predominantes				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta
Es relevante tener la palabra o palabras clave entre las primeras 100 palabras del texto (lo cual podría venir a equivaler a un pequeño párrafo) ya que de este modo le estamos diciendo a Google que son relevantes para la comprensión general del artículo.				

29. Palabras clave en H2, H3 y Tags				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta
La inclusión de las palabras clave en tags los Hn (donde n es un número del 1 al 6), y en algunas etiquetas (, y) es un indicador para los buscadores acerca de la relevancia de la palabra etiquetada.				

30. El orden de las palabras claves es relevante				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta
El orden de las palabras claves importa. Una coincidencia exacta entre el contenido de una página y las palabras buscadas hace que un contenido tenga más probabilidades de aparecer en los primeros resultados. Para seleccionar el orden es recomendable extraer información de búsquedas reales a través de las herramientas de Google Adwords.				

31. Outbound Link Quality o enlaces de calidad

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

Muchos SEO piensan que enlazar a sitios de autoridad ayuda a enviar señales de confianza a Google sobre la calidad y autoría de los contenidos. Este factor fue popularizado a raíz de los algoritmos antispam basados en HITS.

32. Outbound Link Quality Theme o enlaces de calidad relacionados

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Media

Según Moz, los motores de búsqueda pueden utilizar el contenido de las páginas a las que se enlaza como una señal de la relevancia para un tema determinado. Por ejemplo, un blog de dinosaurios que enlaza a películas de dibujos animados puede dar a entender que el blog está más cercano al ocio infantil que a la paleontología.

33. Gramática y ortografía

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

La correcta gramática y ortografía influye en el posicionamiento Seo ya que es una señal para Google acerca de la calidad del contenido del sitio web.

34. Contenido sindicado

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Negativa	Alta	Alta

El contenido copiado de una página ya indexada posicionará peor que la página original y será albergada en bases de datos complementarias con otros contenidos copiados o de baja calidad.

35. Contenido complementario útil

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

El llamado contenido complementario útil (en inglés "helpful supplementary content") es un factor relevante a la hora de posicionar una página. Contenido enriquecido o contenido complementario útil puede incluir: convertidores de divisas, calculadoras, Sliders, datos del tiempo, fichas resumen, etc.

36. Número de enlaces entrantes Do Follow

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Negativa	Media	Alta

Un número demasiado elevado de enlaces entrantes Do Follow es perjudicial para la visibilidad porque pierde PageRank a través de los enlaces salientes.

37. Contenido multimedia				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Media
El contenido enriquecido multimedia es una señal de contenido de calidad. Además, la ventaja más de usar contenido de vídeo es que las SERPs reservan en su parte superior resultados de videos de Youtube, en muchas ocasiones está vacía.				

38. Número de enlaces internos apuntando a una página				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media
El número de enlaces internos a una página indica su importancia en relación con otras páginas en el sitio. Una manera interna de favorecer el posicionamiento de un contenido es apuntar muchos enlaces internos de nuestra propia página indicando que ese contenido es más relevante.				

38. Calidad de enlaces internos apuntando a una página				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta
Los enlaces internos de páginas autoridad tienen un efecto más fuerte que las páginas con nulo o bajo PR.				

40. Enlaces rotos				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Negativa	Media	Media
Tener un número alto de enlaces rotos es una señal para Google de que el sitio web ha sido abandonado. Es recomendable retirarlos lo antes posible.				

41. Nivel de lectura				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Baja
Google tiene en cuenta el nivel de dificultad de lectura de los textos en función a la materia tratada en cuestión. Google hace es una estimación del nivel de lectura de la página (básica, intermedia y avanzada). Esta clasificación es utilizada por Google para distinguir contenidos de calidad de sitios web pertenecientes a granjas de contenidos.				

42. Enlaces de afiliados				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Negativa	Media	Alta
Una pequeña cantidad de enlaces de afiliados no afecta a los rankings de Google, no obstante una cantidad excesiva en relación a la cantidad de contenido puede ser señal de baja calidad.				

43. Errores HTML				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Negativa	Alta	Alta
Los errores de HTML o de codificación pueden ser un signo de un sitio de mala calidad.				

44. Autoridad del hosting del dominio

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

En igualdad de condiciones entre dos páginas, tendrán ventaja las que estén alojadas en host con más autoridad.

45. Page Rank

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

Suele existir una correlación que indica que los sitios mejor posicionados por lo general son también los que tienen mayor PR.

46. La longitud de las Url

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

Una longitud de Url demasiado elevada puede perjudicar el posicionamiento. Se recomienda una longitud entre los 75 y 92 caracteres.

47. Ruta de la Url

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

Una página más cerca de la página de inicio puede obtener un ligero impulso de autoridad. La autoridad de la página principal de un sitio por norma general siempre es la más fuerte ya que la mayoría de los enlaces entrantes apuntan hacia ella.

48. Editores humanos

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

A pesar de que nunca ha sido confirmado, Google ha patentado un sistema donde se permite a editores humanos influenciar en las SERP, lo que sugiere algún tipo de impulso manual a través de inputs humanos de auditores web.

49. Categoría de la página

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

La categoría de página es una señal de la relevancia ya que permite la desambiguación de su contenido. Una página que forma parte de una categoría muy relacionada con una búsqueda debería recibir un impulso de relevancia en comparación a una página que está archivada en una categoría no relacionada o menos relacionada.

50. WordPress Tags

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

Las etiquetas del panel de WordPress son un factor de relevancia clave. Las etiquetas permiten relacionar tanto fragmentos de contenido como grupos de posts.

51. Palabra clave en URL				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta
Es un factor SEO clave. Las palabras clave que optimizan el contenido de una página web deben incluirse en la URL.				

52. Urls en eslabones				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta
Las categorías en incluidas en la URL como palabras clave son un factor que indica a Google la temática de un determinado sitio. Ejemplos de Urls encadenadas las encontramos en sitios de comercio electrónico como Fnac o Amazon y sirven también para que el usuario pueda visualizar la posición de la página dentro del sitio.				

53. Citar páginas de referencia				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Media
En el posicionamiento SEO es útil citar y referenciar los artículos mediante enlaces salientes aquellas fuentes de autoridad relevantes a ojos de Google y que ya forman parte del nicho de búsqueda/temático. Lo ideal es enlazar contenidos de otros sitios web que pertenezcan al mismo sector.				

54. Listas numeradas o con marcadores				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media
Las listas numeradas y los marcadores tipo viñeta ayudan a romper el contenido del sitio facilitando el escaneo visual por parte de los lectores. Google prefiere contenidos con viñetas y números.				

55. Listas y galerías de imágenes				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media
Con el objetivo de dar mayor relevancia a un determinado contenido es importante aportar listas y galerías fotográficas.				

56. La importancia de un Sitemap.xml				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta
La prioridad declarada en los documentos Sitemap puede influenciar el posicionamiento.				

57. Demasiados enlaces salientes				
Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Negativa	Alta	Alta
Tener una cantidad demasiado elevada de enlaces salientes puede perjudicar la usabilidad de la página y distraer de los principales contenidos.				

58. Cantidad de palabras clave para las que rankea tu página

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Media	Media

Los sitios web orientados tan solo a un grupo reducido de palabras clave con el único objetivo de explotar un micro nicho y monetizarlo son menos tenidas en cuenta por Google que las páginas que posicionan para varias palabras clave, por tanto cuanto más palabras clave posicionen en SERPs, más natural se verá el contenido de tu página.

59. La edad de la página

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

Aunque Google prefiere el contenido fresco, una página de más edad que se actualiza regularmente puede superar a una página nueva.

60 Páginas: "user friendly layout"

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Positiva	Alta	Alta

Las "páginas de diseño de uso fácil" con un diseño limpio y orientado al usuario hacen que el contenido principal sea automáticamente visible.

61. Parking Domains

Factor	Ámbito	Influencia	Importancia	Evidencia
de página	on page	Negativa	Media	Alta

El su actualización de 2011 Google disminuyó la visibilidad de los rankings de dominios, retirando automáticamente los sitios expirados de las SERPs.

62. Contenido Útil

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Alta	Alta

El contenido de calidad no siempre es considerado útil por los usuarios. Google es capaz de distinguir la diferencia entre calidad y utilidad

63. Contenido de valor e ideas únicas

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Alta	Alta

Google ha llegado a declarar que está a la caza de sitios web que no aportan absolutamente nada nuevo ni de valor a los usuarios, especialmente en el caso de páginas de afiliados.

64. Página de contacto

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Media

Google tiene en cuenta de manera positiva aquellas Webs que tienen una página de contacto. El impacto positivo es mayor si las páginas de contacto coinciden con los datos de contacto de Whois.

65. Dominios de confianza/Trustrank				
Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Media
Es un factor importante de ranking. El nivel de confianza de un sitio web se mide en función de la distancia de enlaces a sitios web considerados de alta confianza.				

66. Arquitectura del sitio Web				
Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Media
Una página bien organizada en cuanto a su estructura ayuda a Google a categorizar temáticamente el contenido.				

67. Frecuencia de actualización				
Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Alta
Uno de los factores de mayor relevancia. La frecuencia de actualización de contenidos es un factor de freshness.				

68. El número de páginas de la Web				
Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Alta
El número de páginas de un sitio tiene es una señal secundaria de autoridad. La cantidad de páginas es un factor/métrica útil tenida en cuenta por Google para cribar las páginas de afiliados que no aporten valor.				

69. Presencia de un Sitemap				
Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Alta
Un sitemap ayuda indexaciones más fáciles y más profundas, mejorando la visibilidad.				

70. Tiempo de actividad "on"				
Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Negativa	Media	Media
Las caídas del servidor prolongadas en el tiempo puedes provocar pérdidas de posicionamiento y puede incluso dar lugar a desindexación si no se corrige.				

71. Localización del servidor				
Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Alta
Ubicación del servidor puede influir en el posicionamiento del sitio en diferentes regiones geográficas. Es especialmente importante para las búsquedas geográficas específicas.				

72. Certificados SSL (para Ecommerce)

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Alta	Media

Google ha confirmado que es capaz de indexar el certificado SSL de Webs de ecommerce. Es un factor de calidad a tener en cuenta a la hora de rankear tiendas online.

73. Términos de servicio y privacidad

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Media

A través de estas páginas los sitios ganan puntos de confianza a ojos de Google, ya que entiende que son webs sin nada que ocultar y comprometidas con el usuario y con internet en general. Se trata de un factor relevante.

74. Contenido duplicado "on site"

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Negativa	Alta	Alta

La duplicidad de contenidos en páginas se debe evitar, lo mejor es ponerlas en 'no index' para evitarlo.

75. Ruta de navegación clara

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Media

Google valora de forma positiva las webs que tienen una estructura de URL clara que muestra al usuario en qué parte de la web se encuentra.

76. Diseño 100% responsive

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Alta	Alta

Google valora más las páginas con un diseño responsive, es decir, interfaces que se adaptan para ser visualizadas desde dispositivos móviles (tablets o smartphones). Existen plugins que sirven para adaptar las páginas a este tipo de diseño.

77. Youtube

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Alta

El uso de videos de Youtube en una página es un factor positivo a ojos de Google, en especial en medios de prensa online.

78. Usabilidad del sitio

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Alta	Alta

A través del script de Google Analytics, el buscador reconoce la experiencia negativa de navegación a través de las tasas de rebote de todas las páginas de un sitio web. A un rebote alto, Google entenderá que el sitio no es del agrado de los usuarios en el momento en que aterrizan en él.

79. Utiliza Analytics y Webmaster Tool

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Media

Muchos webmasters creen que el uso de los programas de Google son un factor tenido en cuenta a la hora de posicionar mejor una página. Factor controvertido puesto en duda.

80. Comentarios en una página

Factor	Ámbito	Influencia	Importancia	Evidencia
de sitio	on page	Positiva	Media	Alta

Es un factor de relevancia. Un artículo posicionará mejor cuantos más comentarios de diversas IPs tenga.

81. Enlaces desde páginas con antigüedad

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Media

El valor de los enlaces entrantes es mayor cuando proviene de páginas con un largo tiempo de vida, un valor superior a enlaces provenientes de páginas nuevas o relativamente nuevas.

82. El número de dominios raíz que apuntan a una Web

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

Es uno de los factores más decisivos y que más es tenido en cuenta por Google. En términos de posicionamiento es muy relevante que sean varios los dominios que apunten a una página, siendo preferible recibir 10 enlaces de 10 dominios que 5 y 5 de tan solo 2.

83. Enlaces desde diferentes IPs

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

Enlaces desde diferentes IPs sugiere a Google un rango más amplio de sitios que consideran relevante un mismo contenido. Enlaces desde otros sitios albergados en la misma IP pueden carecer completamente de valor por ser considerados enlaces de nepotismo.

84. Número de Enlaces entrantes

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Alta

El número de enlaces es uno de los factores más relevantes. Puede recibir enlaces desde diferentes páginas de un mismo sitio web.

85. Enlaces desde imágenes con texto "Alt"

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

Es relevante recibir enlaces desde imágenes con la palabra clave en el atributo ALT ya que aporta valor semántico.

86. Enlaces desde dominios .Edu o .Gov

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Alta

Los enlaces entrantes desde dominios con la extensión .edu o .gov aportan mayor relevancia que otro tipo de enlaces.

87. El Page Rank del link de la página

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Alta

El PageRank de una página con un enlace saliente apuntando hacia un sitio web es un factor muy relevante.

88. Autoridad de la página que nos enlaza

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Alta

Además del factor PageRank, un enlace entrante también transmite el Page Authority (PA) y Domain Authority (DA).

89. Enlaces de páginas competidoras

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Alta

Enlaces desde otras páginas visibles en el mismo SERP pueden ser valiosos para esa palabra clave en particular. Recibir un enlace de una página que compite bajo un determinado número de palabras claves en común es algo muy tenido en cuenta por Google a la hora de valorar positivamente una página.

90. Número de veces que una página se comparte en redes sociales

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Media

Las acciones Sociales son un factor importante que influye en el valor del enlace. Los "Social Shares" cuando se producen en una página de una manera masiva, los enlaces salientes de esta página cobran más valor y fuerza.

91. Enlaces negativos de baja calidad

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Negativa	Alta	Alta

Google penalizar enlaces entrantes generados desde páginas spam o con contenidos de baja calidad. En estos casos es posible utilizar herramientas como Disavow Tools [link].

92. Guest post

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Baja	Baja

Google tiene la capacidad de detectar los artículos de autor invitado siguiendo una serie de patrones basados en analizar los backlinks y enlaces de 'Autor'.

93. Enlaces desde la home aportan más fuerza				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media
Los enlaces que provienen de la Home Page, por norma general, son backlinks con mayor fuerza, ya que acumulan más PageRank y Page Authority.				

94. Enlaces No Follow				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Baja	Baja
Las palabras textuales de Google al respecto de este tipo de enlaces es: "in general, we don't follow them" es decir, "en general, no los seguimos". Los enlaces No Follow forman parte de un patrón de enlaces mucho más natural que si solo fueran Do follow.				

95. Patrón de enlaces variado				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Media
Una amplia variedad en la tipología de enlace entrantes y en sus textos de anclaje sugiere a Google naturalidad y por lo tanto tenderá a posicionar mejor. Por el contrario, si la gran mayoría de nuestros enlaces provienen de una sola fuente esto puede ser considerado como un indicio de spam.				

96. Enlaces contextuales				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media
Los enlaces con más fuerza están asociados a un contenido semántico y un envoltorio contextual que facilita la desambiguación.				

97. Enlaces patrocinados				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Negativa	Media	Media
Palabras como "sponsors", "link partners" y "sponsored links" disminuyen el valor del enlace y pueden ser considerados como "sospechosos".				

98. Excesivas redirecciones 301				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	on page	Negativa	Media	Media
Los enlaces que proceden de redirecciones 301 pierden fuerza. Recibir un número demasiado alto de estos puede ser perjudicial para el posicionamiento.				

99. Anchor text del enlace				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Alta
Los textos de anclaje proporcionan las descripciones más exactas de las páginas web. Ha perdido valor a raíz de la proliferación de las granjas de enlaces pero todavía envía una fuerte señal fuerte relevancia.				

100. Anchor text de los enlaces internos

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

El texto de anclaje de los enlaces internos es otra señal de relevancia. Una buena práctica consiste en insertar enlaces internos que apunten al post que queremos posicionar con alguna de las palabras clave del post.

101. El título del enlace

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Baja	Baja

Título del enlace de Atribución: El título del enlace (el texto que aparece al pasar el ratón sobre un enlace) también se utiliza como señales de relevancia débiles.

102. Enlaces desde un país determinado

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

Conseguir enlaces desde un país concreto ayuda a posicionar un contenido en búsquedas desde ese país.

103. Localización del enlace en el contenido

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

Los enlaces colocados al principio del contenido en un texto tienen ligeramente más fuerza que un enlace colocado al final del mismo.

104. Localización del enlace en páginas

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

En general, los enlaces incrustados en el contenido de una página rodeados de texto son más poderosos que los enlaces en el pie de página o área de barra lateral.

105. Enlace desde dominio relacionado

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Media

Es mucho más fuerte un enlace desde una página del mismo nicho de búsqueda que de otra diferente. Este tipo de enlaces refuerzan la temática de dominio.

106. Enlaces desde páginas relacionadas

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Media

Similar al anterior pero a nivel página y no de dominio.

107. "Link sentiment" o enlaces con "sentimiento positivo"				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media
Google tiene en cuenta el contexto de las palabras que rodean un enlace y lo valora el sentimiento asociado al mismo, ya sea de aprobación o de descrédito.				

108. palabra clave en el título				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	on page	Positiva	Alta	Alta
Google da más valor a los enlaces de las páginas que contienen la palabra clave de la página en el título.				

109. Velocidad de enlaces creciente				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Media
Una velocidad de crecimiento de enlaces positiva puede incrementar significativamente el posicionamiento, ya que es una señal de aumento de la popularidad.				

110. Velocidad de enlaces decreciente				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Negativa	Alta	Media
Del mismo modo, una página que retrocede en el ritmo de obtención de enlaces con respecto a su historial pasado, es señal de que su popularidad está decreciendo, afectando al posicionamiento.				

111. Links desde "Hub" Pages				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Alta
Conseguir enlaces de páginas que se consideran los mejores recursos (o hubs) en un tema determinado son mucho más potentes.				

112. Enlaces desde sitios de autoridad globales				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Media
Un enlace de un sitio considerado como un "sitio de la autoridad" pasa valor que un enlace de un pequeño sitio de un micro nicho.				

113. Enlaces No Follow desde la Wikipedia				
Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Baja
Es una creencia SEO de lo más extendida que los enlaces desde sitios como Wikipedia, a pesar de ser No Follow, transmiten parte de autoridad a los sitios enlazados.				

114. "Co-ocurrencias"

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

Las palabras alrededor de un enlace le ayudan a Google a determinar la temática del mismo. De tal modo que se pueden considerar otro factor de relevancia. Enlaces contextuales.

115. Edad de los backlinks

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Media

Según Google, los enlaces de más antigüedad transmiten más autoridad que los enlaces más recientes.

116. Links reales VS Links "splogs"

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Alta

Debido a la proliferación de las redes de blogs en los últimos años, Google probablemente da más peso a los enlaces que provienen de "sitios reales" que de blogs falsos. Probablemente utilizan señales de marca y la interacción del usuario para distinguir entre los dos.

117. Perfil de enlaces natural

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Alta

Una página con un perfil natural de enlaces siempre tendrá más probabilidades de posicionarse más fuerte y será más resistente a los updates del algoritmo. Antinatural podría ser exceso de links desde: blogrolls, footers, solo a la Home Page, con mismo anchor y palabra clave, con un crecimiento o de crecimiento natural... etc.

118. Intercambio recíproco de enlaces

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	on page	Negativa	Alta	Alta

Un intercambio excesivo de enlaces entre páginas web es una mala señal ante Google. El "excesivo intercambio de enlaces" es algo a evitar.

119. Enlaces generados por usuarios

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Alta

Google es capaz de diferenciar los enlaces generados de manera natural de aquellos totalmente antinaturales.

120. Links desde redirecciones 301

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Negativa	Baja	Baja

Los enlaces desde redirecciones 301 tienen menos fuerza que enlaces directos. Cuanto más corta es la ruta del enlace más fuerza es capaz de conservar.

121. Microformatos Schema.org

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	on page	Positiva	Alta	Media

Las páginas que contienen microdatos son más propensas a posicionar que aquellas páginas que no los tienen.

122. Lista de páginas en Dmoz

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

Los buscadores otorgan un mejor posicionamiento a los sitios que aparecen listadas en el directorio Dmoz.

123. Lista de páginas en Yahoo!

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

Google otorga una mayor relevancia a las páginas contenidas en directorios como el de Yahoo! O Dmoz.

124. Número de enlaces salientes de la página

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Negativa	Alta	Alta

El Page Rank no es algo infinito. Un número de enlaces salientes demasiado elevado desde una misma página disminuye el PR distribuido.

125. Enlaces desde perfiles de foro

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Negativa	Baja	Alta

Debido a que hace años este tipo de enlaces habían sido una de las mayores causas de spam, Google ha devaluado bastante su fuerza.

126. Número de palabras que acompañan al contenido

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Media	Media

Un enlace de un mensaje de 1000 palabras es más valioso que un enlace dentro de un fragmento de 25 palabras ya que la fuerza contextual semántica y de la indexación en el primer caso será mucho mayor.

127. Enlaces desde contenido cualificado

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Positiva	Alta	Media

Los enlaces desde contenido pobre, con poco texto y con una variedad semántica escasa transmiten menos valor que enlaces desde páginas con textos bien redactados y contenido multimedia.

128. Compresión de enlaces repetidos

Factor	Ámbito	Influencia	Importancia	Evidencia
de enlaces	off page	Negativa	Media	Media

Los artículos que incluyen varios enlaces hacia una misma página solo contabilizan como uno.

129. CTR en los resultados de búsqueda de una palabra clave

Factor	Ámbito	Influencia	Importancia	Evidencia
Interacción del usuario	on page	Positiva	Alta	Alta

Las páginas con un mayor CTR en SERPs de una palabra clave mejoran su posicionamiento SEO en esa palabra clave específica.

130. Un mayor CTR por páginas global

Factor	Ámbito	Influencia	Importancia	Evidencia
Interacción del usuario	off page	Positiva	Alta	Alta

Los CTR orgánicos acumulados de clics en diversas páginas de resultados para diferentes palabras clave puede favorecer el posicionamiento de los mismos en las SERPs frente a otros con un menor CTR.

131. Porcentaje de rebote

Factor	Ámbito	Influencia	Importancia	Evidencia
Interacción del usuario	on page	Positiva	Alta	Alta

El rebote puede afectar negativamente al SEO. Una manera de Google para usar sus usuarios como probadores de calidad es el porcentaje de rebote, ya que en páginas donde el porcentaje de rebote sea mayor hará Google entienda que no tiene demasiado interés para el usuario.

132. El tráfico directo beneficia tu SEO

Factor	Ámbito	Influencia	Importancia	Evidencia
Interacción del usuario	off page	Positiva	Alta	Media

Google utiliza datos de Google Chrome para determinar si las personas visitan un sitio y con qué frecuencia. Una métrica importante es el tráfico directo. Los sitios con mucho tráfico directo son probablemente de mayor calidad que los sitios que reciben muy poco tráfico directo.

133. Tráfico repetido o tráfico recurrente

Factor	Ámbito	Influencia	Importancia	Evidencia
Interacción del usuario	off page	Positiva	Alta	Media

También pueden ver si los usuarios vuelven a una página o sitio después de una primera visita. Los sitios con visitantes recurrentes pueden recibir una bonificación de ranking.

134. Sitios bloqueados

Factor	Ámbito	Influencia	Importancia	Evidencia
Interacción del usuario	off page	Negativa	Baja	Baja

Función del navegador Chrome que permitía bloquear sitios web spam. Se utilizó como señal de calidad para el algoritmo Google Panda.

135. Chrome Bookmarks

Factor	Ámbito	Influencia	Importancia	Evidencia
Interacción del usuario	off page	Positiva	Baja	Media

Google capta información a través de su navegador Chrome. Las páginas incluidas en los bookmarks de los usuarios pueden tener puntos adicionales de posicionamiento.

136. Google Toolbar Data

Factor	Ámbito	Influencia	Importancia	Evidencia
Interacción del usuario	off page	Positiva	Baja	Baja

Google usa su barra de herramientas de Google como una señal de ranking a través de la información recogida a través de todos los usuarios que la han instalado.

137. Número de comentarios

Factor	Ámbito	Influencia	Importancia	Evidencia
Interacción del usuario	on page	Positiva	Media	Alta

Las páginas con un montón de comentarios son una señal de la interacción del usuario y de la calidad de los contenidos.

138. Tiempo en página

Factor	Ámbito	Influencia	Importancia	Evidencia
Interacción del usuario	on page	Positiva	Alta	Alta

En tiempo en página es proporcional a la calidad o al interés que generan los contenidos. A través de los scripts de Google Analytics ubicados en la mayoría de páginas web, Google conoce exactamente cuál es el tiempo de permanencia de un usuario en una página web. Este es uno de los factores considerados como "relevante" por la mayor parte de la comunidad SEO. El tiempo de permanencia indica el interés por los contenidos de la página a escala global y en relación con una/s palabra/s clave.

139. Las consultas merecen sitios recientes

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Media	Media

Google da un plus de posicionamiento a páginas nuevas que traten sobre un tema sin contenidos recientes.

140. Las consultas merecen diversidad

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Media	Media

Google puede añadir diversidad de páginas web en SERPs para grupos de búsquedas ambiguas que puedan tener más de una intencionalidad.

141. Historial de navegación del usuario

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Media	Media

Los sitios que visitan los usuarios con frecuencia mientras están conectados a Google consiguen un extra de visibilidad en SERP para sus búsquedas.

142. Historial de búsquedas del usuario

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Media	Media

Las búsquedas históricas de una sesión son utilizadas por Google para poder ajustar los resultados de las posteriores búsquedas.

143. Geo-targeting

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Alta	Alta

Google da preferencia a las páginas cuya IP se encuentra geolocalizada en un lugar relativamente próximo al usuario. Del mismo modo tiene en cuenta las extensiones de dominio específicas de país.

144. Búsqueda segura

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	on page	Positiva	Media	Media

Resultados de la búsqueda con palabras vulgares o contenido para adultos no aparecerán a usuarios con Búsqueda segura activada.

145. Círculos de Google como factor relevante

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Baja	Media

Google muestra resultados de autores y sitios que han sido añadidos a círculos de Google Plus.

146. Quejas de Copyrights				
Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Negativa	Media	Media

Google degrada sitios que han recibido quejas de Copyrights.

147. Diversidad de dominios				
Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	on page	Positiva	Media	Media

Según Google, la "Bigfoot Update" vino a añadir un mayor número de dominios en cada páginas de las SERPs.

148. Búsquedas transaccionales				
Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Media	Media

Google muestra resultados diferentes para las palabras clave relacionadas transacciones, como búsquedas vuelos.

149. Búsquedas locales				
Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Alta	Alta

Tras el lanzamiento de Google Hummingbird en noviembre de 2013, la geolocalización de los dispositivos móviles desde los que se efectúa la búsqueda son un factor de ranking más. Los resultados de una determinada página web serán más propensos a aparecer en SERPs si la página web/negocio ha sido localizada en Google Places y Google Plus.

150. Cuadro de noticias de Google				
Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Baja	Media

Determinadas palabras clave activan un cuadro de noticias de Google.

151. Preferencia a las marcas				
Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Baja	Baja

Después de cambios de algoritmo pasados como Google Vince, el buscador ha otorgado una mayor relevancia en las SERPs a los nombres de marca de las compañías para un determinado tipo de búsquedas cortas (Apple->la compañía de ordenadores en lugar de la manzana).

152. Resultados comerciales

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Baja	Media

Google es capaz de mostrar resultados de compra en las SERPs.

153. Imágenes en los resultados

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	on page	Positiva	Alta	Media

Uno de los factores que pueden traer tráfico a un sitio web son las imágenes posicionadas en los motores de búsqueda de imágenes. Aparentemente, posicionar en Google Imágenes es más fácil que hacerlo a través del contenido orgánico a través de texto alt y título.

154. Resultados Huevo de Pascua

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	off page	Positiva	Baja	Baja

Google tiene docenas de resultados del tipo huevo de Pascua. Por ejemplo, la búsqueda "Atari Breakout" en Google imágenes se convierte en un juego jugable <https://en.wikipedia.org/wiki/List_of_Google_hoaxes_and_easter_eggs#Search_Engine>

155. Preferencia por sitios específicos para búsquedas de marcas

Factor	Ámbito	Influencia	Importancia	Evidencia
Regla de algoritmos especiales	on page	Positiva	Baja	Baja

Las palabras clave de marca tienden a posicionar sitios web específicos. En España 'ElCorte Inglés', 'FNAC' y 'Amazon España'.

156. Número de Tweets

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Alta	Alta

El número de tweets es una de las señales o validaciones sociales que Google tiene en cuenta a la hora de posicionar.

157. Influencia del perfil de Twitter

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Media	Media

Es probable que los tweets provenientes de perfiles de autoridad con un montón de seguidores tienen más de un efecto de los tweets de cuentas nuevas con un bajo influencia.

158. Número de Likes en Facebook

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Baja	Baja

Aunque Google no puede ver la mayoría de las cuentas de Facebook, lo más probable es que consideran que el número de 'Likes' de Facebook que una página recibe sea una señal de clasificación débil.

159. Facebook Shares

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Alta	Alta

Las acciones sociales de compartir en Facebook pueden tener una influencia más fuerte que los 'Likes' de Facebook.

160. Autoridad de la cuenta de Facebook

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Media	Media

Al igual que en Twitter, las acciones de Facebook tendrán más autoridad cuando sean realizadas por un perfil de Facebook en concreto capaz de potenciar el posicionamiento de un contenido. De este modo los "Me Gusta" de personas influyentes y famosas deberían ser de gran relevancia.

161. Los Pins de Pinterest

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Alta	Alta

Google los considera como una señal social de que un contenido es atractivo para el usuario.

162. Votos en plataformas sociales

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Media	Media

Según una gran cantidad de SEOs, Google tiene en cuenta la difusión que un contenido tiene en las distintas plataformas y agregadores sociales, tales como Digg en EEUU o Menéame en España.

163. Número de +1 en Google Plus

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Baja	Media

Uno de los factores asumidos como más determinantes a la hora de posicionar son los +1 de Google+.

164. Autoridad de la cuenta de Google+

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Baja	Media

La autoridad de la cuenta de Google+ es algo relevante. No es lo mismo un +1 de un perfil con 100 seguidores que de uno con 10.000.

165. Verificación de autor a través de Google Plus

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Baja	Media

Los contenidos verificados por autor son más propensos a aparecer en los resultados de búsqueda con más fuerza. Google puede entender que es un experto en dichas materias.

166. Señales sociales de relevancia

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Alta	Alta

Un conjunto de señales sociales es más potente que tener solo menciones en una sola red social.

167. Señales sociales por niveles/Site Level Social Signals

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Alta	Media

Hace referencia al tanto por ciento de páginas de un sitio que tienen un número de Shares.

168. Nombre de marca en el anchor text

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales sociales	off page	Positiva	Media	Media

Incluir el nombre de marca en el texto anclaje del enlace es una simple, pero potente, señal de cara a Google a la hora de potenciar el ranking de la marca en cuestión.

169. Búsquedas de marca

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales Marca	off page	Positiva	Media	Media

Google valora si una marca es muy utilizada en formulario de búsqueda y si el sitio recibe mucho tráfico directo.

170. Marcas que tienen una página en Facebook

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales Marca	off page	Positiva	Media	Media

Las páginas con gran número de fans activos, menciones, Me Gusta etc. son ligeramente más propensas a posicionar en SERPs.

171. Sitios con gran número de Followers

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales Marca	off page	Positiva	Alta	Alta

Un alto número de followers en Twitter es señal de relevancia de marca, lo cual favorece en los buscadores.

172. Página de empresa con perfil de LinkedIn

Factor	Ámbito	Influencia	Importancia	Evidencia
Señales Marca	off page	Positiva	Media	Media

Tener una página en esta red profesional es otra señal menor de autoridad del sitio.

173. Empleados de LinkedIn en tu compañía

Factor	Ámbito	Influencia	Importancia	Evidencia

Señales Marca	off page	Positiva	Media	Media
Tener a empleados con perfiles profesionales en LinkedIn asociados a la compañía es un factor de relevancia de marca de cara a Google.				

174. Cuentas con interacción potente				
Factor	Ámbito	Influencia	Importancia	Evidencia
Señales Marca	off page	Positiva	Alta	Alta
En Social Media las cuentas con más interacción social tienen más repercusión posterior en el posicionamiento que los contenidos que cuentas con más número de followers, pero con una interacción nula.				

175. Menciones en Google News				
Factor	Ámbito	Influencia	Importancia	Evidencia
Señales Marca	off page	Positiva	Alta	Media
Las páginas que han sido aceptadas en Google News y aparecen en SERPs de noticias, por lo general están mejor vistas por Google.				

176. Co-citaciones				
Factor	Ámbito	Influencia	Importancia	Evidencia
Señales Marca	off page	Positiva	Baja	Baja
Las marcas consiguen menciones no asociados a enlaces. Google tiene en cuenta este factor, el de páginas que son mencionadas, pero no siempre enlazadas.				

177. Número de suscriptores por RSS				
Factor	Ámbito	Influencia	Importancia	Evidencia
Señales Marca	off page	Positiva	Media	Media
El número de suscriptores vía RSS es tenido en cuenta por los motores de búsqueda. Google puede extraer esta información de seguidores de blogs específicos a través de fuentes propias a través de su crawling de Google Blogsearch.				

178. Localización del negocio				
Factor	Ámbito	Influencia	Importancia	Evidencia
Señales Marca	off page	Positiva	Alta	Alta
El algoritmo Hummingbird mejora el posicionamiento de un sitio web según la geolocalización física de la empresa que puede ser declarada en Google Places.				

179. El sitio web depende de una empresa que tributa				
Factor	Ámbito	Influencia	Importancia	Evidencia
Señales Marca	off page	Positiva	Baja	Baja
Moz informa que Google puede ver si un sitio está asociado con una empresa que paga impuestos.				

180. Penalización de Panda

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Media

Google Panda penaliza páginas con contenido pobre y duplicado.

181. Enlaces desde sitios con baja autoridad

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Media	Media

Una cantidad excesiva de enlaces desde sitios de baja calidad como granjas de enlaces perjudica la posición en SERPs.

182. Enlaces desde directorios

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

Salvo el directorio Dmoz, Google asocia directorios con granjas de enlaces creados con el único fin de manipular el posicionamiento de las páginas.

183. Pop-ups y excesiva publicidad

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

Según la guía de directrices de calidad para Webmasters de Google, un elevado número de publicidad y/o Pop Ups indica contenido de peor calidad.

184. Páginas sobreoptimizadas

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

Las páginas que estén sobreoptimizadas para un determinado número de palabras clave pueden ser penalizadas y degradadas en la posición en SERPs.

185. Página sobreoptimizada

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

Incluye factores on-page como relleno de palabras clave, relleno de etiqueta H1, excesiva profusión de palabras clave.

186. "Ads Above the fold"

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

Un exceso de anuncios que se encuentren en la mitad superior de la página indica a Google páginas de baja calidad. El objetivo es evitar un exceso de publicidad en las llamadas "zonas rojas" o mapa de calor, lugares donde el usuario tiene mayor posibilidad de hacer click.

187. Sitios de afiliados

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

Las páginas que utilizan la afiliación como medio de monetización están sometidos a un análisis mayor por parte del motor de búsqueda.

188. Ocultar enlaces de afiliados

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

Esconder enlaces de afiliados es un factor común de penalización.

189. Contenido autogenerado

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

El contenido autogenerado a través de contenido copiado de terceros vía RSS es penalizado por Google.

190. Exceso de No Follow hacia un mismo contenido

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

Un contenido con un exceso de enlaces No Follow es interpretado por Google como que los webmasters no quieren transmitirle autoridad por algún motivo en particular.

191. Direcciones de IP marcadas como Spam

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

Un servidor IP marcado como Spam devalúa todas las páginas y sitios albergadas en él.

192. "Meta tag Spamming"

Factor	Ámbito	Influencia	Importancia	Evidencia
spam on page	on page	Negativa	Alta	Alta

No se debe repetir una misma palabra clave demasiadas veces en las meta etiquetas.

193. Afluencia masiva de enlaces de golpe

Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Alta	Alta

La aparición de enlaces más alta de lo habitual y concentrada en un corto periodo de tiempo indica a Google de que son enlaces fraudulentos.

194. Penalización de Google Penguin

Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Alta	Alta

Los sitios penalizados por Google Penguin perdieron gran parte de su visibilidad en SERPs. Los webmasters deben limpiar sus enlaces entrantes y enviar email de reconsideración.

195. Perfil de enlaces de calidad limitada

Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Alta	Alta

Un Hub con enlaces de baja calidad puede penalizar a los sitios web a los que apunta.

196. Vincular relevancia de dominio

Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Alta	Alta

Un exceso de enlaces de alta calidad de temática no relacionada puede hacer que Google Penguin penalice la página. La composición natural debe ser de sitios de temática relacionada.

197. Advertencias de backlinks antinaturales

Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Alta	Alta

En caso de una caída en SERPs debido a un tema de link-building se deben notificar los enlaces fraudulentos a través de Webmaster Tools.

198. Enlaces desde la misma clase IP

Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Alta	Alta

Un exceso de enlaces desde la misma página IP puede ser identificado como una granja de enlaces que no se corresponde a un patrón de enlaces similar a otros sitios web de la misma categoría.

199. Anchor text "tóxicos"

Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Alta	Alta

La aparición de numerosos anchor text con palabras spam se consideran comportamientos Black-hat SEO.

200. Penalizaciones manuales

Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Alta	Alta

Las páginas penalizadas de manera manual, a pesar de salir de la penalización, es posible que experimenten un mayor escrutinio por parte de los robots del buscador con visitas de comprobación.

201. Venta de enlaces

Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Alta	Alta

Google puede penalizar el PageRank de una página si detecta que el sitio se dedica a la venta de enlace o hace abuso de enlaces de temática no relacionada con el sitio.

202. Google Sandbox				
Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Baja	Baja
Patrones de crecimiento de enlaces excesivo desde sitios asociados a Black Seo como perfiles de foros y comentarios, puede hacer que Google analice el sitio en su Sandbox durante un tiempo, con la respectiva pérdida de visibilidad.				

203. Google Dance				
Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Positiva	Alta	Alta
El término 'Google Dance' puede referirse al cálculo periódico que hace el buscador del PageRank de las páginas, lo que conlleva fluctuaciones de las SERPs. El 'Google Dance' es constante y periódico y tiene como objetivo que sus algoritmos mejoren el posicionamiento de sitios relevantes.				

204. "Ya utilicé Disavow Tool"				
Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Media	Baja
El uso de la herramienta para desautorizar enlaces puede eliminar una penalización manual o algorítmica para los sitios que fueron víctimas de SEO negativo. Otros SEO opinan que el uso de la herramienta es un factor negativo para el posicionamiento a largo plazo.				

205. Solicitud de reconsideración				
Factor	Ámbito	Influencia	Importancia	Evidencia
spam off page	off page	Negativa	Media	Baja
Una solicitud de reconsideración exitosa puede levantar una sanción, pero no se borrará de su historial.				

1.5.4. Web Spam

En otros apartados de este trabajo hemos realizado alguna presentación previa del concepto de spam aplicado a los buscadores. Ahora podemos recordar su significado acudiendo a la definición que proponen en la *Web Spam Taxonomy (Stanford)*, según la cual se trata de cualquier acción humana deliberada que tiene la intención de incrementar de forma no justificada la relevancia o importancia de una página web.

De esta definición se deduce que no toda acción conducente a incrementar la relevancia de una página debe considerarse spam. Las acciones que tienen como objetivo mejorar los contenidos de una página (por ejemplo ampliando los mismos añadiendo componentes audiovisuales o editando con más calidad los componentes textuales) o las acciones que tienen como objetivo mejorar la experiencia de usuario (por ejemplo, mejorando la arquitectura de la información) no son spam. En cambio las acciones que se hacen para confundir a los buscadores haciéndoles creer que la página tiene una autoridad que no posee en realidad sí son spam (por ejemplo, pagando a cambio de obtener enlaces para hacer creer que la página tiene una gran credibilidad).

Las técnicas de spam representan en el contexto de los buscadores la mala praxis perfecta porque perjudica a todos los actores (menos al que lo lleva a cabo): perjudica a los usuarios de los buscadores que obtienen malos resultados en sus búsquedas, perjudica a los buenos sitios web, que se ven desplazados por contenidos sin calidad y perjudican a los buscadores que ven en peligro su audiencia y, con ella, su modelo de negocio. Por último, puede perjudicar incluso a la empresa que lo aplica, si es detectada por Google (o denunciada por su competencia) y es expulsada del índice (beneficia en cambio al falso profesional de SEO que cobró por sus malos servicios a la empresa expulsada del índice). La cuestión importante es que las técnicas de spam permiten incrementar la relevancia para consultas que no están relacionadas con el tema de una página web. Las técnicas concretas de spam incluyen cualquier tipo de acciones encaminadas a aumentar el ranking de una página web, pero sin mejorar el verdadero valor de una página. En definitiva, el "spamming" es una manipulación que no añade valor a la página y que trata de aprovechar los puntos débiles de un algoritmo de búsqueda.

Hay esencialmente dos tipos de técnicas spam, por un lado, las que sirven para promover la posición en los rankings de páginas web, y por otro lado, las técnicas que sirven para ocultar el contenido real de una web.

Técnicas spam para promover un sitio web. Como hemos indicado, estas técnicas se utilizan para aumentar el valor de una página. Las principales técnicas detectadas son:

- **Spam de palabras clave.** Estas técnicas tratan de manipular la posición en rankings a través de elementos tales como el título de la página (etiqueta <title>) o las etiquetas <meta> (en atributos description o keywords, aunque en la actualidad solo el contenido de <title> sigue siendo considerado un factor de ranking debido al abuso de las etiquetas <meta>).
- **Spam de URL.** La URL está considerada como un factor importante de ranking en motores de búsqueda, motivo por el que su manipulación es una técnica de spam frecuentemente utilizada.
- **Spam de enlaces.** Los motores de búsqueda consideran la cantidad de enlaces y el texto de anclaje de los mismos como una de las señales más importantes desde una perspectiva de ranking. Este tipo de spam incluye la manipulación del texto de anclaje de los enlaces y la creación de enlaces artificiales en diferentes páginas web donde es fácil añadir enlaces por parte de terceros (foros, comentarios, libros de visitas, etc) para incrementar el valor de las páginas de destino.

Técnicas spam para ocultar el contenido de un sitio web. Este conjunto de técnicas utilizan métodos no perceptibles con el fin de conseguir incrementar la posición de una página web. La razón es que el texto que se utiliza para influenciar al buscador es anti natural por la excesiva repetición de palabras y no puede mostrarse a los usuarios humanos. Las principales son:

- **Contenido oculto.** Se trata de técnicas donde las palabras clave y los enlaces se ocultan cuando el explorador muestra una página. Las técnicas más comunes incluyen el uso de combinaciones de colores que hacen que los elementos en cuestión sean efectivamente invisibles.

- **Cloaking.** Esta técnica de ocultación se da cuando el rastreador de un motor de búsqueda es identificado y se le muestra una versión diferente de la página de lo que se mostraría a un usuario normal.
- **Redirección.** La página redirige automáticamente a una página diferente a la indexada por los buscadores.

>Aproximaciones a la detección de Spam

Debido a la presión de las técnicas de spam, actualmente, buena parte de la ingeniería de los buscadores está volcada en la detección y anulación del spam, como en una especie de carrera de armamentos. Las aproximaciones actuales en la identificación y neutralización de estas actividades que se indican a continuación se articulan en torno a sus dos grandes ejes, el contenido y los enlaces:

1. Contenido spam

- **Idioma:** a través de tests de búsquedas, los ingenieros detectaron que los idiomas con más spam son el francés, seguido por el alemán y el inglés; por tanto también es en las páginas de estos idiomas dónde los especialistas humanos en detección de spam de los buscadores centran sus esfuerzos.
- **Dominio:** existen TLDs con una mayor incidencia de spam. Los TLDs con mayores tasas de spam son los dominios .BIZ seguido por los dominios .COM y los .US.
- **Palabras por página:** se ha detectado que las páginas con más texto a menudo son las que contienen más spam. Las páginas spam acostumbran a contener entre 750 y 1.500 palabras. La tasa de spam disminuye en páginas que superan las 1.500 palabras.
- **Palabras clave en el título de la página:** los tests han demostrado que las páginas spam tienden a utilizar muchas más palabras clave en el elemento TITLE que las páginas que no son spam.
- **Ratio de texto de anclaje:** la relación entre los textos de anclaje de un sitio o página y el texto estándar es un indicador de spam. Las páginas web con un alto porcentaje de texto de anclaje son más propensos a ser sitios de spam.

- **Ratio de contenido visible:** esta métrica está relacionada con el texto oculto y mide el porcentaje de texto que no se muestra en la página.
- **Compresibilidad:** para combatir el relleno de palabras clave, los motores de búsqueda pueden también aplicar el ratio de contenido repetitivo. Los motores de búsqueda a menudo comprimen una página después de la indexación para guardarla. Las páginas de spam contienen un ratio de compresión característico (no comprimido dividido por el peso comprimido).
- **Global de palabras populares:** para encontrar el relleno palabra clave los buscadores comparan la cantidad de palabras en una página con datos de otras páginas no-spam conocidas. Las páginas spam se caracterizan por un uso más antinatural que las páginas semilla (seed) no-spam utilizadas como modelo.
- **Búsquedas Spam:** conocidos los análisis realizados por buscadores del ratio de clics para la personalización de la SERP, los spammers consultan términos relacionados y hacen clic en sus propios resultados. Al observar el patrón de las consultas en combinación con otras señales, estas tácticas se hacen estadísticamente evidentes.
- **Spam a nivel de host.** Los buscadores cotejan los sitios de los servidores donde se ha localizado spam, añadiendo una referencia a cada servidor según la cantidad del spam localizado en él.
- **Basado en la frase:** bajo este enfoque, los buscadores utilizan modelos de aprendizaje probabilístico utilizando documentos de entrenamiento, a través de los cuáles buscan anomalías textuales en forma de frases relacionadas. Buscando anomalías estadísticas a menudo puede poner de relieve los documentos spam.

2. Spam de enlaces

- **TrustRank:** Este método se basa en el concepto de los "buenos vecinos". La investigación muestra que los sitios fidedignos enlazan con otros de la misma calidad y viceversa.
- **Acumulación de enlaces:** los spammer crean un montón de páginas de baja calidad cuyos enlaces apuntan a la página de destino que quieren optimizar. Los sitios de spam tienden a tener una mayor proporción de este tipo de enlaces en comparación con un conjunto de páginas de entrenamiento no-spam.

- **Enlaces nepotistas:** este grupo incluye los enlaces de pago y los recíprocos. Aunque es una práctica habitual entre los SEOs, los motores de búsqueda consideran los enlaces recíprocos como una forma de spam.
- **Topología de enlaces spam** (granjas de enlaces o link farms): los motores de búsqueda comparan el porcentaje de enlaces aparecido en su grafo con el de sitios web considerados "buenos". Normalmente aquellos sitios que buscan manipular los buscadores tendrán un mayor porcentaje de enlaces de este tipo.
- **Anomalías temporales:** los buscadores comparan los datos históricos de sus páginas y contrastan el promedio de adquisición de enlaces de los sitios "normales" de su índice, lo que permite detectar los sitios spam con índices de crecimiento de enlaces no naturales.

Pese a lo que hemos presentado, en la mayoría de los casos no existe una señal ni enfoque que se considere definitivo a la hora de localizar spam. Los buscadores a menudo emplean una combinación de métodos para localizar spam, estableciendo métricas y umbrales que les permiten evitar falsos positivos.

De hecho, en uno de los artículos que hemos utilizado para elaborar esta sección, titulado *Web Spam: The Definitive Guide for SEOs*¹³⁶ aparece una batería de referencias y enlaces hacia patentes artículos y vídeos que muestran las diferentes herramientas utilizadas estos últimos años por los buscadores para combatir el spam. En este contexto de "carrera de armamentos", y durante el periodo 2010-2012, Google llevó a cabo dos cambios importantes en sus algoritmos de ordenación de resultados, conocidos como algoritmo Panda (2011) y algoritmo Penguin (2012), lo cual afectó al comportamiento del buscador durante los periodos electorales objetos de estudio.

¹³⁶Web Spam: The Definitive Guide for SEOs <<http://www.wordstream.com/blog/ws/2010/04/28/web-spam-guide>>

>Content Farms, Spam y SEO Black Hat

En el contexto de la World Wide Web, el término granja de contenido o content farms¹³⁷ se utiliza para describir a una compañía que emplea a un gran número de escritores freelance para generar grandes cantidades de contenido de texto que está específicamente diseñado para satisfacer los algoritmos de los motores de búsqueda.

Su objetivo principal es generar tráfico a través de visitas de usuarios de buscadores que a su vez se supone que generarán ingresos por algún tipo de conversión o por publicidad. Se trata de un negocio que mueve millones de dólares¹³⁸ anuales y que ha hecho prosperar granjas de contenidos en especial en el mercado de lengua inglesa.

De igual forma, el negocio para obtener ingresos por publicidad también ha multiplicado los casos de sitios que hacen scrapping de contenidos, copiando todo el contenido de un sitio web y reutilizándolo en la creación de otro nuevo. Algunas de estas granjas de contenido albergan millones de artículos¹³⁹, en muchas ocasiones copias idénticas que están mejor optimizadas para el SEO que los artículos originales de otros medios de menor peso.

El objetivo de Google Panda (que presentamos con más detalle un poco más adelante) es, entre otros, penalizar el scrapping de contenidos y hacer caer de sus rankings a sitios web que usen otras técnicas de spam para la construcción de nuevos sitios web. En particular, las técnicas penalizadas por Panda son:

- Duplicar contenidos de otros sitios web a través de sus RSS.
- Duplicar contenidos a mano mediante "cortar y pegar".
- Robar fragmentos de varios contenidos distintos para montar un tercero más amplio.
- Reescribir pequeñas partes y fragmentos de un artículo.
- Automatizar la recogida de fragmentos de contenidos de varias fuentes.
- Crear sitios sociales donde publicar duplicados de otros sitios web.
- Crear contenidos de baja calidad que no aporten valor añadido.
- Añadir contenido multimedia (fotos, videos, etc.) al texto extraído de otro sitio web.

¹³⁷ Content farms en Wikipedia <http://en.wikipedia.org/wiki/Content_farm>

¹³⁸ Beet.tv. Yahoo Harvests "Content Farm" Associated Content for \$90 Million, Report <<http://www.beet.tv/2010/05/yahoo-harvests-content-farm-associated-content-for-90-million-report-.html>>

¹³⁹ Jongales.com. List of content farms <<http://www.jongales.com/blog/2011/02/14/list-of-content-farms/>>

>Google Panda

Por su interés, dedicamos esta sección a discutir con más detalle el significado de Google Panda. Se trata de un cambio en el algoritmo de clasificación de resultados de Google que fue lanzado por primera vez en febrero de 2011. Con el objetivo final de mejorar la experiencia del usuario, Google Panda evalúa algorítmicamente la calidad de los sitios en Internet, localiza sitios de calidad y mejora su posición en el ranking.

Google Panda no es un cambio global en el algoritmo del buscador, es únicamente un nuevo factor de ranking que se ha añadido al ya existente. Como hemos señalado, el objetivo de Google Panda es reducir el posicionamiento de sitios de baja calidad, especialmente los indicados sitios tipo scrapers y 'content farms', devolviendo en cambio a posiciones más altas a sitios y contenidos de alta calidad.

De acuerdo con la empresa, los **objetivos** operativos del algoritmo Panda son los siguientes:

- Mejorar la relevancia en los resultados
- Mejorar la indexación y clasificación de contenidos
- Sancionar la venta y el intercambio de enlaces
- Favorecer el branding¹⁴⁰ y el desarrollo de marcas
- Combatir los contenidos duplicados
- Mejorar los procesos y seguir en constante optimización
- Combatir la falta de originalidad, de calidad y relevancia
- Mejorar los procesos de búsqueda

Tras la aplicación de Google Panda se ha reportado un aumento en el ranking de sitios web de noticias y sitios de redes sociales, y una caída en el ranking de sitios que contienen grandes cantidades de publicidad. En la primera actualización de Panda, Google estimó que había afectado a la clasificación de casi el 12% de sus resultados.

¹⁴⁰ Branding en Wikipedia <<http://es.wikipedia.org/wiki/Branding>>

>> Cronología de las actualizaciones de Google Panda

Durante el año 2011 se observó un aumento global de los contenidos de tipo spam y de baja calidad que rastreaban las arañas del buscador Google. Ante las quejas de muchos sitios web , que veían como los scrapers conseguían un mejor posicionamiento que los sitios con contenido original, Google pidió públicamente datos para ayudar a detectar dichos sitios web. Para conseguir más información, adicionalmente Google lanzó Chrome Site Blocker , una extensión que permitía a los usuarios especificar qué sitios web querían bloquear en sus resultados.

Con esta información, en febrero de 2011 Google lanzó Google Panda con el objetivo de solucionar el problema de los sitios de baja calidad. Tras varias actualizaciones desde el lanzamiento original, el efecto Google Panda se hizo global en agosto de 2011 con el lanzamiento de su versión 2.4.

Las fechas de lanzamiento de las versiones previas de Google Panda fueron:

Panda 1.0 : 24 de Febrero de 2011

En su primera actualización Google afirmó haber afectado a un 12% de los resultados de búsqueda, acabando con sitios de escaso contenido, granjas de contenido, y sitios con alto ratios de anuncio/contenido.

Panda 2.0 : 11 de Abril de 2011

Google lanzó la actualización de Panda a todas sus consultas en lengua inglesa, incluyendo las no-US. Esta versión integraba los datos de sitios web bloqueados por los usuarios a través del navegador Chrome.

Panda 2.1 : 9 de Mayo de 2011

La nueva actualización no fue reconocida por Google y parece ser que se produjeron cambios relativamente menores.

Panda 2.2 : 18 de Junio de 2011

La versión 2.2 fue reconocida oficialmente por Google. La nueva versión de Panda mejoró los filtros de detección de scrapers.

Panda 2.3 : 22 de Julio de 2011

Los foros de webmasters sugirieron que Google lanzó una nueva actualización en esta fecha. No estaba claro si se habían introducido nuevos factores, o se trataba simplemente de una actualización de los datos de Panda y los factores de ranking.

Panda 2.4 : 12 de Agosto de 2011

Google Panda fue lanzado a nivel internacional en todos los idiomas excepto para chino, japonés y coreano. Google informó que había impactado al 6-9% de las consultas en los países afectados.

Panda 2.5 : 28 de Septiembre de 2011

Después de un mes, Google lanzó una nueva actualización de Panda. Los detalles específicos de las modificaciones no estaban claros, pero algunos sitios reportaron grandes pérdidas.

Panda 3.1 – 3.4: Noviembre de 2011 - Marzo de 2012.

Después de la versión 2.5, Google entró en un periodo de "Panda Flux", realizando actualizaciones más frecuentes y de menor impacto. La actualización 3.2 devolvió el posicionamiento a sitios web previamente penalizados. En las siguientes actualizaciones no hubo cambios en el algoritmo, únicamente carga de nuevos datos.

Panda 3.5: 19 de Abril de 2012.

Se aprovecha el lanzamiento de Google Penguin para resolver un bug con un clasificador de sitios expirados. En la versión 3.6 de ese mismo mes se realizan modificaciones menores.

Panda 3.7 – 3.9.2: 8 de Junio – 18 de Septiembre de 2012.

Las actualizaciones de verano de 2012 supusieron cambios moderados en los rankings de Google para menos del 1% de las búsquedas

Panda #20: 27 de Septiembre de 2012

Esta nueva actualización afectó oficialmente al 2,4% de las búsquedas,

rebautizándose en función del número de actualizaciones del algoritmo, en este caso la vigésima. Esta versión se solapó con la actualización EMD (Exact-Match Domain) que reducía la visibilidad de sitios de baja calidad cuyo nombre de dominio coincidía con keywords populares de búsqueda.

Panda #21 - #25: Noviembre de 2012 – Marzo de 2013

Continúa el "Panda Flux", con actualizaciones frecuentes de menor impacto. En marzo de 2013 Matt Cutts anuncia la última actualización de Panda antes de ser integrado en un nuevo algoritmo.

Panda 4.0 (#26): Mayo de 2014

Google confirmó una importante actualización de Panda que probablemente incluía tanto una actualización del algoritmo como una actualización de datos. Oficialmente, alrededor de 7,5% de las consultas en idioma Inglés se vieron afectados

Panda 4.1 (#27): Octubre de 2014

Google anunció la última versión de su Panda Update pocos días antes de finalizar la redacción de esta tesis. Google dijo que dependiendo de la ubicación, entre un 3% y un 5% de las consultas de búsqueda se verán afectadas¹⁴¹.

¹⁴¹ Panda 4.1 — Google's 27th Panda Update — Is Rolling Out <<http://searchengineland.com/panda-update-rolling-204313>>

>>Recomendaciones para Google Panda

Como ya hemos señalado, la primera actualización de Panda (febrero de 2011) supuso una caída drástica de las granjas de contenido en los rankings de Google. Las siguientes actualizaciones de Panda han seguido filtrando sitios de baja calidad y han afectado a una gran variedad de sitios que han visto caer sus resultados de búsqueda. De acuerdo con diversos informes y análisis de especialistas como los de Garry Przyklenk¹⁴² de 2011 o Jeremy Page de 2014¹⁴³, para amortiguar esta caída y adaptar el sitio web a los requerimientos de Panda, se recomienda:

- **Menor cantidad de anuncios.** En la actualidad existen sitios web llenos de Google Adwords. Panda identifica sitios web que están claramente dedicados a la publicidad y los penaliza. Para ajustarse a los requisitos de Google Panda, es recomendable considerar la eliminación de algunos anuncios, especialmente los que no se correspondan claramente con contenido del sitio web. Cabe señalar aquí la relativa contradicción (en este caso para bien del usuario) entre el sistema AdWords que persigue colocar cuantos más anuncios mejor, y el algoritmo Panda que penaliza los sitios que tienen la parte principal de la una página dedicada a publicar anuncios.
- **Contenido más largo y atractivo.** Las recomendaciones clásicas de 350 palabras por página para contenidos escritos con fines SEO son ahora insuficientes. Los estudios de Google dieron a entender que una satisfacción completa de los usuarios requiere de información más completa, con una extensión de los artículos de entre 550 y 600 palabras, con imágenes, información de gráficos y videos (p.e. de Flickr, Creative Commons y YouTube).
- **Bajar el porcentaje de rebote.** Las actualizaciones de Panda indagan también en las métricas de usuario, como tiempo en el sitio y porcentaje de abandonos. Una alta tasa de rebote de un sitio es un claro indicador de que los usuarios no encuentran en el sitio la información que buscaban. Panda toma esta información a través de Google Analytics, donde queda reflejado el porcentaje

¹⁴² Top 5 Google Panda Update SEO Survival Tips <<http://searchenginewatch.com/article/2080032/Top-5-Google-Panda-Update-SEO-Survival-Tips>>

¹⁴³ How to survive an angry Panda Update <<http://www.semrush.com/blog/publications/how-to-survive-an-angry-panda-update/>>

de abandonos del sitio. En la lista de las páginas más visitadas y las palabras clave que llevaron a los visitantes a esas páginas se puede observar el porcentaje de abandonos en relación a esas páginas.

Para optimizar la página y mejorar los índices de rebote podemos disminuir el porcentaje de abandonos de su sitio. Con este objetivo se debe:

- o Ofrecer a los visitantes todo lo que se necesita saber acerca de la palabra clave a través de la cual ha llegado a esa página.
 - o Ofrecer a los visitantes de ese tema una razón para continuar navegando por el sitio.
-
- **Una menor dependencia de tráfico de buscadores.** Otra conclusión que han obtenido los webmasters derivada de la influencia del Panda es que un sitio web que es demasiado dependiente de los resultados de búsqueda orgánica es vulnerable a los cambios de Google y a sus criterios de valoración sobre la calidad de enlaces, redes y directorios de negocios locales a los que estamos enlazados. Por ello, ahora se considera recomendable mantener el tráfico de motores de búsqueda a un 45% o menos para evitar que las actualizaciones de Google Panda tengan un efecto drástico en el tráfico del sitio.

 - **Aumentar el *Social Engagement*.** Con la introducción de Google +1, Google está interesado en la participación social, propagando el *Social Engagement* a los resultados de búsqueda. Las redes sociales, como Twitter y Facebook, influyen en los resultados de búsqueda, y las acciones y los votos que reciba una web son esenciales para indicar a los motores de búsqueda que el contenido es tan bueno que la gente quiere compartirlos con su red. Es recomendable añadir los botones de participación social en la página de inicio y en las páginas de contenido ([AddThis](#) , [Share This](#), [official Tweet Button](#), [Facebook LikeButton](#), y [Google +1](#)).

 - **Obedecer a las Directrices para webmasters.** Chequea regularmente las Webmaster Guidelines para asegurarte de que una actualización de un algoritmo futuro no afecta negativamente a tu sitio.

En la blogosfera SEO podemos encontrar diferentes listas para la realización de una auditoría sobre los parámetros más relevantes de Google Panda. Las *checklist* ofrecen una buena guía sobre los diferentes puntos a inspeccionar si queremos sobrevivir a Google Panda.

La lista que presentamos a continuación es una composición propia presentada en forma de preguntas de chequeo que puede hacerse el webmaster del sitio está y organizada en cuatro secciones, Confianza, Autoridad, Relevancia y Valor.

Confianza¹⁴⁴

- ¿La información que ves en la página que has llegado después de una búsqueda inspira confianza?
- ¿Introducirías tus datos personales y/o el número de tu tarjeta de crédito?
- ¿Recomendarías la página a tus amigos en redes sociales?
- ¿Pondrías la página en tu bookmark?
- ¿El diseño de las páginas está actualizado?
- ¿El sitio posee botones sociales y muestran hits de otros usuarios?

Autoridad

- ¿El autor es un gran conocedor del tema?
- ¿El sitio es una autoridad en su nicho de búsqueda?
- ¿Se trata de una marca reconocida?
- ¿La página presenta información y/o contenido original?
- ¿La información es imparcial?
- ¿Se observa un análisis en profundidad de la información?
- ¿Los contenidos parecen descuidados y producidos de forma precipitada?
- ¿Se revisa regularmente la calidad de los contenidos del sitio?
- ¿El sitio posee una alta calidad y cantidad de enlaces entrantes?

Relevancia

- ¿Los artículos están orientados a usuarios o al volumen de búsquedas en Google?
- ¿El sitio web ofrece una visión completa del tema?

¹⁴⁴ Javier Casares. Determinar la credibilidad de un sitio web <<http://www.durky.com/determinar-la-credibilidad-de-un-sitio-web/>>

- ¿Los artículos web del sitio tienen suficiente contenido?
- ¿Los anuncios destacan más que el contenido?
- ¿Consideras que lo ha escrito un experto?
- ¿Consideras que lo ha escrito un aficionado que conoce del tema?
- ¿Consideras que el tratamiento del tema es de carácter superficial?

Valor añadido

- ¿El contenido del artículo es original o se trata de una copia?
- ¿El artículo no está orientado a SEO y contiene variaciones de palabras clave?
- ¿El sitio web tiene información relevante que otros sitios en el top 10 no tienen?
- ¿Proporciona información con valor añadido respecto a otras páginas de su misma SERP?
- ¿Trata temas similares con variaciones de palabras literalmente diferentes y no repite el mismo contenido y grupos de palabras clave en todas las páginas?
- ¿La información es reducida, carece de características útiles y de valor para finalizar una acción o compra?

>Google Penguin

Google Penguin es otro cambio en el algoritmo de ordenación de resultados de Google que ayuda a establecer el ranking de los sitios web. Fue lanzado por primera vez en abril de 2012. Es un algoritmo optimizado gracias a sus predecesores, capaz de identificar y penalizar acciones de sobreoptimización y bonificar las modificaciones y ajustes a las recomendaciones del propio Google, aumentando el peso de los factores de ranking derivados de la web social.

De acuerdo con la empresa, los **objetivos** operativos del algoritmo Penguin son:

- Combatir la sobreoptimización.
- Penalizar los sitios que muestren métricas no naturales, es decir, anomalías respecto a las métricas SEO que Google tiene como referencia gracias a la enorme cantidad de datos que posee de sitios de todo el mundo.
- Aumentar el peso de los factores relacionados con redes sociales.

El nuevo algoritmo de Google, llamado en primer lugar «Over Optimization Penalty», más tarde «Web spam Algorithm Update», y finalmente Penguin, se centra, como hemos señalado, en detectar y penalizar la sobreoptimización SEO de sitios web, analizando contenidos y enlaces. Los principales aspectos son:

A). Penguin penaliza diferentes **técnicas de sobreoptimización**:

- El nombre de dominio con palabras clave de búsqueda exactas y sin contenido original.
- Repetir palabras clave dentro de las URLs.
- El abuso de negritas y subrayados.
- Existencia de enlaces sin texto relacionado.
- La ausencia de coherencia entre los títulos y el contenido.
- Las técnicas basadas en contenido de baja calidad o contenidos duplicados.

B). Respecto a los **enlaces**, el algoritmo penaliza:

- Disparidades en la tipología de enlaces. La calidad (Page Rank) y el tipo de los enlaces entrantes (web, blog, nofollow, Twitter, Facebook, y otras redes sociales) debe ser proporcional y natural respecto a los datos que Google conoce de otros sitios web similares.
- Incremento de enlaces entrantes demasiado lineal y artificioso a lo largo del tiempo.
- Los bloques sospechosos de enlaces de 'webs amigas'.
- El abuso de palabras clave exactas en el texto de anclaje de enlaces internos de la web.
- El uso repetitivo de palabras clave exactas en el texto de anclaje de los enlaces entrantes.
- Enlaces contextuales con textos de anclaje muy optimizados y provenientes de contenidos de baja calidad.
- Redes privadas de compra-venta de enlaces.
- Grandes cantidades de enlaces artificiales.
- Enlaces entrantes provenientes de granjas de enlaces.
- Existencia de enlaces sin relación contextual entre los sitios.

>> Cronología de las actualizaciones de Google Penguin

Penguin 1 (#1): 24 de Abril 24 de 2012

En su lanzamiento se vieron afectados un 3,1% de los sitios y supuso una caída de visitas que en algunos sitios alcanzó el 50%.

Penguin 1.1 (#2): 26 de Mayo 26 de 2012

Su segunda versión afectó a menos de un 0.1%

Penguin #3: 5 de Octubre de 2012

La siguiente versión del algoritmo afectó al 0.3% de las búsquedas en lengua inglesa e italiana y al 0,4% de las búsquedas en español y francés.

Penguin 2.0 (#4): 22 de Mayo de 2013

La cuarta actualización del algoritmo supuso un moderado impacto en resultados, siendo su objetivo afectar no tanto a nivel de dominio como de página.

Penguin 2.1 (#5): 4 de Octubre de 2013

Actualización menor del algoritmo, probablemente una actualización de datos, con un impacto moderado en resultados.

Penguin 3.0 (#6): 17 de Octubre de 2014

A diferencia de los updates previos, este parece ser sólo una actualización del filtro, sin nuevas señales añadidas. Impactó alrededor de 1% de las consultas en inglés.

>> Recomendaciones para Google Penguin

Una primera recomendación puede parecer sorprendente para personas al SEO: los contenidos deben ser escritos para las personas, no para spiders ni algoritmos. Lo que esto significa es que en el pasado, con el fin de optimizar un sitio se producían contenidos pensando en el efecto que tendrían en el buscador. Pero esta política tenía el resultado perverso de hacer posicionar bien sitios cuyos contenidos podría ser de escaso o nulo interés para las personas. Como resultado, Google estaba favoreciendo sitios con poca calidad para los usuarios, de aquí la recomendación de dejar de escribir para los robots (Google) y escribir para las personas.

Con el objetivo de prevenir y amortiguar la caída de visibilidad y tráfico provocada por posibles penalizaciones del algoritmo Penguin, el propio Google recomienda adaptar el sitio web a un conjunto de requerimientos derivados de la evolución de la red social y del propio spam:

- **Enfatizar las señales sociales.** El SEO de 2012 enfatiza las redes sociales. La importancia del número de Tweets o shares aumenta y se convierten en factores de ranking determinantes. Google Penguin contabiliza la acción social asociada a un contenido y modifica la posición de una página web en función de las señales sociales que acompañan a los contenidos y nuevas

publicaciones: Facebook, Twitter, Google+, LinkedIn, etc. Un estudio de 2012 sobre factores de ranking en sitios web de UK muestra la creciente importancia de este tipo de enlaces.

- **Publicación de nuevos contenidos.** Google es capaz de valorar la calidad de los contenidos de un sitio web y de los enlaces que recibe. Ambos deben ser confiables, útiles e informativos, y deben aparecer de forma constante y regular, tanto los contenidos del sitio como el número de enlaces. Los sitios deben ofrecer el contenido en distintos formatos: pdf, docs, vídeos, infografías, presentaciones, contenido interactivo, etc.
- **Mejorar los contenidos respecto la competencia.** El contenido del sitio debe proporcionar más información significativa que otros resultados del nicho. Un análisis de las tendencias de búsqueda del mercado mostrará las oportunidades, carencias y necesidades de contenidos que se deben ofrecer en un nicho de búsqueda. Los contenidos especializados sobre un tema mejoran la autoridad y visibilidad del sitio.
- **Palabras clave del Anchor Text.** Se debe diversificar los textos de anclaje de los enlaces de entrada. Muchos enlaces con el mismo texto de anclaje es sospechoso a ojos de Google, que puede considerar que ha sido fruto de una campaña del propietario del sitio en lugar de corresponder a un crecimiento espontáneo y no coordinado de los mismos. Cuando es así, lo habitual no es que una misma palabra clave apunte a una o varias páginas de un sitio. Google considera más natural que existan enlaces que incluyan como texto de anclaje la misma URL o textos diferentes entre ellos.
- **Enlaces entrantes.** El algoritmo de Google puede detectar la “naturalidad” de la evolución de los enlaces entrantes a un sitio web, es decir, si siguen patrones de crecimiento naturales o existe algún grado de manipulación. Por ejemplo, la tipología de los enlaces (web, blog, redes sociales, etc) y su Pagerank deben ser proporcional a los registros que Google posee de otros sitios web similares del nicho.

Además, uno de los factores de Google para determinar la posición en los resultados es la relevancia de los enlaces. El valor de la cantidad de enlaces externos disminuye si no existe una relación contextual. El nuevo algoritmo analiza el contexto de cada enlace, evalúa la relación temática entre los sitios enlazados, y penaliza los enlaces spam creados con herramientas automáticas de captación de enlaces o el alta masiva en directorios. Recibir muchos enlaces de webs de temáticas diferentes sin contenido original o tener muchos enlaces salientes hacia webs de diversas temáticas puede ser penalizado.

- **Evitar la sobreoptimización y las técnicas Black hat seo.** Evitar el uso abusivo de un texto de anclaje exacto, es decir, de un texto de enlace que corresponda exactamente a la palabra clave que optimiza la página de destino enlazada. De igual forma, dentro del sitio web deben evitar los textos o enlaces ocultos, las redirecciones sospechosas, las palabras clave irrelevantes o el exceso de las mismas. También debe evitarse la compra e intercambio de enlaces, y los obtenidos de manera automatizada a través de la automatización de comentarios.

Capítulo 2. Sistema de análisis

2.1. Antecedentes

El primer desafío al que tuvo que enfrentarse la investigación fue la necesidad de la creación *ex novo* de un sistema de análisis debido a la falta de investigaciones previas en este mismo campo.

Uno de los problemas que tuvimos que abordar fue que, tras la exposición del modelo por W. L. Bennet (2003), no aparecen citas al modelo en la producción académica que puede consultarse en las bases de datos de referencia o en Google Scholar.

Por su parte, el artículo de J. Peretti titulado *Culture Jamming, Memes, Social Networks, and the Emerging Media Ecology. The "Nike Sweatshop Email" as Object-To-Think-With* alcanza 26 resultados en Google Scholar, mientras que el artículo de W.L. Bennett, *Communicating Global Activism: Strengths and Vulnerabilities of Networked Politics. Information, Communication & Society* ha sido citado por otros 138 artículos académicos durante la última década, pero sin referencias al modelo de Infoesfera.

De este modo, aunque la idea original era prometedora y parecía tener una base sólida, el principal inconveniente fue la falta de un corpus teórico que desarrollase la idea de J.Peretti, lo que obligó a retomar la investigación desde un momento 0, el de su planteamiento, y adaptar el modelo a la realidad actual.

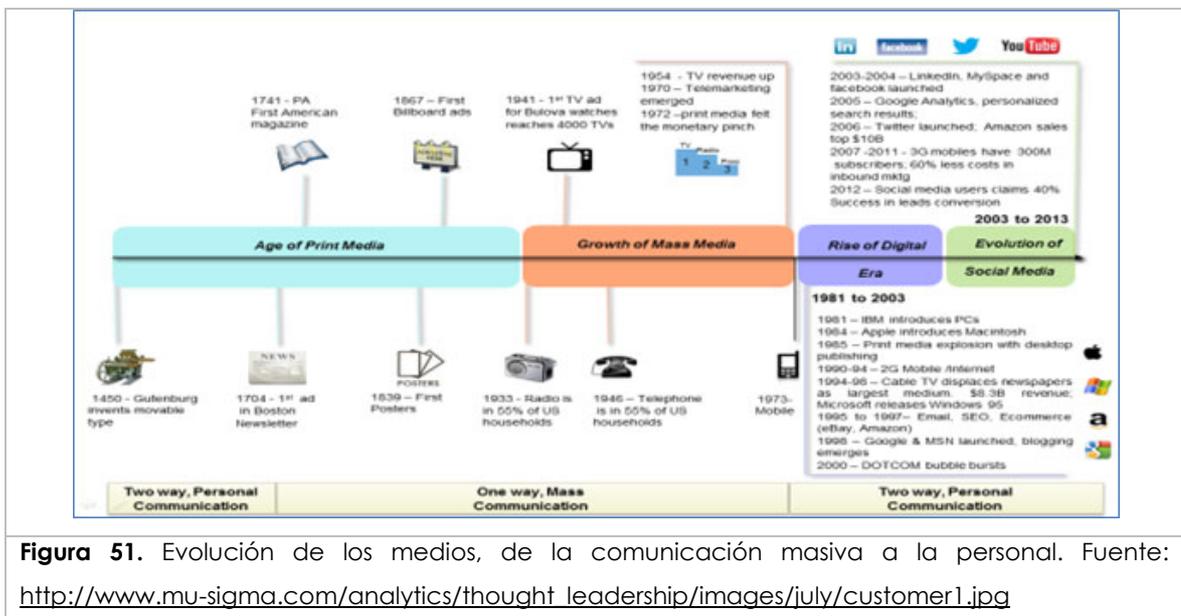


Figura 51. Evolución de los medios, de la comunicación masiva a la personal. Fuente: http://www.mu-sigma.com/analytics/thought_leadership/images/july/customer1.jpg

Por eso decimos que uno de los desafíos de esta tesis ha sido el desarrollo y actualización del modelo. Recordemos que fue ideado a principios de la década de

2001, y desde la síntesis de Bennet en 2003 no ha sido trabajado ni desarrollado por otros autores. Este vacío a partir de la elaboración del modelo ha supuesto que el modelo quedase desfasado. Una década después de la elaboración de la teoría, se da la necesidad de adaptación del modelo a la web actual en la que el espectro de los medios convencionales se ha multiplicado y el acceso a la información es completamente diferente a la de hace diez años. Concretamente, los principales motivos han sido:

1. **Aparición y dominio de los buscadores.** Ambos autores escriben entre 2001 y 2003, y no mencionan el papel de los buscadores. Además de no estar incluido en el modelo, ni siquiera se plantean su importancia como medio en la dispersión de las noticias en el medio Internet, hecho que desarrollamos en esta tesis.

Aunque por aquel entonces el buscador Google ya despuntaba frente a competidores como Altavista o Yahoo¹⁴⁵, los autores seguramente ni llegaron a imaginar el papel preponderante de Google en la red inmediatamente posterior. En todo caso, los buscadores, Google en particular, no fueron incluidos en el modelo.

Según la interpretación del modelo, una primera opción situaría a los buscadores en el Estrato Medio, ya que son ellos los principales “guardianes de las puertas” (del inglés “Gatekeepers”) de los contenidos de Internet. Sin embargo, a pesar de no tener una redacción y no producir informativos propios, una segunda opción nos lleva a situar a los buscadores dentro del estrato **Mass Media** a causa del papel preponderante que han venido ejerciendo en la web actual, y en el caso de Google por ejercer una influencia casi monopolística de acceso a la red.

2. **El crecimiento de la blogosfera y su papel preponderante.** La red de blogs a nivel mundial apenas había comenzado a desarrollarse cuando en el año 2001 se dio el caso de “Nike fábrica de explotación”. En los tres años siguientes, cuando fueron redactados los artículos, los autores empezaban a intuir cuál sería su papel en el desarrollo de la web social, pero todavía estaban muy lejos de la situación actual. Millones de blogs y bloggers profesionales han transformado el embrión que motivó la teoría.

¹⁴⁵En febrero de 2003, AltaVista fue comprada por Overture Services, Inc. En octubre de 2003, Overture fue adquirida por Yahoo! En agosto de 2004, poco después de la adquisición por parte de Yahoo!, el sitio web de AltaVista empezó a utilizar la tecnología del motor de búsqueda de Yahoo!. Con el acuerdo celebrado entre Microsoft y Yahoo! en julio de 2009, AltaVista comenzó a utilizar el motor de búsquedas Bing. <<http://es.wikipedia.org/wiki/AltaVista>>

3. **La revolución tecnológica de la web 2.0.** El término Web 2.0 fue planteado por Tim O'Reilly en una conferencia en 2004, y muchos de los desarrollos de Web 2.0 a los que estamos acostumbrados como las comunidades web, los servicios web, las aplicaciones Web, los servicios de red social, los servicios de alojamiento de videos, las wikis, blogs, mashups y folcsonomías, apenas ninguno de estas plataformas existían cuando fue planteado el modelo en 2001, de ahí la necesidad de incluirlas en un nuevo modelo que abarque toda esta nueva tipología de sitios web.

4. **La revolución social de la Web.** La nueva web social, surgida gracias al desarrollo tecnológicos de la web 2.0 fue apuntado por Peretti en su ensayo "Nike fábrica de explotación". Peretti, en su cita a las "redes sociales" hace referencia a la blogosfera y en 2001 apenas pudo llegar a intuir lo que serían Twitter o Facebook, los sitios web estrella de la web social actual. Al igual que los buscadores, la web social como desarrollo de la web 2.0 tampoco existía, por lo que han tenido que ser agregados al modelo para actualizarlo a la realidad de la web de nuestra década.

2.2. Metodología

El análisis que presentaremos se ha centrado en las informaciones de actualidad que podían encontrarse a través de Google coincidiendo con los procesos electorales que tuvieron lugar en España durante el periodo 2010-2012.

La recolección de datos para la investigación, efectuado durante cada uno de los procesos electorales, fue llevada a cabo a través de la captura de resultados de buscadores como Google, utilizando conjuntos de **palabras claves**. Para cada una de las elecciones se ha delimitado un conjunto específico de tales palabras clave o consultas que hemos intentado que sean representativas de cada proceso electoral.

Estos conjuntos de palabra clave son los que han permitido la extracción, para cada proceso electoral, de diferentes conjuntos de **páginas de resultados**¹⁴⁶ (SERP, a partir de ahora) compuestas a su vez por un total de diez resultados. Como es sabido, estos

¹⁴⁶ Denominadas SERP en la bibliografía técnica internacional, por las siglas Search Engine Result Page.

resultados son enlaces que apuntan hacia diferentes informaciones alojadas o publicadas en sitios web bajo el formato (generalmente) de páginas web, es decir, de documentos codificados con el lenguaje (X)HTML.

Estos resultados son, primero filtrados por las mencionadas palabras clave, y después jerarquizados algorítmicamente con el objetivo de presentarlos ordenados en función de su relevancia. Ya hemos explicado en otros apartados que la relevancia consiste en la adecuación teórica de cada resultado a la consulta formulada por el usuario.

Cabe señalar que esta relevancia se supone que tiene grados (no es del tipo todo o nada) y por tanto, algunos resultados serán muy relevantes (en tanto por ciento, por ejemplo, algunos tendrán un 90 o un 100 por cien de relevancia) pero otros serán muy poco relevantes (tendrán, por ejemplo, un 10 por ciento de relevancia), mientras que otros, por supuesto se situarán en zonas intermedias. El ordenamiento por relevancia entonces sitúa primero los documentos más relevantes, de modo que si estamos viendo la primera SERP estamos viendo los 10 documentos más relevantes según el buscador.

De esta manera, los diez resultados de cada SERP, vienen determinados, por un lado, por las palabras clave que seleccionan los usuarios para representar su necesidad de información¹⁴⁷, y por otro, por el sistema de ranking con el que los buscadores ordenan los resultados.

En nuestro caso, a través de las muestras que han representado las SERP, se ha buscado aplicar el modelo teórico de **Infoesfera** a la web actual, desarrollando el modelo apuntado y convirtiéndolo en un modelo de análisis global capaz de categorizar toda la esfera web, tanto la formada por sitios web y blogs, así como las páginas asociadas a los llamado 'social media'.

Para llevar a cabo esta categorización hemos utilizado los conceptos de **conglomerado** y **sector**. El conglomerado está inspirado en el análisis estadístico de conglomerados, también llamado de clústers. En esta tesis entendemos por 'conglomerado' a un conjunto de sitios web que presenten un nivel de similitud o una vinculación temática muy alta. De forma similar a la técnica estadística, el conglomerado es la unidad que permite ordenar los sitios web en grupos que

¹⁴⁷Parte de esta selección de palabras clave fue se corresponde a búsquedas reales de usuarios de Google, y está respaldada por datos extraídos de la Google Keywordtool que muestran cifras reales de búsqueda.

comparten puntos de asociación, como pueden ser todos aquellos sitios web que dependen de una administración estatal específica (Estatad, municipal, autonómica, foral, etc.) o que están vinculados a un partido político determinado. La yuxtaposición de diferentes conglomerados afines conforman unidades mayores que denominamos **sectores**; de esta manera, todos los conglomerados de las administraciones estatales conformarán el sector de la 'Administración', mientras que todos los conglomerados pertenecientes a cada uno de los diferentes partidos políticos (PP, PSOE, IU, etc) estarán englobados en el sector de la 'Esfera política'.

Este análisis de conglomerados y sectores puede hacerse de forma global abarcando todo el conjunto de palabras clave, o de forma particular en algún subgrupo temático del conjunto mediante muestras de palabras clave con alguna vinculación temática, como por ejemplo búsquedas de tipo electoral que incluyan términos específicos como podrían ser el nombre de una formación política o de un candidato. Una vez separados los resultados de un subconjunto temático se puede estudiar la composición de los sitios web o **conglomerados** que componen los resultados y compararlos con los resultados del conjunto de palabras clave u otros subconjuntos temáticos.

A efectos de nuestro análisis, por encima de las unidades de conglomerado y sector, hemos utilizado el concepto de los tres **Estratos**: superior, medio y micro. El Estrato Superior de acuerdo con el modelo que seguimos es siempre el que está formado por los Medios de Comunicación, mientras que los componentes de los otros dos Estratos pueden corresponder a actores diferentes según el tema de análisis. A continuación los presentamos de forma un poco más detallada:

- **Estrato Superior.** En este grupo consideramos los medios pertenecientes a empresas de comunicación tales como el grupo Prisa, Vocento, Godó, etc. Respecto a este grupo, se estudia por tanto cómo es y qué caracteriza a la visibilidad de los grandes medios convencionales de prensa a través de las SERP, y esto para cada una de las series temporales.
- **Estrato Medio.** Es el actor principal, y por tanto es el grupo principal de sitios web del que se estudia la visibilidad en la Infoesfera, tanto en resultados de lo que denominaremos *web search* (búsquedas en páginas de sitios web) como de *blog search* (búsquedas páginas de blogs). Hemos considerado que forman

parte de este estrato los siguientes actores: entes de la Administración, Universidades y partidos políticos.

- **Estrato Micro.** Para este tercer grupo hemos considerado la conveniencia de agrupar los sitios denominados '**social media**' (p.e. Twitter y Facebook serían los más representativos), y los sitios vinculados a movimientos como el denominado "15M".

2.3. Técnicas e Instrumentos de Recolección de Datos

La recolección de datos ha consistido en un procedimiento automatizado mediante el cual se lanzaron las consultas representadas por grupos de palabras clave y se recopilaban los resultados presentes en las SERP de los buscadores analizados. Para cada proceso electoral se utilizaron en torno a las 44 palabras claves y se recopilaban datos durante 14 días. Los detalles concretos se muestran más adelante.

Como resultado de este proceso de captura automatizado se han compilado más de **1,5 millones** de URL, aparecidas en SERPs de Google, Yahoo y Bing en los procesos electorales del periodo de investigación y que conforman de este modo la **población o universo** original objeto de estudio.

La técnica de extracción de datos utilizada fue el denominado "*web scraping*" de resultados de buscadores. *Web scraping* es el término técnico con el que se denominan las diferentes técnicas de obtención de datos procedentes de páginas web.

Para nuestra investigación, la herramienta utilizada ha sido el programa **Advanced Web Ranking** (AWR)¹⁴⁸. Una vez capturadas las SERPs, éstas fueron posteriormente exportadas desde la herramienta AWR en formato .csv y tratadas con el programa de hoja de cálculo Excel para la identificación de conglomerados.

Los más de 30.000 dominios recuperados fueron clasificados en categorías temáticas dando lugar a **conglomerados homogéneos**, y fueron añadidos a una base de datos **Access** junto al histórico de posiciones de URLs de cada dominio para cada proceso electoral. La tabla siguiente muestra las principales categorías de cada uno de los tres estratos:

¹⁴⁸Advanced Web Ranking <<http://www.advancedwebranking.com/>>

ESTRATO SUPERIOR
1.A.Mass_Media_Prensa.1Nacional_1Estatal.
1.A.Mass_Media_Prensa.1Nacional_2Regional.
1.A.Mass_Media_Prensa.1Nacional_3Económica.
1.A.Mass_Media_Prensa.1Nacional_4Blog.
1.A.Mass_Media_Prensa.1Nacional_5Deportiva.
1.A.Mass_Media_Prensa.2Televisión.
1.A.Mass_Media_Prensa.3Radio.
1.A.Mass_Media_Prensa.4Digital.
1.A.Mass_Media_Prensa.5Internacional.
1.A.Mass_Media_Prensa.5Internacional_Deportiva.
1.A.Mass_Media_Prensa.5Internacional_económica.
1.B.Mass_Media_Internet.Google.
1.B.Mass_Media_Internet.Microsoft.
1.B.Mass_Media_Internet.Terra.
1.B.Mass_Media_Internet.Yahoo.

ESTRATO MEDIO	
2.A.Administración.Admón_Estatal.	2.A.Esferapolítica.EQUO.
2.A.Administración.Admón_Foral.	2.A.Esferapolítica.ERC.
2.A.Administración.Admón_Generalitat.	2.A.Esferapolítica.Humanista.
2.A.Administración.Admón_Generalitat_Valencia.	2.A.Esferapolítica.Izquierda Unida-ICV.
2.A.Administración.Admón_Junta_Andalucia.	2.A.Esferapolítica.Izquierda Unida.
2.A.Administración.Admón_Junta_Extremadura.	2.A.Esfera política.PA.
2.A.Administración.Admón_Municipal.	2.A.Esferapolítica.PNV.
2.A.Administración.Admón_Provincial.	2.A.Esferapolítica.PSC.
2.A.Administración.Admón_Regional.	2.A.Esferapolítica.PSOE.
2.A.Administración.Admón_Xunta_Galicia.	2.A.Esferapolítica.PXC.
2.A.Administración.Organismos oficiales.	2.A.Esferapolítica.Partido castellano.
2.A.Administración.Portal CIS.	2.A.Esferapolítica.Partido_Popular.
2.A.Administración.Portal INE.	2.A.Esferapolítica.Pirata.
2.A.Administración.Portal Junta electoral.	2.A.Esferapolítica.Portal Independencia.
2.A.Educación.Portal educación.	2.A.Esferapolítica.Portal electoral.
2.A.Educación.Universidad.	2.A.Esferapolítica.Portal político.
2.A.Esferapolítica.Animalista.	2.A.Esferapolítica.Portal_Encuestas.
2.A.Esferapolítica.BILDU.	2.A.Esferapolítica.Reagrupament.
2.A.Esferapolítica.BNG.	2.A.Esferapolítica.Sindicatos.
2.A.Esferapolítica.CUP.	2.A.Esferapolítica.Solidaritat.
2.A.Esferapolítica.CiU.	2.A.Esferapolítica.UPN.
2.A.Esferapolítica.Ciudadanos.	2.A.Esferapolítica.UPYD.
2.A.Esferapolítica.Coalición Canaria.	2.A.Esferapolítica.Ultraderecha.
2.A.Esferapolítica.Comunista.	

ESTRATO MICRO	
3.A.Blogosfera.Blog político.	3.A.Movimientoprotesta.Acampada.
3.A.Blogosfera.Blogs CiU.	3.A.Movimientoprotesta.DRY.
3.A.Blogosfera.Blogs Cs.	3.A.Movimientoprotesta.Hipoteca.
3.A.Blogosfera.Blogs ERC.	3.A.Movimientoprotesta.Indymedia.
3.A.Blogosfera.Blogs ICV.	3.A.Movimientoprotesta.Movimiento_protesta.
3.A.Blogosfera.Blogs IU.	3.A.Red_Social.Facebook.
3.A.Blogosfera.Blogs Independentista.	3.A.Red_Social.Portal ciudadano.
3.A.Blogosfera.Blogs PA.	3.A.Red_Social.Portal foro.
3.A.Blogosfera.Blogs PP.	3.A.Red_Social.Twitter.
3.A.Blogosfera.Blogs PSC.	3.A.Red_Social.Web Social.
3.A.Blogosfera.Blogs PSOE.	3.A.Red_Social.Wiki.
3.A.Blogosfera.Elecciones-2010com.	3.A.Red_Social.Wikipedia.
3.A.Blogosfera.Empresarial_Empresas.	3.A.Red_Social.Yahoo.
3.A.Movimiento protesta.15m.	3.A.Red_Social.Youtube.

Figura 52. Tabla de conglomerados homogéneos desarrollada en la fase metodológica

Las fichas anteriores muestran las diferentes categorizaciones en las cuales se agruparon las URL recuperadas. De este modo, para nosotros cada categoría (ie. '3.A.Red_Social.')

y sus subcategorías (ie. '3.A.Red_Social.Facebook.' o '3.A.Red_Social.Youtube.')

constituyen lo que denominamos un **sector** y sus diversos **conglomerados** de páginas web.

Por tanto, una vez los sitios web fueron categorizados en clases de sitios o conglomerados, se cruzó la tabla resultante con la **población** o **universo original** de las 1,5 millones de URLs capturadas en los diversos procesos electorales y se seleccionó una muestra de **784.800** URLs correspondiente al subconjunto de URL aparecidas en el buscador Google en sus opciones de búsqueda Websearch, Blogsearch y Google News. Los resultados de las capturas de Yahoo y Bing se desagregaron y se han utilizado para el estudio introductorio a los resultados de investigación.

De forma paralela, gracias a la reducción del universo de estudio en casi un 50%, se logró limitar el peso de los archivos a menos de 20 MB, lo que redujo los tiempos de análisis.

Ahora bien, aunque el proceso de captura de datos profundizó hasta la tercera SERP, y por lo tanto hasta los primeros 30 resultados (Top30), el análisis para esta investigación se ha llevado a cabo utilizando los resultados de la primera SERP, es decir de los

primeros diez resultados, lo que nos ha permitido reducir más el universo de estudio situándolo en algo más de 250 mil URL (exactamente **253.928** URLs).

El análisis estadístico descriptivo de este subconjunto de URLs revela la evolución de los indicadores de mención del Top10 durante los días de campaña electoral, mostrando la visibilidad de cada conglomerado dentro de la esfera de información, y permitiendo hacer análisis específicos de dominios y URLs específicas.

2.4. Técnicas de Procesamiento y Análisis de Datos

Como hemos señalado, la recolección de datos se ha realizado mediante técnicas de **scraping**¹⁴⁹ de resultados de buscadores utilizando grupos de palabras clave de temática electoral. Las búsquedas se han realizado en los motores **Google**, **Yahoo** y **Bing**. En cada caso, se configuró el buscador con las siguientes opciones:

- Búsqueda: web
- Idioma: español
- País: España

Además se hicieron las búsquedas en los índices:

- Noticias
- Blogs

El instrumento principal de captura ha sido la herramienta **Advanced Web Ranking**, cuyas funcionalidades han permitido exportar los datos a Excel como instrumento para almacenar y tratamiento primario de la información. Para un tratamiento de segundo nivel de los datos se ha utilizado el editor de textos **TextPad**, el programa de hoja de cálculo **Excel** y la base de datos **Access** del paquete ofimático Office 2007, y el paquete estadístico **GNU PSPP**¹⁵⁰.

¹⁴⁹ El Web Scraping se puede considerar como un proceso que analiza la información proporcionada por un sitio web, se enmarca en una técnica llamada *data minning* o minería de datos.

¹⁵⁰GNU PSPP <<https://www.gnu.org/software/pspp/>>

2.5. Selección de nichos temáticos: Uso de palabras clave

A efectos de esta investigación consideramos que una palabra clave, o más frecuentemente, un grupo de palabras clave relacionadas temáticamente, determina lo que denominaremos un nicho temático (sinónimo: nicho de búsqueda).

Por motivos prácticos, a su vez, el nicho temático puede tener una denominación que no necesariamente se corresponde con las palabras clave utilizadas (aunque como es lógico habrá una correspondencia conceptual). Como se ha indicado en el punto 2.2, el nicho de búsqueda principal es el que denominamos 'procesos electorales', vinculado operacionalmente al término 'elecciones', palabra clave recurrente en todo los conjuntos de búsqueda utilizados en las diversas elecciones objeto de análisis.

Para definir los nichos específicos de cada proceso electoral, se han añadido palabras clave específicas propias de cada una de las elecciones, conformando de este modo un conjunto de búsquedas informativas¹⁵¹ relativas a cada proceso en particular.

La tipología de palabras clave utilizadas se puede dividir en cuatro clases principales:

1. Búsquedas **genéricas**, cuando incluyen el término 'Elecciones' junto a algún atributo relativo a cada uno de los procesos electorales estudiados, como puede ser el año o el parlamento
2. Búsquedas relativas o bien a **candidatos** o bien a **partidos políticos**, cuando incluyen el nombre de políticos o siglas de partidos políticos.
3. Búsquedas **geográficas**, si incluían alguna provincia o zona geográfica específica.
4. Búsquedas **específicas**, cuando incluían necesidades informativas específicas relativas a los procesos electorales, como podían ser búsquedas relativas a la intención de encontrar sondeos, encuestas o resultados.

Todos los conjuntos de palabras clave utilizados en las diferentes elecciones son idénticos en su temática, y están compuestos por proporciones similares (idénticas en el caso de las dos elecciones a la Generalitat de Catalunya) lo que nos han permitido establecer las comparativas¹⁵².

¹⁵¹La tipología tradicional de búsquedas se divide en tres grupos: navegacionales, informativas, o transaccionales.

¹⁵²Las proporciones de tipología de palabras clave son idénticas en el caso de las dos elecciones a la Generalitat de Catalunya, lo que permite hacer una comparación directa entre los resultados de 2010 y 2012.

2.6. Acotación y Selección de las unidades de tiempo

Las series temporales de resultados en torno a una palabra clave permiten observar la evolución de las esferas longitudinalmente a lo largo de los 14 días del periodo de campaña, y transversalmente a largo de los seis procesos electorales.

Las unidades de tiempo del estudio se reflejan en la siguiente tabla:

Elecciones estudiadas	Calendario de scraping de días de campaña/paquete/sesión		
Elecciones al Parlamento de Cataluña de 2010	12 días de campaña	+Cita electoral de 28-nov-10	+1 Jornada post electoral
Elecciones municipales de España de 2011	12 días de campaña	+Cita electoral de 22-may-11	+1 Jornada post electoral
Elecciones generales de España de 2011	12 días de campaña	+Cita electoral de 20-nov-11	+1 Jornada post electoral
Elecciones al Parlamento de Andalucía de 2012	12 días de campaña	+Cita electoral de 25-mar-12	+1 Jornada post electoral
Elecciones al Parlamento de Galicia de 2012	12 días de campaña	+Cita electoral de 21-oct-12	+1 Jornada post electoral
Elecciones al Parlamento de Cataluña de 2012	11 días de campaña	+Cita electoral de 25-nov-12	+2 Jornadas post electorales

Figura 53. Tabla de elecciones y calendario de scraping

2.7. Muestreo por conglomerados

Para realizar el análisis de resultados se ha dividido el universo en diversas unidades menores, primero en los tres estratos correspondientes al modelo de Infoesfera definido por W. L. Bennett en *Political Communication, Citizenship and Democracy: From the Mass Society to Personal Information Networks (2004)*.

- 1. Estrato Superior.** Este estrato engloba resultados de sitios web de medios de comunicación. En este estrato se han clasificado un total de **3.014** dominios, tanto españoles como internacionales.
- 2. Estrato Medio.** El segundo estrato engloba sitios web destacados, revistas electrónicas y grupos de presión. En este estrato se han clasificado un total de **11.954** dominios.
- 3. Estrato Micro.** El tercer estrato, correspondiente a la blogosfera, *social media* y comunidades online engloba un total de **15.436** dominios.

Además, y como veremos a continuación, ha sido necesario dividir los estratos en subniveles que denominamos **coronas** y **sectores**, permitiendo categorizar los dominios por afinidades y temáticas, agrupándolos en conglomerados homogéneos.

El análisis de resultados de nuestro estudio se centra en los **conglomerados** más significativos para nuestro estudio, siendo el objeto de investigación principal una selección de conglomerados correspondientes a las tres esferas: **Medios de prensa nacionales** del Estrato Superior, **páginas web de partidos** del Estrato Medio, y webs del **movimiento protesta, blogs políticos** y resultados de **redes sociales** provenientes del Estrato Micro.

Estos tres conglomerados centran el **análisis principal** de la tesis, si bien el resto de conglomerados pueden ser objeto de muestras específicas con el propósito de analizar aspectos propios de cada uno de los conglomerados en relación con los principales.

2.8. Composición de los estratos y principales conglomerados de la muestra

Para tener una visión global de la naturaleza de los datos capturados, recordemos que el total de capturas que conforman el **universo** objeto de estudio son aproximadamente **1.500.000** URL, pertenecientes a un total de **33.085** sitios web diferentes detectados durante el proceso de *scraping* de SERPs.



Cada uno de estos 33.085 sitios web ha sido adscrito dentro de uno de los tres estratos del modelo, y han sido categorizadas en alguno de los conglomerados detectados durante el desarrollo del modelo teórico.

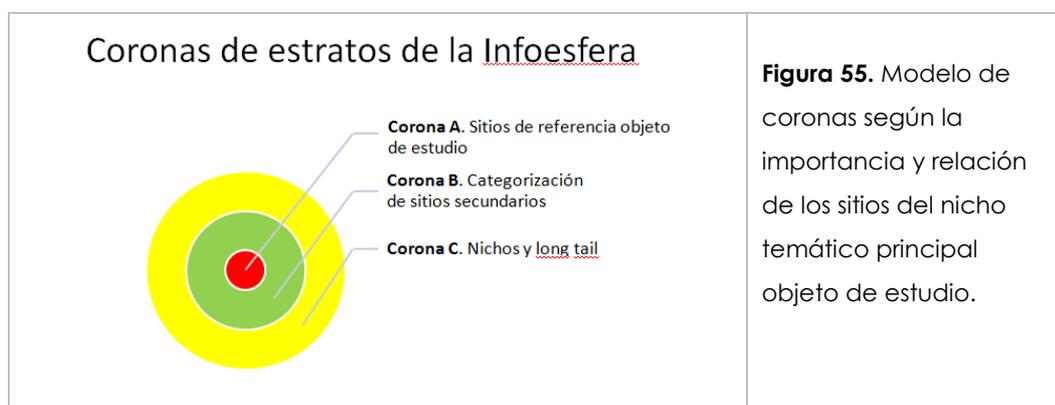
A continuación veremos los principales conglomerados¹⁵³, y datos relativos a sus tamaños, si bien los datos sobre sus proporciones y variación en las diferentes capturas de datos los veremos en el apartado de análisis de resultados.

Además de la clasificación en **estratos** y **conglomerados**, se ha añadido un nivel extra de categorización. A la división inicial en tres esferas se ha añadido una subdivisión en '**coronas**', unidades más pequeñas (menores que los estratos) para facilitar la observación y la representación gráfica.

>>Subdivisión > Estratos - Coronas

Cada uno de los tres estratos del modelo identificados con números (1, 2, 3) está dividido en **dos** o **en tres coronas** que identificamos con letras (A, B, C), de modo que tenemos las coronas **1.A** o **1.B** en el primer estrato y así sucesivamente según la importancia de los conglomerados dentro de la muestra.

De acuerdo con esto, **las coronas A** (ie. las coronas 1.A, 2.A y 3.A) están compuestas siempre por los principales conglomerados objeto de estudio. En el primer estrato son **3.486** medios de comunicación convencionales y digitales. En cambio, la corona A del segundo estrato son **3.039** sitios web (Administración, Educación y Política), mientras que en el tercer estrato son **1.923** sitios de la blogosfera, del movimiento protesta y redes sociales.



¹⁵³Las categorías aquí mostradas pertenecen a una categorización de sitios del Top30. Al aplicar el modelo a resultados del Top10, existen conglomerados que no tuvieron visibilidad en el Top10 aunque sí la tuvieron en posiciones más altas de hasta Top30.

Las coronas B están compuestas por conglomerados de segundo grado de importancia; en el primer estrato, encontramos **128 sitios** de alguno de los principales portales de Internet: Google, Microsoft o Yahoo (+Terra). Por su parte, en el Estrato Medio abarcan el resto de sitios web que no están directamente relacionados con el objeto de estudio. Este nivel incluye diferentes conglomerados que engloban colectivos diversos con un total de **2.375 sitios web**. En el tercer estrato encontramos una amplia clasificación de **4.166 blogs** agrupado en conglomerados que comparten características temáticas o de afinidad.

Las coronas C abarcan el resto de casos. En el primer estrato no existe nivel C. En el segundo estrato abarca diversos **nichos** de búsqueda (**8.040 sitios web**) de estudios preliminares (consultoras, bufetes de abogados empresas del IBEX, comercio electrónico, etc); en el tercer estrato engloba un total de **10.172 blogs** aparecidos en SERPs provenientes de plataformas de publicación de blogs (Wordpress, Blogspot u otras), que por su envergadura se han clasificado juntos.

A continuación presentamos una descripción más detallada de la composición de los estratos según la distribución presentada.

>>Subdivisión > Estratos – Coronas – Conglomerados del Estrato Superior

Las webs de medios aparecidas durante los diferentes comicios electorales se han etiquetado como **1.A.** Este primer estrato dominante es el que más se asemeja a la definición original de J. Peretti.

Este segmento incluye un total de **3.460** sitios los medios de prensa agrupados en conglomerados en función de su nacionalidad y temática, distinguiendo los de cobertura *estatal* y *regional*, la prensa *económica* y *deportiva*, además de los *blogs* publicados bajo el dominio del medio. Los mass media convencionales se completan con los sitios web de *Televisión* y *Radio* y la prensa internacional.

Los diferentes conglomerados de la corona principal del primer estrato son:

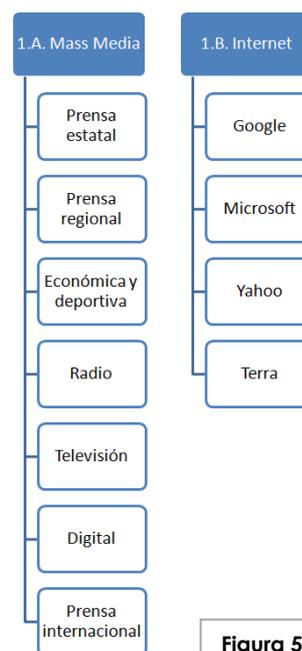


Figura 56.

>1.A.Nacional_1Estatad. Este primer gran conglomerado incluye todos los medios de prensa de ámbito nacional e incluye 73 sitios pertenecientes a 17 medios nacionales. Este subgrupo es el más importante y el que más visibilidad alcanza en Google debido a la cantidad y calidad de los contenidos generados. Ha sido el más fácil de categorizar dado su bajo número. Cada uno de ellos tiene un número determinado de hosts temáticos localizados durante las diferentes capturas de datos.

>1.A.Nacional_2Regional. Este conglomerado incluye 460 medios de prensa regional de los que se han localizado hasta 641 hosts distintos. Abarca todos los medios de prensa regionales españoles; en este grupo se han clasificado las versiones digitales de todos los medios de cabecera provinciales, y que al igual que el conglomerado anterior, además pueden clasificarse por grupos mediáticos (fuera de la clasificación principal).

>1.A.Nacional_3Económica. El conglomerado de prensa económica incluye un total de 44 sitios web pertenecientes a medios de comunicación y cibermedios económicos.

>1.A.Nacional_4Blog. En este grupo se han clasificado un total de 72 blogs, todos los resultados pertenecientes a blogs de periodistas que publican dentro del dominio de un medio convencional, y por tanto comparten la misma IP. Estos blogs podrían haber sido clasificados junto al resto de blogs como medios personales de micromedia, sin embargo se ha optado por agruparlos en este primer estrato ya que los autores son periodistas profesionales.

>1.A.Nacional_5Deportiva. Este conglomerado incluye un total de 14 medios deportivos nacionales o regionales así como cibermedios de temática deportiva.

>1.A.Televisión. El conglomerado de televisión incluye un total de 217 sitios web tanto españoles como internacionales, entre los que se cuentan todos los canales nacionales y regionales, así como canales de emisión en internet.

>1.A.Radio. Este conglomerado incluye 198 sitios web de emisoras de radio tanto españolas como internacionales.

>1.A.Digital. Este es el mayor conglomerado de este estrato y abarca 1.330 cibermedios que tienen redacciones propias y generan producción periodística digital.

>**1.A.Prensa Internacional.** Este conglomerado reúne un total de 789 sitios de prensa internacionales que incluyen 80 medios de prensa económica y 2 de prensa deportiva, pudiendo dividir este conglomerado en tres subgrupos principales: medios de prensa ordinaria, medios de prensa deportiva y medios de prensa económica. También existe la posibilidad de diferenciarlos por nacionalidad a través del dominio de primer nivel (TLD¹⁵⁴), pero no es objeto de este estudio. Por ese motivo todos los medios no españoles aparecen clasificados dentro del mismo grupo.

La corona B. A diferencia de la clasificación inicial propuesta por Peretti, se ha creado un grupo de portales que engloba a los tres principales buscadores de Internet: Google, Microsoft y Yahoo. En este caso no nos referimos a ellos en tanto a buscadores, sino por el hecho de que los tres mantienen portales con servicios y recursos de actualidad, entendiendo el término portal como el sitio compuesto por numerosos recursos procedentes de otros sitios. Aquí se han clasificado **128 sitios** de servicios o de algún portal nacional correspondiente a las grandes compañías de **internet**, incluyendo en este grupo al portal *Terra* como sitio web de referencia a nivel hispano parlante por su historia, relevancia y posicionamiento. En este grupo destaca el papel de los gatekeepers, por eso un apartado separado entre el Estrato Medio y micro y los medios convencionales.

>>Subdivisión > Estratos – Coronas – Conglomerados del Estrato Medio

2. Estrato Medio. El modelo teórico original (y punto de partida de esta tesis) incluiría aquí, junto a las revistas electrónicas y grupos de presión, los blogs destacados de la temática central del análisis, sin embargo se ha decidido agrupar todos los blogs (incluidos los políticos) en el tercer estrato. Por lo tanto, los componentes de este Estrato son una decisión fruto de los análisis e impuesta por las diferentes tipologías web.

De esta manera, en esta corona principal hemos incluido los sitios web directamente relacionados con el objeto de estudio, es decir, de la **administración** y el **sector político** de la Infoesfera, incluyendo sitios web de **educación** como dependientes de la administración. Esta división se comenta a continuación.

¹⁵⁴ TLD son las siglas de Top Level Domain. Corresponde a dominios tipo .es, .fr o .com, .org, .edu; etc.

>**2.A.Administración** incluye **553** páginas web de cualquier dominio de la administración pública española. Con el objetivo de estudiar el comportamiento de este tipo de páginas dentro del este campo temático virtual, se han categorizado de forma exhaustiva cualquiera de los sitios de las administraciones estatales incluyendo portales educativos y de cualquiera de las universidades españolas. Cada una de las administraciones estatales tiene un grupo de páginas web asociadas que le son propias. Existen portales dedicados por parte de las administraciones para el seguimiento de los comicios, hosts dedicados.

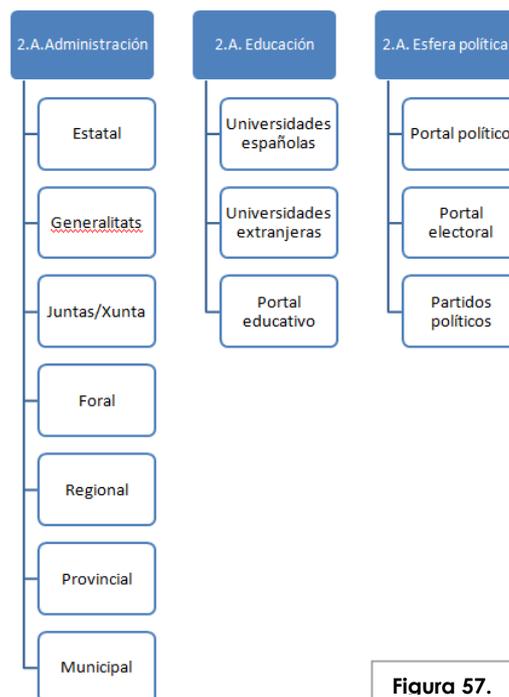


Figura 57.

>**2.A.Educación** abarca **667** universidades nacionales y extranjeras y portales educativos.

>**2.A.Esfera política** incluye **1.819** portales sectoriales de temática política y electoral, tanto sitios web de cualquier partido político español como páginas electorales de campaña. Los Social Media Strategists de cada partido han tenido que ir creando sitios web para sus candidatos, así como para las diferentes elecciones. Algunos de estos sitios web no han sido renovados una vez pasadas las elecciones, como veremos en el análisis de tipología web de partidos.

En una corona secundaria o **Corona B (2.B.)** se han clasificado todas aquellas páginas web relevantes que no pertenecían al campo temático virtual de información de actualidad política pero que sin embargo eran devueltas en respuesta a las preguntas con la que se realizó la toma de datos. Esta corona del segundo estrato abarca un total de **2.168** sitios web. Cada uno de estos conglomerados podría ser la categoría principal, y estar en una corona principal o Corona A de estar directamente relacionados con el tema de su campo temático virtual.



Dentro de la Corona B se han agrupado los sitios que no están directamente relacionados con el estudio principal de la tesis utilizando dos denominaciones principales, 'Colectivos' y 'Nichos'. Figura 58.

En el grupo denominado '**Colectivos**' es un bloque heterogéneo donde se han agrupado todas las páginas web vinculadas por una temática común. Los datos similares para la elaboración de estos listados han sido las categorías de sitios web localizados en diferentes directorios web. Se pueden establecer tantos colectivos como grupos temáticos se puedan distinguir entre los resultados aparecidos.

En un tercer segmento (**2.C.**) de **8.247** sitios web aparecen nichos de búsqueda relacionados con otros ámbitos temáticos. Aquí encontramos los diferentes nichos estudiados a lo largo del perfeccionamiento de la metodología. Son estudios colaterales de identidad corporativa y visibilidad en buscadores específicos¹⁵⁵, o conglomerados que han aparecido a lo largo de todas las capturas de datos y han alcanzado masa crítica para crear su propio grupo temático. Cada uno de estos estudios parciales ha dado lugar a la localización de sitios de poca visibilidad que por su poca frecuencia no han sido clasificados. Estos sitios se engloban en un conglomerado común denominado '**Long tail**' (en origen existían tantos 'long tails' como estudios parciales de la web, sin embargo se han agrupado en un gran conglomerado).

>>Subdivisión > Estratos – Coronas – Conglomerados del micro-estrato

3. Estrato Micro. La definición original de J. Peretti decía que el micro estrato está compuesto por el correo electrónico, listas de distribución y blogs personales. Con el fin de actualizar el modelo, en el que primero Peretti y luego Bennet incluyó la tipología

¹⁵⁵ A lo largo de este proyecto de tesis se han realizado multitud de estudios de visibilidad web que han permitido estudiar determinados campos temáticos virtuales tales como el nicho de búsqueda de **Abogados** (nicho centrado en el estudio de visibilidad de bufetes de abogados para búsquedas transaccionales del campo temático virtual jurídico), **consultoría IT** (búsquedas navegacionales utilizando el nombre de empresas de IT. El objetivo es el seguimiento del portal trabajobasura.info y su visibilidad asociada a estos nombres de empresa), **directorios** (cada uno de los nichos ha dado lugar a un listado de directorios generalistas o específicos para cada nicho; se ha optado por englobar todos los directorios aparecidos en cualquiera de los ensayos en un conglomerado común), **editorial** (diversos estudios centrados en la visibilidad del campo temático virtual del mundo editorial: lanzamiento de novedades editoriales, identidad digital, visibilidad de sitios web de librerías), y **bolsa** (casos de identidad digital corporativa sometidos a seguimiento a lo largo de los últimos tres años).

de contenidos correspondiente a páginas del tipo UGC (del inglés, 'User Generated Content') y entre las que destacaban los blogs personales, hemos añadido las páginas de contenidos que corresponden a comunidades online y redes sociales. Dado que el modelo teórico fue expuesto en el año 2004, cuando el auge de las redes sociales todavía despuntaba tímidamente, Bennet no añadió a este grupo los resultados provenientes de redes sociales como Facebook o Twitter y otros sitios web colaborativos como Wikipedia. En este desarrollo del modelo nosotros sí hemos añadido este tipo de resultados a este estrato (en cambio no hemos considerado el papel del correo electrónico, que queda totalmente fuera de este estudio).

Dentro del micro estrato hemos seguido la misma lógica de segmentación observada hasta ahora y por tanto, en una **primera corona (3.A.)** aislamos los conglomerados correspondientes a:

>**Blogosfera política.** Sitios web relativos a comicios específicos, sitios web de seguimiento político, barómetro, portales políticos, portales de encuestas electorales.

>**Blogosfera vinculada a partidos.** Blogs vinculados con partidos políticos pero sin ser páginas web oficiales. En este apartado han sido detectados 976 blogs. Consiste, por ejemplo, en blogs de autores afines a partidos que aparecen listados a veces en las páginas oficiales de los partidos. Por tanto, incluye blogs de simpatizantes y colaboradores.

>**Eleccions-2010.com.** Este caso solo aparece en el primero de los comicios, siendo una de las hipótesis de trabajo iniciales. Como se podrá observar, la visibilidad alcanzada por este sitio web, lanzado a propósito de las elecciones y con el principal objetivo de aparecer en el ranking, posicionó por encima de cualquier otro blog de político o de partido. Las cotas de visibilidad estuvieron a la altura de medios regionales y superaron a sitios especializados con una mayor visibilidad para palabras clave de amplio uso.

>**Web protesta.** Dentro de este sector encontramos 389 sitios web vinculados a distintos

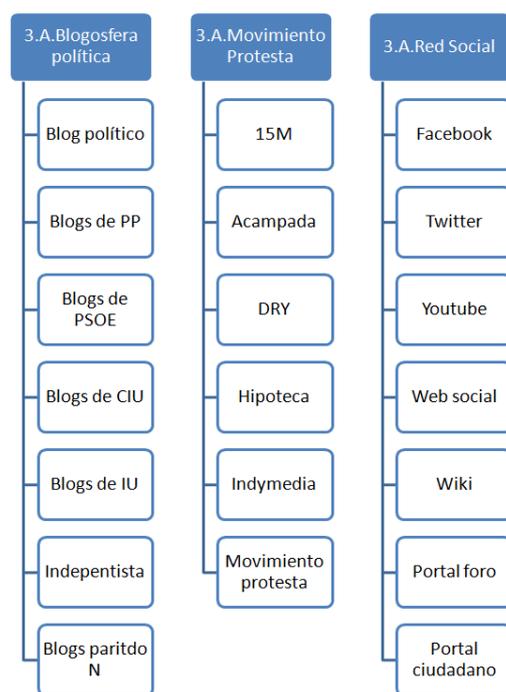


Figura 59.

movimientos populares entre los que destacan el '15M' y 'DemocraciaRealYa'.

>**Redes sociales.** Destacan con nombre propio resultados provenientes de Facebook, Twitter o Youtube, pero como veremos los conglomerados que mayor visibilidad fueron foros, wikis y portales ciudadanos.

En una **segunda corona (3.B.)** se han clasificado blogs de diferente temática, desde blogs de cultura, a fotoblogs personales, pasando por blogs empresariales o de entretenimiento.

En la tercera corona (**3.C.**) se han agrupado todos los blogs pertenecientes a plataformas de publicación de blogs. Los conglomerados más destacados son los de la red Blogspot y la red Wordpress.

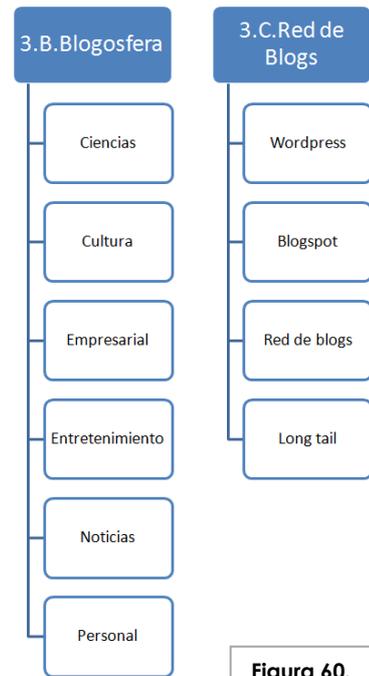


Figura 60.

2.9. Scraping de datos

Con el objetivo de mapear el modelo de Infoesfera, se ha aplicado la categorización anteriormente explicada a los resultados de las siguientes capturas de datos:

Elecciones	Generalitat 2010	Municipales 2011	Generales 2011	Junta Andal. 2012	Xunta de Galicia 2012	Generalitat 2012
Periodo de captura	Noviembre de 2010	10 – 23 de marzo	8 – 21 de noviembre	13 – 26 de Marzo	7 – 22 de Octubre	Noviembre de 2012
Web Search	17.400	17.760	17.640	18.050	18.900	18.900
Blog Search	11.620	5.990	6.191	6.020	6.300	12.600
Google News	12.029	17.918	17.640	18.060	18.850	12.060
Totales	41.049	41.668	41.471	42.130	44.050	43.560

Figura 61. Tabla de resultados de extracción de resultados de SERPs

Web Search. Todos los paquetes de datos son regulares. Incluyen los resultados del Top10 para paquetes de aprox. 42 búsquedas en las tres opciones de búsqueda de Google (web, idioma y país) según se indicó más arriba

Blog Search. Los tamaños de paquetes de datos son proporcionales al número de búsquedas utilizado, si bien en las elecciones catalanas se incluyen dos motores de búsqueda, Blog search Español y Blog search català.

Google News. Estas capturas corresponden a un motor de búsqueda único, pero su tamaño es tres veces mayor al incluir tablas de datos que recogen hasta el Top30 de resultados.

2.9.1. Scraping de datos - Parlament de Catalunya 2010

Periodo de captura		10 – 30 de Noviembre de 2010	
Motores de búsqueda		Búsquedas	
1	Google Blog Catalan (Catalan)	Tipo de búsqueda	Generalitat
2	Google Blog España (Español)		2010 %
3	Google España	1. Genérica	14 32%
4	Google España (Español)	2. Candidatos políticos	0 -
5	Google España (Español)	3. Partido político	16 36%
6	Google España Catalan (España)	4. Geográfica	0 -
7	Google News España	5. Encuestas	12 27%
		6. Resultados	0 -
		7. Temática	2 5%
		Total	44
Conjunto de palabras clave:			
elecciones ERC elecciones ICV elecciones PP Cataluña elecciones catalanas encuesta elecciones cataluña elecciones cataluña 2010 elecciones ciu elecciones ciutadans elecciones encuesta Cataluña elecciones encuesta parlament elecciones encuestas Cataluña elecciones encuestas parlament elecciones generalitat elecciones parlament elecciones parlamento cataluña elecciones parlamento cataluña 2010 elecciones psc elecciones socialistes eleccions Catalunya eleccions Catalunya 2010 eleccions ERC eleccions ICV		eleccions PP Cataluña eleccions catalanes encuesta eleccions ciu eleccions ciutadans eleccions encuesta eleccions encuesta Catalunya eleccions enquestes eleccions enquestes Catalunya eleccions generalitat eleccions parlament eleccions parlamentcatalunya eleccions parlamentcatalunya 2010 eleccions psc eleccions socialistes gencat elecciones gencat eleccions partidos elecciones cataluña 2010 partits eleccions catalunya 2010 sondeig eleccions catalunya sondeos elecciones cataluña votos elecciones cataluña vots elecciones catalunya	
Selección final de resultados en Top10 para análisis			
Web Search		17.400	
Blog Search		11.620	
Google News		12.029	
Figura 62. Tabla de Scraping de datos - Parlament de Catalunya 2010			

2.9.2. Scraping de datos - Municipales de España 2011

Periodo de captura		10 – 23 de marzo de 2011	
Motores de búsqueda		Búsquedas	
1	 Bing España (España)	Tipo de búsqueda	Municipales
2	 Google Blog Catalan (Catalan)		2011
3	 Google Blog España (Español)		%
4	 Google España	1. Genérica	18 42%
5	 Google España (España)	2. Candidatos políticos	3 7%
6	 Google España (Español)	3. Partido político	0 0%
7	 Google España Catalan (España)	4. Geográfico	8 19%
8	 Google News España	5. Encuestas	7 16%
9	 Yahoo España (España)	6. Resultados	4 9%
		7. Temática	3 7%
		Total	43
Conjunto de palabras clave:			
2011 elecciones candidatos PP 2011 candidatos PSOE 2011 elecciones 2011 elecciones 2011 candidatos elecciones 2011 encuestas elecciones 2011 fecha elecciones 2011 municipales elecciones autonomicas elecciones electorales elecciones electorales municipales elecciones España elecciones locales 2011 elecciones municipales elecciones municipales 2007 elecciones municipales 2011 elecciones municipales 2011 fecha elecciones municipales Asturias elecciones municipales Barcelona elecciones municipales Caceres elecciones municipales Santa Cruz de Tenerife		elecciones municipales Sevilla elecciones municipales Toledo elecciones municipales Valencia elecciones municipales Zaragoza elecciones municipales en España elecciones municipales España 2011 elecciones regionales encuesta elecciones 2011 encuestas 2011 encuestas elecciones 2011 encuestas electorales encuestas politicas fecha elecciones municipales 2011 municipales 2011 programa electoral proximas elecciones municipales resultado elecciones municipales resultados elecciones resultados elecciones municipales resultados elecciones municipales 2007 sondeos elecciones municipales 2011 últimas elecciones municipales	
Selección final de resultados en Top10 para análisis			
Web Search		17.760	
Blog Search		5.990	
Google News		17.918	

Figura 63. Tabla de Scraping de datos – Municipales 2011

2.9.3. Scraping de datos - Generales España 2011

Periodo de captura		8 – 21 de noviembre de 2011	
Motores de búsqueda		Búsquedas	
1	Bing España (España)	Tipo de búsqueda	Elecciones
2	Google Blog Catalan (Catalan)		Generales
3	Google Blog España (Español)	1. Genérica	8 18%
4	Google España	2. Candidatos políticos	8 18%
5	Google España (España)	3. Partido político	14 32%
6	Google España (Español)	4. Geográfica	
7	Google España Catalan (España)	5. Encuestas	5 11%
8	Google News España	6. Resultados	1 2%
9	Yahoo España (España)	7. Temática	8 18%
		Total	44
Conjunto de palabras clave:			
2011 elecciones BNG elecciones generales CIU elecciones generales Coalición Canaria elecciones generales ERC elecciones generales ETA elecciones Izquierda Unida elecciones generales PNV elecciones generales PP elecciones PP elecciones generales PSOE elecciones PSOE elecciones generales Partido Pirata elecciones generales Partit Pirata elecciones generales Rajoy elecciones Rubalcaba elecciones UPYD elecciones generales candidatos PP 2011 candidatos PSOE 2011 elecciones 2011 elecciones 2011 candidatos elecciones 2011 encuestas		elecciones 2011 fecha elecciones electorales elecciones España elecciones generales 2011 elecciones generales 2011 20N elecciones generales 2011 España elecciones generales crisis deuda elecciones generales crisis deuda soberana europea elecciones generales crisis económica elecciones generales economia elecciones generales paro encuesta elecciones 2011 encuestas elecciones 2011 encuestas electorales encuestas politicas listas PP elecciones generales listas PSOE elecciones generales ministro José Blanco programa electoral resultados elecciones	
Selección final de resultados en Top10 para análisis			
Web Search		17.640	
Blog Search		6.191	
Google News		17.640	

Figura 64. Tabla de Scraping de datos – Generales 2011

2.9.4. Scraping de datos - Parlamento de Andalucía de 2012

Periodo de captura	13 – 26 de Marzo de 2012		
Motores de búsqueda	Búsquedas		
1  Bing España (España)	Tipo de búsqueda	Andalucía	
2  Google Blog Catalan (Catalan)		2012	%
3  Google Blog España	1. Genérica	16	36%
4  Google España	2. Candidatos políticos	2	5%
5  Google España (España)	3. Partido político	7	16%
6  Google España (Español)	4. Geográfico	6	14%
7  Google España Catalan (España)	5. Encuestas	9	20%
8  Google News España	6. Resultados	2	5%
9  Yahoo España (España)	7. Temática	2	5%
	Total	44	
Conjunto de palabras clave:			
2012 elecciones autonomicas Andalucia 2012 candidatos PP 2012 candidatos PSOE 2012 elecciones Andalucia elecciones Andalucia 2012 elecciones Cordoba elecciones Granada elecciones Huelva elecciones Izquierda Unida Andalucia elecciones Jaen elecciones Junta elecciones Junta Andalucia elecciones Malaga elecciones PP Andalucia elecciones PSOE Andalucia elecciones Partido Andalucista elecciones Sevilla elecciones UPyD Andalucia elecciones andaluzas encuesta elecciones autonomicas Andalucia elecciones autonomicas Andalucia 2008	elecciones autonomicas Andalucia 2012 elecciones autonomicas Andalucia España 2012 elecciones autonomicas 2012 Andalucia españa elecciones electorales autonomicas Andalucia elecciones encuesta Andalucia elecciones encuesta parlamento andaluz elecciones encuestas Andalucia elecciones encuestas parlamento Andalucia elecciones locales 2012 elecciones parlamento Andalucia elecciones parlamento Andalucia 2012 encuestas elecciones 2012 fecha elecciones autonomicas Andalucia 2012 junta Andalucia elecciones partidos elecciones Andalucia 2012 programa electoral Andalucia 2012 proximas elecciones autonomicas Andalucia resultados elecciones autonomicas Andalucia resultados elecciones autonomicas Andalucia 2008 sondeos elecciones Andalucia sondeos elecciones autonomicas Andalucia 2012 votos elecciones Andalucia		
Selección final de resultados en Top10 para análisis			
Web Search	18.050		
Blog Search	6.020		
Google News	18.060		

Figura 65. Tabla de Scraping de datos – Andalucía 2012

2.9.5. Scraping de datos - Xunta de Galicia de 2012

Periodo de captura		7 – 22 de Octubre de 2012	
Motores de búsqueda		Búsquedas	
1	 Bing España	Tipo de búsqueda	Galicia 2012 %
2	 Bing España (España)	1. Genérica	15 34%
3	 Bing España (Español)	2. Candidatos políticos	5 11%
4	 Google Blog España	3. Partido político	6 14%
5	 Google España	4. Geográfica	6 14%
6	 Google España (España)	5. Encuestas	5 11%
7	 Google España (Español)	6. Resultados	3 7%
8	 Google News España	7. Temática	4 9%
9	 Yahoo España (España)	Total	44
Conjunto de palabras clave:			
Alberto Núñez Feijoo autonomicas Galicia 2012 candidato Feijoo PP candidato Jorquera BNG candidato Pachi Vázquez socialista candidatos parlamento Galicia elecciones elecciones 2012 Coruña elecciones 2012 Galicia elecciones 2012 Lugo elecciones 2012 Orense elecciones 2012 Ourense elecciones 2012 Pontevedra elecciones 2012 autonomicas Galicia elecciones BNG Galicia elecciones Galicia elecciones Galicia 2012 elecciones Izquierda Unida Galicia elecciones PSOE Galicia elecciones UPyD Galicia elecciones autonomicas Galicia elecciones autonomicas Galicia 2008 elecciones autonomicas Galicia 2012 elecciones autonomicas Galicia 2012 fecha		elecciones autonomicas gallegas elecciones electorales autonomicas Galicia elecciones encuesta parlamento gallego elecciones encuestas Galicia elecciones encuestas parlamento Galicia elecciones municipales Galicia elecciones parlamento Galicia elecciones parlamento Galicia 2012 elecciones parlamento gallego Galicia fecha elecciones autonomicas Galicia 2012 Galicia elecciones autonomicas Galicia 2012 elecciones autonomicas partidos elecciones Galicia 2012 programa electoral Galicia 2012 proximas elecciones autonomicas Galicia resultado elecciones autonomicas Galicia resultados elecciones autonomicas Galicia 2008 resultados elecciones autonomicas Galicia por provincia sondeos elecciones Galicia sondeos elecciones autonomicas Galicia 2012 votos elecciones Galicia últimas elecciones autonomicas Galicia	
Selección final de resultados en Top10 para análisis			
Web Search		18.900	
Blog Search		6.300	
Google News		18.850	

Figura 66. Tabla de Scraping de datos – Galicia 2012

2.9.6. Scraping de datos - Parlament de Catalunya 2012

Periodo de captura		Noviembre de 2012		
Motores de búsqueda		Búsquedas		
1	Google Blog Catalan (Catalan)	Tipo de búsqueda	2012	%
2	Google Blog España (Español)	1. Genérica	14	31%
3	Google España	2. Candidatos políticos	0	-
4	Google España (España)	3. Partido político	16	36%
5	Google España (Español)	4. Geográfica	0	-
6	Google España Catalan (España)	5. Encuestas	13	29%
7	Google News España	6. Resultados	0	-
		7. Temática	2	4%
		Total	45	
Conjunto de palabras clave:				
elecciones ERC elecciones ICV elecciones PP Cataluña elecciones catalanas encuesta elecciones cataluña elecciones cataluña 2012 elecciones ciu elecciones ciudadanos elecciones encuesta Cataluña elecciones encuesta parlament elecciones encuestas Cataluña elecciones encuestas parlament elecciones generalitat elecciones parlament elecciones parlamento cataluña elecciones parlamento cataluña 2012 elecciones parlamento cataluña 2012 sondeo elecciones psc elecciones socialistes eleccions Catalunya eleccions Catalunya 2012 eleccions ERC eleccions ICV		eleccions PP Cataluña eleccions catalanes encuesta eleccionsciu eleccions ciutadans eleccions encuesta eleccions encuesta Catalunya eleccions enquestes eleccions enquestes Catalunya eleccions generalitat eleccions parlament eleccions parlament catalunya eleccions parlament catalunya 2012 eleccions psc eleccions socialistes gencat elecciones gencat eleccions partidos elecciones cataluña 2012 partits eleccions catalunya 2012 sondeig eleccions catalunya sondeos elecciones cataluña votos elecciones cataluña vots elecciones catalunya		
Selección final de resultados en Top10 para análisis				
Web Search		18.900		
Blog Search		12.600		
Google News		12.060		

Figura 67. Tabla de Scraping de datos – Parlament de Catalunya 2012

2.10. Procedimiento para presentar e interpretar los datos

En el capítulo siguiente se mostrarán los **ratios de visibilidad** en SERPs de diferentes niveles de agrupación de la variable URL a través de gráficos y tablas generales y específicas. Estos ratios se estudiarán para cada uno de los procesos electorales de forma individual o de forma agrupada, es decir, tanto a través de estudios transversales de series temporales que incluyan los seis comicios electorales, como a través de estudios longitudinales centrados en los 14 días de estudio de una serie temporal de unas elecciones concretas.

De acuerdo con los procedimientos de trabajo discutidos hasta ahora, las agrupaciones de URL con las que trabajaremos son:

Estrato.

Este nivel es visualizado mostrando la agregación de todos los sitios web que pertenece a uno de los tres estratos de la Infoesfera, ya sea el **Estrato Superior**, el **Estrato Medio** o el **Estrato Micro**.

Corona.

Este nivel de agrupación se utiliza para diferenciar los sitios objeto de estudio (Corona A) de sitios no relacionados con el tema central de análisis.

Sector.

Es la principal subdivisión objeto de estudio de esta tesis y sirve para dividir los estratos en grupos homogéneos. En el caso del **Estrato Superior**, se han dividido los medios de comunicación de masas en dos sectores, el de empresas propietarias de medios convencionales y de cibermedios, y el de las grandes compañías de Internet.

En el **Estrato Medio**, dado el campo temático de estudio, se analizan los sitios web correspondientes a los sectores de la Administración, educación y partidos políticos.

En el **Estrato Micro**, se estudian en profundidad los sectores correspondientes a la blogosfera, el sector de la web protesta y los sitios de redes sociales.

Conglomerados.

Son agrupaciones de dominios y hosts caracterizados por tener una vinculación

sectorial o temática.

En el caso del **Estrato Superior**, la vinculación es sectorial ya que todos los sitios pertenecen a medios de comunicación. Los diferentes conglomerados muestran diferentes grados de afiliación vinculados a su cobertura o al tema principal tratado.

Dentro **Estrato Medio** los conglomerados por partido incluyen todos los sitios web relativos a un partido político (sitios web oficiales, páginas de campaña, y webs de las juventudes del partido). El tamaño de los conglomerados es variable y puede abarcar desde unos pocos hosts vinculados a unos pocos sitios web como es el caso de medios de internet, o englobar centenares de sitios web diferentes como es el caso de los conglomerados de 'Esfera política', cuya única vinculación es temática.

En el caso del **Estrato Micro**, los conglomerados engloban sitios agrupados por afinidad temática como la web protesta, por tipología de web, como las redes sociales, o por ser todos ellos hosts de un mismo sitio web, como es el caso de los blogs de las redes de Wordpress o Blogspot.

Dominios.

El dominio es una subunidad del conglomerado que engloba todas las URL publicadas bajo un mismo nombre de dominio. Los diferentes hosts que cuelgan dentro de un nombre de dominio se han tomado como unidades diferenciadas del dominio principal, pudiendo estar clasificadas, o no, dentro del mismo conglomerado.

URL.

La URL es la unidad básica y se estudia en casos particulares, en especial en el ciclo de vida y obsolescencia de contenidos.

Los diferentes ratios y distribuciones de frecuencia de **menciones** en SERPs de las diferentes agrupaciones se presentan en forma de **series temporales**. Por regla general estas series se componen de una selección de **catorce días** de datos para cada serie temporal. En todos los casos a excepción de uno, se han seleccionado doce días de campaña, el día de las elecciones y el inmediatamente posterior. El caso de las elecciones a la Generalitat, por falta de uno de esos días de campaña, se han seleccionado dos días después de las elecciones.

Los ratios agrupados en forma de **series temporales** nos permitirán visualizar el **porcentaje de visibilidad** de los diferentes niveles de agrupación, ya sea de los estratos, sectores, conglomerados o dominios; en las series temporales indicaremos las **tasas de crecimiento** o decrecimiento de visibilidad para las mismas variables, y además se visualizarán ratios y niveles de **variables específicas** tales como el número de blogs (Wordpress, Blogspot, de partido o de medio) en SERPs, el formato (pdf, infografía, etc), y los resultados de redes sociales como Facebook o Twitter, entre otras.

En la representación resumen de las series temporales trabajaremos con **números índice** tomando como referencia ratios centrales de muestra o los sitios con valores más altos en sus conglomerados.

En la **representación de datos** utilizaremos gráficos y un sistema de referencia (sombreado, colores, puntos, líneas, símbolos, números) y coordenadas de posicionamiento web que permiten representar la información cuantitativa.

El uso mayoritario de gráficos tiene un objetivo doble, primero como representación gráfica que permite plasmar grandes cantidades de datos que de otra manera serían tablas imposibles de visualizar, y segundo el medio más efectivo para describir y analizar la información.

Los resultados de investigación se muestran en los siguientes tipos de gráficos:

Gráficos de barras verticales o de columnas. Representan valores usando trazos verticales, aislados o agrupados, según muestren la variable/s seleccionada/s en una serie dentro de unos comicios específicos, o a un conjunto de series, caso en el que se usarán barras comparativas de los seis comicios electorales.

Gráficos de barras horizontales. Tienen el mismo uso que las barras verticales pero se utilizan en aquellos casos en el que la longitud de los nombres de sectores o conglomerados altera la visualización en el eje de las X.

Gráficos de barras proporcionales. Se usan cuando lo que se busca es resaltar la representación de los porcentajes de los datos que componen un total. Las barras

pueden ser Verticales u Horizontales y representan proporciones de conglomerados. Se utilizan para comprar estratos o conglomerados.

Gráficos de barras comparativas. Se utilizan para comparar dos o más series de sectores o conglomerados en diferentes elecciones, o para comparar valores de un conglomerado. Las barras pueden ser verticales u horizontales según si los textos correspondientes a cada categoría son muy largos.

Gráficos de barras apiladas. Se usan para mostrar las relaciones entre dos o más series con el total. Las barras pueden ser: verticales u horizontales. Y permiten comparar diferentes conglomerados dentro de un estrato a lo largo de las seis elecciones.

Gráficos de Líneas. En este tipo de gráfico se representan la composición de la Infoesfera y muestra los incrementos de los estratos y de los grandes sectores y conglomerados de cada estrato.

Se usan para representar ratios de estratos, sectores o conglomerados en las seis series temporales de cada una de las elecciones o una comparativa de elecciones; es decir, hay dos niveles de series temporales: los resultados diarios dentro de unas elecciones, o los resultados agrupados de unas elecciones comparados con la serie temporal de las 6 elecciones.

Circulares. Estos gráficos nos permiten ver la distribución interna de los datos de visibilidad en SERPs en forma de porcentajes sobre un total. Se pueden utilizar para visualizar gráficos de estrato o de conglomerado. En Google News y Blog search se agrupan los grandes grupos mediáticos y se muestran en gráficos circulares.

Áreas. En estos tipos de gráficos se busca mostrar la tendencia de la información generalmente en un período de tiempo. Se usarán para representar las series de estratos y conglomerados principales, en especial de medios de prensa y partidos políticos.

Gráficos Mixtos. En estos tipos de gráficos se representan dos o más series de datos, cada una con un tipo diferente de gráfico. Con estos gráficos resaltaremos la evolución de las series temporales.

A continuación pasaremos a ver la aplicación de esta categorización a las capturas de datos y establecer comparativas entre buscadores y entre comicios electorales. Veremos los principales conglomerados, y datos relativos a sus tamaños y proporciones, así como su variación en las diferentes capturas de datos.

El bloque de resultados de la investigación se compone de dos apartados:

1. Mostraremos una **comparativa** de los **tres motores de búsqueda** Google, Bing y Yahoo en las elecciones Generales 2011 y a la Xunta de Galicia 2012.
2. Mostraremos un **estudio monográfico** del motor de búsqueda **Google** y las seis elecciones seleccionadas del periodo 2010-12. En esta segunda parte se incluyen las figuras de las elecciones Generales de 2011 ya mostradas y se integran con las otras cinco elecciones objeto de estudio para formar una serie temporal. En esta parte de los resultados no se incluyen datos de Bing o Yahoo salvo en menciones expresas.

Capítulo 3. Análisis de resultados

3.1. Introducción: la Infoesfera en Google, Yahoo y Bing

En primer lugar vamos a mostrar, en forma de análisis comparativo, los resultados agregados obtenidos de los tres principales buscadores: Google, Bing y Yahoo, correspondientes a las elecciones Generales de 2011 y los comicios a la Xunta de 2012¹.

Como hemos indicado en el Capítulo 2, durante el proceso de captura de datos, los diferentes conjuntos de palabras clave se ejecutaban de manera simultánea en los diferentes motores de búsqueda seleccionados en la misma opción de búsqueda, país 'España', obteniendo datos en paralelo de todos ellos. Esta captura simultánea nos permite mostrar la composición de la Infoesfera en cada uno de los tres principales buscadores.

>Origen de los datos

En los casos de las elecciones **Generales** de **2011** y las autonómicas a la **Xunta** de **2012**, los paquetes de datos resultado del *scraping* de las SERP provienen de una misma serie temporal de 14 días. A diferencia del resto de análisis centrados en resultados de la primera SERP de Google, en esta comparativa de los tres motores de búsqueda los datos utilizados son el resultado de la agregación de las tres primeras SERP (**Top30**).

Por otro lado, para conseguir que la comparativa de los datos de estos dos procesos fuese lo más fidedigna posible, se han ajustado los grupos de SERP a estudiar, seleccionando de entre el total de palabras clave utilizadas en la captura, no solo el mismo número, sino idéntica **tipología temática** de **palabras clave**. Con tal objetivo, se han eliminado algunas de las palabras clave de la captura de datos original de uno u otro grupo, consiguiendo una composición homogénea y proporcional.

¹Se realizaron capturas en todos los buscadores para las elecciones Municipales 2011, las Generales 2011 y las de la Xunta de Galicia 2012. No se capturaron datos de los tres buscadores en el resto de elecciones. No se capturaron datos de Bing o Yahoo en las elecciones a la Generalitat 2010, ni en los comicios a la Junta de Andalucía 2012. El motor Yahoo fue desechado en las elecciones a la Generalitat 2012, cuando tras la compra de Yahoo Search por parte de Bing, los resultados de ambos buscadores empezaron a ser casi idénticos.

37.800		37.112	
URL	Elecciones Generales 2011	URL	Elecciones Xunta Galicia 2012
Tipo de búsqueda 'Genérico': 6			
1260	elecciones 2011	1260	autonomicas Galicia 2012
1260	elecciones españa	1230	elecciones 2012 autonomicas Galicia
1260	elecciones generales 2011	1260	elecciones 2012 Galicia
1260	elecciones generales 2011 20N	1262	elecciones autonomicas galicia
1260	elecciones generales 2011 españa	1260	elecciones autonomicas gallegas
1260	programa electoral	1260	elecciones parlamento Galicia 2012
Tipo de búsqueda 'Partido': 6			
1260	BNG elecciones generales	1260	elecciones BnG Galicia
1260	Coalición Canaria elecciones generales	1260	elecciones Izquierda Unida Galicia
1260	Izquierda Unida elecciones generales	1260	elecciones PSOE Galicia
1260	PP elecciones generales	1260	elecciones UPyD Galicia
1260	PSOE elecciones generales	1260	partidos elecciones Galicia 2012
1260	UPYD elecciones generales	1260	programa electoral Galicia 2012
Tipo de búsqueda 'Candidatos': 5			
1260	candidatos PP 2011	1260	alberto núñez feijoo
1260	candidatos PSOE 2011	1260	candidato feijoo pp
1260	elecciones 2011 candidatos	1260	candidato Jorquera BnG
1260	Rajoy elecciones	1260	candidato Pachi Vázquez socialista
1260	Rubalcaba elecciones	1260	candidatos parlamento Galicia elecciones
Tipo de búsqueda 'Encuestas': 5			
1260	elecciones 2011 encuestas	1260	elecciones encuesta parlamento gallego
1260	encuesta elecciones 2011	1260	elecciones encuestas Galicia
1260	encuestas elecciones 2011	1260	elecciones encuestas parlamento Galicia
1260	encuestas electorales	1260	sondeos elecciones autonomicas Galicia 2012
1260	encuestas politicas	1260	sondeos elecciones Galicia
Tipo de búsqueda 'Temática': 7			
1260	15m elecciones	630	15m elecciones Galicia
1260	elecciones 2011 fecha	1260	elecciones autonomicas Galicia 2008
1260	elecciones generales crisis deuda elecciones generales crisis deuda soberana europea	1260	elecciones autonomicas Galicia 2012 fecha
1260	elecciones generales crisis económica	1260	fecha elecciones autonomicas Galicia 2012
1260	elecciones generales economia	1260	proximas elecciones autonomicas Galicia
1260	elecciones generales paro	1260	últimas elecciones autonomicas Galicia
Tipo de búsqueda 'Resultados': 1			
1260	resultados elecciones	1230	resultado elecciones autonomicas Galicia

Figura 68. Selección de búsquedas de la comparativa Google-Bing-Yahoo

A continuación vamos a analizar los datos de los tres principales buscadores según los siguientes **niveles**:

1. Composición por **estratos**
2. Composición del **Estrato Superior**
3. Composición del **Estrato Medio**
4. Composición del **Estrato Micro**

> La composición por estratos en Google, Bing y Yahoo

Como hemos señalado, los resultados pertenecen a las *elecciones Generales* de noviembre de 2011 y a las elecciones a la *Xunta de Galicia* de octubre de 2012. Estas dos elecciones están separadas por 11 meses, periodo en el que asistimos a la irrupción en resultados de los efectos del algoritmo Penguin, lanzado en abril de 2012.

Los siguientes gráficos circulares nos permiten ver cuál fue la distribución interna de los datos de visibilidad en SERP en estos tres buscadores. El gráfico muestra la proporción de los tres estratos de los que se compone la Infoesfera en forma de porcentajes sobre el total de resultados capturados en cada buscador en la opción de búsqueda por país 'España'.

>Elecciones Generales – Noviembre de 2011

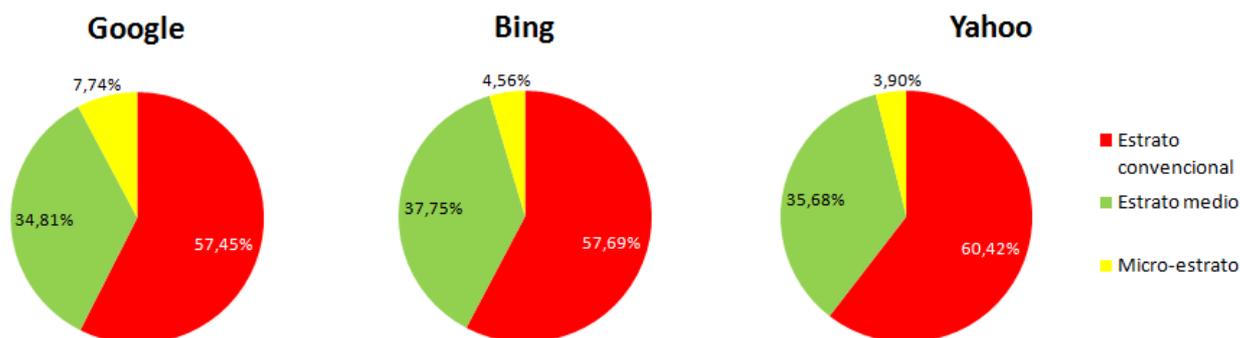


Figura 69. Comparativa Google-Bing-Yahoo – Elecciones Generales

- **Estrato Superior.** En las elecciones Generales, los tres buscadores presentan un **promedio** de visibilidad de resultados procedentes del primer estrato de un **58,5%**. Yahoo es el buscador que muestra más páginas web de medios en el Top10 con un 60,42%, seguido por Bing con un 57,69%, y por Google, con un 57,45%.
- **Estrato Medio.** El promedio de visibilidad del estrato en los tres buscadores es de un **36,1%**. Bing es el buscador que más páginas web de este estrato ofrece en el Top10 con un 37,75%, seguido por Yahoo con un 35,68%, mientras que Google tiene un 34,81%.
- **Estrato Micro.** El promedio de páginas del estrato es del **5,4%**. Google es el buscador que más páginas web de este estrato utiliza para mostrar en SERP con un 7,74%. Le siguen Bing con un 4,56% y Yahoo con un 3,9%.

>Elecciones a la Xunta de Galicia – Octubre de 2012

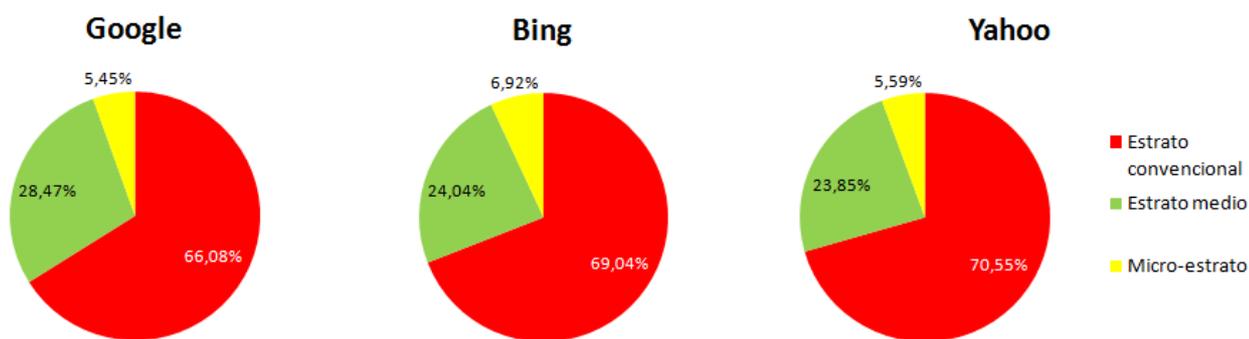


Figura 70. Comparativa Google-Bing-Yahoo – Elecciones Xunta de Galicia

- **Estrato Superior.** En las elecciones gallegas, la visibilidad del Estrato Superior creció en todos los buscadores pasando a representar un **promedio** del **68,6%**, diez puntos más que once meses antes durante las elecciones Generales (58,5%). En el buscador Google pasó a ser un 66,08% de los resultados de SERPs, con un incremento del 15% respecto a las elecciones Generales. De forma similar, en Bing su peso se incrementó un 20%, mientras que en Yahoo los resultados del primer estrato pasaron a un 70,55% del total, lo que supuso un incremento del 17%.
- **Estrato medio.** En las elecciones gallegas el Estrato Medio decreció en los tres buscadores, pasando a representar un **promedio** del **25,4%**. En Google el decremento fue de un 18%, un 36% en Bing y un 33% en Yahoo.
- **Estrato micro.** En estas elecciones la proporción del estrato en los tres buscadores se ajustó en torno a un **promedio** del **6%** del total de resultados. En Google alcanzó un 5,45% con una caída del 30% respecto a las Generales, mientras que la proporción en Bing y en Yahoo aumentó hasta un 6,92% y un 5,59%, suponiendo un incremento del 52% y el 44% respectivamente.

3.1.1. Composición del Estrato Superior

Este apartado muestra los principales sectores y conglomerados objeto de estudio del Estrato Superior, correspondiente al de los medios de comunicación.

Para su mejor visualización, se ha dividido este primer estrato en **dos grandes sectores** que contienen diversos conglomerados.

El **primer sector** engloba los principales conglomerados del estrato aparecidos en los tres buscadores, todos los sitios web de medios convencionales de prensa, medios de televisión y cibermedios. Este primer sector es el compuesto por los siguientes cuatro conglomerados:

- **1.A.Prensa Estatal.**
- **1.A.Prensa Regional.**
- **1.A.Televisión.**
- **1.A.Prensa Digital (Cibermedios).**

En relación a los anteriores, entendemos por Cibermedios a efectos de este análisis, los sitios de noticias que carecen de una versión analógica o impresa.

El **segundo sector** es el compuesto por los cinco conglomerados menores del primer estrato:

- **1.A.Prensa Económica.**
- **1.A.Blog de Prensa.**
- **1.A.Radio.**
- **1.A.Prensa Internacional.**
- **1.B.Internet.**

Para la visualización de los datos **no** se ha optado por la división en coronas debido al bajo peso relativo de los medios internet de la '**corona B**', que finalmente se muestran de forma agregada en el segundo sector. En el caso de '**1.A.Prensa Internacional.**' se ha optado por agregar en un mismo conglomerado la prensa internacional económica y deportiva.

Los datos se visualizan mediante barras verticales comparativas que muestran la proporción de cada conglomerado.

>Elecciones Generales >Estrato Superior>Sector conglomerados principales

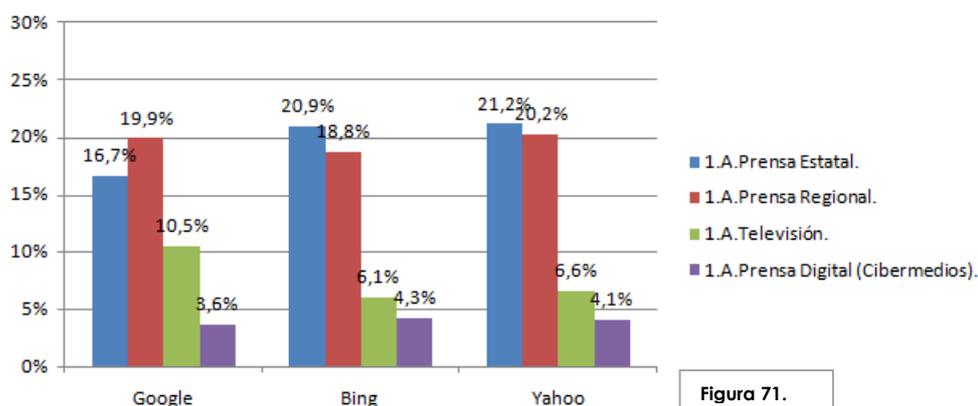


Figura 71.

La suma de estos cuatro conglomerados supuso más de la mitad de resultados aparecidos en las SERPs de Google (50,7%), Bing (50%) y Yahoo (52%)

1.A.Prensa Estatal. El peso de este conglomerado en la composición de SERPs de Google fue menor y se situó en el 16,7%, mientras que en Bing alcanzó el 20,9% y el 21,2% en Yahoo. Los sitios mejor posicionados en SERPs fueron:

- **Google:** elmundo.es (4,9%) | politica.elpais.com (3,12%) | abc.es (2,8%)
- **Bing:** elmundo.es (4,83%) | abc.es (3,3%) | 20minutos.es (2,85%)
- **Yahoo:** elmundo.es (4,64%) | abc.es (3,66%) | 20minutos.es (2,48%)

1.A.Prensa Regional. El 19,9% de resultados de Google pertenecían a este conglomerado frente al 18,8% de Bing y el 20,02% de Yahoo. El Top de medios regionales por buscador fue:

- **Google:** publico.es (1,56%) | lne.es (1,24%) | levante-emv.com (0,82%)
- **Bing:** diariocritico.com (1,82%) | diariodeibiza.es (1,2%) | lne.es (1%)
- **Yahoo:** diariocritico.com (1,67%) | diariodeibiza.es (1,64%) | laopinion.es (1,23%)

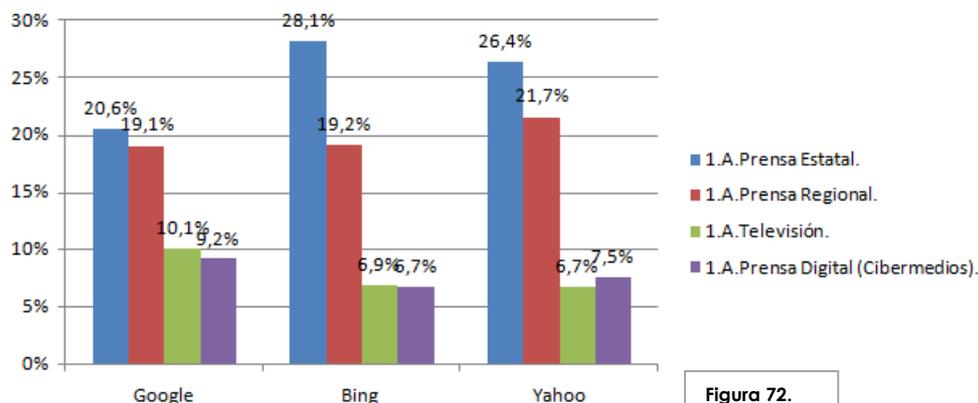
1.A.Prensa Televisión. Google es el buscador que más resultados de televisión mostró con un 10,5%, cifra que baja hasta el 6,1%-6,6% en sus competidores. El Top de medios televisión fue:

- **Google:** rtve.es (4,83%) | eitb.com (1,69%) | antena3.com (1,63%)
- **Bing:** rtve.es (3,77%) | intereconomia.com (0,73%) | telecinco.es (0,63%)
- **Yahoo:** rtve.es (3,79%) | antena3.com (0,78%) | telecinco.es (0,71%)

1.A.Prensa Digital (Cibermedios). Google mostró un 3,6% de cibermedios, frente al 4,3% de Bing y 4,1% de Yahoo. El Top de cibermedios fue:

- **Google:** periodistadigital.com (0,35%) | canarias7.es (0,32%) | noticias.juridicas.com (0,27%)
- **Bing:** canarias7.es (0,73%) | soitu.es (0,33%) | minutodigital.com (0,33%)
- **Yahoo:** canarias7.es (0,74%) | libertaddigital.com (0,29%) | soitu.es (0,28%)

>Elecciones a la Xunta de Galicia >Estrato Superior>Sector conglomerados principales



En las elecciones a la Xunta de Galicia estos conglomerados tienen una visibilidad 10 puntos mayor, pasando a representar la siguiente proporción: Google (59%), Bing (61%) y Yahoo (62,3%).

1.A.Prensa Estatal. Respecto a las Generales Google incrementó un 23,5% los resultados de prensa nacional en SERP, Yahoo un 24,9% y Bing un 34,5%. Este último buscador alcanzó el valor más elevado, con un 28,1% de resultados. El Top por buscador fue:

- **Google:** elmundo.es (4,98%) | abc.es (3,59%) | elpais.com (3,37%)
- **Bing:** elmundo.es (6,49%) | ccaa.elpais.com (5,42%) | 20minutos.es (3,43%)
- **Yahoo:** elmundo.es (6,35%) | ccaa.elpais.com (4,67%) | abc.es (3,48%)

1.A.Prensa Regional. Este conglomerado se mantuvo estable en los tres buscadores. El Top por buscador fue:

- **Google:** lavozdeg Galicia.es (6,56%) | mas.laopinioncoruna.es (1,76%) | elcorreogallego.es (1,71%)
- **Bing:** lavozdeg Galicia.es (5,81%) | elcorreogallego.es (3,61%) | mas.laopinioncoruna.es (1,69%)
- **Yahoo:** lavozdeg Galicia.es (5,46%) | elcorreogallego.es (4,36%) | laopinioncoruna.es (1,76%)

1.A.Prensa Televisión. Este conglomerado se mantuvo estable en los tres buscadores.

- **Google:** rtve.es (4,33%) | antena3.com (2,71%) | lasexta.com (1,12%)
- **Bing:** rtve.es (4,55%) | resultados-elecciones.rtve.es (1,35%) | intereconomia.com (0,52%)
- **Yahoo:** rtve.es (3,91%) | resultados-elecciones.rtve.es (0,86%) | antena3.com (0,51%)

1.A.Prensa Digital (Cibermedios). Los cibermedios fueron el conglomerado que más creció en los tres buscadores. En **Google** creció un 252% en comparación con las elecciones Generales, pasando del 3,2% al **9,2%**. Bing muestra incrementos del 156,8% con un 6,7% del total, mientras que Yahoo alcanza un 7,5% con un incremento del 183,5%. El Top por buscador fue:

- **Google:** periodistadigital.com (1,49%) | lainformacion.com (0,96%) | huffingtonpost.es (0,78%)
- **Bing:** periodistadigital.com (1,56%) | estrelladigital.es (0,68%) | libertaddigital.com (0,6%)
- **Yahoo:** periodistadigital.com (1,17%) | lainformacion.com (1,15%) | estrelladigital.es (0,65%)

>Elecciones Generales >Estrato Superior – Sector conglomerados secundarios

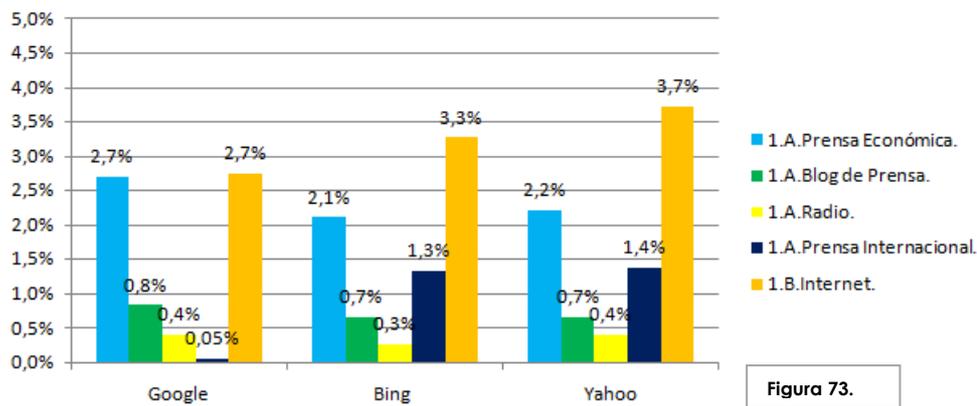


Figura 73.

La proporción en SERPs de estos cinco conglomerados menores fue: Google (6,7%), Bing (7,6%) y Yahoo (8,4%)

1.A.Prensa económica. Google fue el que más visibilidad dio a estos resultados con un 2,7%, mientras que en Bing fue del 2,1% y en Yahoo del 2,2%. El Top por buscador fue:

- **Google:** cincodias.com (1,26%) | eleconomista.es (0,6%) | expansion.com (0,21%)
- **Bing:** ecodiario.eleconomista.es (0,87%) | eleconomista.es (0,53%) | finanzas.com (0,47%)
- **Yahoo:** ecodiario.eleconomista.es (0,8%) | eleconomista.es (0,68%) | finanzas.com (0,45%)

1.A.Blog de prensa. Tuvieron una visibilidad entre el 0,8% en Google 0,7% en Bing/Yahoo. El Top por buscador fue:

- **Google:** blogs.periodistadigital.com (0,29%) | blogs.libertaddigital.com (0,21%)
- **Bing:** blogs.periodistadigital.com (0,41%) | blogs.libertaddigital.com (0,22%)
- **Yahoo:** blogs.periodistadigital.com (0,42%) | blogs.libertaddigital.com (0,22%)

1.A.Prensa Radio. Los sitios web de radio alcanzan una visibilidad muy baja que va desde el 0,3% en Bing hasta el 0,4% en Google y Yahoo. El Top por buscador fue:

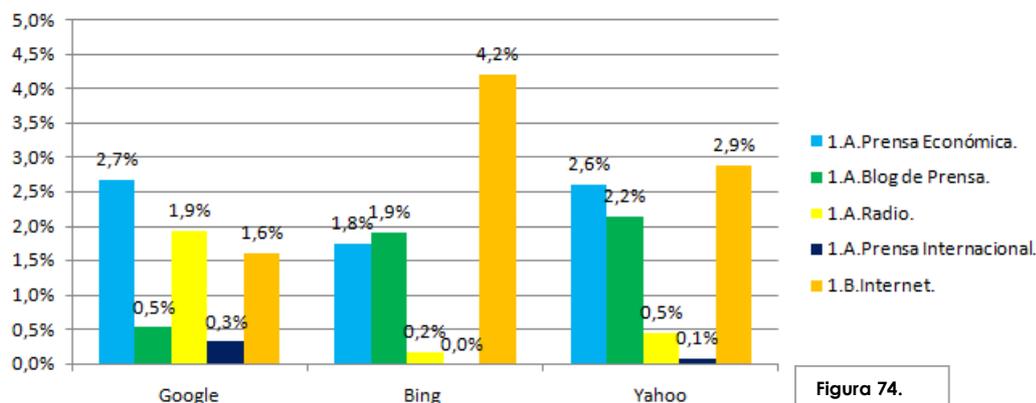
- **Google:** cadenaser.com (0,13%) | cope.es (0,1%) | radioguarena.com (0,05%)
- **Bing:** cope.es (0,12%) | eleccionescuantoantes.cope.es (0,1%) | ondacero.es (0,04%)
- **Yahoo:** eleccionescuantoantes.cope.es (0,18%) | ondacero.es (0,12%) | cope.es (0,1%)

1.A.Prensa internacional. Por tratarse de datos de la opción de búsqueda 'España', el resultado debería haber sido del 0%, sin embargo Bing y Yahoo con un 1,3% y 1,4% mostraron diversos cybermedios con tld '.ar', '.mx' y '.gt'. El 0,05% en Google proviene de resultados de <http://reuters.es>, clasificado como prensa internacional.

1.B.Internet. La proporción de resultados provenientes del portal Terra o de los portales de servicios de Google, Microsoft o Yahoo se situó entre el 2,7% y 3,7%. El Top por buscador fue:

- **Google:** noticias.terra.es (1,47%) | es.noticias.yahoo.com (0,77%) | top.terra.es (0,2%)
- **Bing:** noticias.terra.es (1,47%) | noticias.es.msn.com (0,94%) | es.noticias.yahoo.com (1,06%)
- **Yahoo:** noticias.terra.es (1,49%) | noticias.es.msn.com (1,06%) | es.noticias.yahoo.com (0,83%)

>Elecciones a la Xunta de Galicia >Estrato Superior – Sector conglomerados secundarios



Once meses más tarde estos conglomerados incrementan su visibilidad pasando a representar la siguiente proporción: Google (7,4%), Bing (9,1%) y Yahoo (8,7%).

1.A.Prensa económica. El Top por buscador fue:

- **Google:** expansion.com (0,74%) | labolsa.com (0,68%) | ecodiario.eleconomista.es (0,53%)
- **Bing:** ecodiario.eleconomista.es (0,88%) | labolsa.com (0,39%) | expansion.com (0,34%)
- **Yahoo:** ecodiario.eleconomista.es (1,24%) | expansion.com (0,62%) | labolsa.com (0,55%)

1.A.Blog de prensa. Los blogs de prensa pierden visibilidad en Google, que pasa del 0,61% al 0,47%, pero ven triplicada su visibilidad en Bing y Yahoo pasando a 1,74% y 1,98% respectivamente. El Top por buscador fue:

- **Google:** blogs.periodistadigital.com (0,48%) | blogs.lavozdegalicia.es (0,04%)
- **Bing:** blogs.lavozdegalicia.es (1,12%) | blogs.periodistadigital.com (0,67%)
- **Yahoo:** blogs.lavozdegalicia.es (1,14%) | blogs.periodistadigital.com (0,91%)

1.A.Prensa Radio. Los sitios web de radio siguen con una visibilidad muy baja que baja al 0,47% en Google y al 0,37% en Yahoo mientras que desaparecen por completo de Bing.

1.A.Prensa internacional. Al igual que en los resultados a las elecciones Generales, el resultado debería haber sido del 0% por tratarse de datos de la opción de búsqueda 'España'. El 0,3% de Google proviene del sitio <http://es.reuters.com>, mientras que el 0,1% de Yahoo pertenece al sitio web <http://americaeconomica.com>.

1.B.Internet. La proporción de resultados provenientes de los portales propios disminuyó en Google y Yahoo, creciendo en los resultados de Bing a través de recursos del portal MSN de Microsoft. El Top por buscador fue:

- **Google:** es.noticias.yahoo.com (0,71%) | noticias.terra.es (0,27%) | noticias.es.msn.com (0,31%)
- **Bing:** es.noticias.yahoo.com (1,87%) | noticias.es.msn.com (0,99%) | noticias.terra.es (0,59%)
- **Yahoo:** es.noticias.yahoo.com (1,57%) | noticias.es.msn.com (0,53%) | noticias.terra.es (0,4%)

3.1.2. Composición del Estrato Medio

Este apartado muestra los principales sectores y conglomerados del Estrato Medio. Los datos se visualizan mediante barras verticales comparativas que muestran la composición de cada conglomerado en cada uno de los tres buscadores.

En este apartado se han seleccionado los conglomerados pertenecientes a la '**Corona A**', y se muestran en un formato **serie** que permite su visión comparativa.

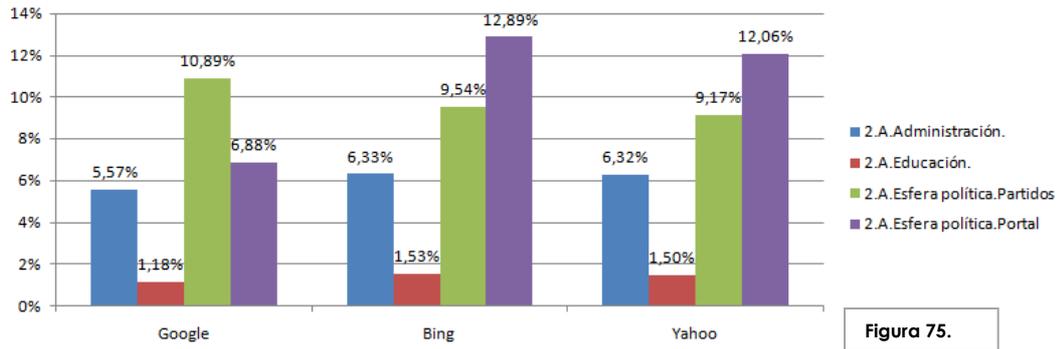
Los principales conglomerados de esta corona son los de 'Administración', 'Educación' y 'Esfera política. El conglomerado 'Administración' incluye todos los sitios web oficiales dependientes de alguna de las administraciones estatales españolas. El caso del conglomerado de 'Educación', que incluye todas las universidades españolas, podría haber sido categorizado como un subgrupo de 'Administración', sin embargo se decidió separarlo y crear un conglomerado propio y diferenciado. Por último, el sector de la 'Esfera política' sí ha sido dividido en dos conglomerados, el primero correspondiente a sitios web de algún partido político, y el segundo que incluye portales de temática política pero sin una vinculación nominal con ningún partido específico.

La **serie** está compuesta por los siguientes conglomerados del segundo estrato:

- **2.A.Administración.**
- **2.A.Educación.**
- **2.A.Esfera política.Partidos**
- **2.A.Esfera política.Portal**

Para la visualización de los datos los conglomerados de partidos políticos se muestran de forma agregada. En el caso de '**2.A.Esfera política.Portal**' se han sumado en el mismo conglomerado los portales políticos sin vinculación partidista, y todos los sitios web electorales.

>Elecciones Generales > Estrato Medio



En las elecciones Generales la suma de los conglomerados de la corona 2.A. supuso un promedio global del **27,95%**. Fue del 24,52% en Google, del 30,29% en Bing y del 29,04% en Yahoo.

2.A.Administración. Dentro del Estrato Medio, las administraciones públicas están representadas con un promedio del **6%**; la menor visibilidad fue en Google con un 5,57%, mientras que Bing y Yahoo fue del 6,3%. El Top por buscador fue:

- **Google:** Admón_Estatal. (2,06%) | Admón_Regional. (1,08%) | Portal CIS.(0,67%)
- **Bing:** Admón_Estatal. (2,40%) | Admón_Regional. (2,06%) | Portal CIS.(0,23%)
- **Yahoo:** Admón_Estatal. (2,48%) | Admón_Regional. (2,17%) | Portal CIS.(0,23%)

2.A.Educación. Los tres buscadores se movieron en una franja de visibilidad entre el 1,18% y el 1,53%.

2.A.Esfera política. Partidos. Esfera política de partido tiene en Google su máxima visibilidad con un 10,89%. Las cifras de Bing y Yahoo son similares y se mueven entre el 9,17% y el 9,54% El Top por buscador fue:

- **Google:** Partido_Popular. (3,60%) | PSOE. (2,91) | Izquierda Unida. (1,53%)
- **Bing:** Partido_Popular. (3,28%) | PSOE. (3,29) | Izquierda Unida. (1,29%)
- **Yahoo:** Partido_Popular. (3,13%) | PSOE. (3,17) | Izquierda Unida. (1,21%)

2.A.Esfera política. Portal. Por el contrario, los portales electorales no alcanzan en Google el 7%, mientras que en Bing y Yahoo la visibilidad está cercana al 12% del total. El Top por buscador fue:

- **Google:** Portal electoral. (5%) | Portal político. (1,87%)
- **Bing:** Portal electoral. (11,37%) | Portal político. (1,52%)
- **Yahoo:** Portal electoral. (10,7%) | Portal político. (1,36%)

>Elecciones a la Xunta de Galicia > Estrato Medio

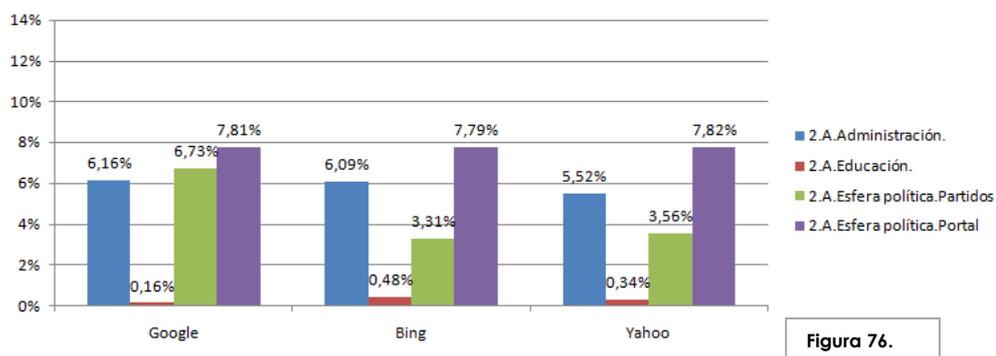


Figura 76.

En las elecciones a la Xunta de Galicia la proporción de los principales conglomerados de la corona 2.A. descendió de un promedio del 27,95% en las Generales, hasta un **18,59%**. Por buscador, las proporciones fueron en Google del 20,86%, en Bing del 17,67%, y en Yahoo del 17,24%.

2.A.Administración. Este sector alcanza mayor visibilidad en Google con un 6,16%. La visibilidad de los mismos conglomerados en los buscadores Bing y Yahoo es ligeramente menor, del 6,09% y del 5,52% respectivamente. El Top por buscador fue:

- **Google:** Admón_Estatal. (1,25%) | Xunta_Galicia. (1,02%) | INE.(1,09%) | CIS.(1,03%)
- **Bing:** Admón_Estatal. (1,9%) | Xunta_Galicia. (1,57%) | Portal INE.(0,28%) | Portal CIS.(0,51%)
- **Yahoo:** Admón_Estatal. (1,7%) | Xunta_Galicia. (1,39%) | Portal INE.(0,26%) | Portal CIS.(0,95%)

2.A.Educación. Los tres buscadores redujeron drásticamente su visibilidad que bajó hasta una franja de visibilidad entre el 0,16% y el 0,48%.

2.A.Esfera política. Partidos. Esfera política de partido sigue teniendo su máxima visibilidad en Google con un 6,73%. Las cifras de Bing y Yahoo son del 3,31% y del 3,56% respectivamente. El Top por buscador fue:

- **Google:** Partido_Popular. (1,26%) | PSOE. (1,41%) | Izquierda Unida. (0,91%)
- **Bing:** Partido_Popular. (0,62%) | PSOE. (0,28%) | Izquierda Unida. (0,82%)
- **Yahoo:** Partido_Popular. (0,48%) | PSOE. (0,38%) | Izquierda Unida. (0,84%)

2.A.Esfera política. Portal. En global, los portales electorales bajaron en visibilidad, que se situó en una franja cercana al 8%. En Google sufrieron un aumento de visibilidad del 13,5% situándose en torno a un 7,81% del total, mientras que en Bing y Yahoo descendió hasta el 7,79% y el 7,82% respectivamente, con descensos entre el 35-40%. El Top por buscador fue:

- **Google:** Portal electoral. (6,60%) | Portal político. (1,21%)
- **Bing:** Portal electoral. (7,72%) | Portal político. (0,07%)
- **Yahoo:** Portal electoral. (7,81%) | Portal político. (0,01%)

3.1.3. Composición del Estrato Micro

Los gráficos de este apartado muestran los principales sectores y conglomerados objeto de estudio del Estrato Micro.

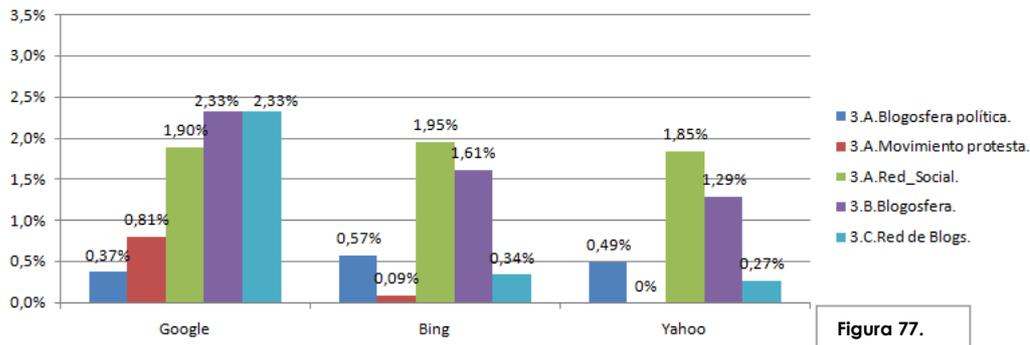
Se muestran los conglomerados pertenecientes a la '**Corona A**', principal objeto de estudio de este estrato, y se agregan los conglomerados secundarios de la '**Corona B**' y '**Corona C**'.

Los principales conglomerados son los blogs de partidos políticos, el conglomerado de 'Movimiento protesta', que incluye todos los sitios web y blogs relacionados con dicho movimiento, en especial sitios relacionados con el 15m, Democracia Real Ya, y afectados por la hipoteca. El último gran integrante de este estrato son las redes sociales, que se muestran de forma agregada.

La **serie** completa de conglomerados del tercer estrato son:

- **3.A.Blogosfera política.**
- **3.A.Movimiento protesta.**
- **3.A.Red_Social.**
- **3.B.Blogosfera.**
- **3.C.Red de Blogs.**

>Elecciones Generales > Estrato Micro



Este estrato es visualizado en forma de cinco conglomerados representativos. En las elecciones Generales, la suma de los conglomerados del tercer estrato supuso un promedio global del **5,4%**. Fue del 7,74% en Google, del 4,56% en Bing y del 3,9% en Yahoo, motor en el que destaca la ausencia de contenidos relacionados con el movimiento protesta.

3.A. Blogosfera política. Es un conglomerado menor. El buscador que mostró más visibilidad fue Bing, seguido por Yahoo. Google aparece en tercera posición con un 0,37% del total. El Top por buscador fue:

- **Google:** Blogs IU. (0,21%) | Blogs PSOE. (0,10%) | Blogs PP. (0,06%)
- **Bing:** Blogs IU. (0,57%)
- **Yahoo:** Blogs IU. (0,49%)

3.A. Movimiento protesta. Este tipo de página alcanza su mayor visibilidad en Google con un 0,81%, es residual en Bing, y no aparecen en las SERPs de Yahoo. El Top por buscador fue:

- **Google:** 15m. (0,12%) | DRY. (0,03%) | Movimiento_protesta. (1,77%)
- **Bing:** 15m. (0,06%) | DRY. (0,03%) | Movimiento_protesta. (0,03%)

3.A. Red_Social. Los tres buscadores muestran una proporción muy similar cercana al 2% del total de resultados. El Top por buscador fue:

- **Google:** Portal ciudadano. (0,26%) | Portal foro. (0,75%) | Web Social. (0,29%) | Wiki. (0,37%)
- **Bing:** Portal ciudadano. (0,45%) | Portal foro. (0,51%) | Web Social. (0,29%) | Wiki. (0,71%)
- **Yahoo:** Portal ciudadano. (0,42%) | Portal foro. (0,44%) | Web Social. (0,29%) | Wiki. (0,7%)

3.B. Blogosfera. Google es el buscador donde más peso tiene con un 2,3%. En Bing y Yahoo las cifras son de 1,6% y 1,3%.

3.C. Red de Blogs. Tiene mayor peso en Google con un 2,3%, mientras que en Bing y Yahoo es prácticamente residual con cifras que apenas alcanzan el 0,3%.

>Elecciones a la Xunta de Galicia> Estrato Micro

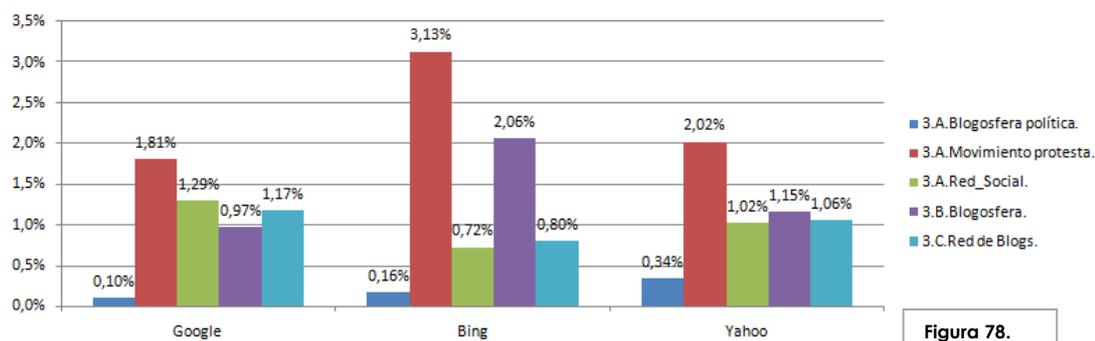


Figura 78.

La suma de los conglomerados del tercer estrato incrementó su visibilidad hasta un promedio global del **5,9%**; mientras que en Google bajó hasta el 5,33%, se incrementó hasta el 6,88% en Bing y el 5,59% en Yahoo.

3.A. Blogosfera política. Este conglomerado menor bajó su visibilidad en los tres buscadores, pasando del 0,48% de promedio en las elecciones Generales a un 0,2% en las gallegas. El Top por buscador fue:

- **Google:** Blogs PP. (0,1%)
- **Bing:** Blogs IU. (0,05%) | Blogs PP. (0,11%)
- **Yahoo:** Blogs IU. (0,13%) | Blogs PP. (0,21%)

3.A. Movimiento protesta. Este conglomerado fue el que más creció, pasando de ser residual a suponer un promedio del 2,37% de las SERPs en las elecciones gallegas.

3.A. Red_Social. Los tres buscadores muestran decrementos, siendo el mayor de ellos en el buscador Bing. El Top por buscador fue:

- **Google:** Portal ciudadano. (0,14%) | Portal foro. (0,69%) | Web Social. (0,1%) | Wiki. (0,2%)
- **Bing:** Portal ciudadano. (0,13%) | Portal foro. (0,48%) | Wiki. (0,11%)
- **Yahoo:** Portal ciudadano. (0,12%) | Portal foro. (0,61%) | Wiki. (0,1%)

3.B. Blogosfera. Google es el buscador donde más peso tiene con un 2,3%. En Bing y Yahoo las cifras son de 1,6% y 1,3%.

3.C. Red de Blogs. Aunque baja un punto porcentual, Google sigue siendo el buscador con mayor peso de este conglomerado. En los otros buscadores su peso se multiplica por dos o por cuatro, alcanzando un peso cercano al 1,17% de Google.

3.1.4. Evolución: de las Generales a la Xunta

La tendencia general observada al comparar los resultados de las elecciones Generales y las de la Xunta de Galicia en los tres buscadores nos permiten identificar una concentración de la mayoría de resultados de SERP en contenidos del Estrato Superior, pertenecientes a medios de comunicación convencionales, que incrementa su visibilidad entre 8 y 12 puntos porcentuales.

En relación a la evolución entre las elecciones Generales y las autonómicas de la Xunta, el *Estrato Medio*, que incluye todos los sitios directamente relacionados con las búsquedas, disminuye entre 8 y 11 puntos dependiendo del buscador.

El *Estrato Micro*, el identificado como "UGC" (por User Generated Content) y que engloba la blogosfera y todos los sitios de redes sociales, únicamente incrementa su visibilidad en Bing y Yahoo, mientras que retrocede 2,3 puntos en Google.

>Estrato Superior. Analizando los diferentes conglomerados vemos como la tendencia general es la de concentrar la mayoría de resultados de SERP en contenidos del *Estrato Superior* pertenecientes a 'Prensa Estatal' o 'Prensa Regional', con incrementos en las elecciones a la Xunta entre el 18% y el 24%. Los resultados del conglomerado 'Televisión' se mantienen estables, mientras que el conglomerado de 'Cibermedios' crece en los tres buscadores, en Google multiplica su visibilidad por 2,5 mientras que en los otros dos buscadores no llega a duplicarse.

Los conglomerados secundarios alcanzan una visibilidad menor del 10% en resultados de SERPs. Esta proporción detectada en las elecciones Generales se repite en las elecciones a la Xunta. Destaca el incremento en Google del conglomerado 'Radio' que multiplica su visibilidad 4,75 veces.

>Estrato Medio. En este estrato la **tendencia general** es la de mostrar entre un 18% y un 28% en resultados de SERPs. Este estrato baja en diez puntos en las elecciones a la Xunta respecto a las Generales. El peso de los resultados del conglomerado 'Educación' alcanza cifras entre el 1,2% y el 1,5% para ser casi residuales en las elecciones gallegas. Destaca el mayor peso en las Generales de los conglomerados de la esfera política, tanto en el de partidos como en portales políticos. Mientras en

Google el decremento para estos conglomerados es del 25%, en Bing y Yahoo el peso se reduce un 50% en las elecciones a la Xunta.

>Estrato Micro. La tendencia de los buscadores respecto a este estrato es la de mostrar entre un 5% y un 6% en resultados de SERP.

La 'Blogosfera política' alcanza una visibilidad baja en las Generales que pasa a residual en las elecciones gallegas, mientras que el conglomerado de 'Movimiento protesta' pasa de resultados residuales en las Generales a dispararse hasta el 3,13% en Bing y el 2% en Yahoo.

Las páginas de redes sociales bajan en visibilidad en los tres buscadores, si bien la caída más destacada es en Bing, perdiendo más de un punto porcentual.

Los resultados de la Blogosfera no política caen en Google de un 4,7% hasta un 2,1%, mientras que Bing y Yahoo incrementan sus resultados en las elecciones gallegas hasta un 2,8% y 2,2% respectivamente.

La comparativa de los tres motores destaca el gran parecido existente en la composición de sus resultados, unas proporciones tan similares que puede hacer pensar que las acusaciones de Google a Bing² respecto a la copia de sus resultados fueron una realidad, por lo menos durante el año 2011.

Volveremos a ver en detalle datos de estas elecciones en el siguiente análisis, centrado en resultados de Google.

² Google acusa a Microsoft de copiar sus resultados en Bing y Microsoft le responde <<http://www.softwarelibre.org.pe/modules.php?name=News&file=article&sid=2309>> | Bing se encara a Google <<http://www.abc.es/20110203/medios-redes/abci-bing-respuesta-google-201102031210.html>>

3.2. Google Web Search

Mientras en el primer apartado de este capítulo hemos presentado una comparativa inicial de los buscadores Google, Bing y Yahoo en dos elecciones con el fin de mostrar los principales elementos y conceptos del análisis, en este segundo apartado del análisis visualizaremos únicamente los datos de Google Websearch y los resultados de los tres estratos a lo largo de la serie temporal de las seis elecciones que componen nuestro estudio global.

En particular, analizaremos los diferentes sectores y los principales conglomerados objeto de estudio y cómo se comportaron en Google Websearch durante todo el periodo de estudio. Los aspectos objeto de estudio son:

- **La serie temporal de elecciones.** Se mostrarán las comparativas de series temporales de 14 días, cada una de ellas correspondientes a las búsquedas vinculadas con los seis comicios electorales seleccionados. El cuadro resumen de la evolución de los estratos es el siguiente:

	Catalunya 2010	Municipales 2011	Generales 2011	Andalucía 2012	Galicia 2012	Catalunya 2012
Estrato convencional	39,78%	32,25%	51,80%	42,79%	61,57%	57,67%
Estrato medio	32,41%	50,68%	34,08%	46,06%	26,82%	31,17%
Micro-estrato	27,82%	17,08%	14,12%	11,15%	11,61%	11,16%

Figura 79.

- **Los estratos y conglomerados principales.** Como es sabido, la visibilidad de los principales estratos, conglomerados y páginas web específicas son el objeto de estudio. Las series temporales muestran el comportamiento y evolución de la visibilidad de estos conglomerados a lo largo de cada una de las elecciones, y de todas en su conjunto.

El siguiente diagrama muestra la evolución de las proporciones de los estratos:

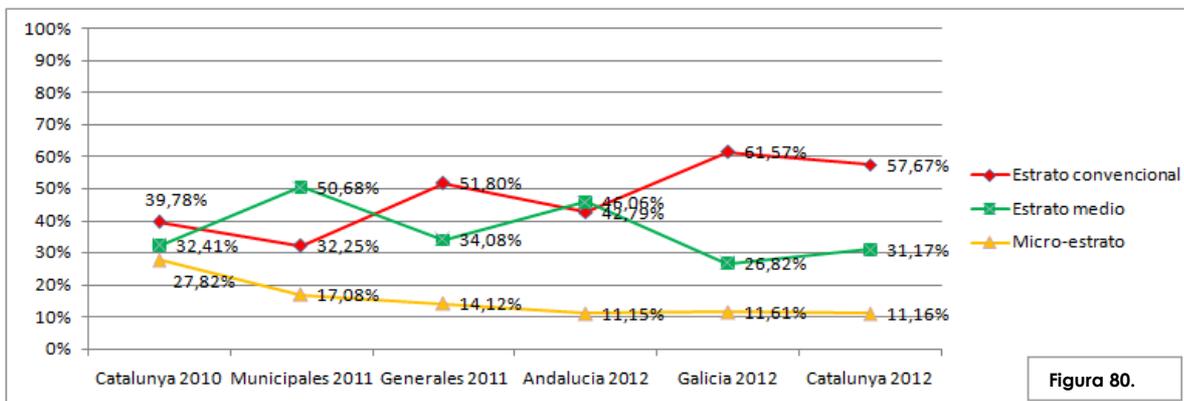


Figura 80.

En el gráfico de líneas³ podemos apreciar:

- **Estrato Superior** o **convencional**, correspondiente a los medios de comunicación: en este caso, la evolución de los sitios web clasificados dentro del Estrato Superior ven incrementada su proporción a lo largo del periodo, pasando del **40%** en las elecciones a la Generalitat 2010 a cerca del **60%** en las dos elecciones del año 2012.

Las elecciones donde el ratio es **menor** son las elecciones municipales de 2011 en las que los medios del Estrato Superior solo representan el **32,25%** de los resultados.

Las elecciones donde su proporción de visibilidad es **mayor** son las elecciones de Galicia 2012, cuando alcanza cotas del **61,57%**, si bien esta tendencia puede deberse a un mayor número de tipo de búsquedas candidatos.

- **Estrato Medio**: los resultados clasificados dentro de este estrato son en torno al 36%, con un valor máximo de un **50,7%** en las elecciones municipales y mínimos en las elecciones a la Xunta de Galicia con un **26,8%**⁴.
- **Estrato Micro**: el promedio del estrato inferior es del **15,5%**, sin embargo la evolución de 2010 a 2012 muestra una tendencia descendente perdiendo un 16% de visibilidad. Los máximos los encontramos en las elecciones a la Generalitat de 2010 con un **27,8%** y los mínimos en las elecciones a la Xunta de 2012 con un **11,16%**.

³Utilizamos los gráficos de líneas para representar los ratios de visibilidad en las series temporales de cada una de las elecciones. En este gráfico se representa la composición de la Infoesfera en Web Search y muestra la variación de visibilidad de los estratos.

⁴ La diferencia en proporciones podría estar causada por la selección de palabras clave y el peso de las diferentes tipologías de palabras clave, sin embargo las proporciones han sido muy similares cuando se han calculado los porcentajes con proporciones diferentes de tipo de palabra clave.

3.2.1. Generalitat de Catalunya – 28/11/2010

>Elecciones Generalitat 2010> Estrato Superior > Conglomerados principales

La **primera serie** engloba los principales conglomerados del estrato aparecidos en Google, e incluye todos los sitios web de medios convencionales de prensa, medios de televisión y cibermedios. El promedio de la serie que alcanza esta serie es del **36,6%**.

En el gráfico de área podemos observar la evolución de las proporciones durante los días de campaña electoral:

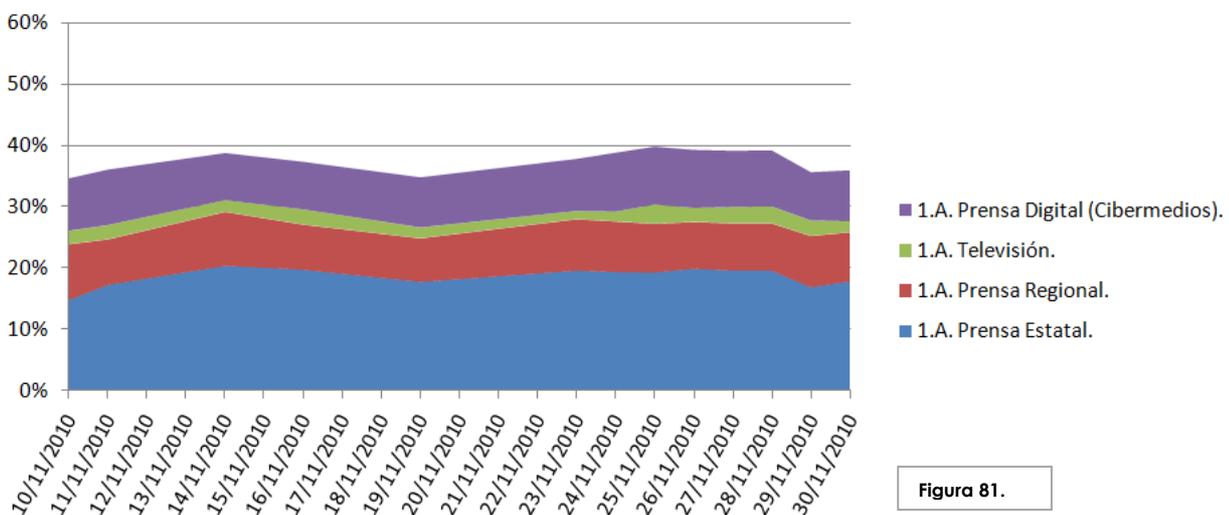


Figura 81.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Estatal.** El promedio del conglomerado fue del **18,87%**. El día de menor visibilidad fue el 10-nov con un 14,84%, y el de máxima el 14-nov con un 20,4%.
- **Prensa Regional.** El promedio de visibilidad fue del **7,91%**. El día de menor visibilidad fue el 19-nov con un 7,06%, y el de máxima el 10-nov con un 9,05%.
- **Televisión.** El promedio de este conglomerado fue del **2,25%**. El día de menor visibilidad fue el 23-nov con un 1,35%, y el de máxima el día 25 con un 3,11%.
- **Prensa Digital (Cibermedios).** El promedio de visibilidad fue del **8,66%**. El día de menor visibilidad fue el 14-nov con un 7,7%, y el de máxima el día 24 con un 9,62%.

>Elecciones Generalitat 2010> Estrato Superior > Conglomerados secundarios

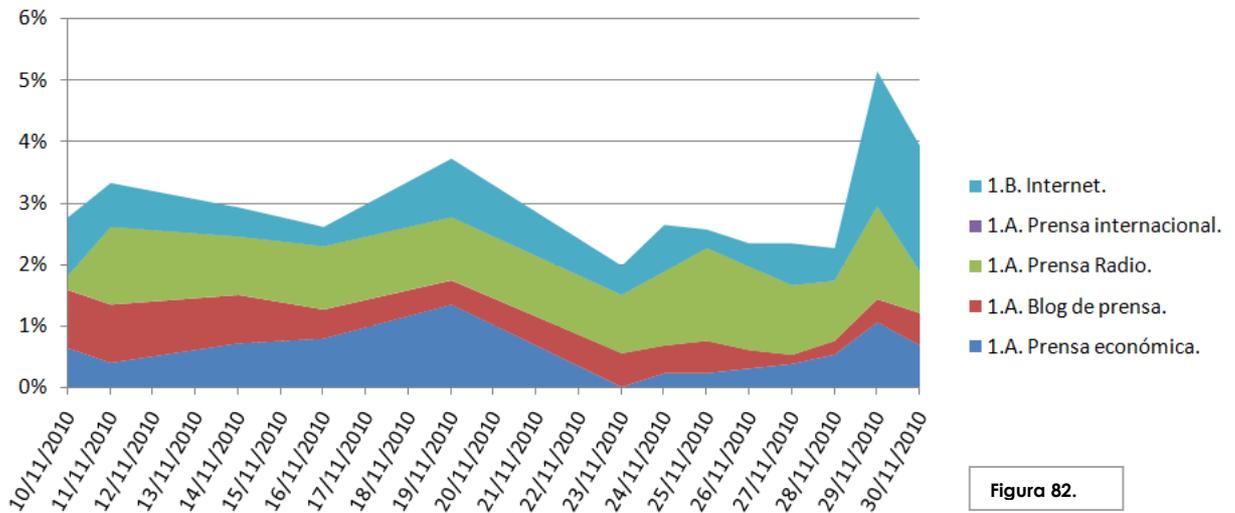


Figura 82.

La **segunda serie** es la compuesta por conglomerados menores del primer estrato supone un promedio del **3,17%** de resultados de SERPs:

1.A.Mass_Media_Prensa.1Nacional

- **1.A.Prensa Económica.** El promedio del conglomerado fue del **0,56%**. El día de mayor visibilidad fue el día posterior a las elecciones, el 29 de noviembre, con un 1,06%. Como veremos a lo largo de la serie temporal, ésta es una tendencia que se repite en todas las elecciones y que puede deberse a un mayor CTR en páginas económicas durante los días más cercanos a las elecciones; este hecho connotaría un interés creciente en prensa económica por parte de los usuarios conforme se acercan los días de los comicios.
- **1.A.Blog de Prensa.** El promedio de visibilidad fue del **0,52%**. El día de menor visibilidad fue el 27-nov con un 0,15%, y el de máxima el 10-nov con un 0,95%.
- **1.A.Radio.** El promedio del conglomerado fue del **1,07%**. El día de menor visibilidad fue el 10-nov con un 0,24%, y el de máxima el día posterior a las elecciones, el 29 de noviembre, con un con un 1,52%.

1.A.Prensa Internacional. Este conglomerado no alcanzó visibilidad en SERPs.

1.B.Internet. El conglomerado de medios de internet alcanzó una visibilidad promedio de **0,83%**. El día de menos visibilidad fue el 25-nov con un 0,30%, y el de mayor el día el 29-nov con un 2,2%.

>Elecciones Generalitat 2010> Estrato Medio

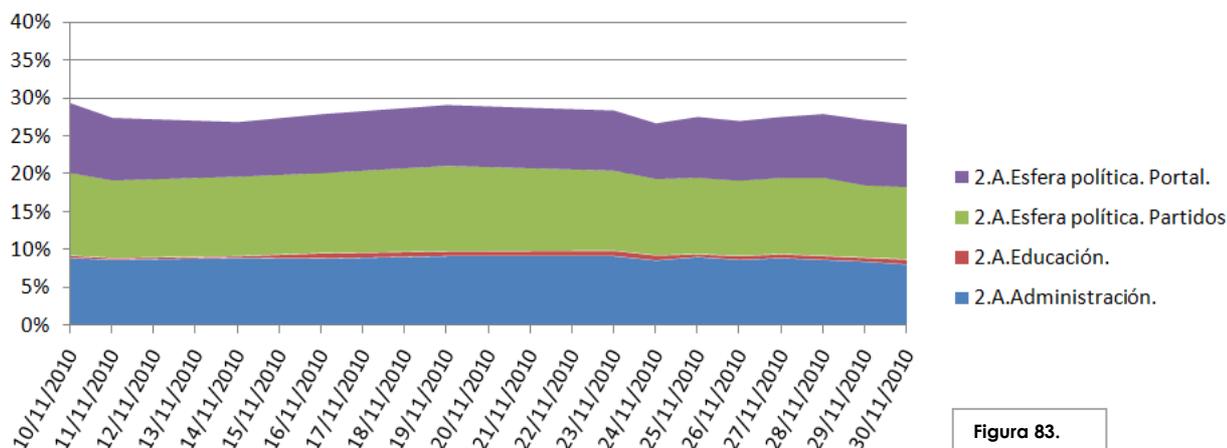


Figura 83.

Los conglomerados del segundo estrato objeto de estudio obtuvieron una visibilidad global promedio del **27,22%**:

2.A.Administración. El promedio de los conglomerados fue del **8,11%**. El día de mayor visibilidad fue el día 19 con un 9,21%, mientras que el de menor visibilidad fue el día posterior a las elecciones con un 8,11%⁵.

2.A.Educación. La visibilidad de este conglomerado es residual. El promedio del conglomerado fue del **0,61%**. El día de mayor visibilidad fue el día 23 con un 0,63%.

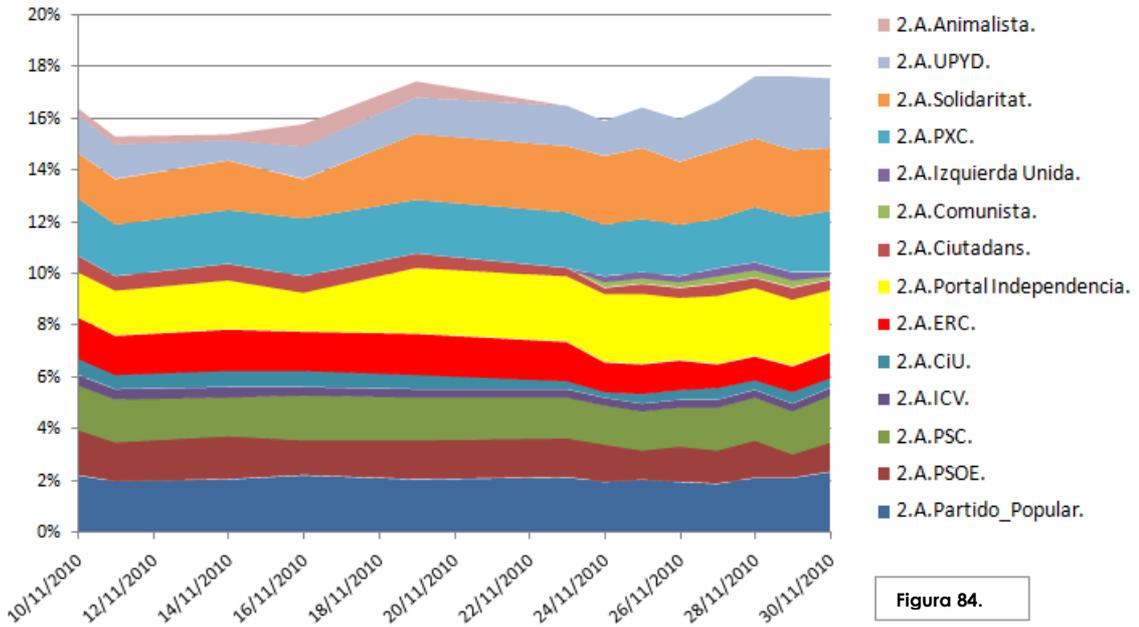
2.A.Esfera política.Portal. El promedio del conglomerado fue del **8,06%**. El día de mayor visibilidad fue el primer día de captura con un 9,13%.

- **Portal electoral.** El promedio de este conglomerado es en torno al 6,3%. Destacan los portales electometro.es (3%), portalelectoral.es (0,82%) y gutierrez-rubi.es (0,57%).
- **Portal político.** El promedio de este conglomerado es en torno al 1,7%.

2.A.Esfera política.Partidos. El promedio del conglomerado fue del **9,62%**. El día de mayor visibilidad fue el primer día de captura con un 10,95%, y el de menor, el día de las elecciones con un 9,62%. Vemos el detalle de la serie a continuación.

⁵ Los conglomerados más destacados fueron el de la 'Administración estatal' que acumuló el 7,21% del total y el de la administración de la Generalitat que acumuló el 73,7% de los resultados provenientes de las administraciones.

>Elecciones Generalitat 2010> Estrato Medio> Partidos



2.A.Esfera política.Partidos

- **Partido_Popular.** El conjunto de páginas del PP tiene una visibilidad promedio del **2,09%** durante toda la serie. El día de las elecciones alcanza su máximo de la serie es el día posterior a los comicios con un 2,35%.
- **PSOE-PSC.** El conjunto de páginas del Partido Socialista Obrero Español tiene una visibilidad promedio del **1,39%** durante toda la serie, mientras que las páginas del Partido Socialista Catalán alcanzan el **1,62%**. La suma de ambas es del **3%** de la visibilidad total. La mayor visibilidad se registra la primera semana mientras que las cifras más bajas de visibilidad se dan los días centrales de la muestra para ambas formaciones. En el caso del PSC su máximo lo encontramos el día posterior a las elecciones con un 1,82%.
- **Izquierda Unida-ICV.** El conjunto de páginas de Izquierda Unida tiene una visibilidad promedio del **0,33%**. Tiene más visibilidad la primera semana de datos.
- **Resto de partidos.** Los datos de visibilidad de las demás agrupaciones políticas son: CiU con un 0,46%, ERC con un 1,29%, Ciutadans con un 0,46%, **Solidaritat** con 2,33%, **PXC** con 2,09% y **UPyD** con 1,71%. Destaca el conglomerado 'Portal Independencia' con un 2,33%.

>Elecciones Generalitat 2010> Estrato Micro

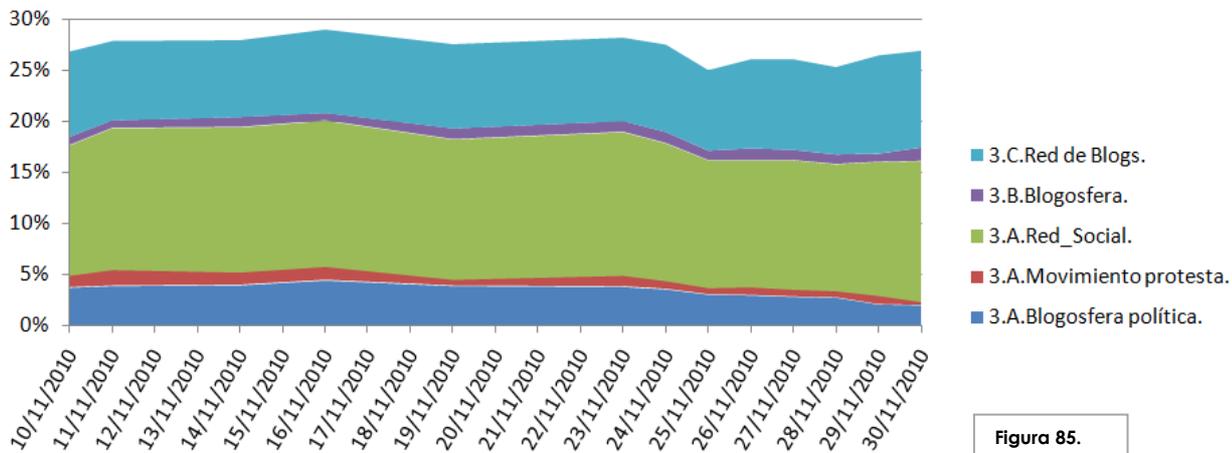


Figura 85.

La serie completa de conglomerados del tercer estrato son:

3.A.Blogosfera política. El promedio del conglomerado fue del **3,29%**, el mayor registrado de toda la serie de elecciones. El día de mayor visibilidad fue el día 16 con un 4,44%. Los dos días de menor visibilidad fueron los días inmediatamente posteriores a las elecciones con un 2,05% y un 1,97%. Conglomerados destacados:

- **Eleccions-2010.com**⁶: el sitio web ideado para la realización de esta tesis alcanzó **514** posiciones de Top10 en la muestra, un **91%** del total de resultados de blogs políticos.
- **Blogs políticos**: los blogs políticos sin adscripción partidista coparon un total de **14 posiciones** en Top10, entre los que destacó el consultor político '*albertmedran.com*'.
- **Blogs de CiU, PSC, PP**: el acumulado de blogs políticos de partido únicamente logró **33** posiciones en Top10, un 5,8% de los resultados de la blogosfera política.

3.A.Movimiento protesta. El promedio del conglomerado fue del **0,94%**. El día de mayor visibilidad fue el día 11 con un 1,59%. El día de menor visibilidad fue el 30 con un 0,38%. Sitios web destacados:

⁶ Se ha creado un conglomerado separado del resto de sitios web de la esfera política con el objetivo de destacar la visibilidad del blog 'eleccions-2010.com', ideado como parte del proyecto de tesis para testear la posibilidad de posicionar contenidos políticos y competir con el resto de sitios web oficiales.

- **Nacionred.com:** este portal acumuló el 61,2% de los resultados del conglomerado.
- **Indymedia.org:** el host 'barcelona' de este popular portal protesta supuso el 20,6% de los resultados.
- **Otros:** *Rebellion.org* con un 15,9% y *Kaosenlared.info* con un 3%.

3.A.Red_Social. El promedio del conglomerado fue del **13,33%**. El día de mayor visibilidad fue el día 23 con un 14,05%. Se contabilizaron **40 sitios** distintos. Sitios web más destacados fueron:

- **Wikipedia:** éste fue el principal sitio web del conglomerado con resultados en español y catalán que acumularon el **50,1%** de los UGC de las elecciones.
- **Youtube:** los resultados del portal de vídeo fueron un **3,22%** del total de UGC.
- **Facebook:** las páginas de esta red social fueron el **3,18%**.
- **Twitter:** los resultados de micro-blogging supusieron un **0,96%** del total.

3.B.Blogosfera. El promedio del conglomerado fue del **0,94%**. El día de mayor visibilidad fue el día 30 con un 1,29%.

3.C.Red de Blogs. El promedio del conglomerado fue del **8,46%**. El día de mayor visibilidad fue el día 29 con un 9,62%.

3.2.2. Elecciones Municipales – 22/5/2011

>Elecciones Municipales 2011> Estrato Superior > Conglomerados principales

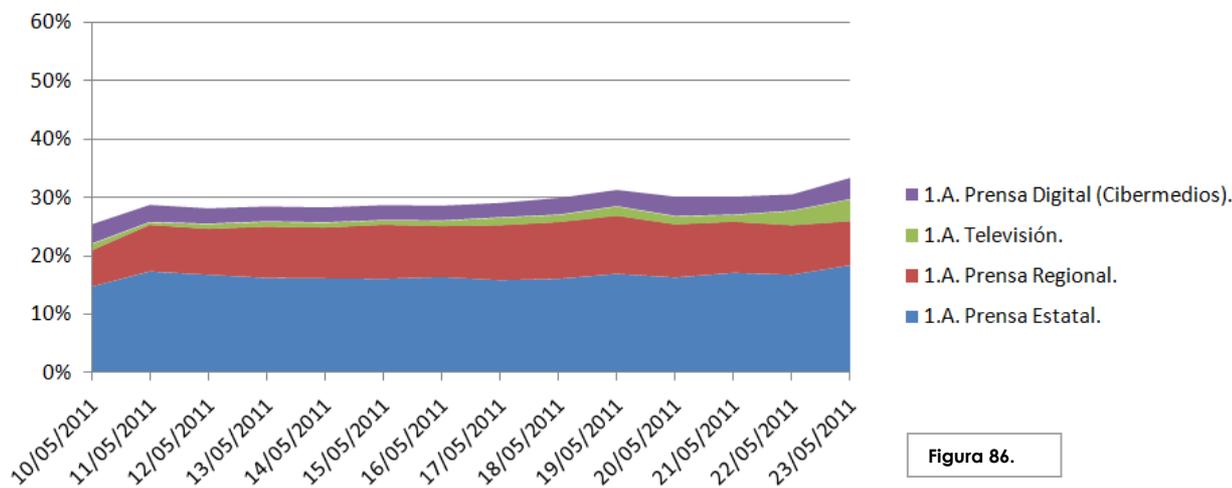


Figura 86.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Estatal.** El promedio del conglomerado fue del **16,55%**. El día de mayor visibilidad fue el día posterior a los comicios, el 23 de mayo, con un 18,45%, y el de menor el día 10 con un 14,76%.
- **Prensa Regional.** El promedio de visibilidad fue del **8,58%**. El día de más visibilidad fue el día 19 con un 9,92%, y el de menor el día 10 con un 6,19%.
- **Televisión.** El promedio del conglomerado fue del **1,35%**. El día de mayor visibilidad fue el día posterior a los comicios con un 3,8%, y el de mínima el 11 de mayo con un 0,48%.
- **Prensa Digital (Cibermedios).** El promedio de visibilidad fue del **2,87%**. El día de mayor visibilidad fue el día posterior a los comicios con un 3,64%, y el de menor visibilidad, el 17 de mayo con un 2,48%.

>Elecciones Municipales 2011> Estrato Superior > Conglomerados secundarios

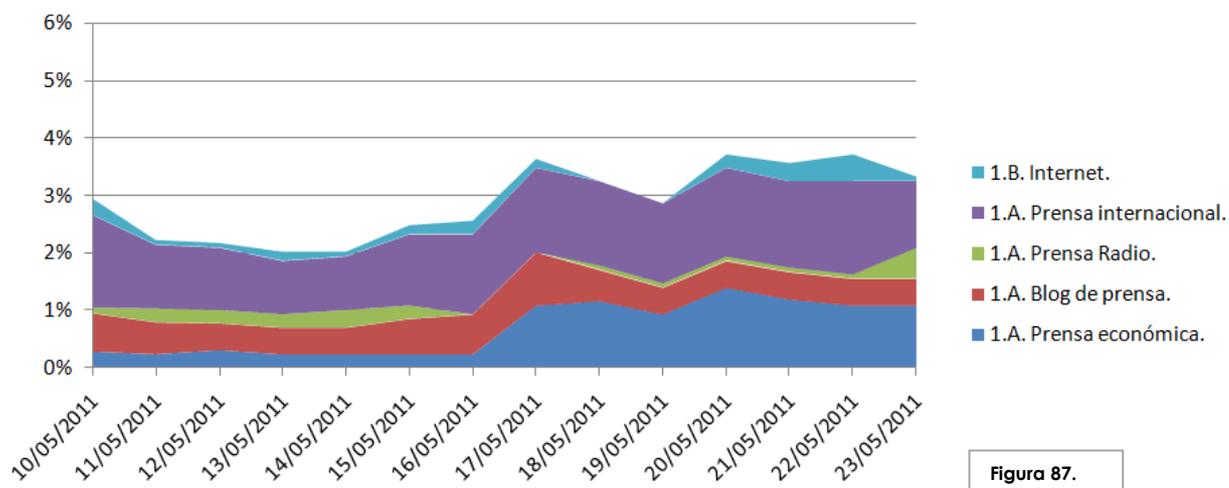


Figura 87.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Económica.** El promedio del conglomerado fue del **0,69%**. El día de menor visibilidad fue el 11 de mayo con un 0,24%, y el de máxima el 20 con un 1,4%. El interés creciente en prensa económica detectado en las elecciones a la Generalitat se repite en estos comicios de una manera más clara.
- **Blog de Prensa.** El promedio de visibilidad fue del **0,55%**. El día de mayor visibilidad fue el 17 de mayo con un 0,93%.
- **Radio.** El promedio del conglomerado fue del **0,16%**. El día de mayor visibilidad fue el día posterior a los comicios con un 0,54%.

1.A.Prensa Internacional. El promedio de visibilidad fue del **1,32%**. El día de menor visibilidad fue el 14 de mayo con un 0,93%, y el de máxima el día de las elecciones con un 1,63%.

1.B.Internet. El conglomerado de medios de internet alcanzó una visibilidad promedio de **0,17%**. El día de máxima visibilidad fue el día de las elecciones con un 0,47%.

>Elecciones Municipales 2011> Estrato Medio

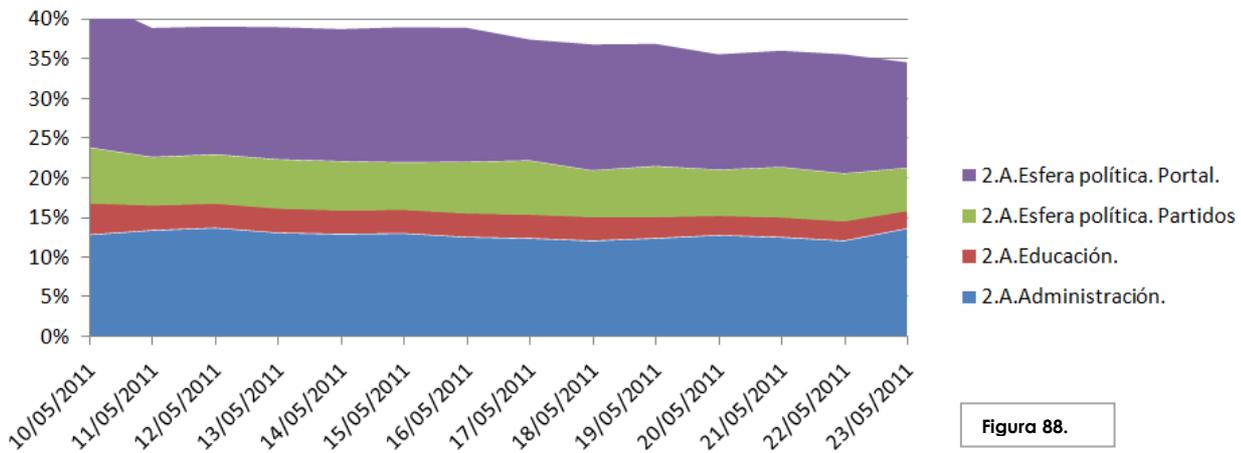


Figura 88.

2.A.Administración. El promedio del conglomerado fue del **12,83%**. El día de mayor visibilidad fue el día posterior a las elecciones con un 13,64%.

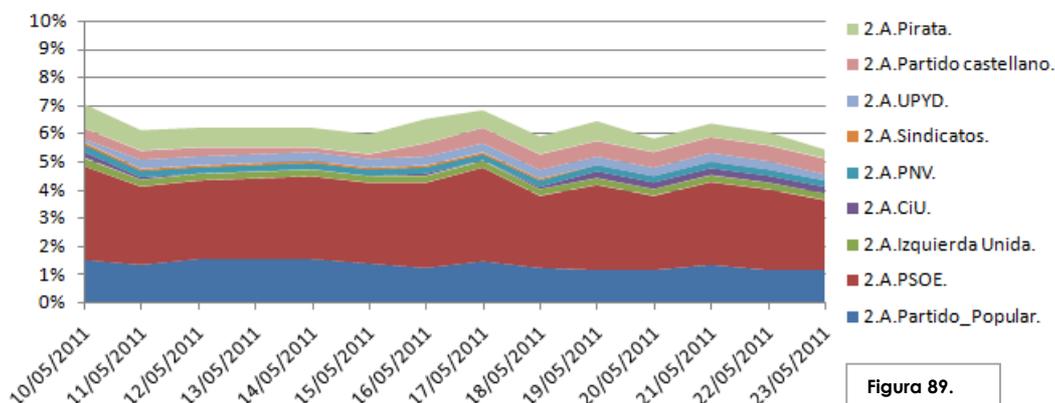
2.A.Educación. El promedio del conglomerado fue del **2,93%**. El día de mayor visibilidad fue el primer día de captura con un 4%.

2.A.Esfera política.Portal. El promedio del conglomerado fue del **15,8%**. El día de mayor visibilidad fue el primer día de captura con un 18,67%. El día de menor visibilidad fue el posterior a las elecciones con un 13,26%.

- **Esfera política.Portal electoral.** El promedio de este conglomerado es del **15,18%**. Destacan los portales electometro.es (5,5%), portalelectoral.es (3,4%) y municipales2011.com (1,21%).
- **Esfera política.Portal político.** El promedio de este conglomerado es del 0,64%.

2.A.Esfera política.Partidos. El promedio del conglomerado fue del **6,22%**. El día de mayor visibilidad fue el primer día de captura con un 7,05%. El día de menor visibilidad fue el posterior a los comicios con un 5,43%. Vemos el detalle en la siguiente figura.

>Elecciones Municipales 2011> Estrato Medio> Partidos



2.A.Esfera política.

- **Partido_Popular.** El conjunto de páginas del PP tiene una visibilidad promedio del **1,35%** durante toda la serie. Durante la primera semana alcanza cifras de 1,55% para bajar al 1,2% la segunda semana.
- **PSOE.** El conjunto de páginas del Partido Socialista Obrero Español tiene una visibilidad promedio del **2,89%** durante toda la serie. Las cifras más altas de visibilidad son el primer día de captura (3,33%) y el anterior a las elecciones (2,94%).
- **Izquierda Unida-ICV.** El conjunto de páginas de Izquierda Unida tiene una visibilidad promedio del **0,24%** y se mantiene constante durante toda la serie.
- **Resto de partidos.** Las demás agrupaciones políticas logran los siguientes datos de visibilidad: CiU un 0,11%, PNV un 0,24%, UPYD un 0,29% y el Partido Castellano un 0,42%.

>Elecciones Municipales 2011> Estrato Micro

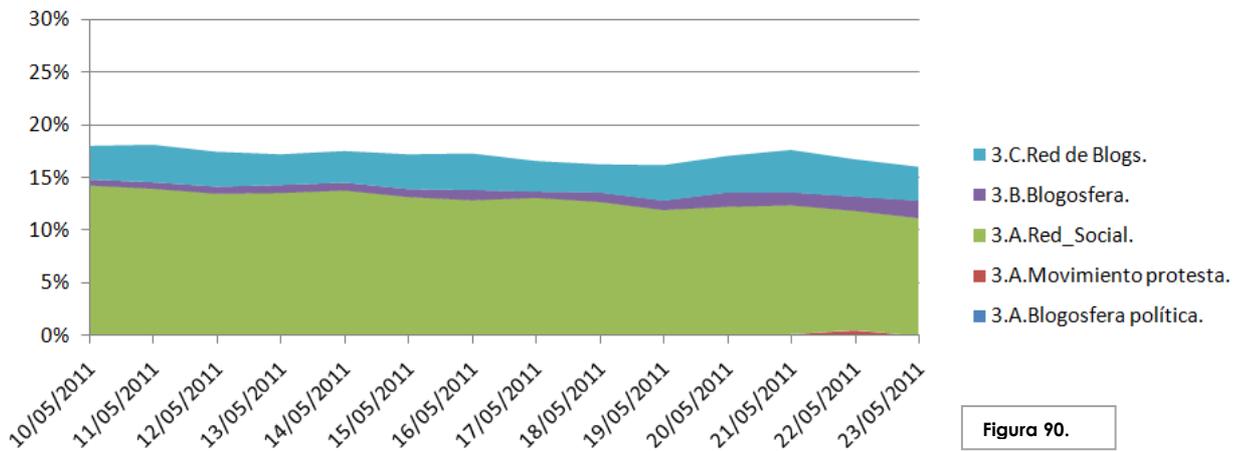


Figura 90.

3.A.Blogosfera política. El promedio del conglomerado fue residual del **0,06%**. El día de mayor visibilidad fue el día anterior a los comicios y el día de las elecciones con un 0,16%.

3.A.Movimiento protesta. El promedio del conglomerado fue del **0,03%**. El día con más apariciones en SERP fue la jornada electoral con un 0,39%.

3.A.Red_Social. El promedio fue del **12,81%**. El día de mayor visibilidad fue el primer día de captura con un 14,19%.

- **Wikipedia:** este sitio acumuló el **63,85%** de los UGC de estas elecciones.
- **Web social:** las redes sociales secundarias fueron el **5,37%** del total de resultados del conglomerado, entre las que destacan 'practicopedia.com' y 'slideshare.net'.
- **Youtube:** los resultados del portal de vídeo fueron un **3,92%** del total destacando el posicionamiento de vídeos de los candidatos del PSOE y PP.
- **Portal ciudadano:** estas páginas fueron el **6,21%** del conglomerado.
- **Portal foro:** los foros fueron el **9,34%** del conglomerado.
- **Facebook:** sus páginas supusieron el **4,32%** del conglomerado.
- **Twitter:** los resultados de micro-blogging no aparecieron en SERPs.

3.B.Blogosfera. El promedio fue del **0,96%**. El día con más visibilidad fue el día posterior a las elecciones con un 1,71%.

3.C.Red de Blogs. El promedio fue del **3,23%**. El día más visible en SERP fue el anterior a las elecciones con un 3,97%.

3.2.3. Elecciones Generales – 20/11/2011

>Elecciones Generales > Estrato Superior > Conglomerados principales

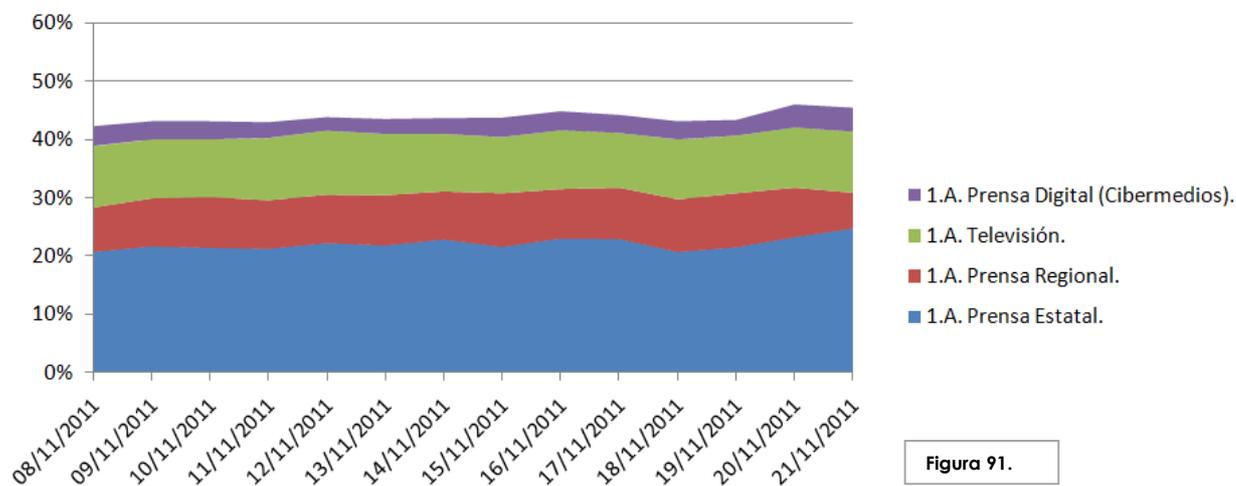


Figura 91.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Estatal.** El promedio del conglomerado fue del **22,21%**. El día de menor visibilidad fue el primer día de captura, el 8 de noviembre, con un 20,79%, y el de máxima el día posterior a los comicios con un 24,84%.
- **Prensa Regional.** El promedio de visibilidad fue del **8,40%**. El día de menor visibilidad fue el día posterior a las elecciones con un 6,11%, y el de máxima el 15-nov con un 9,21%.
- **Televisión.** El promedio del conglomerado fue del **10,08%**. El día de menor visibilidad fue el 17 con un 9,29%, y el de máxima el día 12 con un 10,87%.
- **Prensa Digital (Cibermedios).** El promedio de visibilidad fue del **3,18%**. El día de menor visibilidad fue el 12 con un 2,38%, y el de máxima el día posterior a las elecciones con un 4,21%.

>Elecciones Generales > Estrato Superior > Conglomerados secundarios

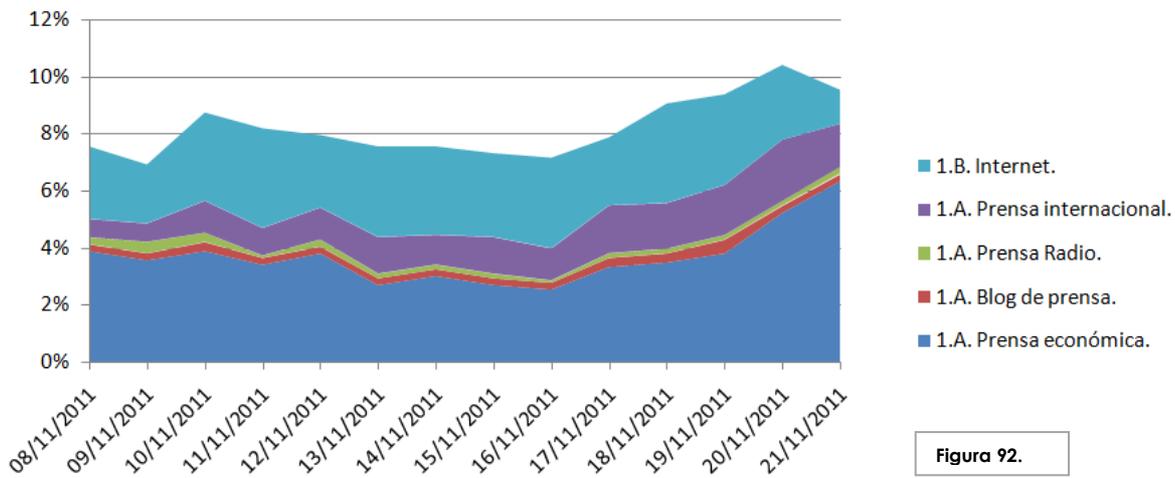


Figura 92.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Económica.** El promedio del conglomerado fue del **3,70%**, el mayor en toda la serie temporal de datos de Web Search. El día de menor visibilidad fue el 16 con un 2,54%, y el de máxima el día posterior a las elecciones con un 6,35%. Al igual que en las elecciones precedentes, destaca el incremento de este conglomerado a partir de la segunda semana de datos, alcanzando sus cotas máximas los días más cercanos a los comicios.
- **Blog de Prensa.** El promedio de visibilidad fue del **0,27%**. El día de máxima visibilidad fue el día anterior a las elecciones, el 19 de noviembre, con un 0,48%.
- **Radio.** El promedio del conglomerado fue del **0,19%**. El día de máxima visibilidad fue el 9-nov con un 0,4%.

1.A.Prensa Internacional. El promedio de visibilidad fue del **1,27%**. El día de menor visibilidad fue el primer día de captura, 8 de noviembre, con un 0,63%, y el de máxima el día de las elecciones con un 2,14%.

1.B.Internet. El conglomerado de medios de internet alcanzó una visibilidad promedio de **2,78%**. El día de menos visibilidad fue el posterior a las elecciones con un 1,19%, y el de mayor el día 11 con un 3,49%.

>Elecciones Generales > Estrato Medio

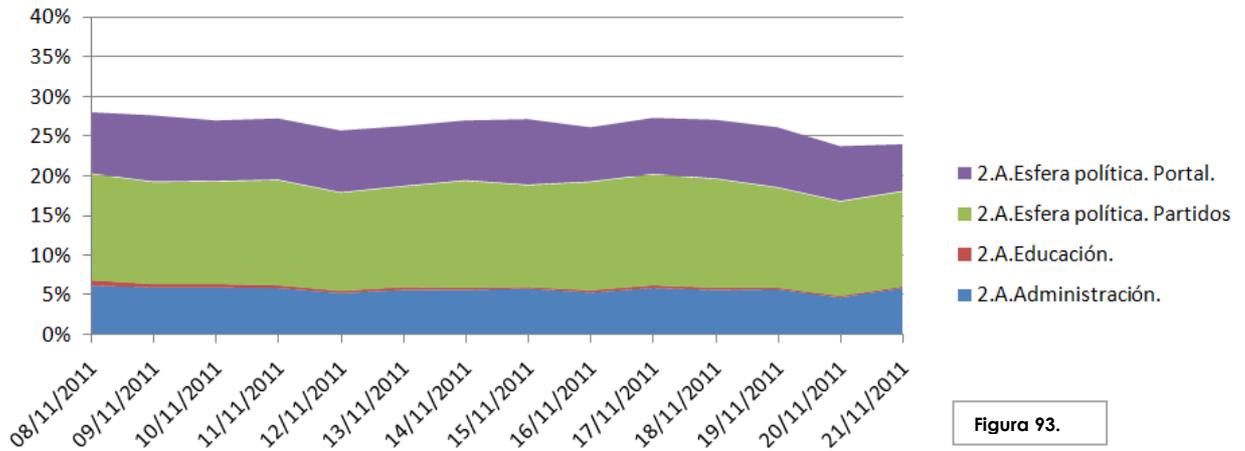


Figura 93.

2.A.Administración. El promedio del conglomerado fue del **5,67%**. El día de mayor visibilidad fue el día 8 con un 6,19%.

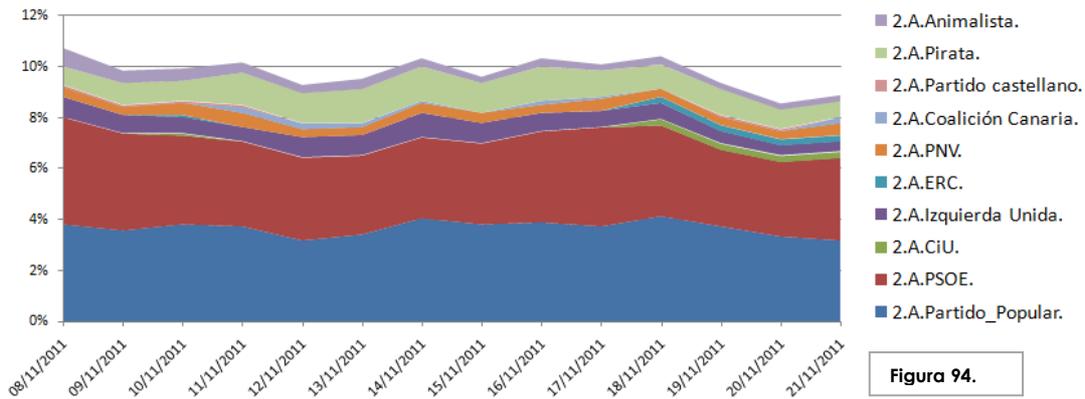
2.A.Educación. El promedio de contenidos posicionados en SERP es residual y fue del **0,36%**. El día de mayor visibilidad fue el día 8 con un 0,71%.

2.A.Esfera política.Portal. El promedio del conglomerado fue del **7,13%**. El día de mayor visibilidad fue el día 15 con un 7,86%.

- **Esfera política.Portal electoral.** El promedio de este conglomerado es en torno al 5,6%. Destacan los portales portalelectoral.es (1,88%), electometro.es (1,85%) y municipales2011.com (1,21%).
- **Esfera política.Portal político.** El promedio de este conglomerado es en torno al 1,5%, destacando politicaenelmundo.com (0,71%).

2.A.Esfera política.Partidos. El promedio del conglomerado fue del **12,98%**. El día de mayor visibilidad fue el día 17 con un 13,97%. Lo vemos en más detalle a continuación.

>Elecciones Generales > Estrato Medio> Partidos



2.A.Esfera política.

- **Partido_Popular.** El conjunto de páginas del PP tiene una visibilidad promedio del **3,67%** durante toda la serie. Su visibilidad se mantiene constante y el día de las elecciones alcanza un 3,33% y su mínimo el día posterior con un 3,17%.
- **PSOE.** El conjunto de páginas del Partido Socialista Obrero Español tiene una visibilidad promedio durante la serie del **3,41**. Las cifras más bajas de visibilidad es el día de las elecciones con un 2,94%.
- **Izquierda Unida-ICV.** El conjunto de páginas de Izquierda Unida tiene una visibilidad promedio del **0,66%**. Sigue la tendencia del resto de partidos y tiene sus días de menor visibilidad alrededor de los días del comicio bajando hasta el 0,4%.
- **Resto de partidos.** Destaca la formación PNV con un 0,39% y el Partido Pirata con un 1%. Las demás agrupaciones políticas logran una visibilidad residual menor al 0,1%.

>Elecciones Generales > Estrato Micro

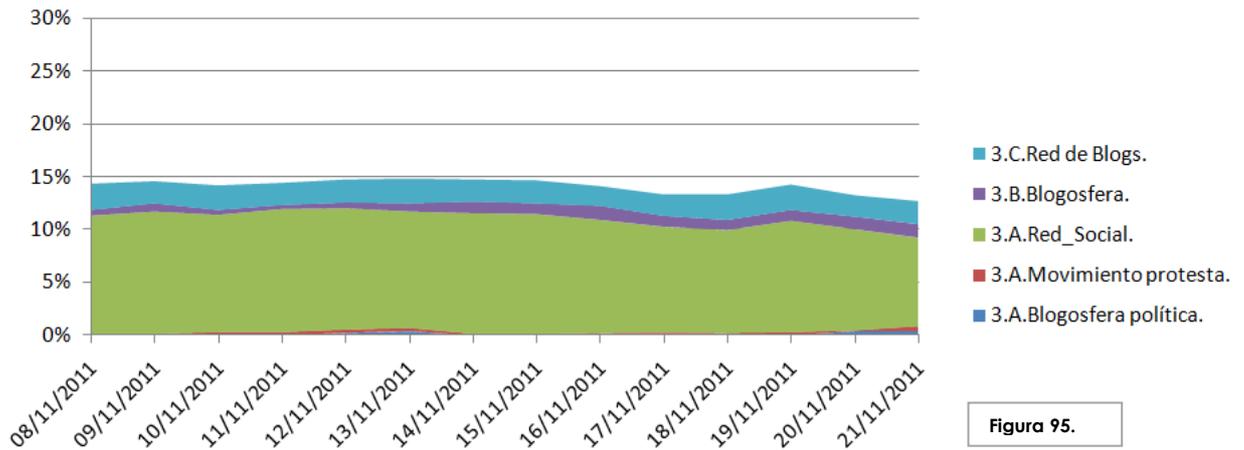


Figura 95.

3.A.Blogosfera política. El promedio del conglomerado fue del **0,15%**. Los días de mayor visibilidad fueron los días 13 y 20 con un 0,4%.

3.A.Movimiento protesta. El promedio fue del **0,2%**. El día de mayor visibilidad fue el día 21 con un 0,48%. El único sitio aparecido en SERPs fue 'nacionred.com'.

3.A.Red_Social. El promedio fue del **10,63%**. El día de mayor visibilidad fue el día 11 con un 11,59%.

- **Wikipedia:** este sitio acumuló el **59,84%** de los UGC de estas elecciones.
- **Web social:** las redes sociales secundarias fueron el **3,68%** del total de resultados del conglomerado, entre las que destacan 'encuestafacil.com', 'scribd.com' y 'slideshare.net'.
- **Youtube:** los resultados del portal de vídeo fueron un **22,67%** del total destacando el posicionamiento de vídeos de los candidatos del PSOE y PP.
- **Portal ciudadano:** estas páginas fueron el **4,81%** del conglomerado con entradas de los sitios web 'internautas.org' y 'hazteoir.org'.
- **Portal foro:** los foros fueron el **1,17%** del conglomerado con páginas de 'forociudad.com' y 'forocoches.com'.
- **Facebook:** sus páginas supusieron el **6,08%** del conglomerado. Las únicas páginas vinculadas partidos políticos fueron creadas por PP y PSOE para las elecciones municipales de Menorca y Canarias. No aparecieron en SERP ninguna de las páginas oficiales de partido y/o candidato.
- **Twitter:** los resultados de micro-blogging son residuales con un 0,11% del total de contenidos UGC provenientes de la cuenta oficial de Rubalcaba.

3.B.Blogosfera. El promedio fue del **0,9%**. El día con más apariciones en SERP fue el día 16 con un 1,35%.

3.C.Red de Blogs. El promedio fue del **2,24%**. El día de mayor visibilidad fue el día 8 con un 2,54%.

3.2.4. Junta de Andalucía – 12/5/2012

>Elecciones Andalucía 2012> Estrato Superior > Conglomerados principales

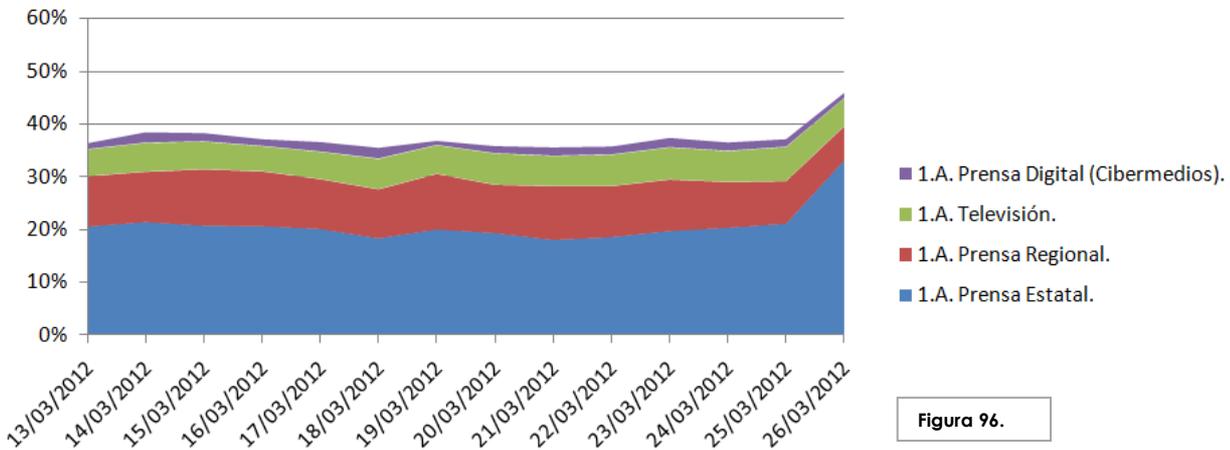


Figura 96.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Estatal.** El promedio del conglomerado fue del **20,91%**. El día de menor visibilidad fue el 18 con un 18,37%, y el de máxima el 26 con un 33,2%.
- **Prensa Regional.** El promedio de visibilidad fue del **9,39%**. El día de menor visibilidad fue el 25 con un 6,48%, y el de máxima el 19 con un 10,54%.
- **Televisión.** El promedio del conglomerado fue del **5,54%**. El día de menor visibilidad fue el 16 con un 4,73%, y el de máxima el 25 con un 6,43%.
- **Prensa Digital (Cibermedios).** El promedio de visibilidad fue del **1,45%**. El día de menor visibilidad fue el 19 con un 0,78%, y el de máxima el 18 con un 2,02%.

>Elecciones Andalucía 2012> Estrato Superior > Conglomerados secundarios

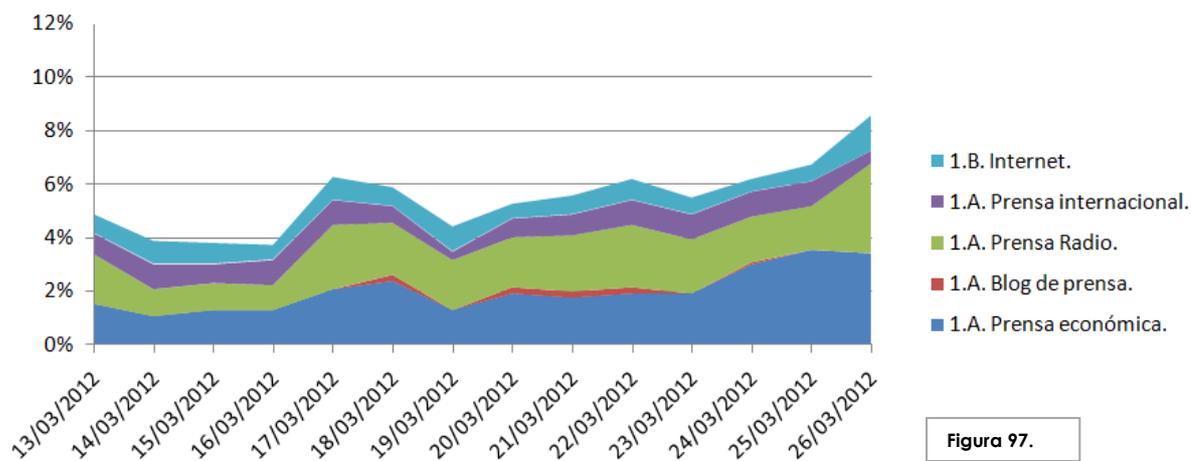


Figura 97.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Económica.** El promedio del conglomerado fue del **2,05%**. El día de menor visibilidad fue el 14 con un 1,09%, y el de máxima el día de las elecciones con un 3,57%. Al igual que en el resto de comicios, destaca por su incremento en visibilidad los días más cercanos a los comicios.
- **Blog de Prensa.** El promedio de visibilidad fue del **0,07%**. Alcanzó una visibilidad del 0,23% en cuatro jornadas.
- **Radio.** El promedio del conglomerado fue del **1,86%**. El día de menor visibilidad fue el 14 con un 1,01%, y el de máxima el día posterior a las elecciones con un 3,36%.

1.A.Prensa Internacional. El promedio de visibilidad fue del **0,78%**. Los días de más visibilidad alcanzó fue del 0,93%.

1.B.Internet. El conglomerado de medios de internet alcanzó un promedio de visibilidad en el Top10 del **0,74%**. El día de máxima visibilidad fue el día posterior a las elecciones con un 1,33%.

>Elecciones Andalucía 2012> Estrato Medio

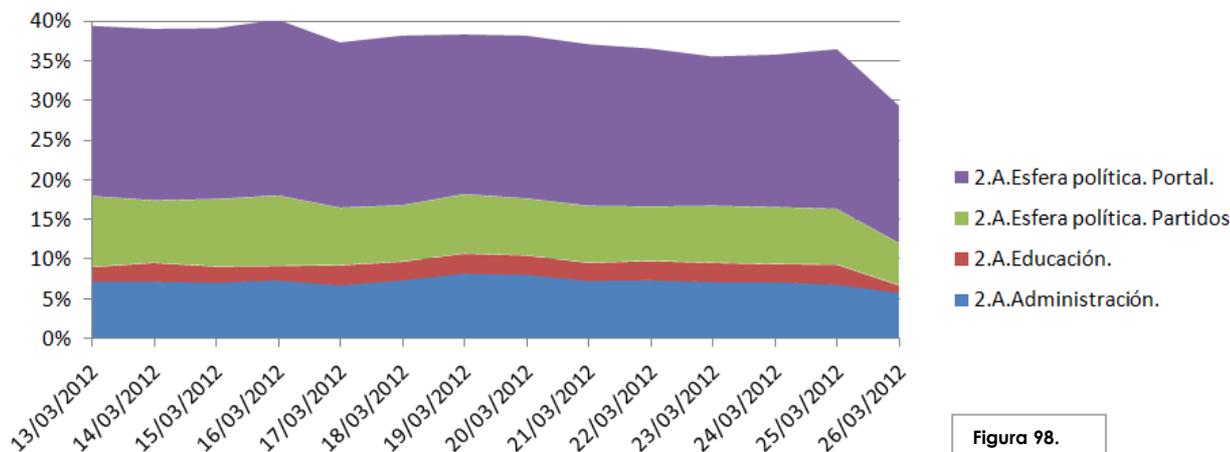


Figura 98.

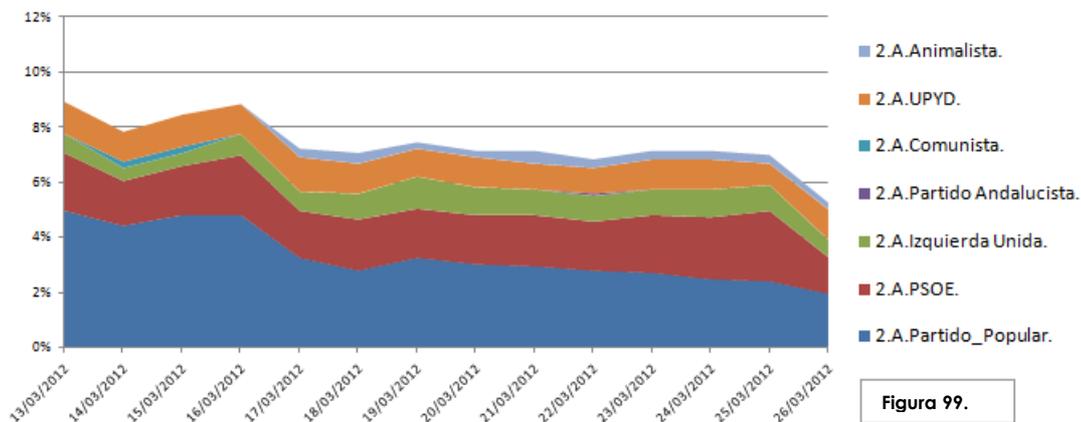
2.A.Administración. El promedio del conglomerado fue del **7,27%**. El día de mayor visibilidad fue el día 19 con un 8,29%.

2.A.Educación. El promedio del conglomerado fue del **2,17%**. Los días de mayor visibilidad fueron el 17, 19 y 25 con un 2,48%.

2.A.Esfera política.Portal. El promedio del conglomerado fue del **20,4%**. El día de mayor visibilidad fue el día 16 con un 22,17%. Los portales políticos tienen una visibilidad marginal, mientras que destacan los portales electorales eleccionesandalucia.es (8,45%) y electometro.es (4,47%).

2.A.Esfera política.Partidos. El promedio del conglomerado fue del **7,38%**. El día de mayor visibilidad fue el día 13 con un 8,91%. Vemos el detalle a continuación.

>Elecciones Andalucía 2012> Estrato Medio> Partidos



2.A.Esfera política.

- **Partido_Popular.** El conjunto de páginas del PP tiene una visibilidad promedio del **3,33%** durante toda la serie. Su tendencia es decreciente desde un 4,96% el primer día hasta un 1,95% el día posterior a las elecciones.
- **PSOE.** El conjunto de páginas del Partido Socialista Obrero Español tiene una visibilidad promedio del **1,91%** constante durante toda la serie para caer al 1,33% el día posterior a los comicios.
- **Izquierda Unida.** El conjunto de páginas de Izquierda Unida tiene una visibilidad promedio del **0,83%**. La tendencia es opuesta a PP-PSOE y tiene su peor día a principio de la serie para llegar al 1% el día de las elecciones.
- **UPYD.** Las páginas de este partido tienen una visibilidad promedio del **1,06%** y se mantienen constante en toda la serie.
- **Resto de partidos.** Las demás agrupaciones políticas logran una visibilidad prácticamente simbólica, entre las que destaca el Partido Animalista con un 0,22%.

>Elecciones Andalucía 2012> Estrato Micro

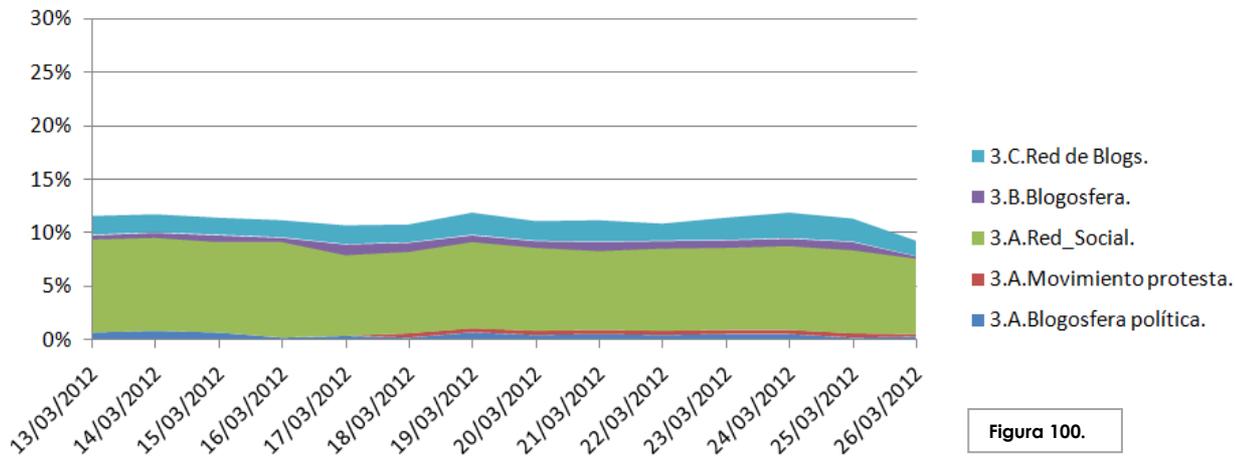


Figura 100.

3.A.Blogosfera política. El promedio del conglomerado fue del **0,49%**. El día de mayor visibilidad fue el día 14 con un 0,85%.

3.A.Movimiento protesta. El promedio del conglomerado fue del **0,24%**. El día de mayor visibilidad alcanzó un 0,39%. Sus únicos representantes en SERPs fueron <http://democraciarealyajaen.blogspot.com> y <http://wiki.nolesvotes.org>.

3.A.Red_Social. El promedio del conglomerado fue del **8%**. El día de mayor visibilidad fue el día 16 con un 8,99%.

- **Wikipedia:** este sitio acumuló el **67,41%** de los UGC de estas elecciones.
- **Youtube:** los resultados del portal de vídeo fueron residuales con un **0,14%** del total del conglomerado con dos apariciones de vídeos de UPyD.
- **Facebook:** sus páginas supusieron el **13,31%** del conglomerado.
- **Twitter:** los resultados de micro-blogging fueron el 0,20% del conglomerado con tres únicas apariciones de la cuenta del Partido Popular.

3.B.Blogosfera. El promedio del conglomerado fue del **0,64%**. El día de mayor visibilidad fue el día 17 con un 1%.

3.C.Red de Blogs. El promedio del conglomerado fue del **1,78%**. El día de mayor visibilidad fue el día 24 con un 2,33%.

3.2.5. Xunta de Galicia – 21/10/2012

>Elecciones Galicia 2012> Estrato Superior > Conglomerados principales

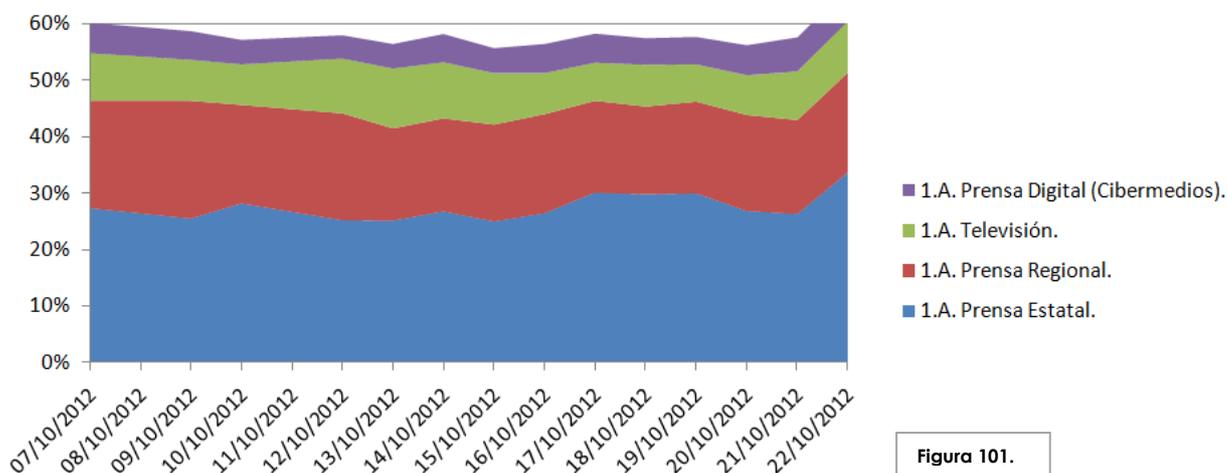


Figura 101.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Estatal.** El promedio del conglomerado fue del **27,64%**. El día de menor visibilidad fue el 15 con un 25%, y el de máxima el posterior a las elecciones con un 33,7%.
- **Prensa Regional.** El promedio de visibilidad fue del **17,32%**. El día de menor visibilidad fue 18 con un 15,5%, y el de máxima el 9 con un 20,8%.
- **Televisión.** El promedio del conglomerado fue del **8,22%**. El día de menor visibilidad fue el 19 con un 6,6%, y el de máxima el 13 con un 10,6%.
- **Prensa Digital (Cibermedios).** El promedio de visibilidad fue del **5,02%**. El día de menor visibilidad fue el 12 con un 4,15%, y el de máxima el día de las elecciones con un 6,07%.

>Elecciones Galicia 2012> Estrato Superior > Conglomerados secundarios

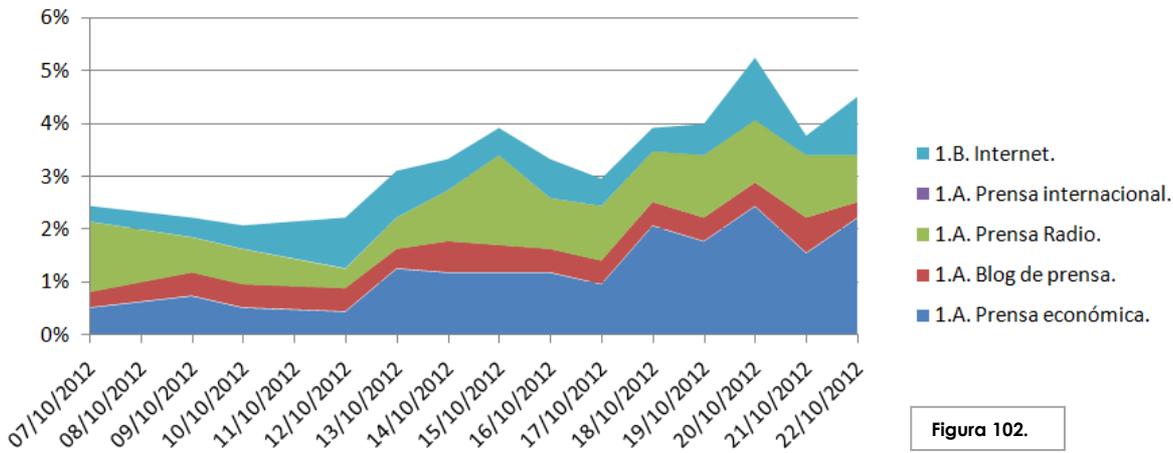


Figura 102.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Económica.** El promedio del conglomerado fue del **1,29%**. El día de menor visibilidad fue el 12 con un 0,44%, y el de máxima el día anterior a las elecciones con un 2,44%. En estos comicios también destaca el incremento en visibilidad los días más cercanos a los comicios.
- **Blog de Prensa.** El promedio de visibilidad fue del **0,45%**. El día de menor visibilidad fue el primer día de captura con un 0,3%, y el de máxima el día de los comicios con un 0,67%.
- **Radio.** El promedio del conglomerado fue del **0,98%**. El día de menor visibilidad fue el 12 con un 0,37%, y el de máxima el 15 con un 1,7%

1.A.Prensa Internacional. Este conglomerado no alcanzó visibilidad en el Top10.

1.B.Internet. El conglomerado de medios de internet alcanzó una visibilidad promedio de **0,65%**. El día de menos visibilidad fue el primer día de captura con un 0,3%, y el de mayor el día post-electoral con un 1,11%.

>Elecciones Galicia 2012> Estrato Medio

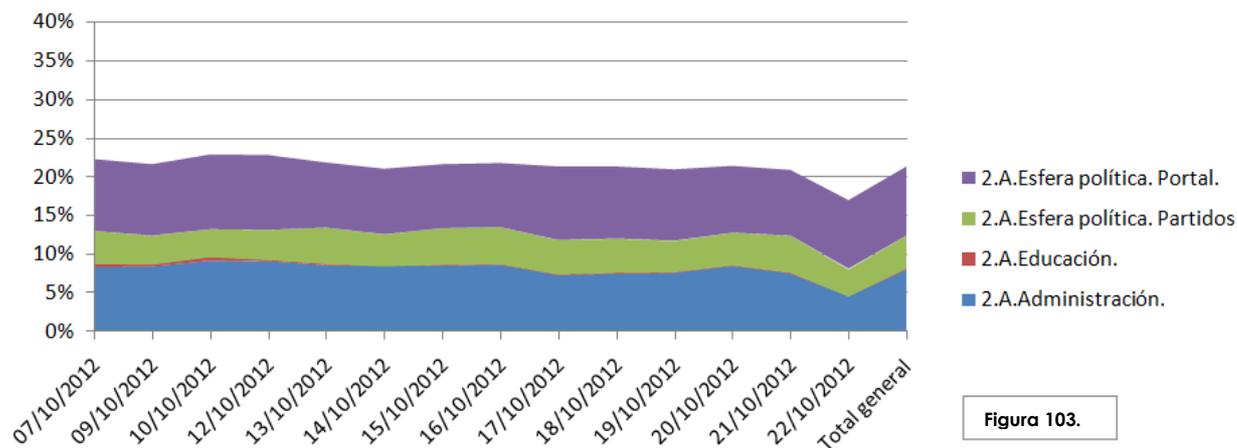


Figura 103.

2.A.Administración. El promedio del conglomerado fue del **7,89%**. El día de mayor visibilidad fue el día 10 con un 9,11%.

2.A.Educación. El promedio del conglomerado fue del **0,2%**. El día de mayor visibilidad fue el día 10 con un 0,52%.

2.A.Esfera política.Portal. El promedio del conglomerado fue del **8,98%**. El día de mayor visibilidad fue el día 10 con un 9,7%.

- **Esfera política.Portal electoral.** El promedio de este conglomerado es del 7,15. Destacan los portales electometro.es (2,61%). portalelectoral.es (1,77%).
- **Esfera política.Portal político.** El promedio de este conglomerado es en torno al 1,5%, destacando politicaxxi.net (1,42%).

2.A.Esfera política.Partidos. El promedio del conglomerado fue del **4,19%**. El día de mayor visibilidad fue el día 16 con un 4,74%. Los detalles de los partidos aparecen en la siguiente figura.

>Elecciones Galicia 2012> Estrato Medio> Partidos

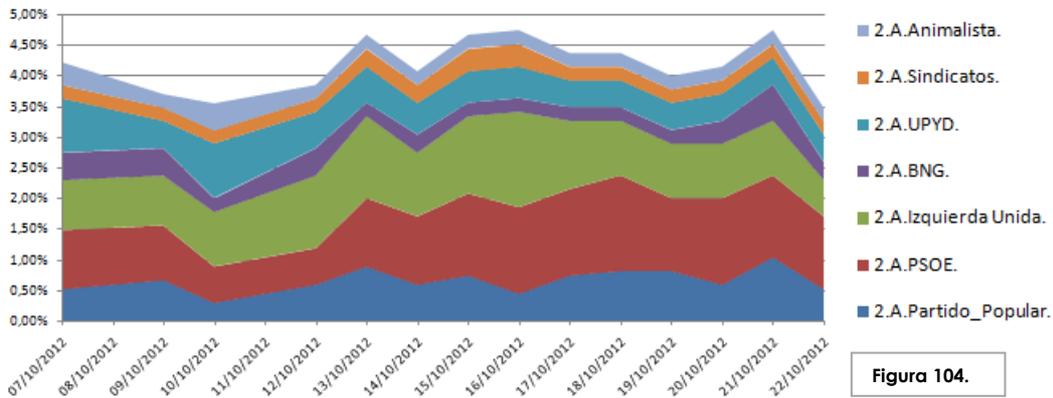
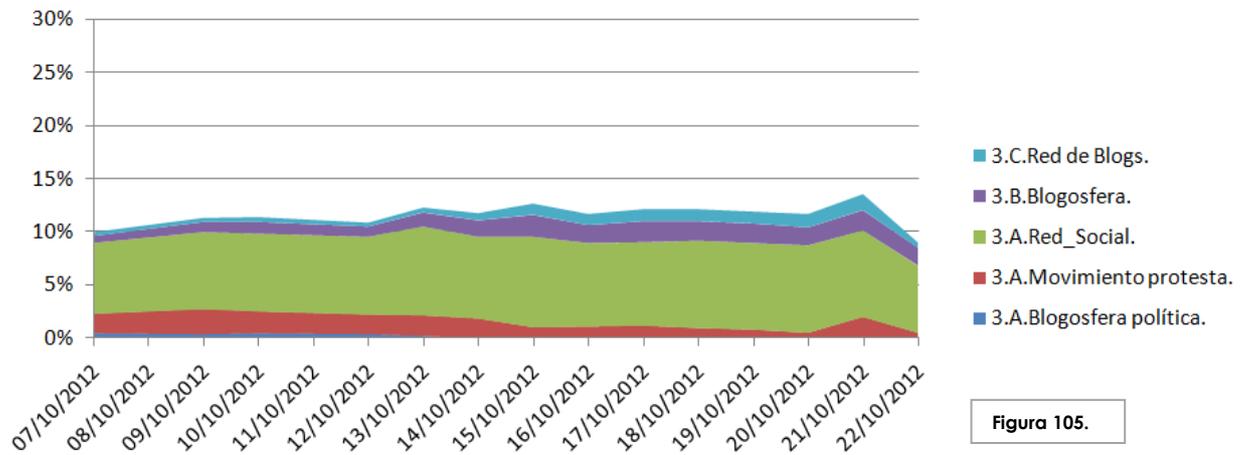


Figura 104.

2.A.Esfera política.

- **Partido_Popular.** El conjunto de páginas del PP tiene una visibilidad promedio del **0,66%** durante toda la serie. El día 13 alcanza un 0,89% y el día de los comicios, el 21 de octubre, alcanza su máximo de la serie con un 1,04%.
- **PSOE.** El conjunto de páginas del Partido Socialista Obrero Español tiene una visibilidad promedio del **1,15%** durante toda la serie. Las cifras más bajas de visibilidad pertenecen a la primera semana, alcanzando su máximo el día 20 con un 1,41%.
- **Izquierda Unida-ICV.** El conjunto de páginas de Izquierda Unida tiene una visibilidad promedio del **1%**. El día 16 alcanza su máximo con un 1,56% de resultados.
- **Resto de partidos.** Las demás agrupaciones políticas logran una visibilidad menor entre las que destaca UPYD con un 0,54% y el BNG con un 0,32%, seguidos por el Partido Animalista con un 0,25%.

>Elecciones Galicia 2012> Estrato Micro



3.A.Blogosfera política. El promedio del conglomerado fue del **0,17%**. El día de mayor visibilidad fue el día 7 con un 0,44%.

3.A.Movimiento protesta. El promedio del conglomerado fue del **1,39%**. El día de mayor visibilidad fue el día 9 con un 2,37%. Destacó la página 'nodo50.org'

3.A.Red_Social. El promedio del conglomerado fue del **7,71%**. El día de mayor visibilidad fue el día 15 con un 8,52%.

- **Wikipedia:** este sitio acumuló el **84%** de los UGC de estas elecciones.
- **Youtube:** los resultados del portal de vídeo fueron un **5,22%** del total destacando el posicionamiento de vídeos del partido animalista PACMA y UPyD.
- **Portal ciudadano:** estas páginas fueron el **2,88%** del conglomerado.
- **Portal foro:** los foros fueron el **2,06%** del conglomerado.
- **Facebook:** sus páginas supusieron el **3,02%** del conglomerado.
- **Twitter:** los resultados de esta red fueron el **2,75%** con resultados de las cuentas del candidato del PP y de la Falange de las Jons.

3.B.Blogosfera. El promedio del conglomerado fue del **1,53%**. El día de mayor visibilidad fue el día 15 con un 2,07%.

3.C.Red de Blogs. El promedio del conglomerado fue del **0,81%**. El día de mayor visibilidad fue el día 21 con un 1,48%.

3.2.6. Generalitat de Catalunya – 25/11/2012

>Elecciones Generalitat 2012> Estrato Superior > Conglomerados principales

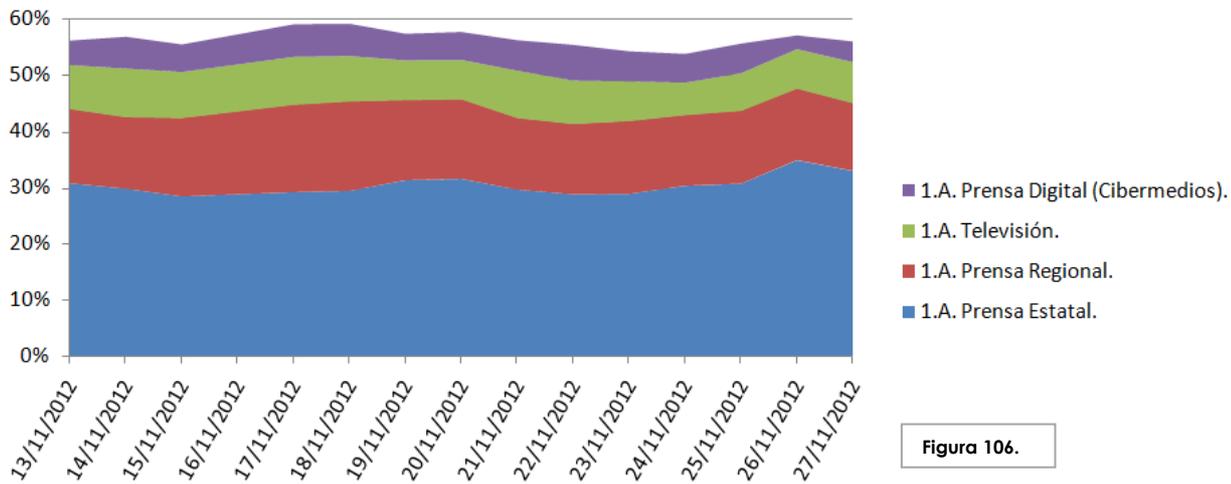


Figura 106.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Estatal.** El promedio del conglomerado fue del **30,53%**. El día de menor visibilidad fue el 15 con un 28,48%, y el de máxima el 26 con un 34,92%.
- **Prensa Regional.** El promedio de visibilidad fue del **13,44%**. El día de menor visibilidad fue el 27 con un 12%, y el de máxima el 18 con un 15,91%.
- **Televisión.** El promedio del conglomerado fue del **7,58%**. El día de menor visibilidad fue el 24 con un 5,83%, y el de máxima el 14 con un 8,71%.
- **Prensa Digital (Cibermedios).** El promedio de visibilidad fue del **4,94%**. El día de menor visibilidad fue el 26 con un 2,42%, y el de máxima el 22 con un 6,29%.

>Elecciones Generalitat 2012> Estrato Superior > Conglomerados secundarios

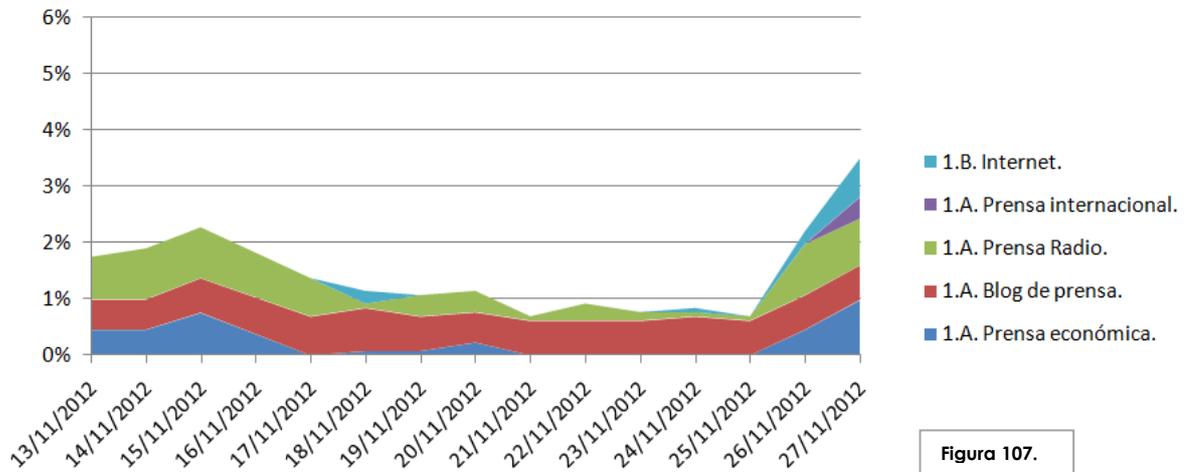


Figura 107.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Económica.** El promedio del conglomerado fue del **0,25%**. El día de mayor visibilidad fue el 27 con un 0,98%. Aunque menor, al igual que en el resto de comicios destaca el incremento en visibilidad los días posteriores a los comicios.
- **Blog de Prensa.** El promedio de visibilidad fue del **0,61%**. El día de mayor visibilidad fue el 18 con un 0,76%.
- **Radio.** El promedio del conglomerado fue del **0,47%**. Los días de mayor visibilidad fueron 14-15 y el 26 con un 0,91%.

1.A.Prensa Internacional. El promedio de visibilidad fue del **0,03%**. El único día de visibilidad fue el 27 con un 0,38%.

1.B.Internet. El conglomerado de medios de internet alcanzó una visibilidad promedio de **0,09%**. El día de mayor visibilidad fue el 27 con un 0,68%.

>Elecciones Generalitat 2012> Estrato Medio

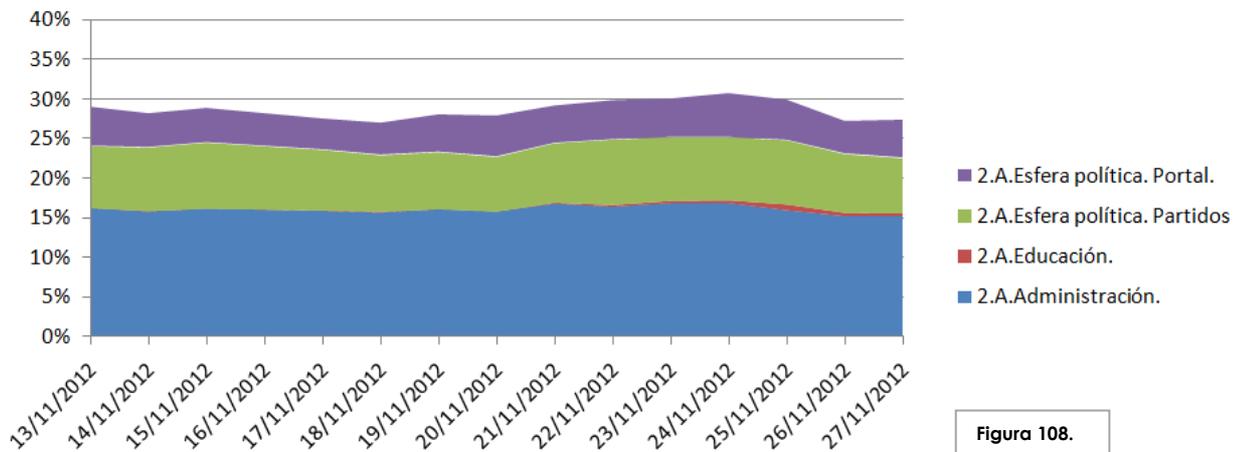


Figura 108.

2.A.Administración. El promedio del conglomerado fue del **16,1%**. El día de mayor visibilidad fueron los días 23-24 con un 16,89%⁷.

2.A.Educación. El promedio del conglomerado fue del **0,21%**. El día de mayor visibilidad fue el día 25 con un 0,76%.

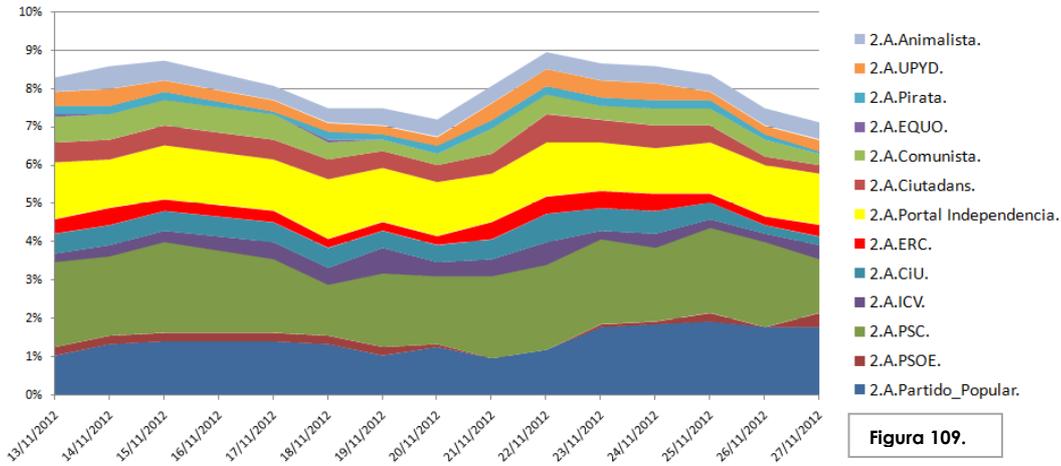
2.A.Esfera política.Portal. El promedio del conglomerado fue del **4,99%**. El día de mayor visibilidad fue el día 24 con un 5,61%.

- **Esfera política.Portal electoral.** El promedio de este conglomerado es del 4,56%. Destacan los portales electometro.es (3,36%) y portalelectoral.es (0,52%).
- **Esfera política.Portal político.** El promedio de este conglomerado es residual.

2.A.Esfera política.Partidos. El promedio del conglomerado fue del **7,66%**. El día de mayor visibilidad fue el día 15 con un 8,33%. Vemos el detalle en la siguiente figura.

⁷ 'Gencat.cat', el principal sitio de la administración catalana acumuló el 67,8% de los resultados provenientes de la administración, bajando la proporción de 2010 pero incrementando el número de resultados de 1125 en 2010 a 2017 resultados en 2012. Los diversos sitios de 'Administración estatal' pasaron del 7,21% en 2010 al 18,35% en 2012.

>Elecciones Generalitat 2012> Estrato Medio> Partidos



2.A.Esfera política.

- **Partido_Popular.** El conjunto de páginas del PP tiene una visibilidad promedio del **1,43%** durante toda la serie. El día de las elecciones alcanza el máximo de la serie con un 1,93%.
- **PSOE-PSC.** El conjunto de páginas del Partido Socialista Obrero Español tiene una visibilidad promedio del **0,15%** durante toda la serie, mientras que las páginas del Partido Socialista Catalán alcanzan el **2%**. La suma de ambas es del **2,15%** de la visibilidad total. La visibilidad del PSC es constante a lo largo de la serie.
- **Izquierda Unida-ICV.** El conjunto de páginas de Izquierda Unida tiene una visibilidad promedio del **0,37%**. Tiene más visibilidad la semana central de datos.
- **Resto de partidos.** Las demás agrupaciones políticas logran una visibilidad menor: CiU con un 0,49%, ERC con un 0,33%, Ciudadans con un 0,49% y Comunista con un 0,49%. Destaca el conglomerado Portal Independencia con un 1,35%.

>Elecciones Generalitat 2012> Estrato Micro

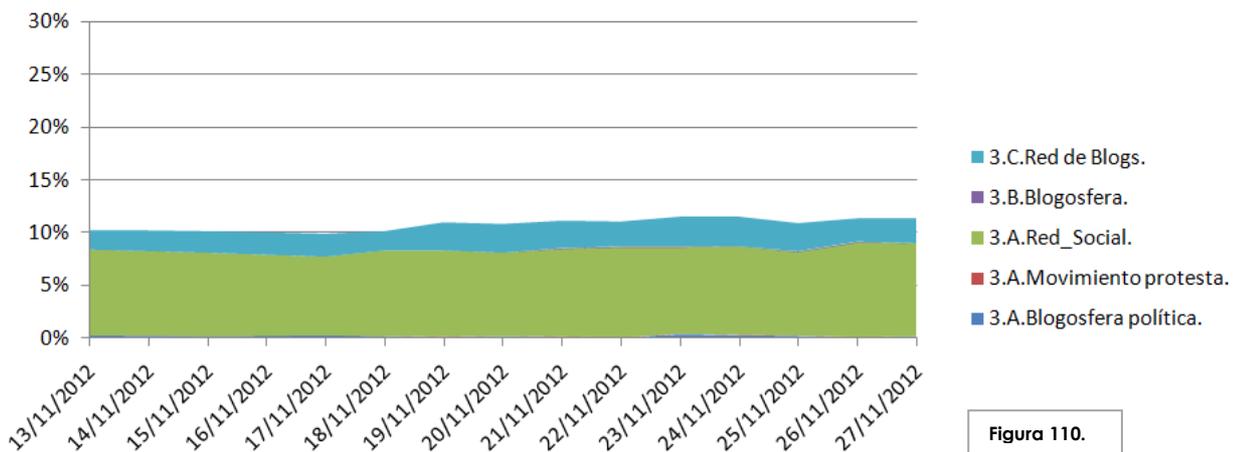


Figura 110.

3.A.Blogosfera política. El promedio del conglomerado fue del **0,23%**. El día de mayor visibilidad fue el día 23 con un 0,38%. Conglomerados y sitios web destacados:

- **Blogs políticos:** los blogs políticos sin adscripción partidista lograron posicionarse **9** veces en Top10
- **Blogs de PP y ICV:** los blogs políticos de partido que en 2010 habían logrado visibilidad quedaron reducidos dos años después a 7 menciones de la blogosfera del Partido Popular ('pp exterior.wordpress.com'), y 25 de Iniciativa Catalunya Verds ('ricardgoma.blogspot.com').

3.A.Movimiento protesta. El promedio del conglomerado fue residual, del **0,04%** frente a un 0,94% dos años antes. El día de mayor visibilidad fue del 0,08%. Los sitios web protesta desaparecieron del Top10, apareciendo únicamente resultados del host 'old.kaosenlared.net'.

3.A.Red_Social. El promedio del conglomerado fue del **8,17%**. El día de mayor visibilidad fue el día 26 con un 8,94%. Se contabilizaron **9 sitios** distintos frente a los 40 de las elecciones de 2010. Sitios web destacados:

- **Wikipedia:** el principal sitio web fue Wikipedia (host 'es' y 'ca') acumuló el **91,5%** de los UGC aparecidos en Top10 frente al 50,1% de 2010.
- **Youtube:** los resultados del portal de vídeo fueron un **5,3%** del total de UGC frente al 3,22% de dos años antes.
- **Facebook:** las páginas de esta red social fueron el **1,13%** frente al 3,18% en 2010
- **Twitter:** los resultados de micro-blogging desaparecieron del Top10

3.B.Blogosfera. El promedio del conglomerado fue del **0,03%** frente al 0,94% de 2010. El día de mayor visibilidad fue un 0,08%.

3.C.Red de Blogs. El promedio del conglomerado pasó a ser del **2,35%** mientras que en 2010 era del 8,46%. Los días de mayor visibilidad fueron el 23 y 24 con un 2,8%.

3.2.7. Tendencias Google Web Search

Si en el punto 3.2. mostramos cuál había sido la evolución de la composición de los tres estratos en el motor de búsqueda web de Google, en este apartado hacemos una recapitulación de los datos vistos en cada una de las elecciones, destacando las tendencias detectadas dentro de las series temporales de cada uno de los comicios.

Estrato Superior. El promedio global en la serie de seis elecciones es del 47,6%. El Estrato Superior aumenta su visibilidad del 39,8% en las elecciones a la Generalitat de 2010 a un 57,67% dos años más tarde.

Estrato Medio. Durante el periodo de estudio de 2010 a 2012 el estrato mantiene su visibilidad en torno a un 36%. Su visibilidad máxima la alcanza en las elecciones a la Junta de Andalucía con un 46%, y la menor en las de Catalunya 2012 con un 31,2%.

Estrato Micro. El Estrato Micro se ve reducido durante toda la serie del periodo de estudio de 2010 a 2012, pasando del 27,8% en las elecciones a la Generalitat de 2010 a apenas el 11,2% dos años más tarde.

Generalitat Catalunya – 28 Noviembre 2010.

Estrato Superior. El conjunto de páginas de medios convencionales a visibilidad del se mantiene estable todo el periodo en torno a un 39,5%. Las variaciones internas del estrato más significativas las encontramos en los días más cercanos a los comicios, pudiendo observar que el día de las elecciones cae la visibilidad de la prensa 'Estatal' siendo compensada por un aumento de los medios de prensa 'Regional. El conglomerado de medios de 'Televisión' sube los dos días anteriores a las elecciones y se mantiene el día de los comicios.

En la serie conglomerados secundarios destaca el aumento de los conglomerados de prensa 'Económica' y 'Radio' los días más cercanos a los comicios y la entrada en resultados de medios de 'Internet' como Terra y Yahoo la semana de las elecciones, en especial el día de los comicios y posterior. Este patrón puede ser debido al factor de posicionamiento del CTR de SERP, es decir, al comportamiento de los usuarios durante esos días específicos

destacando una mayor tendencia al clic en los sitios web de dichos conglomerados.

Estrato Medio. La visibilidad se mantiene estable todo el periodo en torno al 32,32% en un intervalo entre el 30,17% y el 35,56%. El día de los comicios es ligeramente superior al promedio llegando al 33,5% el día posterior. Los sitios de la 'Esfera política' concentran una mayor visibilidad la primera semana de la serie y caen en los días post-electorales.

Estrato Micro. La visibilidad se mantiene estable en torno al 27,5%. La Blogosfera política concentra su máxima visibilidad la primera semana de la serie y cae en los días post-electorales al igual que los sitios de la 'Esfera política' del Estrato Medio.

Elecciones Municipales-22 Mayo 2011.

Estrato Superior. La visibilidad de los medios convencionales tiene un promedio del 32,2%, que se ve incrementada progresivamente a lo largo de la serie con máximos los días de las elecciones y post-electoral. A diferencia de las elecciones a la Generalitat, todos los medios de prensa incrementan su visibilidad el día post-electoral.

Estrato Medio. La visibilidad es entorno al 50% todo el periodo. Destaca la pérdida de visibilidad progresiva a lo largo de toda la serie temporal hasta llegar a mínimos el día post-electoral con un 47,36%. Los sitios de la administración 'Estatal' logran máximos el día posterior a los comicios. Los sitios de la 'Esfera política' sufren una caída y alcanzan mínimos el día post-electoral, probablemente debido a una pérdida de interés por parte de los usuarios reflejada en el CTR de SERPs.

Estrato Micro. La visibilidad se mantiene estable todo el periodo en una franja del 16-18%. Comienza en máximos a inicios de la serie y alcanza su mínimo (16%) el día posterior a las elecciones. Los blogs de partido tienen una visibilidad residual, y se detectan incrementos en la visibilidad de 'Redes de blogs' a finales de la serie.

Elecciones Generales- 20 Noviembre 2011.

Estrato Superior. La visibilidad de los medios de prensa tiene un promedio del 51,8%. Se observa una mayor visibilidad del estrato los días centrales de los comicios con máximos el día de las elecciones con un 56%. El repunte de visibilidad viene de los conglomerados de 'Prensa Estatal', 'Televisión' y 'Digital', mientras que 'Prensa Regional' se mantiene estable.

En la serie conglomerados secundarios, los sitios de 'Radio' alcanzan su máxima visibilidad el día de las elecciones y posterior. La 'Prensa económica' logra su máxima visibilidad el día posterior a los comicios con un 6,35%, tendencia que denota una preocupación por la economía. De igual forma, la 'Prensa internacional', tanto genérica como económica, suben y alcanzan máximos de la serie el día de los comicios y posterior.

Estrato Medio. La visibilidad tiene un promedio del 34%, con mínimos de la serie el día de las elecciones. El PP tiene una visibilidad media (3,6%) superior a la del PSOE (3,4%). Destaca la caída los días precedentes a las elecciones de los partidos políticos con mínimos el día de las elecciones y un tímido repunte el día posterior. Los portales electorales se mantienen toda la serie en torno a un 4-5%.

Estrato Micro. La visibilidad se mantiene estable todo el periodo entre un 12,5-15%. Los sitios de la blogosfera política y el 'Movimiento protesta' apenas aparecen en el Top10.

Junta de Andalucía - Marzo 2011.

Estrato Superior. La visibilidad de este estrato se mantiene estable todo el periodo en torno a un 42% y se dispara el día posterior a las elecciones alcanzando un 54,5%. El conglomerado que más visibilidad logra después del día de las elecciones es el de 'Prensa estatal', probablemente debido a un aumento de consultas (CTR) en SERPs.

En la serie conglomerados secundarios la visibilidad fluctúa entre el 4% y el 8%, y al igual que hemos observado en las elecciones precedentes, el día post-electoral la prensa económica repunta junto a la radio.

Estrato Medio. La visibilidad se mantiene estable todo el periodo sobre un 46% y se hunde el día posterior a los comicios bajando a un 36,25%. Los portales políticos pierden casi por completo su visibilidad alcanzada durante el periodo electoral, probablemente debido a un bajo CTR de los usuarios que clican en sitios web de prensa.

Estrato Micro. La visibilidad se mantiene estable alrededor del 11% y cae ligeramente el día posterior a las elecciones. El 'Movimiento protesta' alcanza visibilidad la segunda semana de la serie.

Xunta Galicia – 21 Octubre 2012.

Estrato Superior. Esta serie temporal es la de máxima visibilidad para este estrato. El promedio de visibilidad de medios convencionales es del 61%, y sube hasta el 71% el día posterior debido principalmente a la 'Prensa estatal'.

Al igual que las otras elecciones, los medios de prensa económica y de Internet tienen una visibilidad estable durante toda la serie en torno al 2,5%, pero que se dobla el día de las elecciones.

Estrato Medio. La visibilidad se mantiene estable todo el periodo sobre el 26%. Cae el día posterior a las elecciones, en especial administración y partidos políticos a excepción de PP y PSOE, con máximos el día de los comicios.

Estrato Micro. La visibilidad se mantiene estable todo el periodo sobre el 10%, alcanza su máximo el día de las elecciones y cae al 9% el día posterior. El 'Movimiento protesta' alcanza visibilidad durante toda la serie con un promedio del 1,4%.

Generalitat Catalunya - 25 Noviembre 2012.

Estrato Superior. La visibilidad de este estrato es del 60,57% y alcanza máximos los días post-electorales. De manera similar a lo sucedido dos años antes, el repunte de visibilidad viene principalmente del conglomerado de 'Prensa Estatal' que sube cinco puntos respecto a la media de la serie (del 30% al 35%) y

de 'Prensa económica'. 'Prensa Regional' se mantiene estable todo el periodo mientras que 'Televisión' y prensa 'Digital' caen después de los comicios.

Estrato Medio. La visibilidad promedio es del 31,2% y cae hasta el 29% los días post-electorales. La visibilidad de los portales electorales fluctúa entre el 4% y el 6%, logrando máximos el día de los comicios y el precedente. Destaca el conglomerado 'Partido Popular' que dobla su visibilidad en día de las elecciones respecto al primer día de captura.

Estrato Micro. La visibilidad promedio es del 11,2% y tiene mayor visibilidad durante la segunda semana de la serie. Las redes sociales incrementan su visibilidad y suponen un 8%, mientras que el movimiento protesta apenas es visible.

3.3. Google News Search

Antes de presentar los resultados de este apartado conviene recordar algunas características de Google News como índice independiente. Como ya sabemos, los buscadores y entre ellos Google de forma muy especial mantienen una serie de índices independientes a despecho de que en la SERP suelen presentar una combinación de resultados de los distintos índices. Tal cosa, que sucede desde que, históricamente, Google decidió innovar en su página de resultados y presentarlos en forma combinada, hace que el público entienda de forma espontánea que todos los resultados proceden de un solo índice. Esto no sería ningún problema si los distintos índices siguieran la misma política o la misma tecnología de indización; pero con el caso de Google News no es así. De aquí la necesidad de esta primera aclaración.

Lo más significativo en el contexto de nuestro trabajo es que, a diferencia del índice de sitios web, no cualquier sitio puede formar parte de Google News. En primer lugar, la inclusión de un sitio en Google News no tiene lugar de forma automática, sino que debe solicitarse expresamente a través de un formulario que Google pone a disposición de los editores de sitios de noticias (suponemos que esto no fue necesario en su momento para sitios de grandes empresas de comunicación). A través de ese formulario Google quiere saber si el sitio candidato a formar parte de Google News dispone de una estructura redaccional mínima (varios redactores diferentes identificados y al menos un director o un responsable editorial). Además, en sus directrices de calidad para sitios de Google News, establece que las noticias deben ser originales, deben citar fuentes y estar elaboradas con criterios periodísticos.

Todo lo anterior hace que, en el caso de este índice, la proporción de sitios que aparecen en las primeras páginas vinculados con el Estrato Superior sea mucho mayor. Esto, por un lado, resta interés teórico a priori a considerar los otros dos Estratos, pero aún así era necesario hacer los análisis también referidos a Google News, aunque solamente fuera para corroborar la hipótesis más predecible. Por otro lado, aquí se hace especialmente significativo, en cambio, el análisis de la composición de niveles del Estrato Superior.

Por todo ello, en este tercer apartado del análisis veremos los resultados de Google News y visualizaremos la evolución de los estratos a lo largo de la serie temporal de

elecciones. Como en los casos anteriores, analizaremos las diferentes coronas y los principales conglomerados así como su comportamiento en el buscador durante todo el periodo de estudio 2010-12.

Los aspectos objeto de estudio son:

La serie temporal de elecciones. Se mostrarán las comparativas de series temporales de 14 días correspondientes a las búsquedas en Google News de los seis comicios electorales analizados. El cuadro resumen de la evolución de los estratos es el siguiente (donde puede apreciarse claramente el peso casi marginal, y predecible, de los Estratos medio y micro):

	Catalunya 2010	Municipales 2011	Generales 2011	Andalucía 2012	Galicia 2012	Catalunya 2012
Estrato convencional	95,54%	94,03%	94,93%	97,61%	98,92%	97,93%
Estrato medio	3,81%	5,75%	4,91%	2,31%	0,84%	1,84%
Micro-estrato	0,72%	0,22%	0,15%	0,08%	0,24%	0,23%

Figura 111.

Los estratos y conglomerados principales. El objeto de estudio es la visibilidad de los principales conglomerados y sitios web de prensa específicos.

Los grupos mediáticos. La clasificación de los medios de prensa por grupo mediático permite observar la evolución de visibilidad de cada uno de ellos en el Top10 del buscador de noticias.

La evolución de proporciones⁸ de los diferentes estratos ha sido:

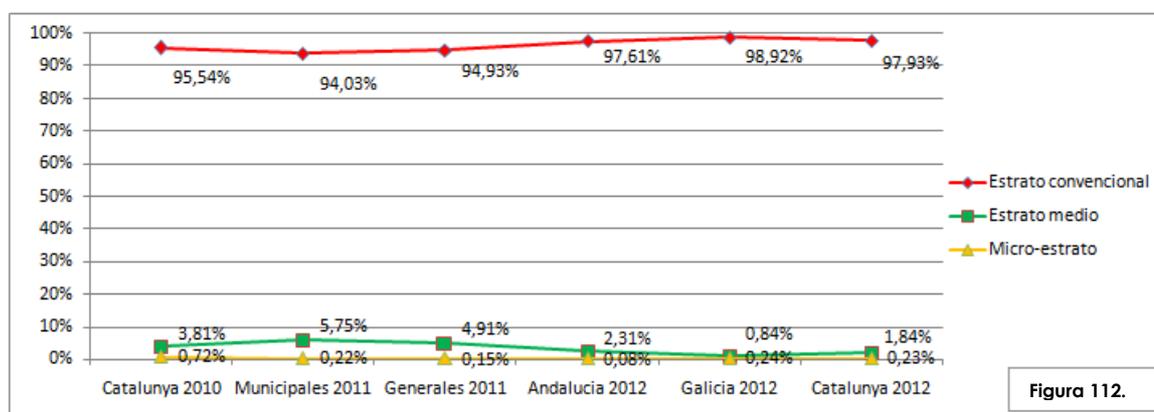


Figura 112.

En el gráfico de líneas podemos apreciar:

⁸ Blogsearch - Gráficos de Líneas. Utilizamos este gráfico para representar los ratios de visibilidad en las seis series temporales de cada una de las elecciones. En este gráfico se representa la composición de la Infoesfera en Web Search y muestra la variación de visibilidad de los estratos.

- **Estrato Superior:** la evolución de los sitios web del Estrato Superior ven incrementada su proporción a lo largo del periodo, pasando del **95,54%** en las elecciones a la Generalitat 2010 a cerca del **98%** en las dos elecciones del año 2012.

Las elecciones donde el ratio es **menor** son las elecciones municipales de 2011 en las que el Estrato Superior solo representa el **94,03%** de los resultados. Las elecciones donde su proporción es **mayor** son las elecciones de Galicia 2012, cuando alcanza cotas del **98,92%**.

- **Estrato Medio:** el promedio del estrato es del **3,25%**. La evolución de los sitios web clasificados en el Estrato Medio tienen una evolución negativa, con máximo en las elecciones municipales con un **5,75%** y mínimos en las elecciones a la Xunta de Galicia con un **0,84%**.
- **El micro estrato:** el promedio de este estrato es del **0,27%** en una tendencia descendente en la que ven disminuida su proporción con máximos en las elecciones a la Generalitat de 2010 con un **0,72%** y los mínimos en las elecciones a la Generalitat de 2012 con un **0,24%**.

3.3.1. Generalitat de Catalunya – 28/11/2010

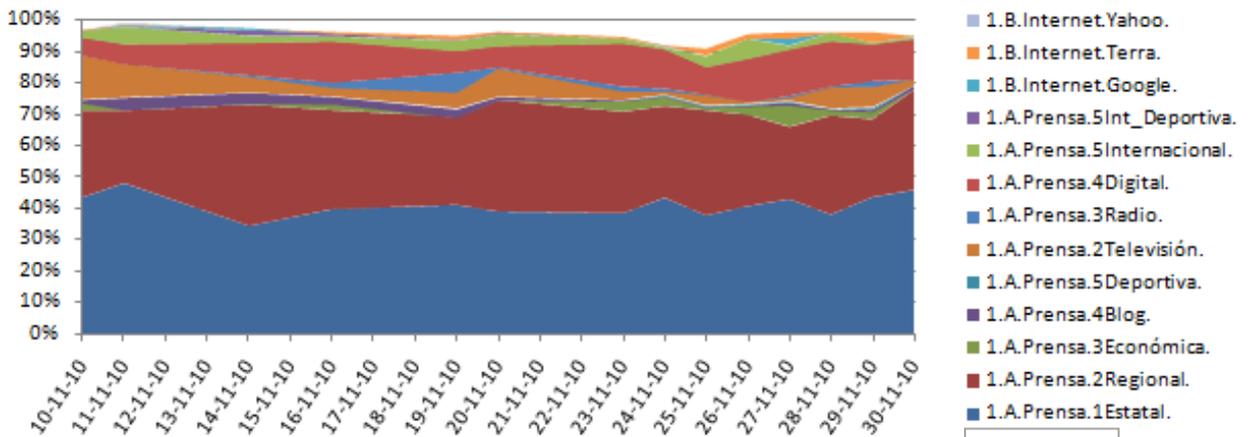


Figura 113.

>Elecciones Generalitat 2010> Estrato Superior > Conglomerados principales⁹

Prensa Estatal. El promedio del conglomerado fue del **41,18%**. El día de menor visibilidad fue el 14-nov con un 34,35%, y el de máxima el 11-nov con un 47,95%. Los días posteriores a las elecciones alcanzó sus otras dos mejores marcas con un 43,53% y un 45,77%.

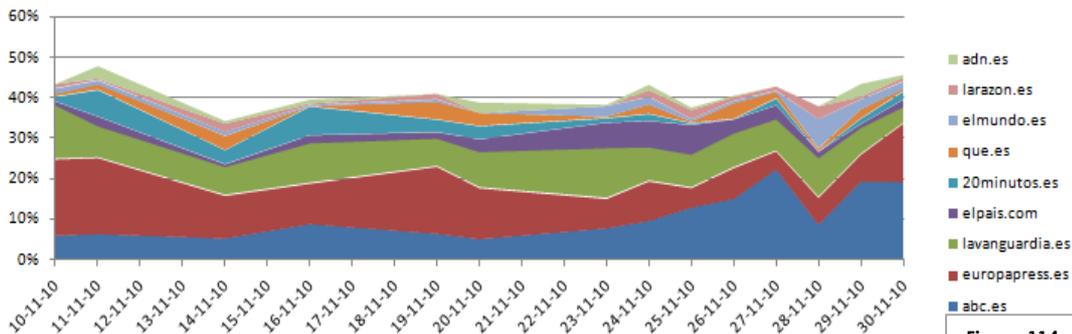


Figura 114.

Prensa Regional. El promedio de este conglomerado fue del **29,61%**. El día de menor visibilidad fue el 27-nov con un 22,86%, y el de máxima el 14-nov con un 38,44%.

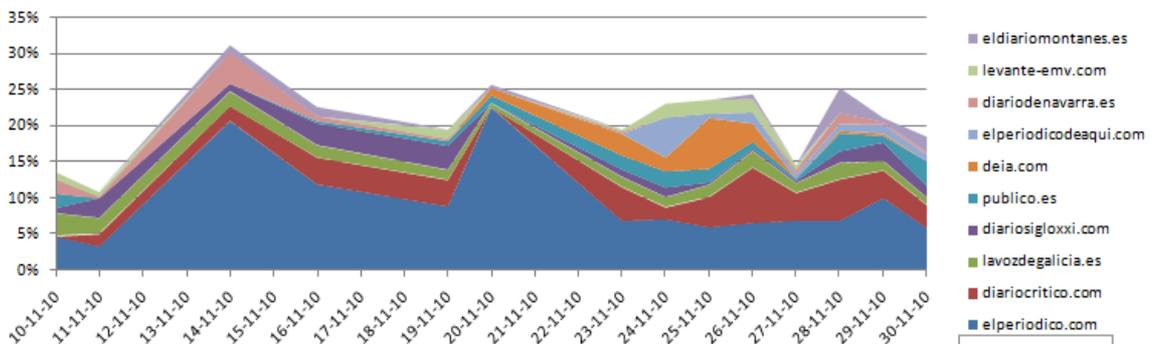


Figura 115.

⁹ El día 20-nov de 2010, durante la sesión de captura de datos, Google bloqueó la IP y se perdió la mitad de las búsquedas, haciendo scraping de 20 de las 44 búsquedas. En algunas de las figuras de estas elecciones el día 20-nov ha sido sustituido por la media de los días circundantes.

Televisión. El promedio del conglomerado fue del **5,03%**. El día de menor visibilidad fue el 26-nov con un 0,95%, y el de máxima el primer día de captura, el 10 de noviembre, con un 14,04%.

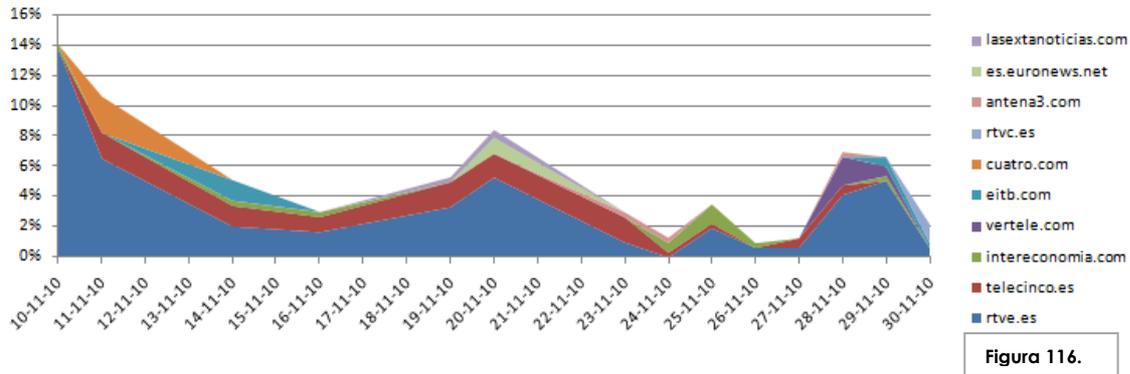


Figura 116.

Prensa Digital (Cibermedios). El promedio de este conglomerado fue del **10,91%**. El día de menor visibilidad fue el 10-nov con un 5,82%, y el de máxima el 27-nov con un 14,6%. El segundo mejor día fue el de las elecciones con un 14,24%.

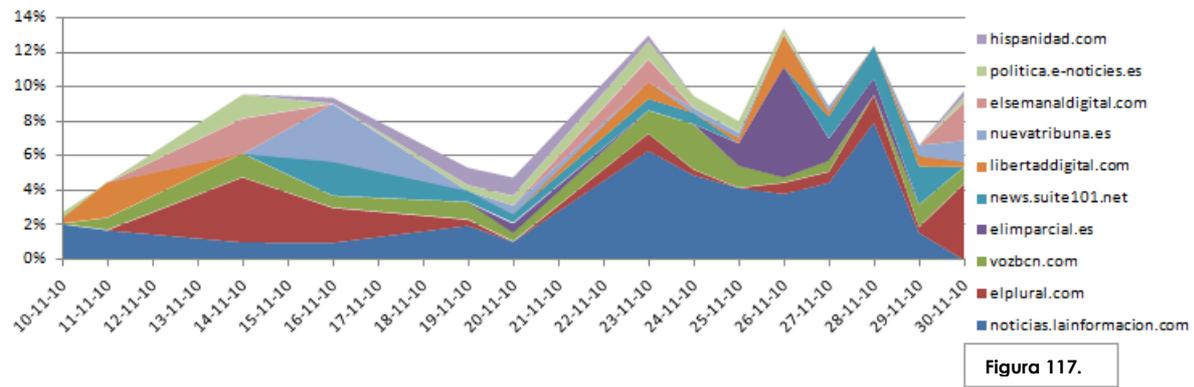


Figura 117.

>Elecciones Generalitat 2010> Estrato Superior > Conglomerados secundarios

Prensa Económica. El promedio de este conglomerado fue del **1,96%**. El día de mayor visibilidad fue el día previo a las elecciones, el 27 de noviembre, con un 6,98%.

Blog de Prensa. El promedio del conglomerado fue del **1,32%**. El día de mayor visibilidad fue el día 11 de noviembre, con un 3,77%.

Radio. El promedio de este conglomerado fue del **1,08%**. El día de mayor visibilidad fue el día 19 de noviembre, con un 6,29%.

Prensa Internacional. El promedio del conglomerado fue del **2,8%**. El día de mayor visibilidad fue el día 26 de noviembre, con un 6,35%.

Internet. El promedio del conglomerado fue del **1,39%**. El día de mayor visibilidad fue el día 29 de noviembre, con un 3,47% del portal Terra.

>Elecciones Generalitat 2010 > Estrato Medio

La visibilidad promedio del segundo estrato fue del **3,81%**. Destaca el conglomerado '2.A.Esfera política.Portal electoral.' con un 2,88% el día 25 de noviembre, y el '2.B.Colectivo.Portal Temático.' que alcanzó un 3,91% el día 24 de noviembre.

Provenientes del Estrato Medio aparecieron 60 portales como <http://animanaturalis.org>, <http://canalpda.com>, <http://republica.es>, <http://www.larepublica.es>, <http://www.iustel.com> o <http://kaosenlared.net>.

>Elecciones Generalitat 2010 > Estrato Micro

La visibilidad promedio del tercer estrato fue del 0,72%. El 'Movimiento protesta' no alcanzó ninguna posición de Top10, mientras que el portal **Youtube** llegó a alcanzar un 1,99% el día 23 de noviembre.

En el Estrato Micro ganaron visibilidad 6 sitios, <http://youtube.com>, <http://hazteoir.org>, <http://internautas.org>, el portal de prensa rosa <http://lacosarosa.com>, <http://gurusblog.com> y <http://rebellion.org>.

>Elecciones Generalitat 2010>Conglomerados

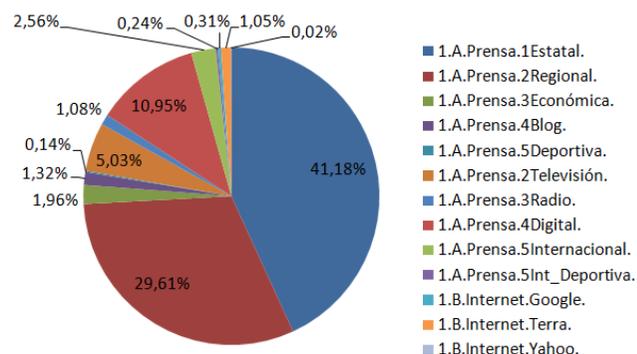


Figura 118.

***Nota explicativa**

El **formato condicional con escalas de color** muestra el color de fondo de una celda de acuerdo a su cercanía con los valores máximo y mínimo dentro del rango. El objetivo es permitir identificar visualmente los valores numéricos que se aproximan tanto al valor máximo como al valor mínimo dentro de un rango. La tonalidad de cada celda dependerá de su cercanía con cualquiera de estos dos extremos.

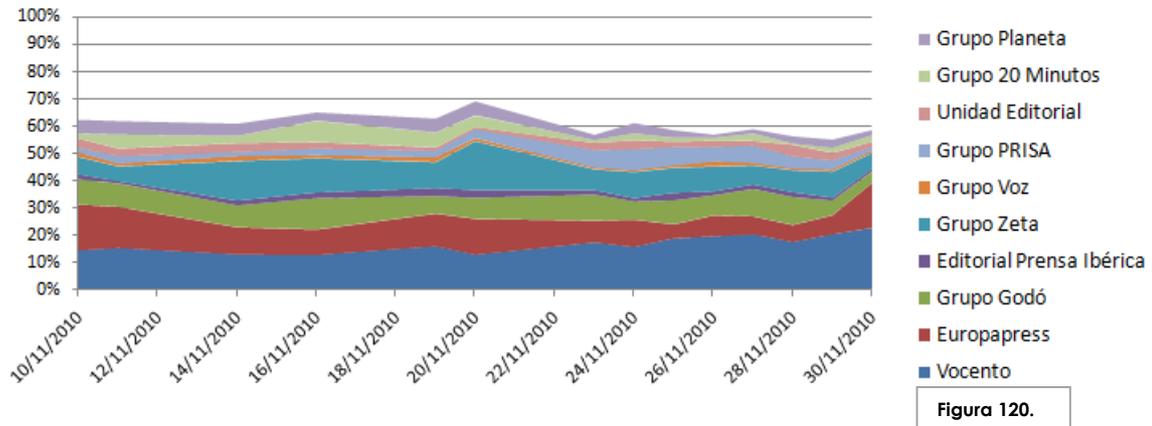
	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Barra roja]	[Barra amarilla]	[Barra verde]

>Elecciones Generalitat 2010>Conglomerado>Tabla de Serie temporal*

Figura 119.

Conglomerados	10-11	11-11	14-11	16-11	19-11	20-11	23-11	24-11	25-11	26-11	27-11	28-11	29-11	30-11	Promedio
1.A.Prensa.1Estatal.	43,49%	47,95%	34,35%	39,67%	41,06%	38,95%	38,41%	43,32%	37,70%	40,63%	42,86%	37,97%	43,53%	45,77%	41,18%
1.A.Prensa.2Regional.	27,40%	22,95%	38,44%	31,33%	27,81%	35,26%	32,12%	28,99%	33,23%	28,89%	22,86%	31,33%	24,61%	31,66%	29,61%
1.A.Prensa.3Económica.	2,40%	0,34%	0,34%	2,00%	0,00%	0,00%	3,31%	2,93%	0,96%	2,86%	6,98%	1,90%	2,52%	0,00%	1,96%
1.A.Prensa.4Blog.	1,03%	3,77%	3,40%	2,33%	2,65%	1,05%	0,33%	0,33%	0,32%	0,32%	0,95%	0,32%	0,63%	1,25%	1,32%
1.A.Prensa.5Deportiva.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,33%	0,32%	0,32%	0,32%	0,00%	0,63%	0,00%	0,14%
1.A.Prensa.2Televisión.	14,04%	10,62%	5,10%	3,00%	5,30%	8,95%	2,98%	1,30%	3,51%	0,95%	1,27%	6,96%	6,62%	2,19%	5,03%
1.A.Prensa.3Radio.	0,00%	0,00%	0,68%	1,67%	6,29%	0,53%	1,66%	0,98%	0,32%	0,00%	0,63%	0,32%	1,89%	0,00%	1,08%
1.A.Prensa.4Digital.	5,82%	6,51%	10,20%	13,00%	6,95%	6,84%	13,58%	12,38%	8,63%	13,65%	14,60%	14,24%	11,67%	12,85%	10,95%
1.A.Prensa.5Internacional.	2,40%	5,48%	2,38%	1,67%	3,31%	3,68%	1,66%	0,65%	3,51%	6,35%	1,27%	2,85%	0,63%	0,63%	2,56%
1.A.Prensa.5Int_Deportiva.	0,00%	0,00%	1,70%	1,00%	0,33%	0,53%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,24%
1.B.Internet.Google.	0,00%	0,68%	0,68%	0,00%	0,00%	0,00%	0,00%	0,33%	0,00%	0,00%	2,22%	0,00%	0,00%	0,31%	0,31%
1.B.Internet.Terra.	0,34%	0,34%	0,00%	0,67%	1,32%	0,53%	0,66%	0,33%	2,24%	1,59%	2,22%	0,32%	3,47%	0,31%	1,05%
1.B.Internet.Yahoo.	0,00%	0,34%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,02%
Total	96,9%	99,0%	97,3%	96,3%	95,0%	96,3%	94,7%	91,9%	90,7%	95,6%	96,2%	96,2%	96,2%	95,0%	95,5%

>Elecciones Generalitat 2010>Grupos mediáticos>Serie temporal



- **Vocento.** El promedio de visibilidad de este grupo fue del **16,88%** sobre el total de resultados aparecidos en Top10 durante el periodo de captura de datos. El **75,07%** del total de resultados de este grupo fueron del diario **abc.es** seguido por el diario **que.es** con un 12,9% y por **elcorreo.com** con un 2%.
- **Europapress.** Su promedio de visibilidad fue del **10,22%**. El día posterior a las elecciones subió hasta el 16,44%.
- **Grupo Godó.** El promedio de visibilidad de este grupo fue del **8,14%**. El día anterior a las elecciones su visibilidad alcanzó un 10,24%.
- **Editorial Prensa Ibérica.** Su promedio de visibilidad fue del **1,77%**.
- **Grupo Zeta.** Su promedio de visibilidad fue del **9,4%**. El día de las elecciones su visibilidad fue del 8,02%.
- **Grupo Voz.** El promedio de visibilidad de este grupo durante el periodo analizado fue del **1,23%**.
- **Grupo PRISA.** El promedio de visibilidad de este grupo en el periodo analizado fue del **4,01%**. Los días centrales de la muestra alcanzó una visibilidad del 7,61%.
- **Unidad Editorial.** El promedio de visibilidad durante el periodo analizado fue del **2,45%**. El día de las elecciones alcanzó su máximo con un 4,23%.
- **Grupo 20 Minutos.** El promedio de visibilidad de este grupo fue del **2,87%**. El día de los comicios bajó al 0,45%.
- **Grupo Planeta.** Su promedio de visibilidad fue del **3,22%**. El día de los comicios alcanzó un 2,56%.

>Elecciones Generalitat 2010>Grupos mediáticos

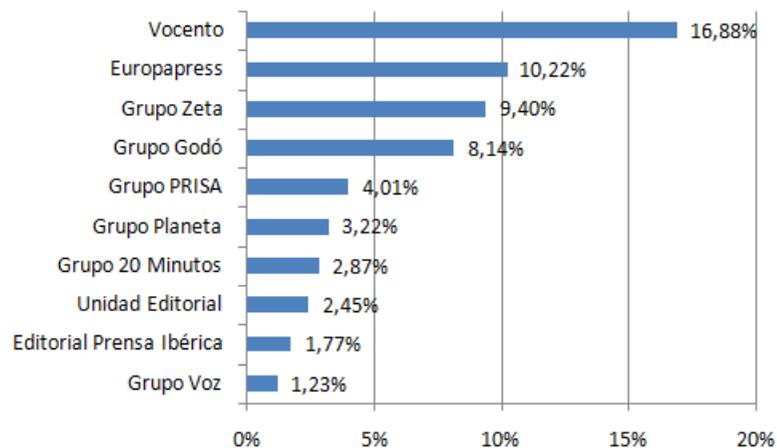


Figura 121.

	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Barra roja]	[Barra amarilla]	[Barra verde]

>Elecciones Generalitat 2010>Top 10 Grupos mediáticos>Tabla de Serie temporal

Figura 122.

Grupos Mediáticos	10-11	11-11	14-11	16-11	19-11	20-11	23-11	24-11	25-11	26-11	27-11	28-11	29-11	30-11	Promedio
Vocento	14,44%	15,26%	12,95%	12,73%	15,92%	12,81%	17,32%	15,68%	18,82%	19,58%	20,33%	17,48%	20,35%	22,71%	16,88%
Europapress	16,78%	15,14%	9,92%	9,26%	12,00%	13,16%	7,97%	9,97%	5,12%	7,63%	6,52%	6,24%	6,93%	16,44%	10,22%
Grupo Godó	9,27%	8,57%	8,05%	11,57%	6,46%	7,72%	9,58%	6,72%	8,80%	7,41%	10,06%	10,24%	5,39%	4,14%	8,14%
Editorial Prensa Ibérica	1,76%	0,82%	1,75%	2,08%	2,88%	2,81%	1,62%	1,23%	2,78%	1,44%	1,66%	1,78%	1,10%	1,01%	1,77%
Grupo Zeta	6,34%	5,40%	14,47%	12,50%	9,46%	17,89%	7,62%	9,52%	9,13%	9,07%	6,96%	8,02%	9,46%	5,82%	9,40%
Grupo Voz	1,76%	1,17%	1,87%	1,16%	1,85%	1,05%	0,81%	0,90%	1,00%	1,99%	1,10%	0,78%	0,99%	0,78%	1,23%
Grupo PRISA	1,76%	2,82%	1,63%	2,55%	2,54%	3,33%	6,35%	7,61%	6,68%	5,31%	6,30%	4,45%	3,30%	1,57%	4,01%
Unidad Editorial	3,40%	2,70%	3,15%	2,31%	1,15%	0,88%	2,66%	3,25%	2,00%	2,21%	1,66%	4,23%	2,75%	1,90%	2,45%
Grupo 20 Minutos	1,76%	5,05%	2,68%	7,75%	5,42%	4,21%	1,04%	2,46%	1,45%	1,11%	2,65%	0,45%	1,87%	2,24%	2,87%
Grupo Planeta	4,93%	4,81%	4,32%	2,89%	4,96%	4,91%	1,85%	3,70%	2,67%	1,11%	1,55%	2,56%	2,97%	1,90%	3,22%
Total	62,2%	61,7%	60,8%	64,8%	62,6%	68,8%	56,8%	61,0%	58,5%	56,9%	58,8%	56,2%	55,1%	58,5%	60,2%

>Elecciones Generalitat 2010>Resto de Grupos mediáticos>Serie temporal

Figura 123.

Grupos Mediáticos	10-11	11-11	14-11	16-11	19-11	20-11	23-11	24-11	25-11	26-11	27-11	28-11	29-11	30-11	Promedio
Grupo Joly	1,17%	0,35%	0,82%	0,23%	0,23%	1,05%	1,15%	0,11%	0,11%	0,00%	0,11%	0,33%	0,11%	1,12%	0,49%
Mediapro	2,70%	2,46%	1,17%	1,50%	2,08%	1,40%	0,69%	1,12%	1,34%	1,11%	0,11%	2,23%	2,31%	2,68%	1,64%
Editorial Iparragirre Corporate Communicator.	0,94%	0,70%	0,23%	0,12%	0,00%	1,58%	1,62%	2,02%	5,23%	1,88%	1,22%	1,89%	1,76%	1,12%	1,45%
Elplural.com	0,00%	0,00%	1,28%	2,08%	0,35%	0,18%	0,81%	0,56%	0,22%	0,77%	0,66%	1,22%	0,22%	2,24%	0,76%
Grupo Begar	0,47%	0,00%	1,63%	1,50%	0,58%	0,18%	1,62%	1,23%	0,67%	0,44%	0,55%	0,00%	0,33%	1,23%	0,75%
Libertad digital S.A	0,35%	2,35%	1,75%	0,81%	0,69%	1,05%	0,81%	0,22%	0,33%	2,32%	0,66%	0,22%	0,77%	0,45%	0,91%
Mediaset España	0,12%	1,53%	1,17%	0,58%	1,27%	1,40%	0,69%	0,11%	0,11%	0,11%	0,22%	0,33%	0,88%	0,89%	0,67%
Radio Popular S.A. - COPE	0,12%	0,00%	0,35%	0,58%	1,38%	0,53%	1,04%	0,67%	0,67%	0,33%	0,44%	0,45%	1,87%	0,00%	0,60%
Grupo Intereconomía	0,47%	0,35%	0,70%	0,46%	0,12%	0,18%	0,00%	0,56%	0,67%	1,22%	0,22%	0,33%	0,11%	0,22%	0,40%
Prensa internacional	2,70%	3,29%	2,92%	3,36%	3,00%	3,86%	2,54%	1,01%	2,23%	4,54%	2,54%	3,56%	2,75%	1,12%	2,81%
Grupo Correo Gallego	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,11%	0,33%	0,55%	0,33%	0,22%	1,23%	0,20%
EKHE (izquierda abertzal)	0,00%	0,00%	0,12%	0,12%	0,23%	0,18%	0,00%	0,00%	0,00%	0,00%	0,22%	0,67%	0,22%	0,89%	0,19%
Grupo Promecal	0,00%	0,23%	0,12%	0,00%	0,00%	0,00%	0,00%	0,00%	1,11%	0,44%	0,00%	0,00%	0,11%	0,11%	0,15%
Sakkia Hiferpress (inmobiliaria)	0,12%	0,00%	0,00%	0,00%	0,35%	0,00%	0,12%	0,00%	0,00%	0,00%	0,00%	0,67%	0,11%	0,00%	0,10%
Grupo publicaciones del sur	0,00%	0,12%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,45%	0,11%	0,00%	0,05%
Grupo La Región.	0,00%	0,12%	0,00%	0,00%	0,00%	0,00%	0,00%	0,11%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,02%
Grupo Serra	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Grupo Alfonso Gallardo	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Otros medios	23,71%	24,06%	21,12%	20,02%	22,61%	17,19%	27,14%	23,96%	21,83%	23,56%	28,73%	27,28%	29,04%	24,27%	23,89%
2.Estrato Medio	4,58%	2,46%	3,38%	2,78%	3,00%	1,93%	4,04%	5,60%	6,01%	5,31%	3,98%	3,23%	3,74%	3,36%	3,81%
3.Estrato Micro	0,35%	0,23%	2,45%	1,04%	1,50%	0,53%	0,92%	1,68%	0,89%	0,77%	0,99%	0,56%	0,22%	0,56%	0,91%
	37,8%	38,3%	39,2%	35,2%	37,4%	31,2%	43,2%	39,0%	41,5%	43,1%	41,2%	43,8%	44,9%	41,5%	39,8%

3.3.2. Elecciones Municipales – 22/5/2011

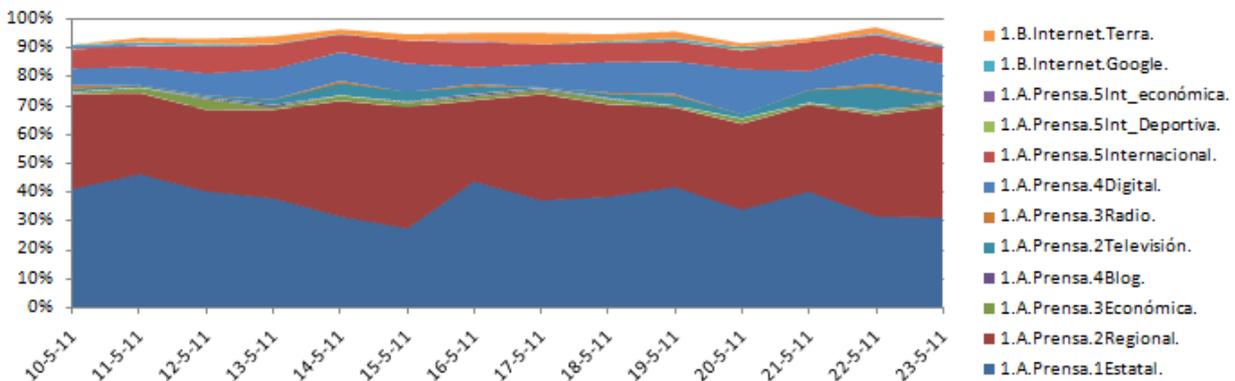


Figura 124.

>Elecciones Municipales 2011> Estrato Superior > Conglomerados principales

Prensa Estatal. El promedio del conglomerado fue del **34%**. El día de menor visibilidad fue el 15 de mayo con un 29,15%, y el de máxima el 10 de mayo con un 40,38%.

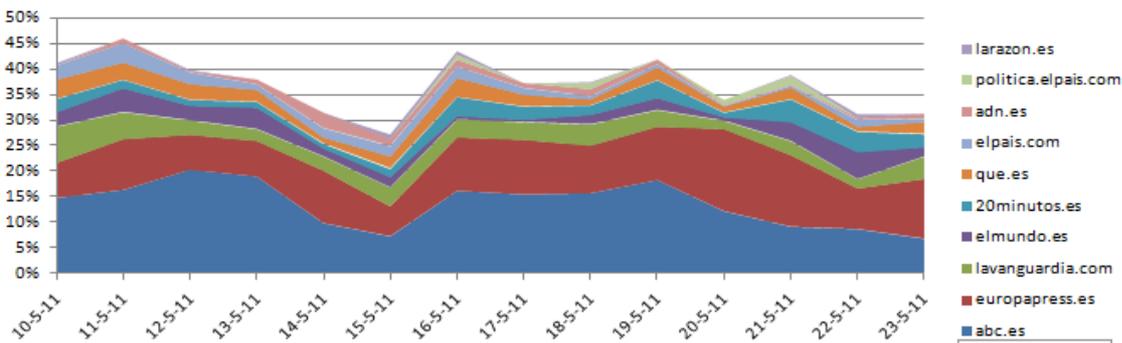


Figura 125.

Prensa Regional. El promedio de este conglomerado fue del **34,61%**. El día de menor visibilidad fue el 11 de mayo con un 31,06%, y el de máxima el 15 con un 41,09%.

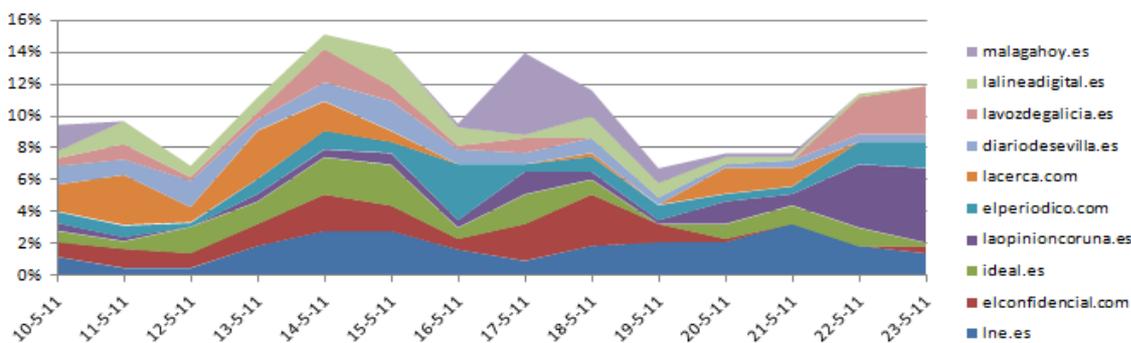


Figura 126.

Televisión. El promedio del conglomerado fue del **2,63%**. El día de menor visibilidad fue el 10 de mayo con un 1,4%, y el de máxima el día de las elecciones con un 5,89%.

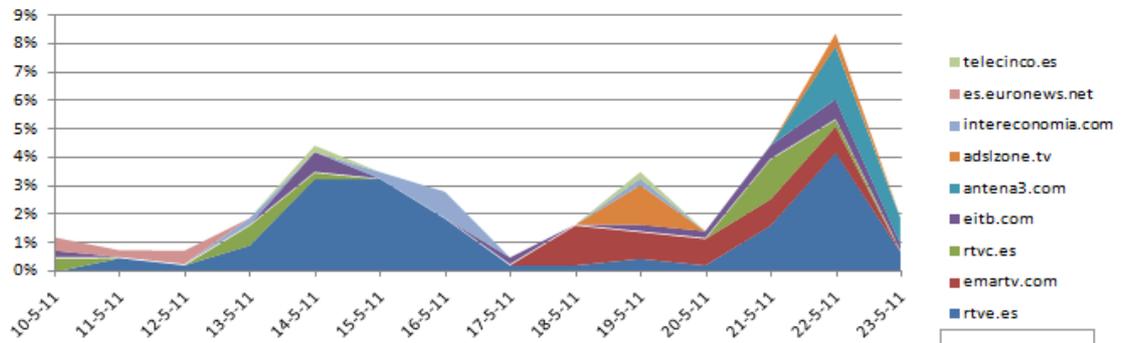


Figura 127.

Prensa Digital (Cibermedios). El promedio de este conglomerado fue del **9,84%**. El día de menor visibilidad fue el 16 de mayo con un 7,52%, y el de máxima el 20 con un 15,66%.

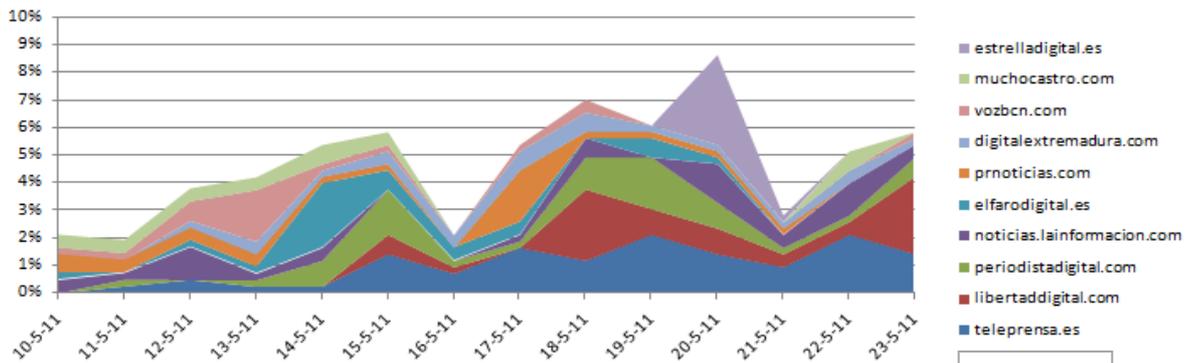


Figura 128.

>Elecciones Municipales 2011> Estrato Superior > Conglomerados secundarios

Prensa Económica. El promedio de este conglomerado fue del **1,61%**. El día de menor visibilidad fue previo a las elecciones, el 22 de mayo, con un 0,78%, y el de mayor el día 12 con un 2,54%.

Blog de Prensa. El promedio del conglomerado fue del **0,41%**. El día de mayor visibilidad fue el 13 de mayo con un 1,01%.

Radio. El promedio de este conglomerado fue del **0,76%**. El día de mayor visibilidad fue el día 22 de mayo, con un 1,4%.

1.A.Prensa Internacional. El promedio fue del **8,23%**. El día de mayor visibilidad fue el día 21 de mayo, con un 10,16%.

1.B.Internet. El promedio del conglomerado de los medios Internet fue del **1,67%**. El día de mayor visibilidad fue el día 17 de mayo, con un 2,64% del portal Terra.

>Elecciones Municipales 2011> Estrato Medio

Este estrato supuso un **5,75%** de los resultados. Del Estrato Medio aparecieron 187 portales que no pertenecían al estrato de prensa convencional. Destacaron sitios web como <http://impulsonegocios.com>, <http://animanaturalis.org>, <http://kaosenlared.net>, <http://trome.pe> o <http://suite101.net>.

>Elecciones Municipales 2011> Estrato Micro

Este estrato supuso un **0,22%** de los resultados. Los resultados del Estrato Micro ganaron visibilidad sitios de social media como <http://youtube.com> y portales como <http://puromarketing.com>.

>Elecciones Municipales 2011>Conglomerados

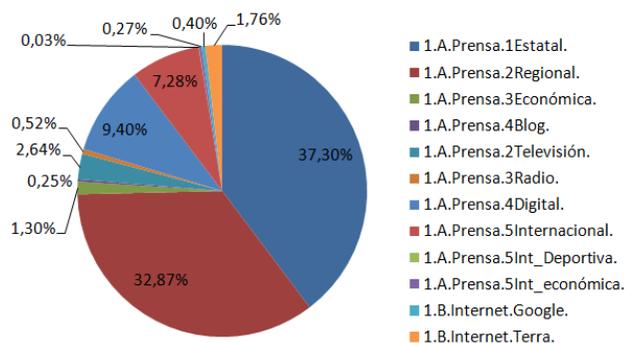


Figura 129.

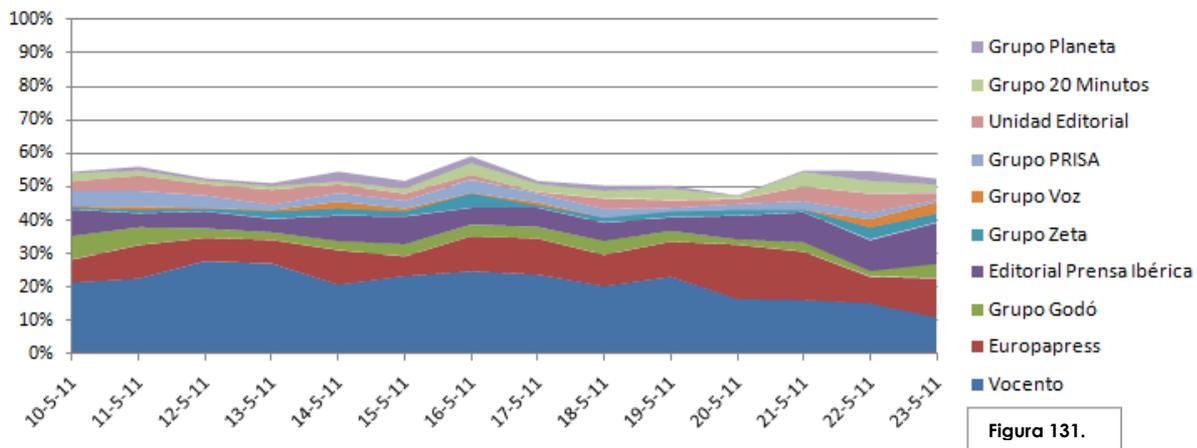
	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Red]	[Yellow]	[Green]

>Elecciones Municipales 2011>Conglomerado>Tabla de Serie temporal

Figura 130.

Conglomerados	10-5	11-5	12-5	13-5	14-5	15-5	16-5	17-5	18-5	19-5	20-5	21-5	22-5	23-5	Promedio
1.A.Prensa.1Estatal.	41,13%	46,25%	40,28%	37,91%	31,63%	27,44%	43,72%	37,21%	38,37%	41,86%	33,95%	40,24%	31,63%	31,16%	37,30%
1.A.Prensa.2Regional.	32,86%	28,09%	28,20%	30,70%	40,00%	42,09%	28,14%	36,74%	32,09%	27,44%	29,77%	30,00%	35,12%	38,60%	32,87%
1.A.Prensa.3Económica.	0,47%	1,69%	3,55%	0,70%	1,63%	1,40%	1,16%	1,40%	1,63%	0,47%	1,63%	0,48%	0,93%	1,16%	1,30%
1.A.Prensa.4Blog.	0,00%	0,00%	0,47%	0,70%	0,00%	0,23%	0,70%	0,47%	0,23%	0,00%	0,00%	0,00%	0,23%	0,47%	0,25%
1.A.Prensa.2Televisión.	1,18%	0,73%	0,71%	1,86%	4,42%	3,49%	2,79%	0,47%	1,63%	3,49%	1,40%	4,52%	8,37%	1,86%	2,64%
1.A.Prensa.3Radio.	1,42%	0,24%	0,24%	0,23%	0,93%	0,00%	0,93%	0,00%	0,47%	0,70%	0,00%	0,24%	1,16%	0,70%	0,52%
1.A.Prensa.4Digital.	5,91%	6,54%	7,82%	10,70%	10,00%	10,00%	6,05%	8,14%	10,70%	11,40%	16,05%	6,67%	10,70%	10,70%	9,40%
1.A.Prensa.5Internacional.	6,62%	7,02%	9,24%	8,37%	6,05%	7,91%	8,37%	6,98%	6,74%	6,98%	6,28%	10,00%	6,28%	5,12%	7,28%
1.A.Prensa.5Int_Deportiva.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,47%	0,00%	0,00%	0,00%	0,03%
1.A.Prensa.5Int_económica.	0,71%	0,73%	0,24%	0,00%	0,00%	0,00%	0,70%	0,00%	0,00%	0,00%	0,47%	0,00%	0,23%	0,70%	0,27%
1.B.Internet.Google.	0,95%	0,73%	0,71%	0,23%	0,23%	0,00%	0,00%	0,00%	0,47%	0,93%	0,47%	0,00%	0,47%	0,47%	0,40%
1.B.Internet.Terra.	0,00%	1,45%	1,66%	2,56%	1,40%	2,09%	2,56%	3,72%	2,33%	2,33%	1,16%	1,19%	1,86%	0,23%	1,76%
Total	91,3%	93,5%	93,1%	94,0%	96,3%	94,7%	95,1%	95,1%	94,7%	95,6%	91,6%	93,3%	97,0%	91,2%	94,0%

>Elecciones Municipales 2011>Grupos mediáticos>Serie temporal



- **Vocento.** El promedio de visibilidad de este grupo fue del **20,91%** sobre el total de resultados aparecidos en Top10 durante el periodo de captura de datos. El **64,4%** del total de resultados de este grupo fueron del diario **abc.es** seguido por el diario **que.es** con un 11,12% y por el **ideal.es** con un 5,52%.
- **Europapress.** Su promedio de visibilidad fue del **9,82%**. El día posterior a las elecciones subió hasta el 11,63%.
- **Grupo Godó.** El promedio de visibilidad de este grupo fue del **3,51%**. El día posterior a las elecciones su visibilidad alcanzó un 4,42%.
- **Editorial Prensa Ibérica.** Su promedio de visibilidad fue del **6,66%**. El día de las elecciones subió hasta un 9,07% y el día posterior hasta el 12,09%.
- **Grupo Zeta.** Su promedio de visibilidad fue del **1,96%**. El día de las elecciones su visibilidad fue del 3,72%.
- **Grupo Voz.** El promedio de visibilidad de este grupo fue del **0,89%**. Su mayor visibilidad fueron los días de las elecciones y el posterior con 2,56% y un 3,49%.
- **Grupo PRISA.** El promedio de visibilidad de este grupo fue del **2,64%**. Los dos primeros días de captura fueron los de más visibilidad con un 4,5% de media.
- **Unidad Editorial.** Su promedio de visibilidad fue del **2,88%**. El día de las elecciones alcanzó el 5,58%.
- **Grupo 20 Minutos.** El promedio de visibilidad de este grupo fue del **2,38%**. El día de los comicios subió hasta el 3,95%, siendo su máximo el día anterior con un 4,52%.
- **Grupo Planeta.** Su promedio de visibilidad fue del **1,42%**. El día de los comicios alcanzó su máximo con un 3,02%.

>Elecciones Municipales 2011>Grupos mediáticos

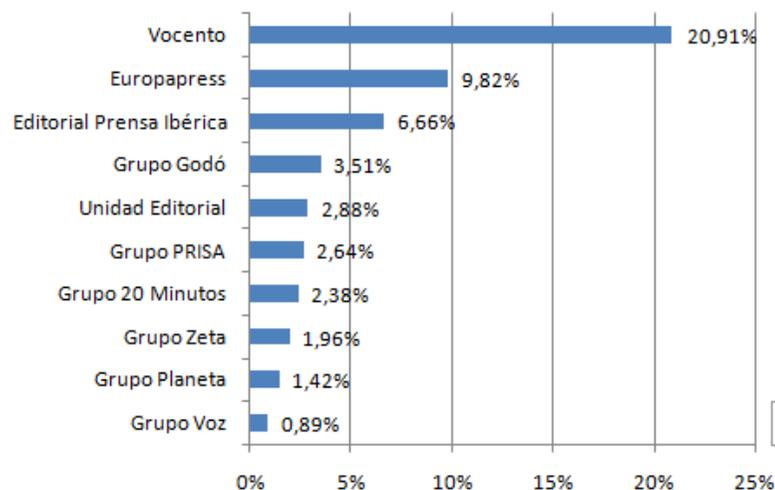


Figura 132.

	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Red]	[Yellow]	[Green]

Figura 133.

>Elecciones Municipales 2011> Top 10 Grupos mediáticos>Tabla de Serie temporal

Grupos Mediáticos	10-5	11-5	12-5	13-5	14-5	15-5	16-5	17-5	18-5	19-5	20-5	21-5	22-5	23-5	Promedio
Vocento	21,28%	22,52%	27,73%	26,98%	20,70%	23,26%	24,65%	23,72%	20,23%	23,02%	16,51%	16,19%	15,12%	10,93%	20,91%
Europapress	6,86%	9,93%	6,87%	6,98%	10,23%	5,81%	10,47%	10,70%	9,30%	10,47%	16,05%	14,29%	7,91%	11,63%	9,82%
Grupo Godó	7,09%	5,33%	2,84%	2,33%	2,79%	3,72%	3,49%	3,49%	4,19%	3,26%	1,63%	2,86%	1,86%	4,42%	3,51%
Editorial Prensa Ibérica	7,80%	4,12%	4,98%	3,95%	7,44%	8,14%	4,88%	5,58%	5,35%	3,95%	6,98%	8,81%	9,07%	12,09%	6,66%
Grupo Zeta	0,71%	1,21%	0,95%	2,33%	2,33%	1,63%	4,42%	0,93%	1,63%	1,86%	1,86%	0,95%	3,72%	2,79%	1,96%
Grupo Voz	0,47%	0,97%	0,24%	0,47%	2,09%	0,93%	0,23%	0,93%	0,00%	0,00%	0,00%	0,00%	2,56%	3,49%	0,89%
Grupo PRISA	4,49%	4,60%	3,79%	1,63%	2,56%	2,33%	3,95%	2,56%	2,79%	1,16%	1,86%	2,62%	2,09%	0,70%	2,64%
Unidad Editorial	2,84%	4,60%	3,32%	4,42%	2,56%	2,09%	1,40%	0,47%	3,02%	2,33%	1,40%	4,29%	5,58%	2,09%	2,88%
Grupo 20 Minutos	2,60%	1,69%	1,18%	1,16%	0,93%	1,63%	3,72%	2,56%	2,33%	3,49%	0,93%	4,52%	3,95%	2,56%	2,38%
Grupo Planeta	0,47%	1,21%	0,71%	0,93%	3,02%	2,33%	2,09%	0,93%	1,63%	0,93%	0,23%	0,48%	3,02%	1,86%	1,42%
	54,6%	56,2%	52,6%	51,2%	54,7%	51,9%	59,3%	51,9%	50,5%	50,5%	47,4%	55,0%	54,9%	52,6%	53,1%

>Elecciones Municipales 2011>Resto de Grupos mediáticos>Serie temporal

Figura 134.

Grupos Mediáticos	10-5	11-5	12-5	13-5	14-5	15-5	16-5	17-5	18-5	19-5	20-5	21-5	22-5	23-5	Promedio
Grupo Joly	3,07%	1,69%	2,13%	0,93%	1,40%	2,09%	2,33%	5,81%	2,56%	3,02%	0,70%	1,19%	2,33%	0,47%	2,12%
Mediapro	0,47%	0,48%	0,71%	0,00%	0,00%	0,93%	0,70%	0,93%	0,70%	0,00%	0,00%	0,48%	0,23%	0,23%	0,42%
Editorial Iparraguirre	2,60%	0,97%	0,00%	0,70%	0,93%	0,00%	0,00%	0,00%	0,00%	0,70%	1,63%	1,67%	0,47%	3,49%	0,94%
Grupo Begar	0,24%	0,00%	1,66%	0,93%	0,47%	0,47%	0,00%	0,00%	0,00%	0,23%	0,23%	0,24%	0,23%	0,00%	0,33%
Libertad digital S.A	0,00%	0,00%	0,00%	0,00%	0,00%	0,70%	0,23%	0,00%	2,56%	0,93%	0,93%	0,48%	0,47%	2,79%	0,65%
Mediaset España	0,00%	0,00%	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	0,00%	0,03%
Radio Popular S.A. - COPE	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	0,70%	0,47%	0,10%
Grupo Intereconomía	0,00%	0,00%	0,00%	0,23%	0,00%	0,23%	0,93%	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	0,00%	0,12%
Prensa internacional	7,33%	7,75%	9,48%	8,37%	6,05%	7,91%	9,07%	6,98%	6,74%	6,98%	7,21%	10,00%	6,51%	5,81%	7,58%
Grupo Correo Gallego	0,24%	0,24%	0,00%	0,00%	0,00%	0,00%	0,23%	0,23%	0,00%	0,00%	0,23%	0,00%	0,00%	0,47%	0,12%
Grupo Promecal	1,89%	1,45%	0,24%	0,00%	0,93%	0,47%	0,00%	0,00%	0,70%	0,00%	0,00%	0,48%	0,00%	0,23%	0,45%
Grupo publicaciones del sur	0,00%	0,00%	0,71%	0,23%	0,00%	0,47%	0,47%	0,00%	0,00%	0,70%	0,23%	0,48%	0,93%	0,93%	0,37%
Grupo La Región.	0,00%	0,00%	0,71%	0,00%	0,00%	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,07%
Grupo Serra	0,00%	0,00%	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,02%
Grupo Alfonso Gallardo	0,00%	0,00%	0,24%	0,47%	0,23%	0,23%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,23%	0,10%
Otros medios	20,80%	24,70%	24,88%	30,93%	31,16%	29,30%	21,86%	29,07%	31,16%	32,09%	33,02%	23,57%	30,23%	23,49%	27,62%
2.Estrato Medio	8,04%	6,54%	6,64%	6,05%	3,72%	3,72%	4,65%	4,65%	4,88%	4,42%	8,14%	6,43%	3,02%	8,84%	5,69%
3.Estrato Micro	0,71%	0,00%	0,00%	0,00%	0,00%	1,63%	0,23%	0,23%	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	0,22%
	45,4%	43,8%	47,4%	48,8%	45,3%	48,1%	40,7%	48,1%	49,5%	49,5%	52,6%	45,0%	45,1%	47,4%	46,9%

3.3.3. Elecciones Generales – 20/11/2011

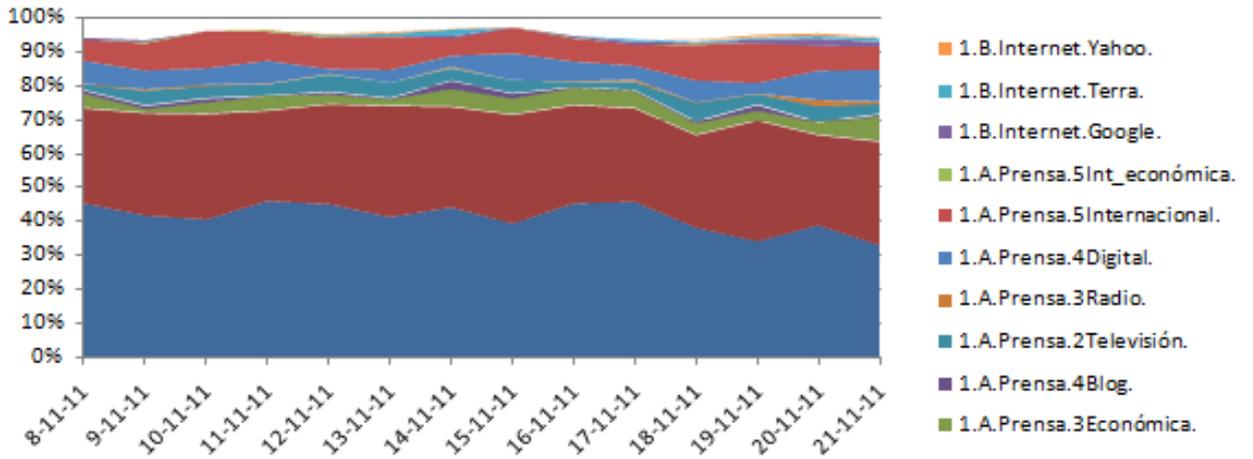


Figura 135.

>Elecciones Generales > Estrato Superior > Conglomerados principales

1.A.Prensa Estatal. El promedio del conglomerado fue del **39,9%**. El día de menor visibilidad fue el 21-nov con un 35,32% y el de máxima el 8 con un 43,73%. La evolución de visibilidad del Top10 de medios fue:

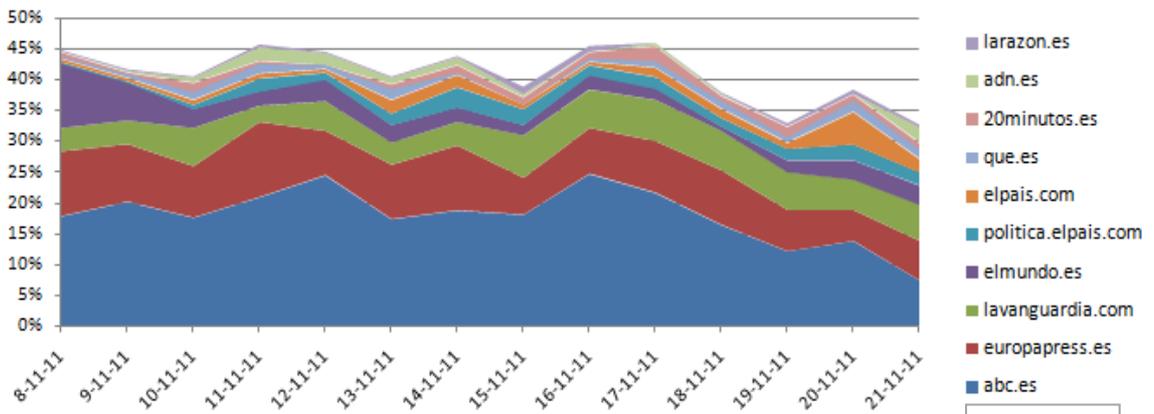


Figura 136.

1.A.Prensa Regional. El promedio de este conglomerado fue del **30,64%**. El día de menor visibilidad fue el 8-nov con un 27,3%, y el de máxima el 13 con un 33,97%.

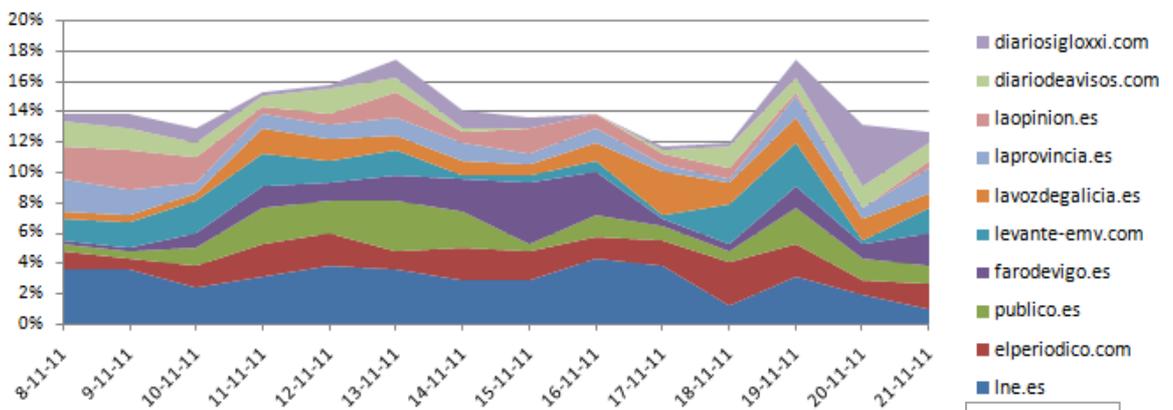


Figura 137.

1.A.Televisión. El promedio del conglomerado fue del **3,31%**. El día de menor visibilidad fue el 16-nov con un 1,67%, y el de máxima el día de las elecciones con un 4,44%.

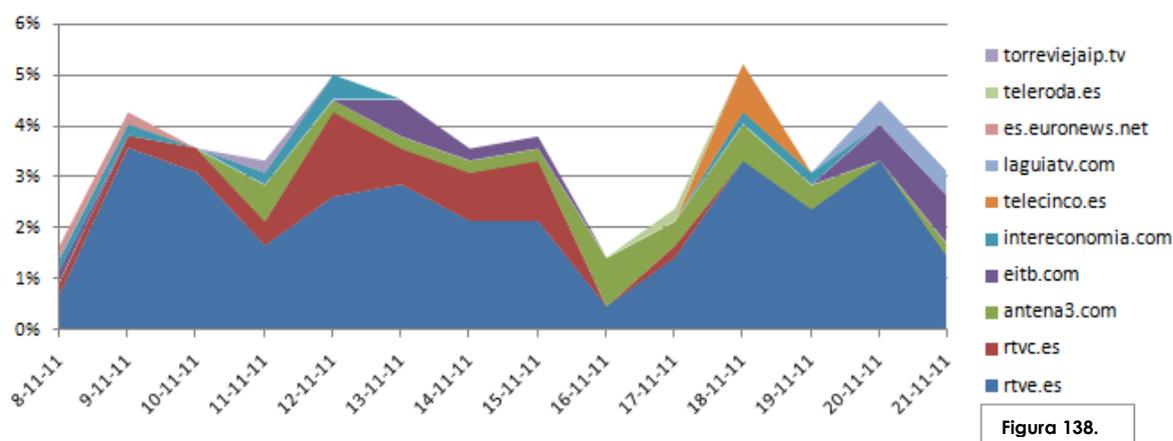


Figura 138.

1.A.Prensa Digital (Cibermedios). El promedio de este conglomerado fue del **5,97%**. El día de menor visibilidad fue el 13-nov con un 4,29%, y el de máxima el 21 con un 8,17%.

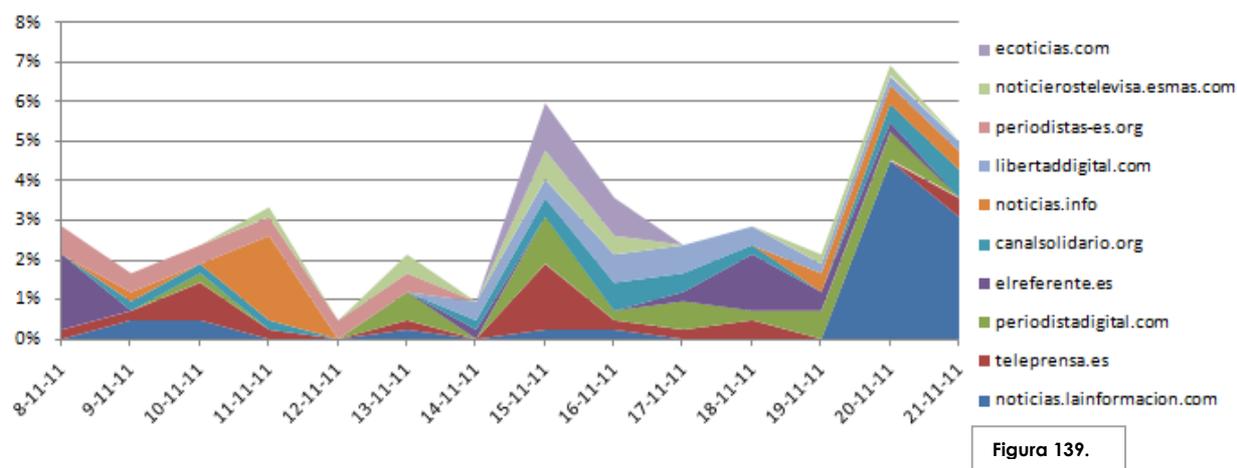


Figura 139.

>Elecciones Generales > Estrato Superior > Conglomerados secundarios

1.A.Prensa Económica. El promedio de este conglomerado fue del **3,42%**. El día de menor visibilidad fue el 9 de noviembre, con un 1,59%, y el de mayor el día 21 con un 5,87%.

1.A.Blog de Prensa. El promedio del conglomerado fue del **0,92%**. El día de mayor visibilidad fue el 17 de noviembre, con un 1,98%.

1.A.Radio. El promedio de este conglomerado fue del **0,55%**. El día de mayor visibilidad fue el día 20 de noviembre, con un 1,17%.

1.A.Prensa Internacional. El promedio del conglomerado fue del **8,25%**. El día de mayor visibilidad fue el día 10 de noviembre, con un 9,84%.

1.B.Internet. El promedio del conglomerado fue del **1,23%**. El día de mayor visibilidad fue el día 21 de noviembre, con un 2,3% del portal Terra.

>Elecciones Generales > Estrato Medio

Este estrato supuso un **4,91%** de los resultados. Provenientes del Estrato Medio aparecieron 156 portales entre los que destacaron <http://suite101.net>, <http://animanaturalis.org>, <http://educaweb.com>, <http://larepublica.es> y <http://kaosenlared.net>.

>Elecciones Generales > Estrato Micro

Este estrato supuso un **0,15%** de los resultados. En el Estrato Micro consiguieron visibilidad sitios protesta como <http://rebellion.org> y portales como <http://puomarketing.com>, <http://trecebits.com>, o <http://es.globalvoicesonline.org>.

>Elecciones Generales 2011>Conglomerados

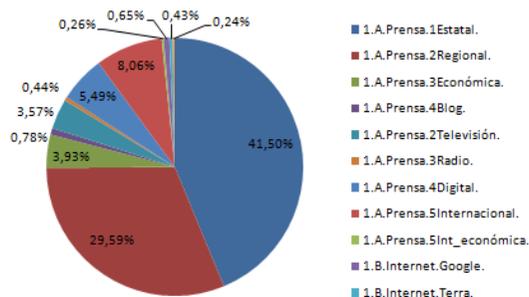


Figura 140.

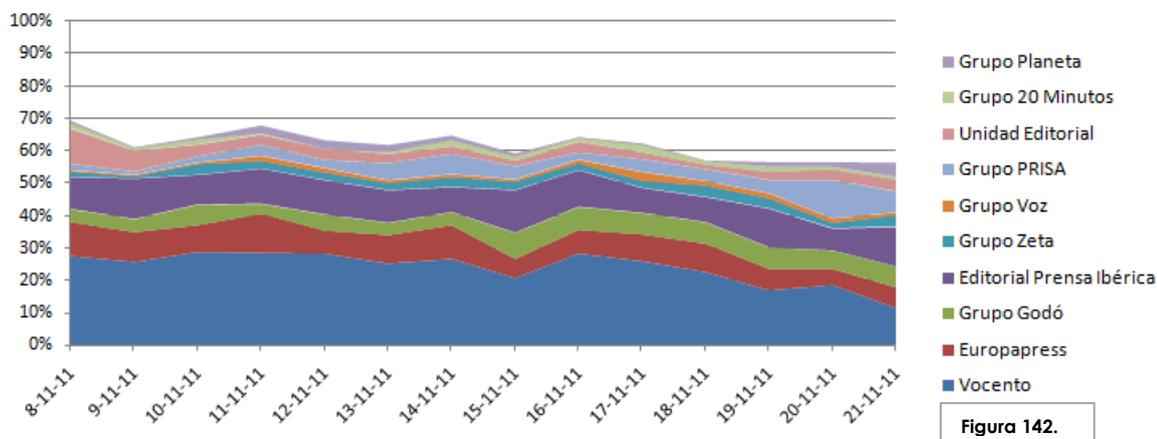
	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Red]	[Yellow]	[Green]

>Elecciones Generales 2011>Conglomerado>Tabla de Serie temporal

Figura 141.

Conglomerados	8-11-11	9-11-11	10-11-11	11-11-11	12-11-11	13-11-11	14-11-11	15-11-11	16-11-11	17-11-11	18-11-11	19-11-11	20-11-11	21-11-11	Promedio
1.A.Prensa.1Estatal.	45,48%	41,90%	40,71%	46,19%	45,24%	41,43%	44,29%	39,52%	45,48%	45,95%	38,33%	34,29%	39,05%	33,10%	41,50%
1.A.Prensa.2Regional.	27,86%	30,00%	30,95%	26,43%	29,29%	32,62%	29,52%	31,90%	28,81%	27,38%	27,14%	35,48%	26,43%	30,48%	29,59%
1.A.Prensa.3Económica.	4,29%	1,19%	3,33%	4,29%	2,86%	1,90%	5,24%	4,76%	5,00%	5,00%	3,33%	2,86%	3,57%	7,38%	3,93%
1.A.Prensa.4Blog.	0,95%	0,95%	1,19%	0,00%	0,71%	0,24%	2,38%	1,43%	0,24%	0,24%	0,48%	1,67%	0,00%	0,48%	0,78%
1.A.Prensa.2Televisión.	1,67%	4,29%	3,57%	3,33%	5,00%	4,52%	3,57%	3,81%	1,43%	2,38%	5,48%	3,10%	4,76%	3,10%	3,57%
1.A.Prensa.3Radio.	0,24%	0,48%	0,48%	0,24%	0,24%	0,00%	0,48%	0,00%	0,24%	0,71%	0,24%	0,24%	1,90%	0,71%	0,44%
1.A.Prensa.4Digital.	6,67%	5,48%	4,76%	6,67%	1,43%	3,81%	3,10%	7,86%	5,71%	4,05%	6,43%	3,10%	8,57%	9,29%	5,49%
1.A.Prensa.5Internacional.	6,19%	7,86%	10,71%	8,33%	9,05%	9,29%	5,24%	7,38%	6,67%	6,19%	10,24%	11,67%	7,38%	6,67%	8,06%
1.A.Prensa.5Int_económica.	0,00%	0,71%	0,00%	0,48%	0,48%	0,48%	0,00%	0,00%	0,48%	0,00%	0,71%	0,00%	0,00%	0,24%	0,26%
1.B.Internet.Google.	0,71%	0,48%	0,00%	0,00%	0,24%	0,24%	0,71%	0,24%	0,48%	0,95%	0,24%	1,19%	2,14%	1,43%	0,65%
1.B.Internet.Terra.	0,00%	0,00%	0,00%	0,00%	0,24%	0,71%	1,90%	0,00%	0,00%	0,71%	0,24%	0,48%	0,71%	0,95%	0,43%
1.B.Internet.Yahoo.	0,00%	0,00%	0,00%	0,24%	0,24%	0,48%	0,24%	0,24%	0,00%	0,00%	0,24%	0,71%	0,71%	0,24%	0,24%
Total	94,0%	93,3%	95,7%	96,2%	95,0%	95,7%	96,7%	97,1%	94,5%	93,6%	93,1%	94,8%	95,2%	94,0%	94,9%

>Elecciones Generales 2011>Grupos mediáticos>Serie temporal



- **Vocento.** El promedio de visibilidad de este grupo fue del **23,96%** sobre el total de resultados aparecidos en Top10 durante el periodo de captura de datos. El **75%** del total de resultados de este grupo fueron del diario **abc.es** seguido por el diario **que.es** con un **4,54%**.
- **Europapress.** Su promedio de visibilidad fue del **31,86%**, el mejor resultado de todas las elecciones estudiadas. El día de las elecciones bajó hasta el 13,26%.
- **Grupo Godó.** El promedio de visibilidad de este grupo fue del **2,87%**. El día de las elecciones su visibilidad alcanzó un 7,91%, mientras que el día posterior llegó al 11,86%.
- **Editorial Prensa Ibérica.** Su promedio de visibilidad fue del **3,11%**. El día de las elecciones bajó a un 1,63%, cayendo el día posterior hasta el 0,23%.
- **Grupo Zeta.** Su promedio de visibilidad fue del **0,56%**. El día de las elecciones triplicó su visibilidad con un 1,63%.
- **Grupo Voz.** El promedio de visibilidad de este grupo fue del **0,32%**. El día anterior a las elecciones alcanzó el 1,63%.
- **Grupo PRISA.** El promedio de visibilidad de este grupo fue del **4,8%**. El día de los comicios bajó hasta el 2,09% recuperándose el día posterior con un 7,44%.
- **Unidad Editorial.** Su promedio de visibilidad fue del **4,55%**. El día de las elecciones subió hasta el 3,11%, cayendo el día posterior hasta un 1,33%.
- **Grupo 20 Minutos.** El promedio de visibilidad de este grupo fue del **1,46%**. El día de los comicios subió hasta el 3,26%.
- **Grupo Planeta.** Su promedio de visibilidad fue del **1,26%**.

>Elecciones Generales 2011>Grupos mediáticos

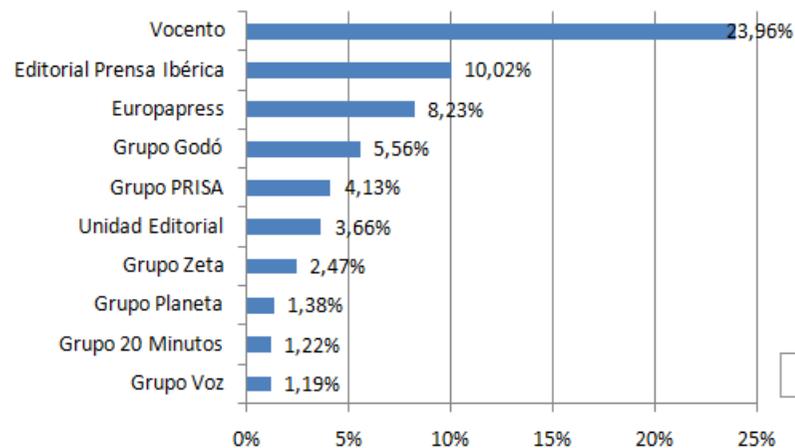


Figura 143.

	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Barra roja]	[Barra amarilla]	[Barra verde]

>Elecciones Generales 2011>Top10 Grupos mediáticos>Serie temporal

Figura 144.

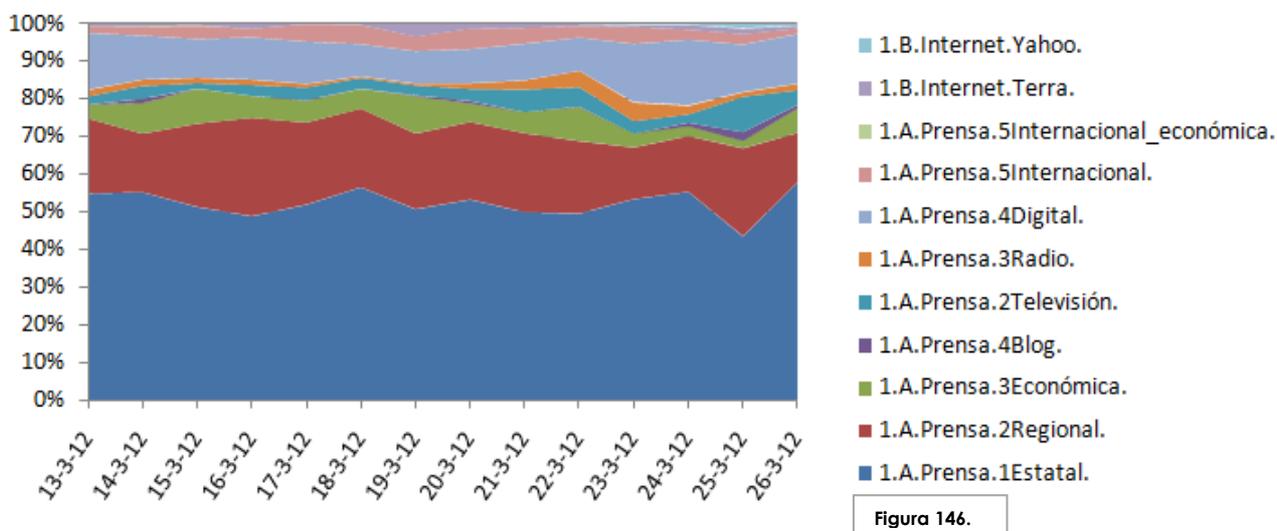
Grupos Mediáticos	8-11-11	9-11-11	10-11-11	11-11-11	12-11-11	13-11-11	14-11-11	15-11-11	16-11-11	17-11-11	18-11-11	19-11-11	20-11-11	21-11-11	
Vocento	27,62%	25,71%	28,81%	28,57%	28,33%	25,24%	26,67%	20,71%	28,33%	25,95%	22,62%	16,90%	18,57%	11,43%	23,96%
Europapress	10,48%	9,29%	8,33%	12,14%	7,14%	8,81%	10,48%	5,95%	7,38%	8,33%	8,81%	6,67%	5,00%	6,43%	8,23%
Grupo Godó	4,05%	4,05%	6,43%	3,10%	5,00%	3,81%	4,05%	8,10%	7,14%	6,67%	6,67%	6,67%	5,71%	6,43%	5,56%
Editorial Prensa Ibérica	9,76%	12,38%	9,05%	10,71%	10,48%	10,00%	7,62%	13,10%	11,19%	7,62%	7,62%	11,90%	6,67%	12,14%	10,02%
Grupo Zeta	1,67%	0,71%	3,33%	2,38%	2,38%	2,14%	3,10%	2,62%	2,14%	2,14%	3,57%	3,10%	1,67%	3,57%	2,47%
Grupo Voz	0,48%	0,48%	0,48%	1,67%	1,43%	0,95%	0,95%	0,71%	1,19%	2,86%	1,43%	1,67%	1,43%	0,95%	1,19%
Grupo PRISA	1,90%	0,95%	1,90%	3,33%	2,38%	5,24%	6,19%	4,05%	2,14%	3,81%	3,57%	4,05%	11,90%	6,43%	4,13%
Unidad Editorial	10,95%	6,43%	3,57%	3,10%	3,57%	2,86%	2,38%	1,67%	3,33%	2,14%	1,43%	2,62%	3,33%	3,81%	3,66%
Grupo 20 Minutos	1,67%	0,95%	1,67%	0,48%	0,00%	0,71%	1,90%	1,43%	1,19%	2,38%	1,19%	1,67%	0,95%	0,95%	1,22%
Grupo Planeta	0,71%	0,24%	0,71%	2,38%	2,62%	2,14%	1,43%	0,95%	0,24%	0,48%	0,24%	1,43%	1,43%	4,29%	1,38%
Total	69,3%	61,2%	64,3%	67,9%	63,3%	61,9%	64,8%	59,3%	64,3%	62,4%	57,1%	56,7%	56,7%	56,4%	61,8%

>Elecciones Generales 2011>Resto de Grupos mediáticos>Serie temporal

Figura 145.

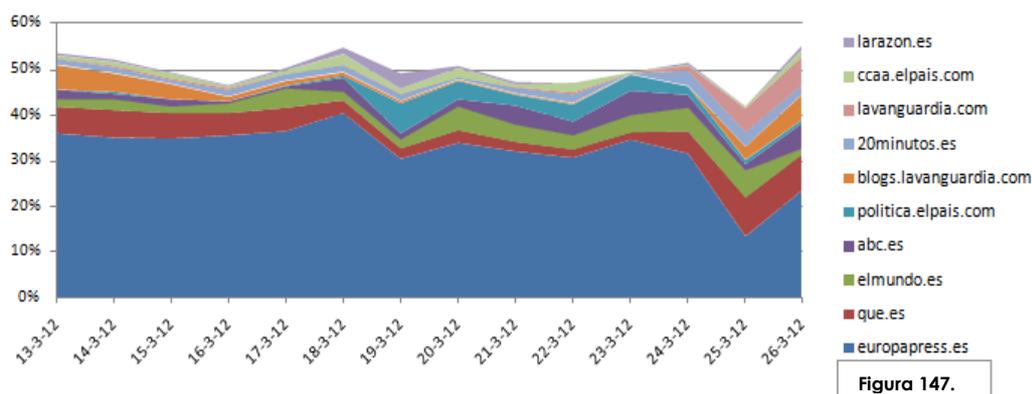
Grupos Mediáticos	8-11	9-11	10-11	11-11	12-11	13-11	14-11	15-11	16-11	17-11	18-11	19-11	20-11	21-11	
Grupo Joly	0,95%	2,38%	0,95%	0,48%	0,71%	0,95%	0,00%	0,95%	0,71%	0,24%	0,48%	1,19%	0,48%	0,24%	0,77%
Mediapro	0,48%	0,48%	1,19%	2,38%	2,14%	3,33%	2,38%	0,71%	1,43%	0,95%	1,19%	2,38%	1,43%	1,19%	1,55%
Editorial Iparraguirre	0,00%	0,00%	0,00%	0,48%	1,90%	0,95%	0,71%	1,19%	1,90%	1,19%	1,19%	0,48%	0,48%	0,48%	0,78%
Grupo Begar	0,00%	0,24%	0,48%	0,48%	0,24%	0,24%	0,24%	0,24%	0,24%	0,00%	0,95%	0,00%	0,00%	0,00%	0,24%
Libertad digital S.A	0,00%	0,00%	0,00%	0,24%	0,00%	0,48%	0,00%	0,71%	0,48%	0,00%	0,00%	0,24%	0,24%	0,00%	0,17%
Mediaset España	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,48%	0,48%	0,07%
Radio Popular S.A. - COPE	0,00%	0,24%	0,24%	0,24%	0,24%	0,00%	0,24%	0,00%	0,00%	0,24%	0,24%	0,00%	0,24%	0,00%	0,14%
Grupo Intereconomía	0,24%	0,24%	0,00%	0,24%	0,48%	0,00%	0,00%	0,00%	0,00%	0,00%	0,24%	0,24%	0,00%	0,00%	0,12%
Prensa internacional	6,19%	8,57%	10,71%	8,81%	9,52%	9,76%	5,24%	7,38%	7,14%	6,19%	10,24%	11,19%	7,14%	6,90%	8,21%
Grupo Correo Gallego	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,24%	0,00%	0,00%	0,00%	0,00%	0,24%	0,95%	0,00%	0,10%
EKHE (izquierda abertzal	0,00%	0,24%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,24%	0,00%	0,00%	0,00%	0,03%
Grupo Promecal	0,00%	0,71%	0,24%	0,48%	1,19%	1,67%	1,43%	1,43%	0,48%	1,90%	0,24%	0,48%	0,00%	0,00%	0,73%
Sakkia Hiferpress (inmobiliaria)	0,00%	0,24%	0,00%	0,00%	0,24%	0,24%	0,00%	0,95%	0,00%	0,00%	0,00%	0,00%	0,24%	0,00%	0,14%
Grupo publicaciones del sur	0,48%	0,24%	0,00%	0,00%	2,38%	1,19%	0,71%	0,00%	0,00%	0,71%	0,00%	0,00%	0,00%	1,19%	0,49%
Grupo La Región.	0,24%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,24%	0,00%	0,00%	0,00%	0,03%
Otros medios	16,43%	18,57%	17,62%	14,52%	12,62%	15,00%	20,71%	24,29%	17,86%	19,76%	20,95%	21,67%	26,90%	27,38%	19,59%
2.Estrato Medio	5,48%	6,19%	4,29%	3,81%	5,00%	4,29%	3,33%	2,86%	5,48%	5,71%	6,67%	5,00%	4,52%	5,71%	4,88%
3.Estrato Micro	0,24%	0,48%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,71%	0,00%	0,24%	0,24%	0,00%	0,14%
Total	30,7%	38,8%	35,7%	32,1%	36,7%	38,1%	35,2%	40,7%	35,7%	37,6%	42,9%	43,3%	43,3%	43,6%	38,2%

3.3.4. Junta de Andalucía – 25/3/2012

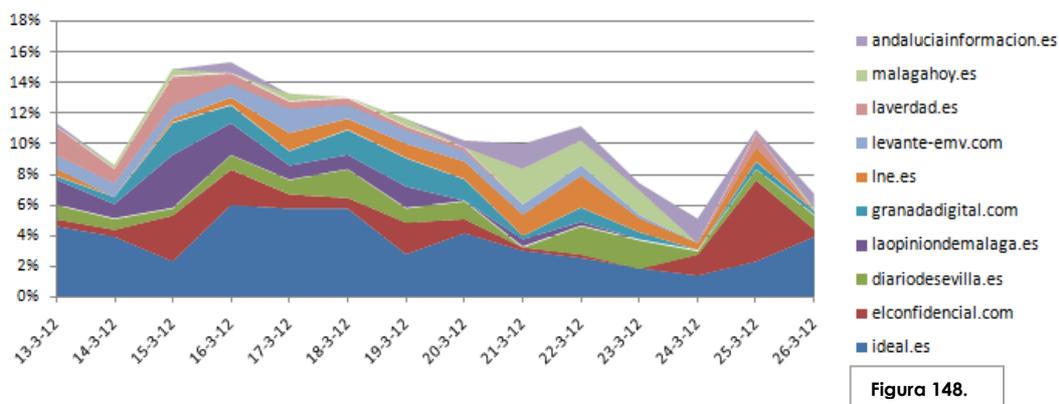


>Elecciones Andalucía 2012> Estrato Superior > Conglomerados principales

1.A.Prensa Estatal. El promedio del conglomerado fue del **43,75%**. El día de menor visibilidad fue el 25 de marzo con un 39,22%, el día de las elecciones, y el de máxima el 14 con un 46,67%.



1.A.Prensa Regional. El promedio de este conglomerado fue del **23,82%**. El día de menor visibilidad fue el 24 de marzo con un 18,6%, y el de máxima el 15 y 19 con un 26,59%.



1.A.Televisión. El promedio del conglomerado fue del **3,52%**. El día de menor visibilidad fue el 15 de marzo con un 2,17%, y el de máxima el día de las elecciones con un 6,28%.

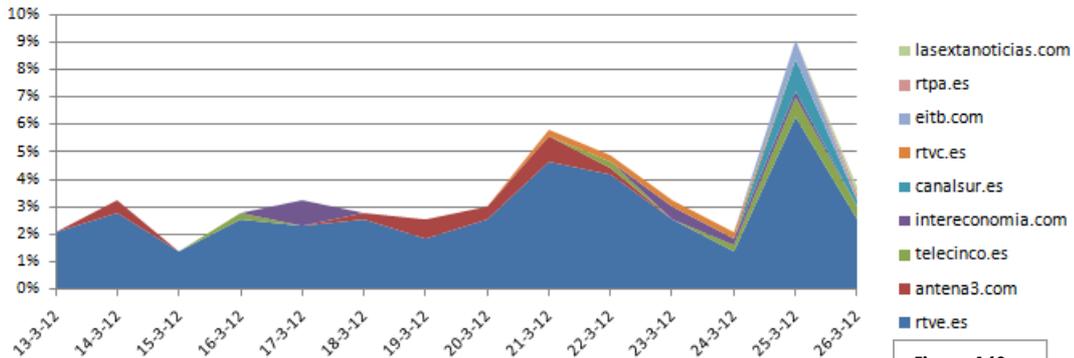


Figura 149.

1.A.Prensa Digital (Cibermedios). El promedio fue del **13,57%**. El día de menor visibilidad fue el 19 de marzo con un 8,76%, y el de máxima el 24 con un 17,6%.

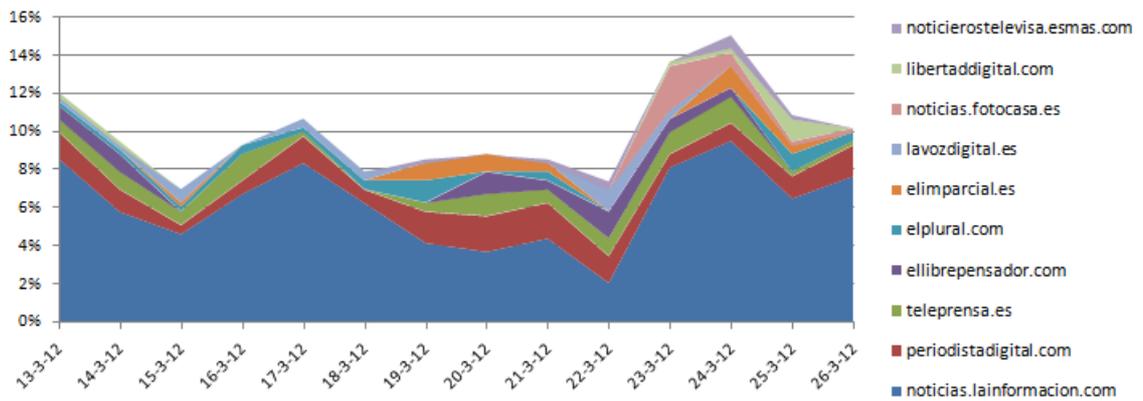


Figura 150.

>Elecciones Andalucía 2012> Estrato Superior > Conglomerados secundarios

1.A.Prensa Económica. El promedio de este conglomerado fue del **5,43%**. El día de menor visibilidad el 25 de marzo, con un 2,71%, y el de mayor el día 19 con un 8,37%.

1.A.Blog de Prensa. El promedio del conglomerado fue del **0,38%**. El día de mayor visibilidad fue el día 26 de marzo, con un 1,09%.

1.A.Radio. El promedio de este conglomerado fue del **1,41%**. El día de mayor visibilidad fue el 23 de marzo, con un 3,33%.

1.A.Prensa Internacional. El promedio del conglomerado fue del **3,95%**. El día de mayor visibilidad fue el día 17 de marzo, con un 5,74%.

1.B.Internet. El promedio del conglomerado fue del **1,07%**. El día de mayor visibilidad fue el día 21 de marzo, con un 1,63% del portal Terra.

>Elecciones Andalucía 2012> Estrato Medio

Este estrato supuso un **2,31%** de los resultados. Provenientes del Estrato Medio aparecieron 112 portales como <http://cronicasdelaemigracion.com> o <http://teinteresa.es> o la revista <http://valenciaplaza.com>.

>Elecciones Andalucía 2012> Estrato Micro

Este estrato tuvo una visibilidad residual y supuso un **0,08%** de los resultados. En el Estrato Micro ganaron visibilidad 7 sitios, entre los que destacaron <http://ecologistasenaccion.org> y el 'Protesta' <http://rebellion.org>.

>Elecciones Andalucía 2012>Conglomerados

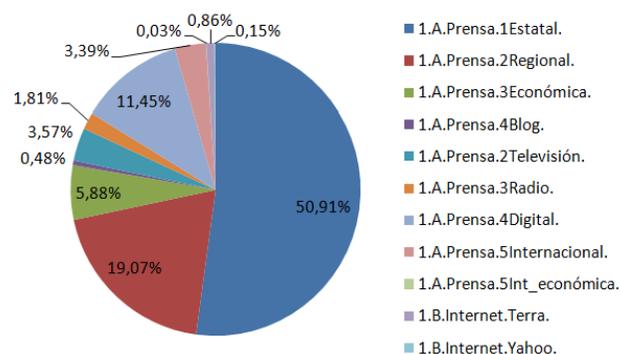


Figura 151.

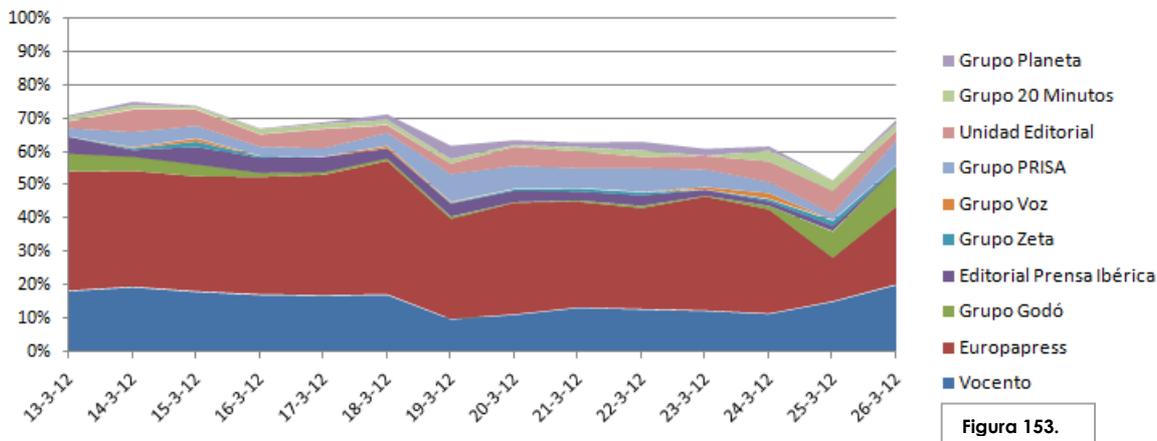
	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Barra roja]	[Barra amarilla]	[Barra verde]

Figura 152.

>Elecciones Andalucía 2012>Conglomerado>Tabla de Serie temporal

Conglomerados	13-3-12	14-3-12	15-3-12	16-3-12	17-3-12	18-3-12	19-3-12	20-3-12	21-3-12	22-3-12	23-3-12	24-3-12	25-3-12	26-3-12	Promedio
1.A.Prensa.1Estatal.	53,49%	54,19%	50,47%	47,91%	51,16%	56,28%	50,23%	51,40%	48,37%	47,67%	51,40%	51,86%	42,09%	56,28%	50,91%
1.A.Prensa.2Regional.	19,53%	15,35%	21,86%	25,81%	21,63%	20,93%	20,00%	20,00%	20,47%	18,60%	13,26%	13,95%	22,79%	12,79%	19,07%
1.A.Prensa.3Económica.	3,72%	8,14%	9,30%	5,81%	5,81%	5,35%	10,00%	4,88%	5,58%	8,84%	3,72%	2,56%	2,09%	6,51%	5,88%
1.A.Prensa.4Blog.	0,00%	1,16%	0,00%	0,00%	0,23%	0,00%	0,23%	0,70%	0,00%	0,23%	0,00%	0,93%	2,33%	0,93%	0,48%
1.A.Prensa.2Televisión.	2,09%	3,26%	1,40%	2,79%	3,26%	2,79%	2,56%	3,02%	5,81%	4,88%	3,26%	2,09%	9,07%	3,72%	3,57%
1.A.Prensa.3Radio.	1,63%	1,63%	1,40%	1,40%	0,93%	0,47%	0,47%	1,40%	2,33%	4,19%	4,65%	2,09%	1,16%	1,63%	1,81%
1.A.Prensa.4Digital.	14,88%	11,63%	10,23%	11,16%	11,16%	8,60%	8,60%	8,84%	9,53%	8,60%	15,12%	16,51%	12,33%	13,02%	11,45%
1.A.Prensa.5Internacional.	1,86%	2,33%	3,49%	2,33%	4,65%	5,12%	3,95%	5,35%	4,19%	3,02%	4,42%	2,56%	2,79%	1,40%	3,39%
1.A.Prensa.5Int_económica.	0,00%	0,23%	0,23%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,03%
1.B.Internet.Terra.	0,47%	0,47%	0,23%	1,16%	0,00%	0,23%	3,26%	1,16%	0,93%	0,47%	0,47%	1,16%	1,40%	0,70%	0,86%
1.B.Internet.Yahoo.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,23%	0,23%	1,16%	0,47%	0,15%
Total	97,7%	98,4%	98,6%	98,4%	98,8%	99,8%	99,3%	96,7%	97,2%	96,5%	96,5%	94,0%	97,2%	97,4%	97,6%

>Elecciones Junta de Andalucía 2012>Grupos mediáticos>Serie temporal



- **Vocento.** El promedio de visibilidad de este grupo fue del **15,02%** sobre el total de resultados aparecidos en el Top10 durante el periodo de captura de datos. El día posterior a las elecciones alcanzó un **20%** de visibilidad. El **29,65%** del total de resultados de este grupo fueron del diario **que.es**, seguido por el **ideal.es** con un **24,12%** y por el **abc.es** con un **16,26%**
- **Europapress.** Su promedio de visibilidad fue del **31,86%**, el mejor resultado de todas las elecciones estudiadas. El día de las elecciones bajó hasta el 13,26%.
- **Grupo Godó.** El promedio de visibilidad de este grupo fue del **2,87%**. El día de las elecciones su visibilidad alcanzó un 7,91%, mientras que el día posterior llegó al 11,86%.
- **Editorial Prensa Ibérica.** Su promedio de visibilidad fue del **3,11%**. El día de las elecciones bajó a un 1,63%, cayendo el día posterior hasta el 0,23%.
- **Grupo Zeta.** Su promedio de visibilidad fue del **0,56%**. El día de las elecciones triplicó su visibilidad con un 1,63%.
- **Grupo Voz.** El promedio de visibilidad de este grupo fue del **0,32%**. El día anterior a las elecciones alcanzó el 1,63%.
- **Grupo PRISA.** El promedio de visibilidad de este grupo fue del **4,8%**. El día de los comicios bajó hasta el 2,09% recuperándose el día posterior con un 7,44%.
- **Unidad Editorial.** Su promedio de visibilidad fue del **4,55%**. El día de las elecciones subió hasta el 3,11%, cayendo el día posterior hasta un 1,33%.
- **Grupo 20 Minutos.** El promedio de visibilidad de este grupo fue del **1,46%**. El día de los comicios subió hasta el 3,26%.
- **Grupo Planeta.** Su promedio de visibilidad fue del **1,26%**.

>Elecciones Andalucía 2012>Grupos mediáticos

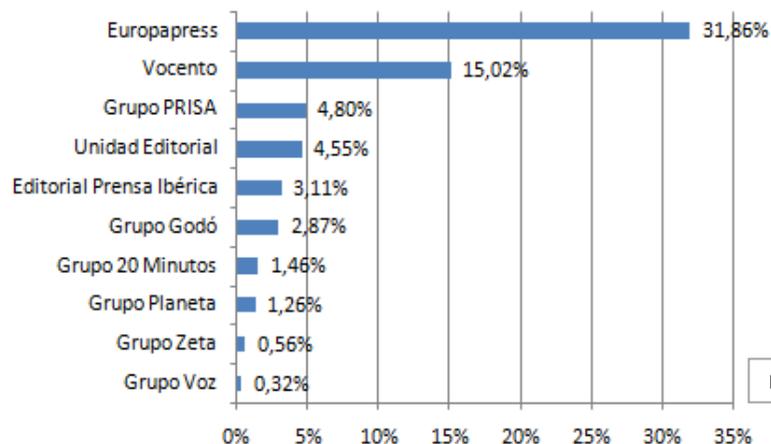


Figura 154.

	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Barra roja]	[Barra amarilla]	[Barra verde]

>Elecciones Andalucía 2012>Top 10 Grupos mediáticos>Tabla de Serie temporal

Figura 155.

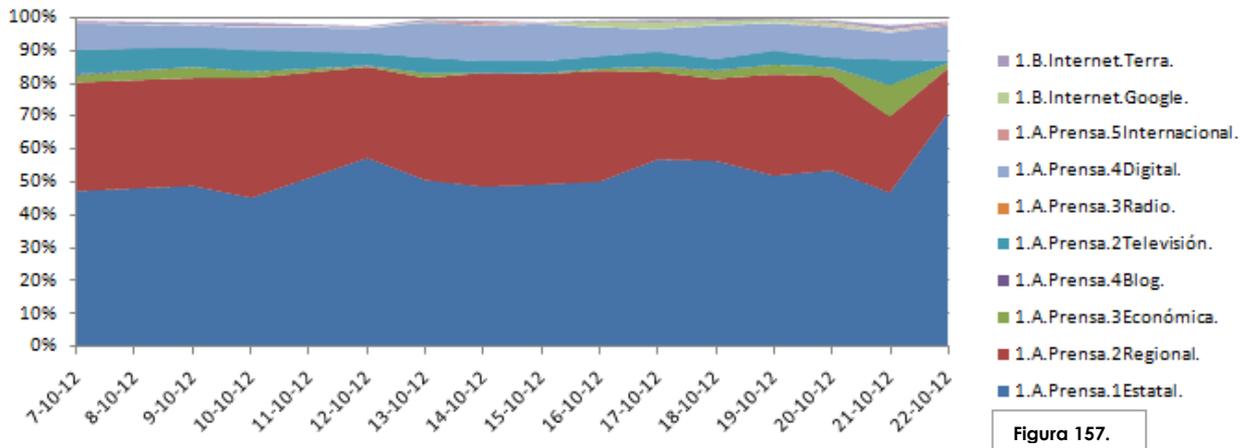
Grupos Mediáticos	13-3-12	14-3-12	15-3-12	16-3-12	17-3-12	18-3-12	19-3-12	20-3-12	21-3-12	22-3-12	23-3-12	24-3-12	25-3-12	26-3-12	Promedio
Vocento	18,14%	19,30%	17,91%	16,98%	16,74%	16,98%	9,53%	10,93%	13,02%	12,56%	12,09%	11,16%	14,88%	20,00%	15,02%
Europapress	35,81%	34,88%	34,65%	35,35%	36,28%	40,23%	30,23%	33,72%	31,86%	30,47%	34,42%	31,40%	13,26%	23,49%	31,86%
Grupo Godó	5,58%	4,42%	3,72%	1,40%	0,93%	0,93%	0,70%	0,23%	0,47%	0,70%	0,23%	1,16%	7,91%	11,86%	2,87%
Editorial Prensa Ibérica	5,12%	2,09%	5,35%	4,65%	4,65%	3,02%	3,95%	3,49%	2,79%	3,26%	1,86%	1,40%	1,63%	0,23%	3,11%
Grupo Zeta	0,00%	0,47%	1,63%	0,47%	0,00%	0,00%	0,23%	0,47%	0,93%	0,93%	0,00%	0,70%	1,63%	0,47%	0,56%
Grupo Voz	0,00%	0,23%	0,93%	0,00%	0,00%	0,70%	0,23%	0,00%	0,00%	0,00%	0,70%	1,63%	0,00%	0,00%	0,32%
Grupo PRISA	2,56%	4,65%	3,72%	2,79%	2,56%	3,95%	8,37%	6,98%	6,05%	7,21%	5,35%	3,49%	2,09%	7,44%	4,80%
Unidad Editorial	2,09%	6,74%	4,88%	3,72%	5,81%	2,33%	3,26%	5,81%	5,12%	3,49%	4,19%	6,28%	6,98%	3,02%	4,55%
Grupo 20 Minutos	1,16%	1,16%	0,70%	1,40%	1,40%	1,40%	1,40%	0,47%	1,16%	1,86%	0,00%	3,26%	3,26%	1,86%	1,46%
Grupo Planeta	0,47%	0,93%	0,23%	0,23%	0,47%	1,63%	3,95%	1,40%	1,40%	2,56%	2,09%	1,16%	0,00%	1,16%	1,26%
Total	70,9%	74,9%	73,7%	67,0%	68,8%	71,2%	61,9%	63,5%	62,8%	63,0%	60,9%	61,6%	51,6%	69,5%	65,8%

>Elecciones Generalitat 2010>Resto de Grupos mediáticos>Serie temporal

Figura 156.

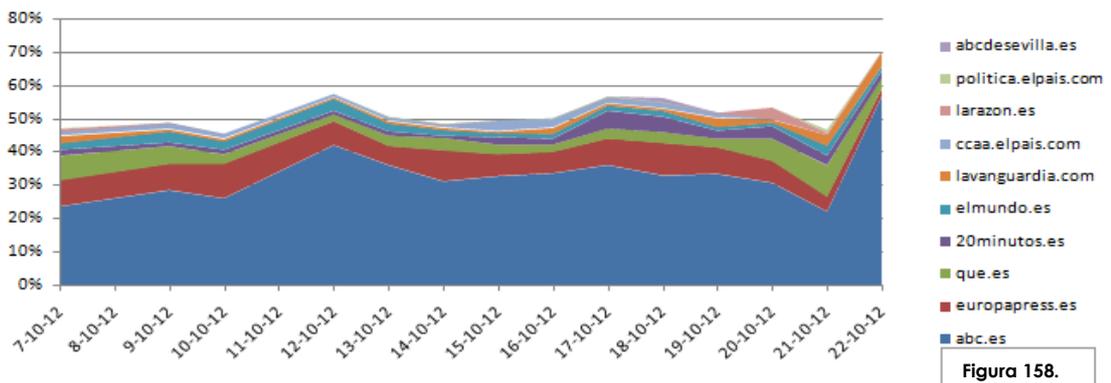
Grupos Mediáticos	13-3-12	14-3-12	15-3-12	16-3-12	17-3-12	18-3-12	19-3-12	20-3-12	21-3-12	22-3-12	23-3-12	24-3-12	25-3-12	26-3-12	Promedio
Grupo Joly	1,16%	2,33%	1,40%	2,56%	3,95%	3,26%	3,02%	3,02%	3,26%	3,72%	3,95%	1,16%	1,63%	1,63%	2,57%
Mediapro	0,23%	0,00%	0,00%	0,00%	0,47%	1,16%	0,23%	0,23%	0,23%	0,70%	0,47%	0,00%	1,86%	0,93%	0,47%
Editorial Iparraguirre Corporate Communicator S.L. - Elplural.com	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	1,86%	0,47%	0,47%	0,47%	0,00%	0,00%	1,16%	0,93%	0,40%
Grupo Begar	0,23%	0,23%	0,23%	0,47%	0,23%	0,47%	1,16%	0,00%	0,47%	0,00%	0,00%	0,00%	0,93%	0,47%	0,35%
Libertad digital S.A	1,63%	0,00%	0,23%	0,00%	0,70%	0,23%	0,93%	0,23%	0,70%	0,47%	0,47%	0,70%	0,23%	0,00%	0,47%
Mediaset España	0,23%	0,23%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,23%	0,23%	1,16%	0,00%	0,15%
Radio Popular S.A. - COPE	0,00%	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	0,00%	0,00%	0,23%	0,00%	0,23%	0,70%	0,47%	0,13%
Grupo Intereconomía	0,23%	0,23%	0,00%	0,00%	0,00%	0,47%	0,47%	0,47%	0,23%	0,47%	1,40%	0,70%	0,47%	0,47%	0,40%
Prensa internacional	0,00%	0,00%	0,00%	0,00%	0,93%	0,00%	0,00%	0,00%	0,00%	0,00%	0,47%	0,23%	0,23%	0,00%	0,13%
Grupo Correo Gallego	1,86%	2,56%	3,72%	2,33%	4,65%	5,12%	3,95%	5,35%	4,19%	3,02%	4,42%	2,56%	2,79%	1,40%	3,42%
EKHE (izquierda abertzal)	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,08%
Grupo publicaciones del sur	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,23%	0,00%	0,02%
Grupo La Región.	0,23%	0,00%	0,00%	0,70%	0,00%	0,00%	0,00%	0,47%	1,63%	0,93%	0,47%	1,63%	0,23%	0,93%	0,51%
Otros medios	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	0,00%	0,02%
2.Estrato Medio	20,93%	17,91%	19,07%	25,12%	19,07%	17,91%	25,81%	22,09%	23,02%	23,26%	23,72%	24,88%	33,95%	20,70%	22,67%
3.Estrato Micro	2,33%	1,40%	1,40%	1,63%	0,93%	0,00%	0,47%	3,26%	2,79%	3,49%	3,26%	6,05%	2,79%	2,56%	2,31%
	0,00%	0,23%	0,00%	0,00%	0,23%	0,23%	0,23%	0,00%	0,00%	0,00%	0,23%	0,00%	0,00%	0,00%	0,08%
	29,1%	25,1%	26,3%	33,0%	31,2%	28,8%	38,1%	36,5%	37,2%	37,0%	39,1%	38,4%	48,4%	30,5%	34,2%

3.3.5. Xunta de Galicia – 21/10/2012

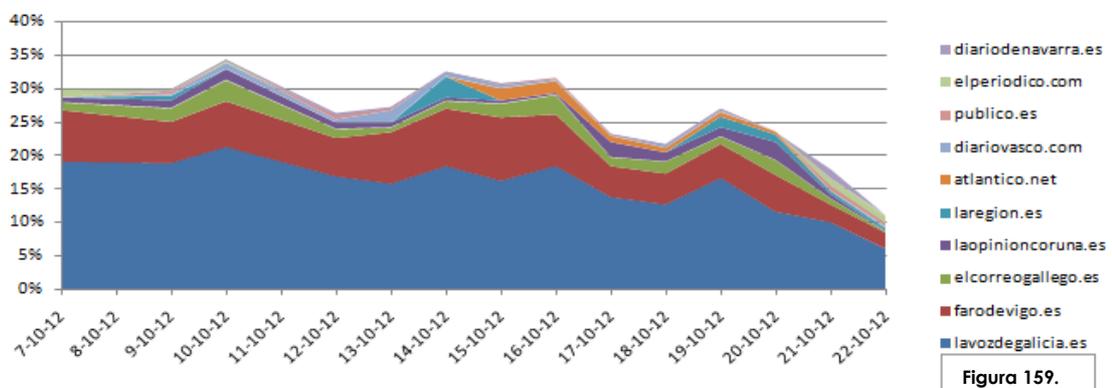


>Elecciones Galicia 2012> Estrato Superior > Conglomerados principales

1.A.Prensa Estatal. El promedio del conglomerado fue del **43,81%**. El día de menor visibilidad fue el 7 de octubre con un 36,3%, el día de las elecciones, y el de máxima el 22 con un 69,63%.



1.A.Prensa Regional. El promedio de este conglomerado fue del **36,06%**. El día de menor visibilidad fue el 22 de octubre con un 16,44%, y el de máxima el 7 con un 42,69%.



1.A.Televisión. El promedio del conglomerado fue del **4,16%**. El día de menor visibilidad fue el 15 de octubre con un 2,17%, y el de máxima el día de las elecciones con un 6,28%.

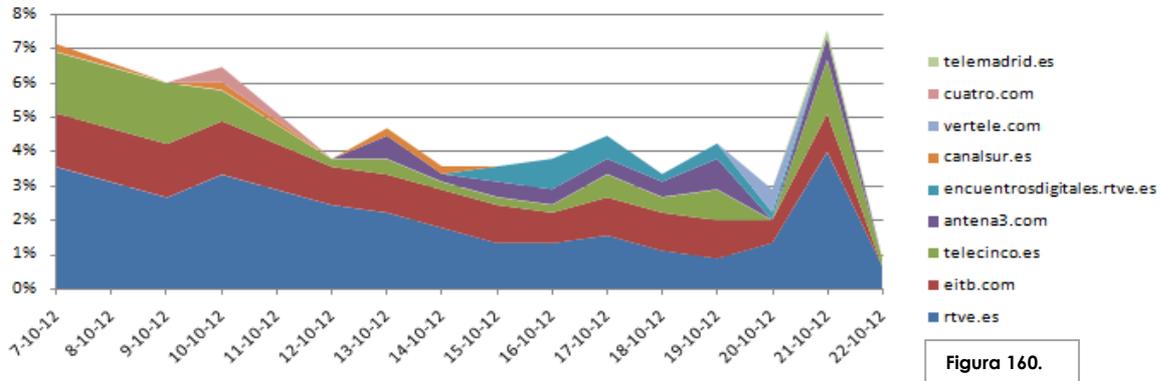


Figura 160.

1.A.Prensa Digital (Cibermedios). El promedio fue del **10,14%**. El día de menor visibilidad fue el 19 de octubre con un 8,76%, y el de máxima el 24 con un 17,6%.

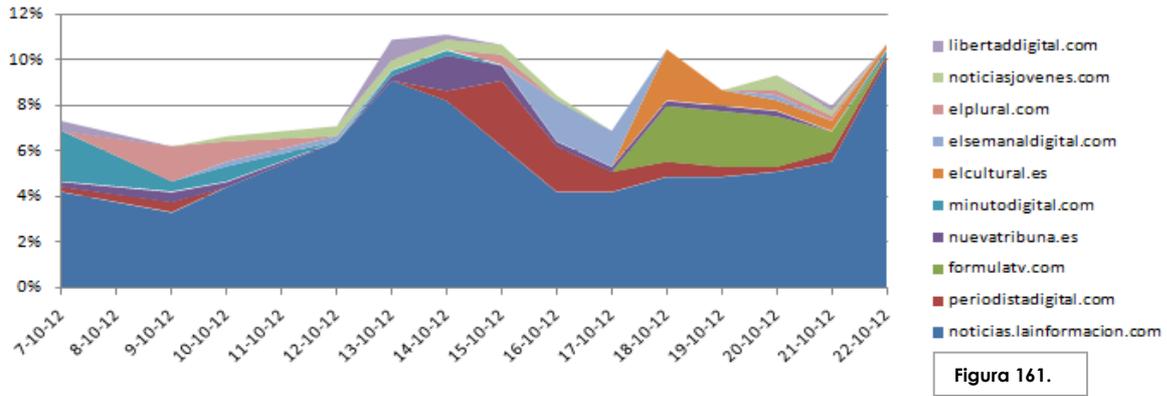


Figura 161.

>Elecciones Galicia 2012> Estrato Superior > Conglomerados secundarios

1.A.Prensa Económica. El promedio de este conglomerado fue del **2,31%**. El día de mayor visibilidad fue el día 21 con un 5,2%.

1.A.Blog de Prensa. El promedio del conglomerado fue del **0,23%**. El día de mayor visibilidad fue el 7 de octubre, con un 0,52%.

1.A.Radio. El promedio de este conglomerado fue del **0,02%**. El día de mayor visibilidad fue el 9 de octubre, con un 0,15%.

1.A.Prensa Internacional. El promedio del conglomerado fue del **0,28%**. Los días de mayor visibilidad fueron los días 14 y 21 de octubre, con un 0,74%.

1.B.Internet. El promedio del conglomerado fue del **1,1%**. El día de mayor visibilidad fue el día 19 de octubre, con un 1,79%.

>Elecciones Galicia 2012> Estrato Medio

Este estrato supuso un **0,84%** de los resultados. Del Estrato Medio surgieron portales como <http://republica.com>, <http://teinteresa.es> o <http://bi-spain.com>

>Elecciones Galicia 2012> Estrato Micro

Este estrato supuso un **0,24%** de los resultados. En el Estrato Micro solo aparecieron en el Top10 la página de Wikipedia de Alberto Núñez Feijó e <http://internautas.org>

>Elecciones Galicia 2012>Conglomerados

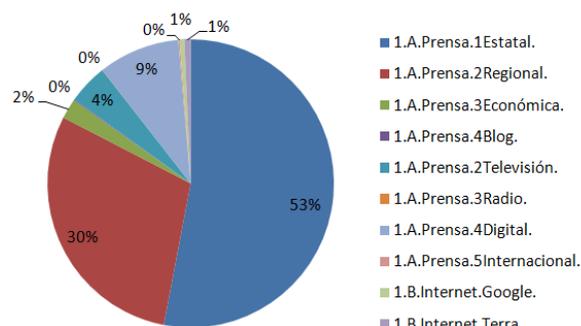


Figura 162.

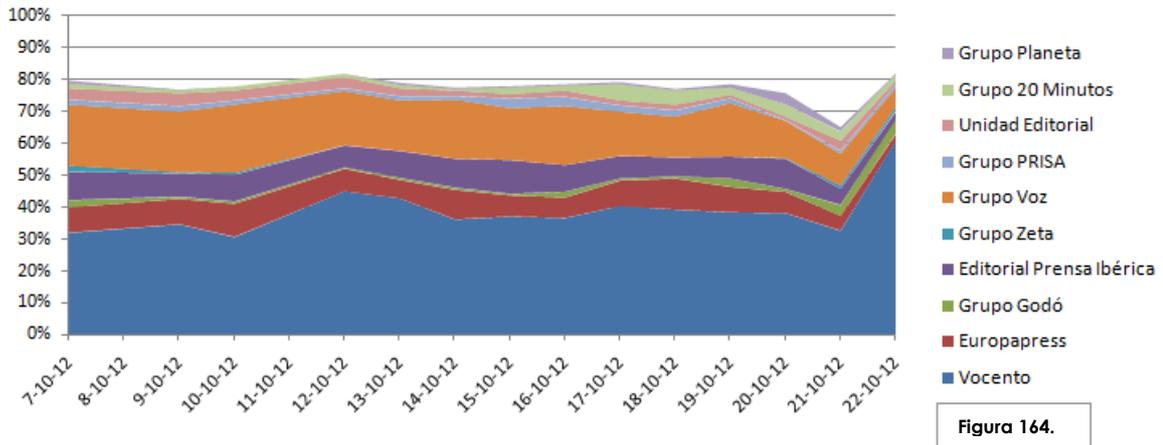
	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Red]	[Yellow]	[Green]

>Elecciones Galicia 2012>Conglomerado>Tabla de Serie temporal

Figura 163.

Conglomerados	7-10-12	9-10-12	10-10-12	12-10-12	13-10-12	14-10-12	15-10-12	16-10-12	17-10-12	18-10-12	19-10-12	20-10-12	21-10-12	22-10-12	Promedio
1.A.Prensa.1Estatal.	47,33%	48,89%	45,33%	57,33%	50,67%	48,67%	49,33%	50,22%	56,89%	56,44%	52,00%	53,56%	46,89%	70,89%	52,46%
1.A.Prensa.2Regional.	32,89%	32,67%	36,22%	27,56%	30,89%	34,44%	33,33%	33,56%	26,44%	24,89%	30,67%	28,44%	23,11%	13,56%	29,19%
1.A.Prensa.3Económica.	2,22%	3,33%	2,00%	0,22%	1,56%	0,00%	0,22%	0,67%	1,56%	2,44%	2,89%	2,67%	9,56%	1,56%	2,21%
1.A.Prensa.4Blog.	0,44%	0,00%	0,00%	0,22%	0,00%	0,00%	0,22%	0,00%	0,22%	0,22%	0,00%	0,22%	0,00%	0,00%	0,11%
1.A.Prensa.2Televisión.	7,11%	6,00%	6,44%	3,78%	4,67%	3,56%	3,56%	3,78%	4,44%	3,33%	4,22%	2,89%	7,56%	0,89%	4,44%
1.A.Prensa.3Radio.	0,22%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,02%
1.A.Prensa.4Digital.	8,00%	6,89%	7,11%	7,78%	10,89%	11,11%	11,56%	8,89%	7,11%	10,44%	8,67%	9,56%	8,44%	10,89%	9,10%
1.A.Prensa.5Internacional.	0,00%	0,00%	0,00%	0,00%	0,00%	0,89%	0,00%	0,00%	0,00%	0,00%	0,00%	0,44%	0,22%	0,67%	0,16%
1.B.Internet.Google.	0,00%	0,00%	0,00%	0,00%	0,22%	0,00%	0,00%	1,78%	2,00%	1,33%	1,11%	0,89%	0,22%	0,00%	0,54%
1.B.Internet.Terra.	0,89%	0,67%	1,33%	0,44%	0,67%	0,44%	0,44%	0,44%	0,89%	0,67%	0,22%	0,67%	1,56%	0,44%	0,70%
Total	99,1%	98,4%	98,4%	97,3%	99,6%	99,1%	98,7%	99,3%	99,6%	99,8%	99,8%	99,3%	97,6%	98,9%	98,9%

>Elecciones Xunta de Galicia 2012>Grupos mediáticos>Serie temporal



- **Vocento.** El promedio de visibilidad de este grupo fue del **38,73%** sobre el total de resultados aparecidos en Top10 durante el periodo de captura de datos. Este resultado es el mejor de Vocento en las seis elecciones que componen el estudio. El día posterior a las elecciones alcanzó un **60,67%** de visibilidad, el máximo registrado en todo el estudio. El 85% del total de resultados de este grupo fueron del diario abc.es.
- **Europapress.** Su promedio de visibilidad fue del **7,22%**, estable durante todo el periodo de estudio pero cayendo el día de las elecciones al 4,67% y al 2,22% el día posterior.
- **Grupo Godó.** El promedio de visibilidad de este grupo fue del **1,46%**. A diferencia del anterior, el día de las elecciones dobló su visibilidad y alcanzó un 3,33%, mientras que el día posterior llegó al 4,67%.
- **Editorial Prensa Ibérica.** Su promedio de visibilidad fue del **7,52%**. El día de las elecciones bajó a un 5,11%, cayendo el día posterior hasta el 2,44%.
- **Grupo Zeta.** Su promedio de visibilidad fue del **0,40%**. Sus mejores días fueron el día de las elecciones y el posterior, cuando logró un 1,1% de visibilidad.
- **Grupo Voz.** El promedio de visibilidad de este grupo gallego fue del **15,41%**. Los días de menor visibilidad fueron el día de los comicios con un 10% y el día posterior con un 6%.
- **Grupo PRISA.** El promedio de visibilidad de este grupo fue del **1,56%**. El día de los comicios bajó hasta el 0,89% recuperándose ligeramente el día posterior con un 1,1%.
- **Unidad Editorial.** Su promedio de visibilidad fue del **2,27%**. El día de las elecciones subió hasta el 3,11%, cayendo el día posterior hasta un 1,33%.
- **Grupo 20 Minutos.** El promedio de visibilidad de este grupo fue del **2,35%**. El día de los comicios subió hasta el 2,89%.
- **Grupo Planeta.** Su promedio de visibilidad fue del **0,75%**. Su mejor día fue el anterior a las elecciones con un 3,56%, bajando al 0,22% el día posterior a los comicios.

>Elecciones Xunta de Galicia 2012>Grupos mediáticos

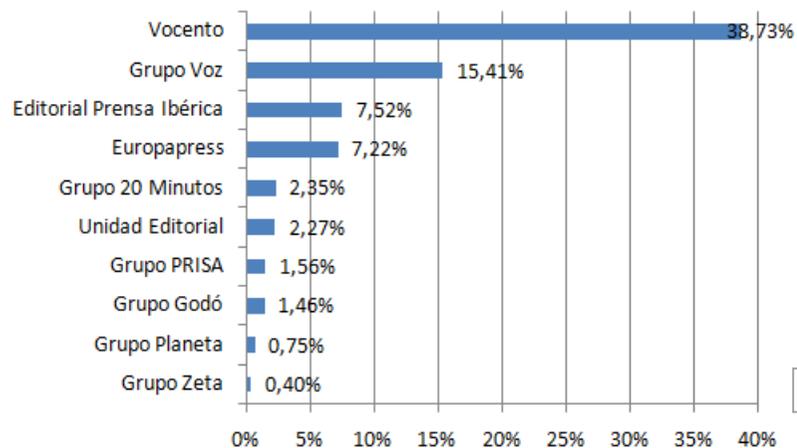


Figura 165.

	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Barra roja]	[Barra amarilla]	[Barra verde]

>Elecciones Xunta de Galicia 2012> Top10Grupos mediáticos>Serie temporal

Figura 166.

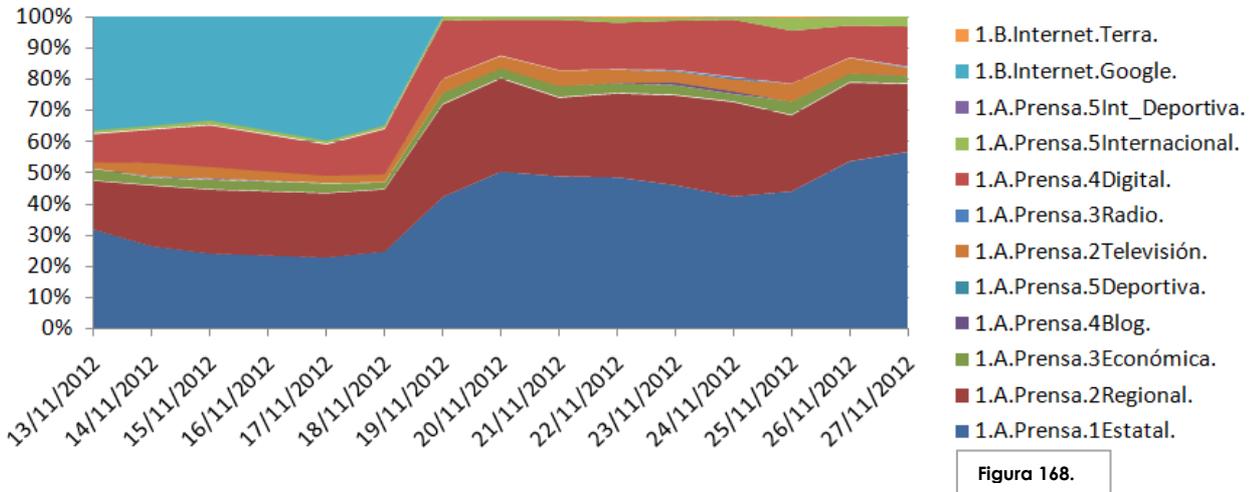
Grupos Mediáticos	7-10-12	9-10-12	10-10-12	12-10-12	13-10-12	14-10-12	15-10-12	16-10-12	17-10-12	18-10-12	19-10-12	20-10-12	21-10-12	22-10-12	Promedio
Vocento	31,78%	34,44%	30,44%	44,89%	42,67%	36,00%	36,89%	36,44%	40,22%	39,11%	38,22%	38,00%	32,44%	60,67%	38,73%
Europapress	8,00%	8,00%	10,44%	7,11%	5,78%	9,33%	6,67%	6,44%	8,00%	9,78%	8,00%	6,67%	4,67%	2,22%	7,22%
Grupo Godó	2,22%	0,67%	0,67%	0,44%	0,67%	0,67%	0,44%	1,78%	0,67%	0,67%	2,67%	0,89%	3,33%	4,67%	1,46%
Editorial Prensa Ibérica	8,89%	7,33%	8,44%	6,89%	8,44%	9,11%	10,67%	8,44%	7,11%	6,00%	6,89%	9,56%	5,11%	2,44%	7,52%
Grupo Zeta	2,00%	0,44%	0,67%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,22%	1,11%	1,11%	0,40%
Grupo Voz	19,11%	18,89%	21,33%	16,89%	15,78%	18,44%	16,22%	18,44%	13,78%	12,67%	16,67%	11,56%	10,00%	6,00%	15,41%
Grupo PRISA	1,56%	1,78%	1,33%	0,89%	1,33%	1,11%	3,11%	2,89%	2,00%	2,00%	1,56%	0,22%	0,89%	1,11%	1,56%
Unidad Editorial	3,56%	4,00%	3,11%	3,56%	2,44%	1,78%	1,33%	2,00%	1,56%	1,78%	1,11%	1,11%	3,11%	1,33%	2,27%
Grupo 20 Minutos	1,78%	1,11%	1,33%	1,11%	1,33%	0,67%	2,22%	1,78%	5,33%	4,67%	2,44%	4,00%	2,89%	2,22%	2,35%
Grupo Planeta	0,89%	0,22%	0,00%	0,00%	0,67%	0,44%	0,44%	0,44%	0,67%	0,44%	1,11%	3,56%	1,33%	0,22%	0,75%
Total	79,8%	76,9%	77,8%	81,8%	79,1%	77,6%	78,0%	78,7%	79,3%	77,1%	78,7%	75,8%	64,9%	82,0%	77,7%

>Elecciones Xunta de Galicia 2012> Resto de Grupos mediáticos>Serie temporal

Figura 167.

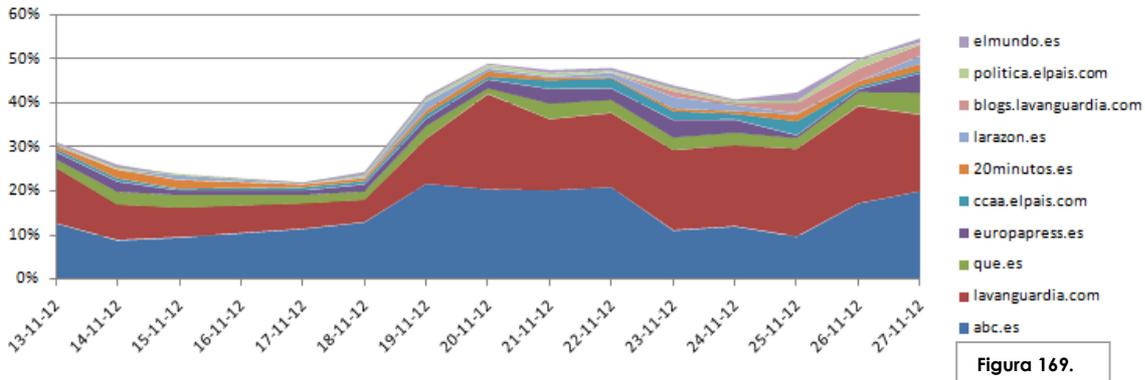
Grupos Mediáticos	7-10-12	9-10-12	10-10-12	12-10-12	13-10-12	14-10-12	15-10-12	16-10-12	17-10-12	18-10-12	19-10-12	20-10-12	21-10-12	22-10-12	Promedio
Grupo Joly	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,89%	0,67%	0,44%	0,14%
Mediapro	0,00%	0,44%	0,00%	0,67%	0,44%	0,00%	0,22%	0,22%	0,22%	0,00%	0,22%	0,00%	0,89%	0,67%	0,29%
Editorial Iparraguirre	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,44%	0,22%	0,00%	0,05%
Corporate Communicator - Eplural.com	0,00%	1,56%	0,89%	0,00%	0,00%	0,00%	0,44%	0,00%	0,00%	0,00%	0,00%	0,22%	0,22%	0,00%	0,24%
Grupo Begar	0,22%	0,22%	0,22%	0,22%	0,00%	0,22%	0,00%	1,78%	1,56%	0,22%	0,00%	0,22%	0,22%	0,00%	0,37%
Libertad digital S.A	0,44%	0,00%	0,00%	0,00%	0,89%	0,22%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,22%	0,00%	0,13%
Mediaset España	1,78%	1,78%	1,33%	0,22%	0,44%	0,22%	0,22%	0,22%	0,67%	0,44%	0,89%	0,00%	1,56%	0,22%	0,71%
Radio Popular S.A. - COPE	0,22%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,02%
Prensa internacional	0,00%	0,00%	0,00%	0,00%	0,00%	0,89%	0,00%	0,00%	0,00%	0,00%	0,00%	0,44%	0,22%	0,67%	0,16%
Grupo Correo Gallego	1,11%	2,00%	3,11%	1,33%	0,67%	1,11%	2,00%	2,89%	1,33%	1,78%	1,11%	2,22%	0,89%	0,22%	1,56%
Grupo Promecal	0,22%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,44%	0,44%	0,22%	0,22%	0,00%	0,11%
Grupo La Región.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	1,78%	1,78%	0,89%	0,67%	0,67%	0,44%	0,00%	0,00%	0,44%
Grupo Martínez Núñez	0,00%	0,00%	0,00%	0,00%	0,44%	0,00%	0,00%	0,00%	0,44%	0,00%	0,00%	0,00%	0,00%	0,00%	0,06%
Otros medios	15,56%	15,56%	15,11%	13,11%	17,56%	18,89%	16,00%	13,78%	15,11%	19,11%	17,78%	18,44%	27,33%	14,67%	17,00%
2.Estrato Medio	0,44%	1,33%	1,33%	2,44%	0,22%	0,67%	1,11%	0,44%	0,22%	0,00%	0,00%	0,22%	2,22%	0,89%	0,83%
3.Estrato Micro	0,22%	0,22%	0,22%	0,22%	0,22%	0,22%	0,22%	0,22%	0,22%	0,22%	0,22%	0,44%	0,22%	0,22%	0,24%
Total	20,2%	23,1%	22,2%	18,2%	20,9%	22,4%	22,0%	21,3%	20,7%	22,9%	21,3%	24,2%	35,1%	18,0%	22,3%

3.3.6. Generalitat de Catalunya – 25/11/2012

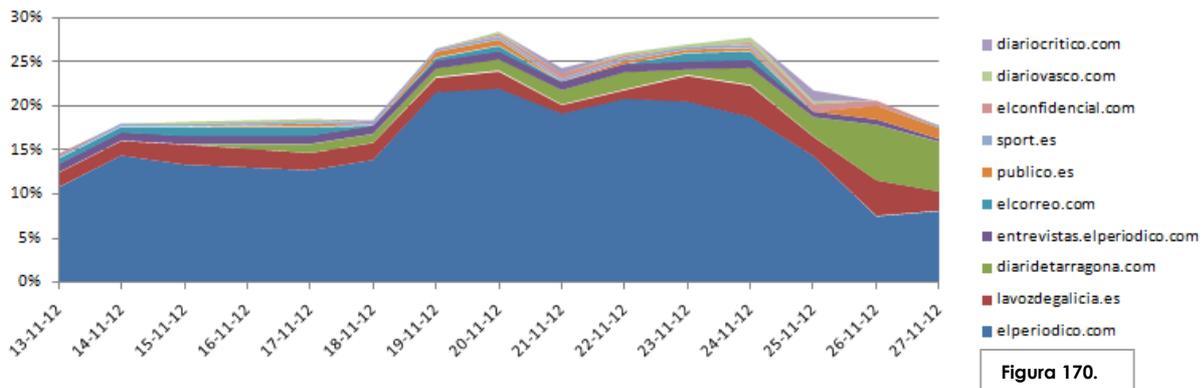


>Elecciones Generalitat 2012> Estrato Superior > Conglomerados principales

1.A.Prensa Estatal. El promedio del conglomerado fue del **39,55%**. El día de menor visibilidad fue el 17-nov con un 22,48%. Los días posteriores a las elecciones alcanzó sus dos mejores marcas con un 52,81% y un 55,45%.



1.A.Prensa Regional. El promedio de este conglomerado fue del **23,64%**. El día de menor visibilidad fue el 13-nov con un 15,31%, y el de máxima el 20-nov con un 29,45%.



1.A.Televisión. El promedio del conglomerado fue del **3,89%**. El día de menor visibilidad fue el 13-nov con un 2,28%, y el de máxima el día de las elecciones con un 5,61%.

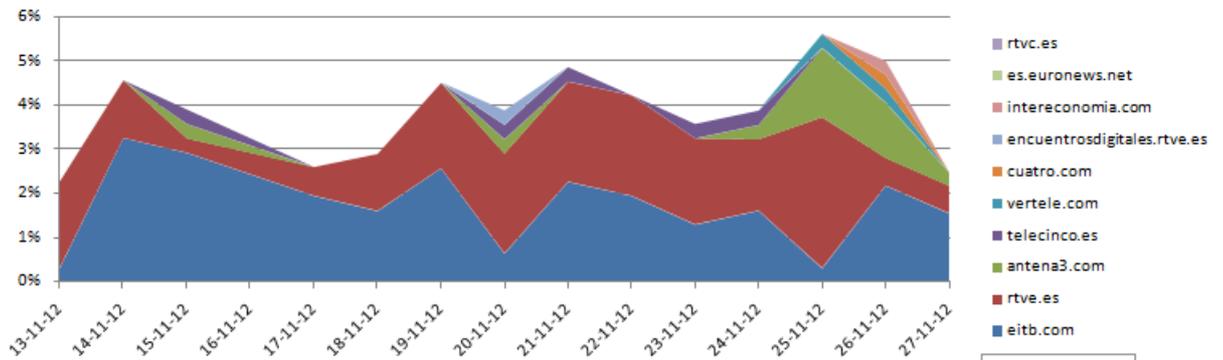


Figura 171.

1.A.Prensa Digital (Cibermedios). El promedio fue del **13,43%**. El día de menor visibilidad fue el 10-nov con un 8,79%, y el de máxima el 24-nov con un 17,8%.

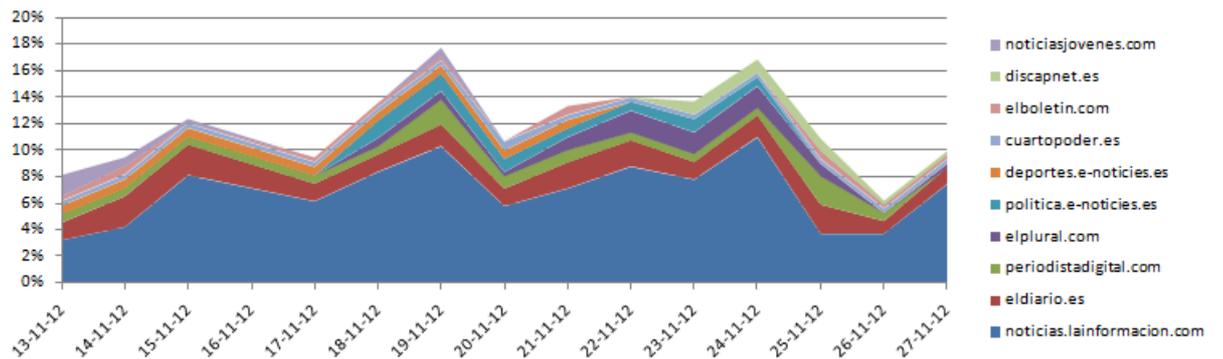


Figura 172.

>Elecciones Generalitat 2012> Estrato Superior > Conglomerados secundarios

1.A.Prensa Económica. El promedio de este conglomerado fue del **2,94%**. El día de mayor visibilidad fue el día de las elecciones, el 25 de noviembre, con un 4,05%.

1.A.Blog de Prensa. El promedio del conglomerado fue del **0,23%**.

1.A.Radio. El promedio de este conglomerado fue del **0,09%**. El día de mayor visibilidad fue el día 24 de noviembre, con un 0,65%.

1.A.Prensa Internacional. El promedio del conglomerado fue del **1,7%**. El día de mayor visibilidad fue el día 25 de noviembre, con un 4,05%.

1.B.Internet. El promedio del conglomerado fue del **12,46%**. El día de mayor visibilidad fue el día 17 de noviembre, cuando las noticias servidas a través de Google fueron del 38,76%¹⁰.

>Elecciones Generalitat 2012> Estrato Medio

Este estrato supuso un **1,84%** de los resultados. Provenientes del Estrato Medio aparecieron 31 portales como <http://teinteresa.es>, <http://pagina66.com> o <http://republica.com>.

>Elecciones Generalitat 2012> Estrato Micro

Este estrato supuso un **0,23%** de los resultados. En el Estrato Micro ganaron visibilidad 3 sitios: <http://rebellion.org>, <http://es.paperblog.com> y <http://ca.wikipedia.org>.

¹⁰ Las URLs recuperadas para esos días provenientes del dominio Google.com expiraron, no pudiendo recuperar el nombre de los medios que publicaban bajo el dominio del buscador.

>Elecciones Generalitat 2012>Conglomerados

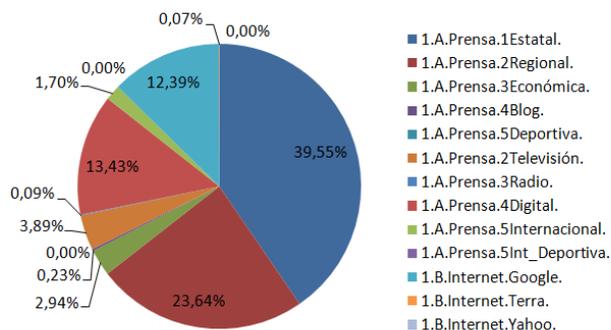


Figura 173.

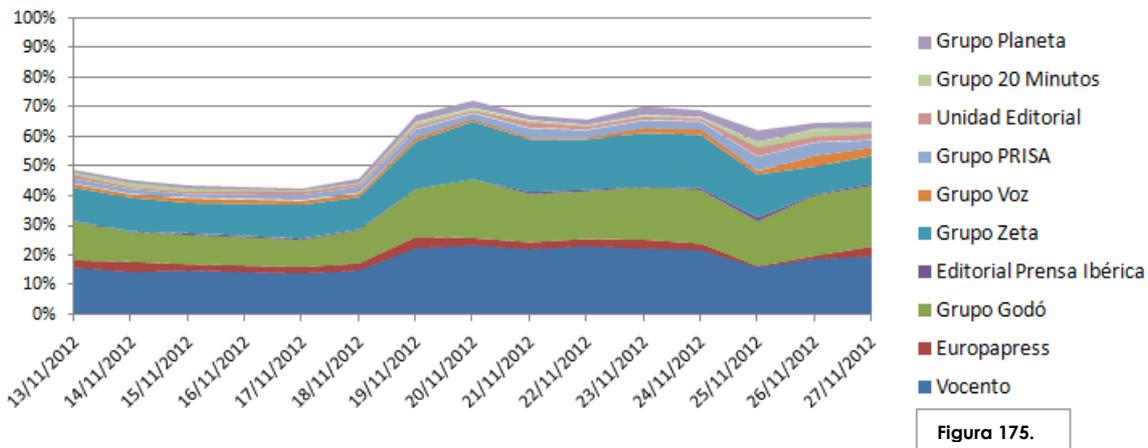
	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Barra roja]	[Barra amarilla]	[Barra verde]

Figura 174.

>Elecciones Generalitat 2012>Conglomerado>Tabla de Serie temporal

Conglomerados	13-11	14-11	15-11	17-11	18-11	19-11	20-11	21-11	22-11	23-11	24-11	25-11	26-11	27-11	Promedio
1.A.Prensa.1Estatal.	31,27%	26,14%	23,78%	22,48%	24,52%	41,94%	49,84%	47,40%	47,88%	44,95%	41,42%	42,37%	52,81%	55,45%	39,55%
1.A.Prensa.2Regional.	15,31%	19,28%	20,20%	20,20%	19,68%	29,35%	29,45%	24,35%	26,38%	28,01%	29,45%	23,36%	24,69%	21,18%	23,64%
1.A.Prensa.3Económica.	3,58%	2,29%	2,93%	2,93%	1,94%	3,55%	3,24%	3,57%	2,93%	2,93%	2,27%	4,05%	2,81%	2,18%	2,94%
1.A.Prensa.4Blog.	0,00%	0,33%	0,33%	0,00%	0,00%	0,00%	0,00%	0,00%	0,33%	0,98%	0,97%	0,00%	0,00%	0,31%	0,23%
1.A.Prensa.5Deportiva.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
1.A.Prensa.2Televisión.	2,28%	4,58%	3,91%	2,61%	2,90%	4,52%	3,88%	4,87%	4,23%	3,58%	3,88%	5,61%	5,00%	2,49%	3,89%
1.A.Prensa.3Radio.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,33%	0,65%	0,00%	0,00%	0,31%	0,09%
1.A.Prensa.4Digital.	8,79%	10,46%	13,03%	9,77%	14,19%	18,39%	11,33%	15,58%	14,66%	15,31%	17,80%	16,20%	10,00%	12,46%	13,43%
1.A.Prensa.5Internacional.	1,30%	1,31%	1,63%	1,30%	1,29%	1,29%	0,97%	0,97%	1,63%	0,98%	0,97%	4,05%	2,81%	3,12%	1,70%
1.A.Prensa.5Int_Deportiva.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
1.B.Internet.Google.	35,50%	34,31%	32,57%	38,76%	34,19%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	12,39%
1.B.Internet.Terra.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,33%	0,33%	0,00%	0,31%	0,00%	0,00%	0,07%
1.B.Internet.Yahoo.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Total	98,0%	98,7%	98,4%	98,0%	98,7%	99,0%	98,7%	96,8%	98,4%	97,4%	97,4%	96,0%	98,1%	97,5%	97,9%

>Elecciones Generalitat 2012>Grupos mediáticos>Serie temporal



- **Vocento.** El promedio de visibilidad de este grupo fue del **18,9%** sobre el total de resultados aparecidos en Top10 durante el periodo de captura de datos, dos puntos porcentuales por encima respecto a la muestra de las elecciones a la Generalitat de 2010.
- **Europapress.** Su promedio de visibilidad fue del **2,46%**. El día de los comicios apenas alcanzó un 0,46%. El segundo día posterior a las elecciones subió hasta el 3,35%.
- **Grupo Godó.** El promedio de visibilidad de este grupo fue del **15,15%**. Los días posteriores a las elecciones su visibilidad alcanzó una media del 20,3%.
- **Editorial Prensa Ibérica.** Su promedio de visibilidad fue del **0,49%**.
- **Grupo Zeta.** Su promedio de visibilidad fue del **13,83%**. El día de las elecciones su visibilidad fue del 14,37%. Los días posteriores a las elecciones su visibilidad alcanzó una media del 9,4%.
- **Grupo Voz.** El promedio de visibilidad de este grupo fue del **1,5%**.
- **Grupo PRISA.** El promedio de visibilidad de este grupo fue del **2,59%**. El día de las elecciones alcanzó una visibilidad del 4,79%.
- **Unidad Editorial.** Su promedio de visibilidad fue del **1,4%**. El día de las elecciones alcanzó su máximo con un 3,19%.
- **Grupo 20 Minutos.** El promedio de visibilidad de este grupo fue del **1,07%**. El día de los comicios bajó al 1,82% y el día posterior su máximo con un 2,49%.
- **Grupo Planeta.** Su promedio de visibilidad fue del **1,88%**. El día de los comicios alcanzó un 3,88%.

>Elecciones Generalitat 2012>Grupos mediáticos

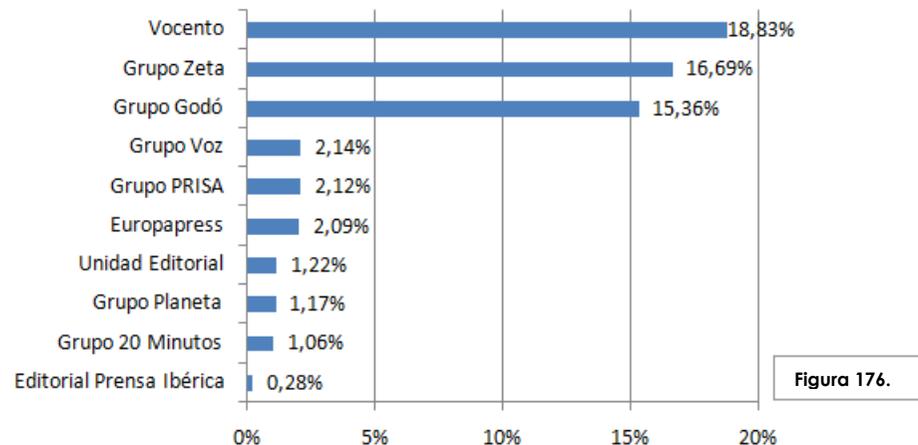


Figura 176.

	Mínima	Punto medio	Máxima
Tipo:	Valor más bajo	Percentil	Valor más alto
Valor:	(Valor más bajo)	50	(Valor más alto)
Color:	[Red]	[Yellow]	[Green]

>Elecciones Generalitat 2012>Top 10 Grupos mediáticos>Tabla de Serie temporal

Figura 177.

Grupos Mediáticos	13-11	14-11	15-11	17-11	18-11	19-11	20-11	21-11	22-11	23-11	24-11	25-11	26-11	27-11	Promedio
Vocento	15,64%	13,07%	14,66%	15,31%	14,84%	25,48%	23,95%	23,70%	24,43%	15,64%	16,50%	14,33%	20,94%	24,92%	18,83%
Europapress	1,63%	2,29%	0,98%	0,98%	1,61%	1,61%	1,94%	3,25%	2,61%	3,91%	2,91%	0,62%	0,63%	4,36%	2,09%
Grupo Godó	12,70%	8,17%	6,84%	5,86%	5,16%	10,65%	22,33%	16,23%	16,94%	19,54%	18,77%	22,12%	27,81%	20,87%	15,36%
Editorial Prensa Ibérica	0,00%	0,00%	0,65%	0,98%	0,32%	0,32%	0,00%	0,00%	0,00%	0,00%	0,32%	0,00%	0,31%	0,93%	0,28%
Grupo Zeta	12,05%	15,69%	14,66%	14,01%	15,16%	23,23%	23,30%	20,45%	22,15%	21,82%	20,06%	14,95%	8,13%	8,72%	16,69%
Grupo Voz	1,63%	1,63%	2,28%	1,95%	1,94%	1,61%	1,94%	0,97%	0,98%	2,93%	3,56%	2,18%	4,06%	2,18%	2,14%
Grupo PRISA	0,98%	1,31%	0,65%	0,98%	1,29%	2,26%	1,94%	2,92%	2,93%	3,58%	2,59%	4,05%	3,13%	0,93%	2,12%
Unidad Editorial	0,98%	0,98%	0,65%	0,98%	0,97%	0,97%	0,32%	2,27%	1,95%	1,30%	0,65%	3,12%	0,94%	0,93%	1,22%
Grupo 20 Minutos	0,65%	1,96%	1,95%	0,65%	0,65%	0,97%	1,29%	0,65%	0,33%	0,65%	0,65%	1,56%	1,25%	1,56%	1,06%
Grupo Planeta	0,33%	0,00%	1,30%	0,00%	0,32%	1,94%	0,97%	0,32%	0,98%	2,93%	1,62%	1,87%	1,25%	2,49%	1,17%
Total	46,58%	45,10%	44,63%	41,69%	42,26%	69,03%	77,99%	70,78%	73,29%	72,31%	67,64%	64,80%	68,44%	67,91%	60,96%

>Elecciones Generalitat 2012>Resto de Grupos mediáticos>Serie temporal

Figura 178.

Grupos Mediáticos	13-11	14-11	15-11	17-11	18-11	19-11	20-11	21-11	22-11	23-11	24-11	25-11	26-11	27-11	Promedio
Grupo Joly	0,00%	0,33%	0,33%	0,65%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,09%
Mediapro	0,00%	0,00%	0,00%	0,33%	0,00%	0,65%	0,65%	0,00%	0,33%	0,33%	0,32%	0,00%	1,56%	1,56%	0,41%
Editorial Iparraguirre	0,00%	0,00%	0,33%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,32%	0,31%	2,19%	0,93%	0,30%
Corporate Communicator S.L. - Elplural.com	0,00%	0,00%	0,00%	0,00%	0,65%	0,65%	0,32%	0,97%	1,63%	1,63%	1,62%	0,93%	0,00%	0,31%	0,62%
Grupo Begar	0,65%	0,65%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,32%	0,00%	0,00%	0,00%	0,11%
Libertad digital S.A	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,31%	0,31%	0,05%
Mediaset España	0,00%	0,00%	0,33%	0,00%	0,00%	0,00%	0,32%	0,32%	0,00%	0,33%	0,32%	0,00%	0,31%	0,00%	0,14%
Radio Popular S.A. - COPE	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Grupo Intereconomía	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,31%	0,00%	0,02%
Prensa internacional	1,30%	1,31%	1,63%	1,30%	1,29%	1,29%	0,97%	0,97%	1,63%	0,98%	0,97%	4,05%	2,81%	3,12%	1,70%
Grupo Correo Gallego	0,33%	0,33%	0,33%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,07%
EKHE (izquierda abertzal)	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Grupo Promecal	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Sakkia Hiferpress (inmobiliaria)	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Grupo publicaciones del sur	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Grupo La Región.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Otros medios	49,19%	50,98%	50,81%	54,07%	54,52%	27,42%	18,45%	23,70%	21,50%	21,82%	25,89%	25,86%	22,19%	23,36%	33,46%
2.Estrato Medio	1,63%	0,98%	1,30%	1,63%	1,29%	0,97%	1,29%	1,95%	1,63%	2,28%	2,59%	3,74%	1,88%	2,49%	1,84%
3.Micro-estrato	0,33%	0,33%	0,33%	0,33%	0,00%	0,00%	0,00%	1,30%	0,00%	0,33%	0,00%	0,31%	0,00%	0,00%	0,23%
Total	53,42%	54,90%	55,37%	58,31%	57,74%	30,97%	22,01%	29,22%	26,71%	27,69%	32,36%	35,20%	31,56%	32,09%	39,04%

3.3.7. Tendencias Google News

En este apartado mostramos a modo de resumen cuál ha sido la evolución de la composición de los tres estratos en el motor de búsqueda web de Google.

Estrato Superior, convencional o de medios de comunicación. Google News se alimenta del Estrato Superior un promedio del **96,5%** durante el periodo de estudio de 2010 a 2012.

Estrato Medio. El estrato alcanza una visibilidad promedio del **3,24%** y a lo largo de la serie temporal de elecciones muestra una tendencia decreciente.

Estrato Micro. Este estrato tiene una visibilidad promedio del **0,26%**. Las dos elecciones de referencia muestran una evolución decreciente de visibilidad durante el periodo de estudio de 2010 a 2012.

Generalitat Catalunya – 28 Noviembre 2010.

En estas elecciones la visibilidad del conglomerado de 'Prensa.1Estatl' fue de un **41,18%**, mientras que 'Prensa.2Regional' alcanzó una media de **29,61%**. Los sitios web de <http://abc.es> y <http://elperiodico.com> son los líderes de sus respectivos conglomerados, destacando el incremento del medio de Vocento los días previos a las elecciones. En estos comicios el conglomerado 'Prensa.2Televisión' tiene una visibilidad del 5,03%. El canal de televisión preferido por Google fue <http://rtve.es>. El cibermedio con más visibilidad fue <http://noticias.lainformacion.com>.

Elecciones Municipales-22 Mayo 2011.

Al igual que pocos meses antes, la visibilidad de los conglomerados principales de 'Prensa.1Estatl' y 'Prensa.2Regional' lograron el **70%** del total de resultados. Destacaron <http://abc.es> y <http://lne.com> como prensa líder de sus respectivos conglomerados. Aún así, el conglomerado 'Prensa.1Estatl' alcanzó la menor visibilidad registrada (37,3%), probablemente debido a la tipología de palabra clave utilizada en el test. El canal de televisión preferido por Google fue de nuevo rtve.es. El cibermedio con más visibilidad fue libertaddigital.com y teleprensa.es. El conglomerado '1.A.Prensa.3Económica.' sólo consiguió un 1,3% de los resultados. Por el contrario, '1.A.Prensa.5Internacional.' logró un 7,28% de visibilidad, la mayor cifra de todas las elecciones, debido al mayor peso de las búsquedas genéricas.

Elecciones Generales- 20 Noviembre 2011.

Los conglomerados principales de 'Prensa.1Estatal' y 'Prensa.2Regional' superaron el **71%** de resultados, destacando <http://abc.es> y <http://lne.es> como líderes de sus conglomerados. El grupo mediático Vocento fue el que alcanzó más visibilidad con un 24% del total de resultados mostrados en Top10. El canal de televisión preferido por Google fue de nuevo rtve.es. El cibermedio con más visibilidad fue lainformacion.com y teleprensa.es.

Junta de Andalucía - Marzo 201.

En estas elecciones el conglomerado de 'Prensa.1Estatal' tuvo una visibilidad global del **50,9%**, sólo superada en las elecciones gallegas. El sitio web de prensa con más visibilidad del conglomerado fue <http://europapress.es> con un 32% del total de resultados en Top10, el mayor registrado en toda la serie temporal. La 'Prensa.2Regional' alcanzó la menor visibilidad registrada sobre el total de resultados, alcanzando un **19,07%** con el <http://ideal.es> como líder del conglomerado. El canal de televisión preferido por Google fue rtve.es. El cibermedio con más visibilidad fue lainformacion.com y periodistadigital.es.

Xunta Galicia – 21 Octubre 2012.

La visibilidad del conglomerado de '**Prensa.1Estatal.**' fue la mayor registrada con un 52,46% del total de resultados, y los conglomerados principales superaron el 81,56% de resultados, cifra máxima registrada. '**1.A.Prensa.3Radio.**' logró su menor visibilidad en estos comicios el 0,02%. Los sitios web de prensa con más visibilidad líderes de los conglomerados fueron <http://abc.es> y <http://lavozdegalicia.es>. El grupo Vocento alcanzó su máxima visibilidad en serie de elecciones con un **39%** del total de resultados mostrados en Top10 seguido por el Grupo Voz.

El canal de televisión preferido por Google fue rtve.es. Los cibermedios con más visibilidad fueron lainformacion.com y periodistadigital.es.

Generalitat Catalunya - 25 Noviembre 2012.

Los segundos comicios catalanes mostraron variaciones significativas. El conglomerado de '**Prensa.1Estatal**' descendió dos puntos hasta un 39,55%, mientras que '**Prensa.2Regional**' perdió un 6% y bajó hasta el 23,64%. Destacó el incremento de visibilidad de '**1.A.Prensa.3Económica.**' de un punto, y los cibermedios que subieron un 2,5%.

Los sitios web de prensa con más visibilidad líderes de los conglomerados fueron <http://abc.es> y la <http://lavanguardia.com> en 'Estatad' y en 'Regional' <http://elperiodico.es>. El grupo Vocento fue el más visible en Top10 con casi un 20% del Top10, seguidos por el Grupo Godó (16,2%) y por el Grupo Zeta (14,6%). En el caso de estos dos últimos grupos cabe destacar que casi duplicaron su visibilidad respecto a las elecciones de dos años antes. El canal de televisión preferido por Google para estas elecciones fue <http://eitb.com> y <http://rtve.es> quedó relegado a segunda posición. Al igual que dos años antes el cibermedio con más visibilidad fue <http://noticias.lainformacion.com> seguido por eldiario.es.

3.4. Google Blog Search

En este tercer bloque de análisis se visualizarán los resultados de los tres estratos en el índice específico centrado en la blogosfera, y que denominamos Blog Search, así como su evolución a lo largo de la serie temporal compuesta por las seis elecciones del periodo 2010-2012. Analizaremos los diferentes sectores y los principales conglomerados objeto de estudio y cómo se comportaron en Google Blog Search durante todo el periodo de estudio. Los aspectos objeto de estudio son:

- **La serie temporal de elecciones.** Al igual que hemos realizado en Web Search y Google News, mostraremos las comparativas de series temporales de 14 días correspondientes a las búsquedas de los seis comicios electorales seleccionados. El cuadro resumen de la evolución de los estratos es el siguiente:

	Catalunya 2010	Municipales 2011	Generales 2011	Andalucía 2012	Galicia 2012	Catalunya 2012
Estrato convencional	16,07%	27,53%	32,47%	25,00%	51,79%	43,01%
Estrato medio	27,26%	32,74%	31,65%	35,60%	20,59%	27,74%
Micro-estrato	56,67%	39,73%	35,88%	39,40%	27,62%	29,25%

Figura 179.

- **Los sectores y conglomerados principales.** Son objeto de estudio la visibilidad en Blog Search de los principales sectores, conglomerados y páginas web específicas que hemos analizado en las otras opciones de búsqueda. Las series temporales muestran el comportamiento y evolución de la visibilidad de estos conglomerados a lo largo de cada una de las elecciones, y de todas en su conjunto.

La evolución de proporciones de los diferentes estratos se muestran en el siguiente diagrama:

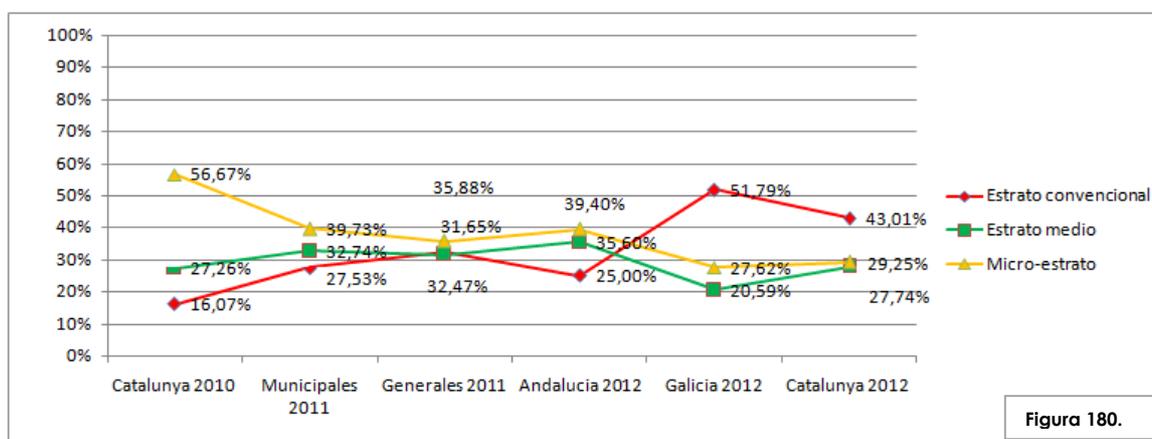


Figura 180.

En el gráfico de líneas¹¹ podemos apreciar:

- **Estrato Superior**, convencional o de medios de comunicación: en este apartado los conglomerados del Estrato Superior se muestran agrupados y no discriminados entre principales y secundarios como se ha realizado en los apartados previos. La evolución de los sitios web del Estrato Superior muestra una trayectoria claramente ascendente a lo largo del periodo, pasando del 16% en las elecciones a la Generalitat 2010 a cifras del 51% en las elecciones a la Xunta y del 43% en las dos elecciones a la Generalitat del año 2012.
- **Estrato Medio**: la evolución de los sitios web del Estrato Medio tienen una evolución entorno al 30%, con máximos en las elecciones municipales con un 32% y mínimos en las elecciones a la Xunta de Galicia con un 20,6%.
- **Estrato Micro**: A pesar de que la naturaleza de la opción de búsqueda en blog sería mostrar resultados de la blogosfera, nos encontramos que Google mezcla en SERPs páginas de las otras esferas, y que a lo largo del tiempo va disminuyendo la visibilidad de los sitios web considerados del Estrato Micro que pierden de 2010 a 2012 el 27,4% de su visibilidad a favor del Estrato Superior.

¹¹ Blogsearch - Gráficos de Líneas. Utilizamos este gráfico para representar los ratios de visibilidad en las seis series temporales de cada una de las elecciones. En este gráfico se representa la composición de la Infoesfera en Blogsearch y muestra la variación de visibilidad de los estratos.

3.4.1. Generalitat de Catalunya – 28/11/2010

>Elecciones Generalitat 2010> Estrato Superior

A causa de las características de Universal Search, Google combina resultados propios del canal de búsqueda Blog Search y resultados relevantes provenientes de otros canales de búsqueda, en este caso webs y noticias. El promedio de este estrato fue del **16,07%**. La serie de datos de visibilidad del Estrato Superior y los datos de visibilidad de los conglomerados objeto de estudio fueron:

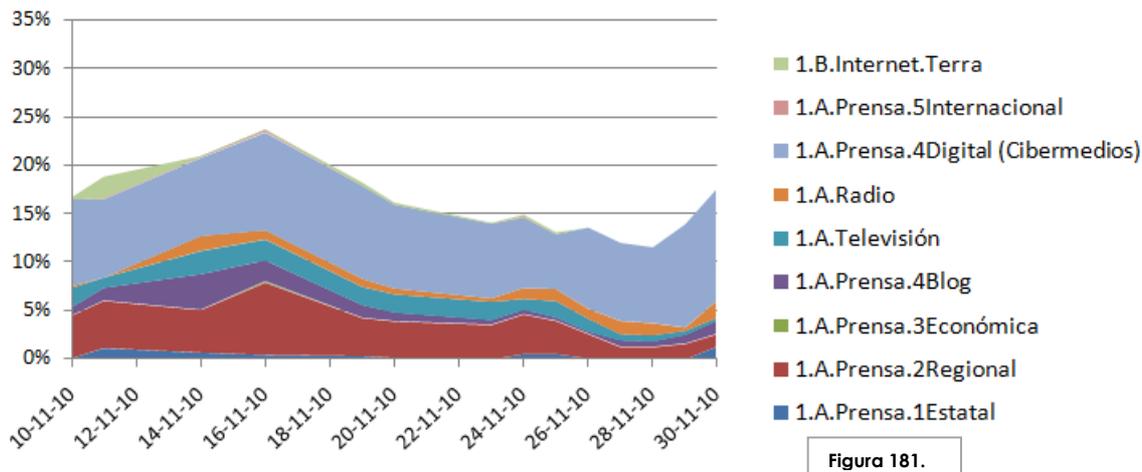


Figura 181.

1.A.Mass_Media_Prensa.1Nacional

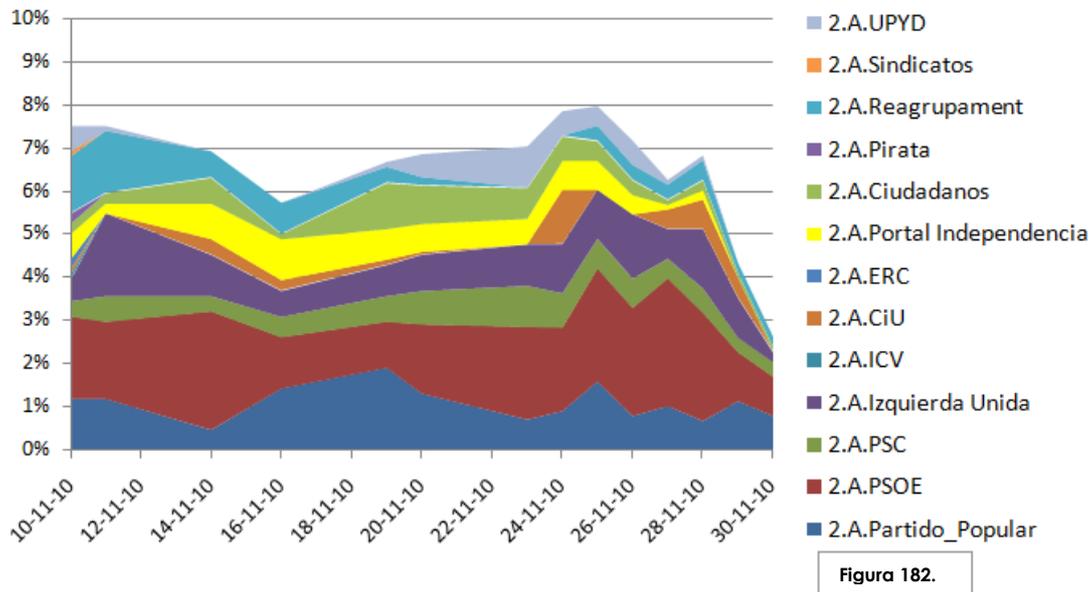
- **Prensa Estatal.** El promedio es del **0,44%** con un repunte el día 30-nov posterior a las elecciones. Es la peor cifra del conglomerado en toda la serie de elecciones.
- **Prensa Regional.** Este conglomerado tiene una visibilidad media del **3,23%**. Destaca la primera semana con una visibilidad máxima del 7,5% el día 16-nov.
- **Prensa Económica.** Solo aparece en Top10 el día 16-nov con un 0,12%.
- **Blog de Prensa.** El promedio es del **1%**. El día 14-nov alcanza un 3,69%.
- **Televisión.** Este conglomerado tiene una media del **1,42%**. Destaca la primera semana de captura con medias del 1,94%
- **Radio.** Con una visibilidad promedio del **0,9%**, destaca la activación del conglomerado los días de las elecciones y posteriores, alcanzando un 1,82% el día 30 de noviembre.
- **Prensa Digital (Cibermedios).** Su visibilidad media del **8,7%** durante toda la serie temporal. Los días de más visibilidad son los posteriores a las elecciones, los días 29 y 30-nov con un 10,68% y un 11,59%.

1.A.Mass_Media_Prensa.Internacional. El conglomerado logra visibilidad tres días de la primera semana de capturas.

1.B.Internet.Terra. Este conglomerado es visible los primeros diez días de la muestra. El día 11-nov alcanza un 2,26% y desaparece de SERPs a partir del 26-nov.

>Elecciones Generalitat 2010> Estrato Medio

El promedio de este estrato fue del **27,26%**. La serie de datos de conglomerados y sus datos de visibilidad fueron:



2.A.Esfera política.

- **Partido_Popular.** El conjunto de páginas del PP tiene una visibilidad promedio del **1,06%** durante toda la serie. El día de las elecciones alcanza su mínimo de la serie con un 0,68%.
- **PSOE-PSC.** El conjunto de páginas del Partido Socialista Obrero Español tiene una visibilidad promedio del **1,95%** durante toda la serie, mientras que las páginas del Partido Socialista Catalán alcanzan el 0,58%. La suma de ambas es del **2,53%** de la visibilidad total. Las cifras más bajas de visibilidad son los días posteriores a las elecciones.
- **Izquierda Unida-ICV.** El conjunto de páginas de Izquierda Unida tiene una visibilidad promedio del **1,08%**. Sigue la misma tendencia y tiene su peor día el 30 de noviembre.
- **Resto de partidos.** Las demás agrupaciones políticas logran la siguiente visibilidad: CiU un 0,26%, ERC un 0,02%, Ciudadans un 0,37% y UPYD con un 0,26%. En la esfera independentista catalana destacan Reagrupament con un 0,50%, una visibilidad mayor al conglomerado 'Portal Independencia' que alcanza un 0,45%.

>Elecciones Generalitat 2010 > Estrato Micro

El promedio de este estrato fue del **56,67%** (un 41,6% pertenecía a Redes de blogs). La serie de datos objeto de estudio y su visibilidad en Top10 fue:

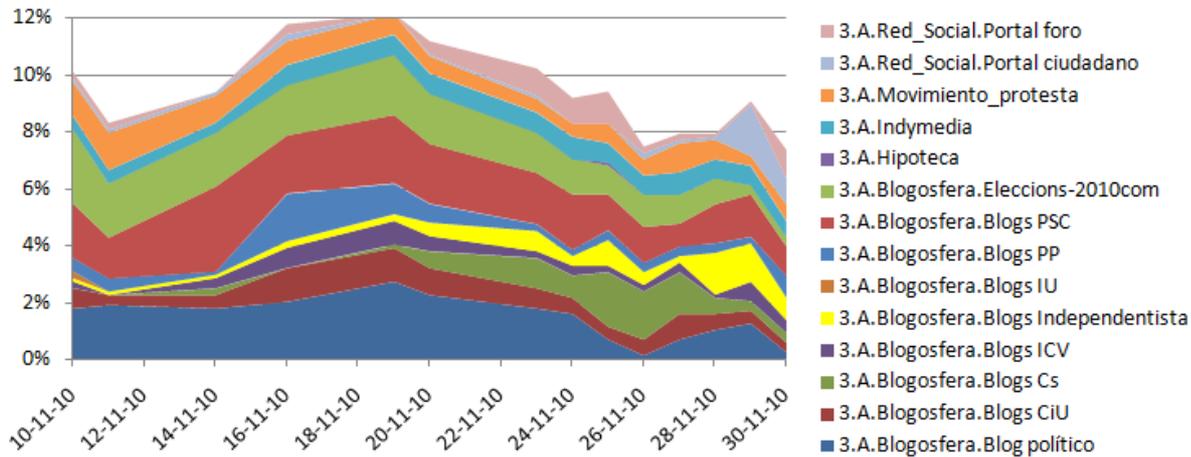


Figura 183.

3.A.Blogosfera política.

- **Blog político.** El promedio de los blogs políticos sin adscripción partidista fue del **1,49%**. El día de mayor visibilidad fue el día 20 con un 3,33%. El segundo día posterior a las elecciones baja hasta el 0,23%. Su visibilidad se concentra los ocho primeros días de captura de datos.
- **Blogs CiU.** El promedio es del 0,61%. La visibilidad de este conglomerado se concentra la primera semana de captura.
- **Blogs Cs.** El promedio es del 0,61%. Su visibilidad se concentra los cuatro días anteriores a las elecciones.
- **Blogs ICV.** El promedio es del 0,34%. La visibilidad se concentra la primera semana de capturas.
- **Blogs Independentista.** El promedio es del 0,51%. La visibilidad de los blogs independentistas se concentra el día de las elecciones y el inmediatamente posterior.
- **Blogs PP.** El promedio es del 0,58%. La visibilidad es mayor la primera semana de datos.
- **Blogs PSC.** El promedio es del 1,75% con una visibilidad constante a lo largo del periodo.
- **Eleccions-2010com.** El promedio de esta página experimental es del **1,37%**. El primer día de captura registra un **2,62%** del total y pierde visibilidad paulatinamente hasta su peor registro el 30-nov con un **3,34%**.

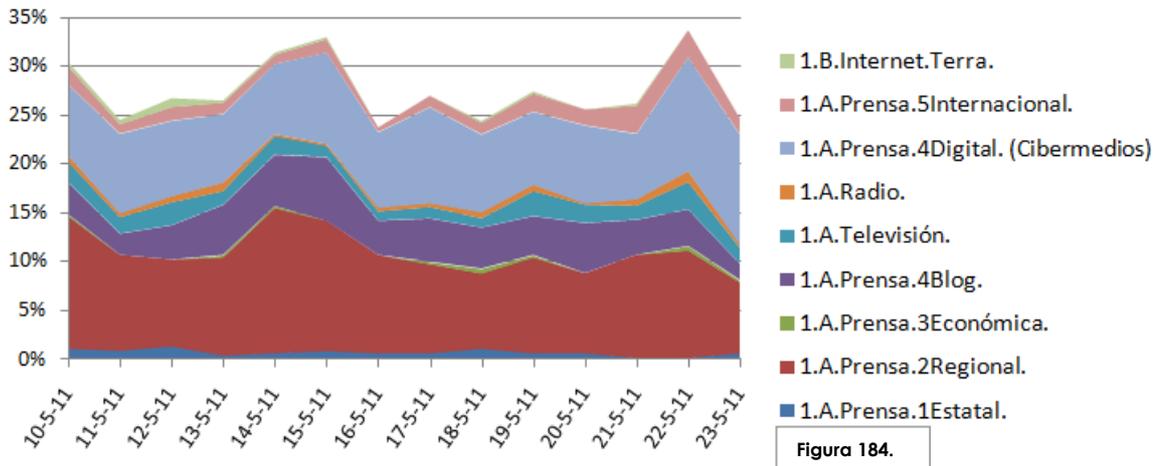
3.A.Movimiento_protesta. El promedio del conglomerado fue del **0,78%**. El día de mayor visibilidad fue el día 20 con un 1,19%.

Resto de conglomerados. Blogs IU. Alcanza un residual 0,02%, **3.A.Hipoteca.** un 0,01%, **3.A.Red_Social.Portal ciudadano.** El 0,28%, y **3.A.Red_Social.Portal foro.** un 0,4%.

3.4.2. Elecciones Municipales – 22/5/2011

>Elecciones Municipales 2011> Estrato Superior

El promedio de este estrato fue del 27,53%. A continuación mostramos la **serie** de datos de conglomerados de este estrato durante las elecciones y los datos de visibilidad de sus conglomerados:



1.A.Mass_Media_Prensa.1Nacional

- **Prensa Estatal.** El promedio es del **0,77%**. Los peores días de visibilidad son el día previo a las elecciones y la jornada electoral con un 0,24% y un 0,23%.
- **Prensa Regional.** El promedio es del **10,24%**. La visibilidad se mantiene constante durante todo el periodo.
- **Prensa Económica.** El promedio es residual del 0,17%.
- **Blog de Prensa.** El promedio es del 4%. Los mejores días son el final de la primera semana de captura.
- **Televisión.** El promedio es del **1,7%**. La visibilidad se mantiene constante durante todo el periodo y sube hasta el **2,79%** el día de las elecciones.
- **Radio.** El promedio es del 0,6% y sube hasta el 1,16% el día 22.
- **Prensa Digital (Cibermedios).** El promedio es del **8,33%**. La visibilidad se mantiene constante durante todo el periodo y alcanza sus mejores cifras el día de las elecciones y consecutivo con un **11,4%**.

1.A.Prensa.Internacional. El promedio es del 1,52%. Tiene un repunte el día de las elecciones y el posterior.

1.B.Internet.Terra. El promedio es de un residual 0,23%.

>Elecciones Municipales 2011> Estrato Medio

El promedio de este estrato fue del **32,74%** y los conglomerados que alcanzaron visibilidad en el Top10 fueron:

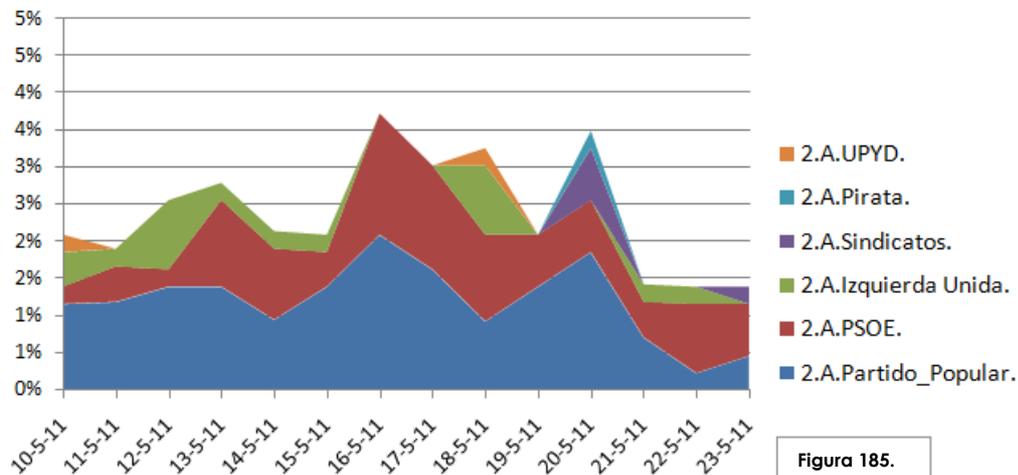


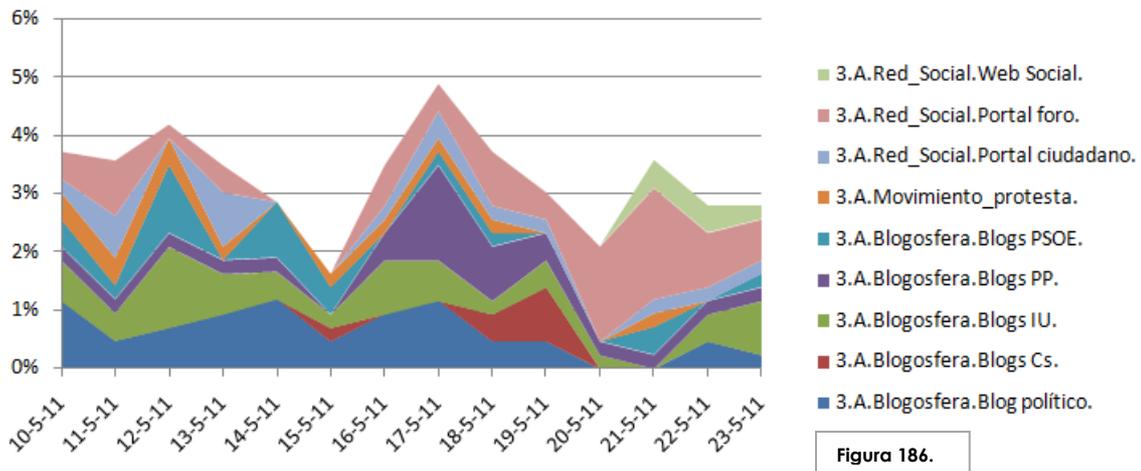
Figura 185.

2.A.Esfera política.

- **Partido_Popular.** El promedio de visibilidad del conjunto de páginas del PP es del **1,20%**. Sus mejores días son la semana central de captura, cayendo el día de las elecciones hasta el 0,23%.
- **PSOE.** El promedio es del **0,8%**. Sus mejores días son el 16-18 de mayo.
- **Izquierda Unida.** El promedio es del **0,27%**. Su visibilidad se concentra en la primera semana.
- **Resto de partidos.** La visibilidad del resto de grupos políticos es residual y solo alcanzan visibilidad los conglomerados de Sindicatos, Partido Pirata y UPYD que juntos suman un 0,12%.

>Elecciones Municipales 2011> Estrato Micro

El promedio de este estrato fue del **39,73%** y los conglomerados objeto de estudio visibles en el Top10 fueron:



3.A.Blogosfera.

- **Blog político.** El promedio es del **0,62%**.
- **Blogs Cs.** El promedio es residual del **0,12%** con visibilidad la semana central de captura.
- **Blogs IU.** El promedio es del **0,57%**, el segundo conglomerado con más visibilidad.
- **Blogs PP.** El promedio es del **0,42%**. Despunta el primer día de la segunda semana para caer hasta el 0,23% el día de las elecciones.
- **Blogs PSOE.** El promedio es del **0,32%**. Desaparece el día de las elecciones.

3.A.Movimiento_protesta. El promedio es del **0, 2%**. Aparece los nueve primeros días de captura.

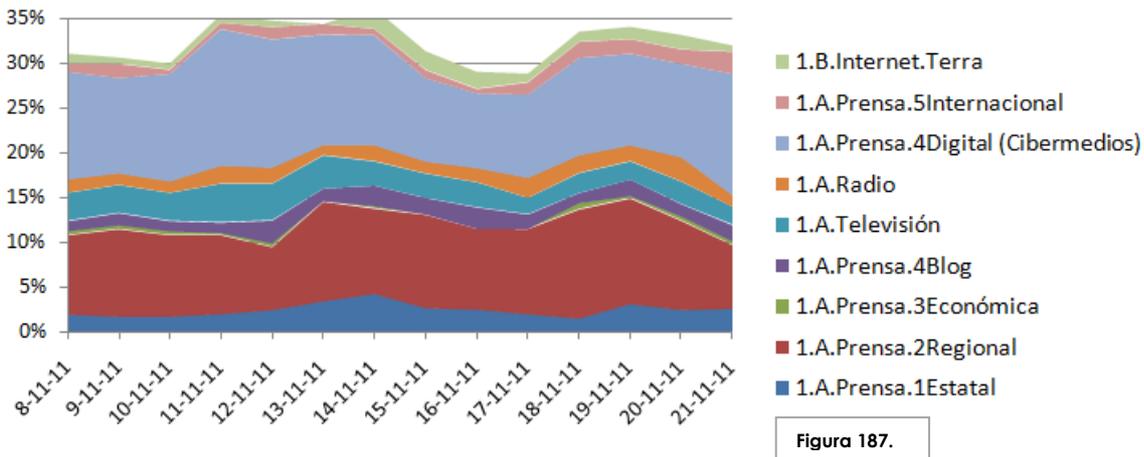
3.A.Red_Social.

- **Portal ciudadano.** El promedio es del **0,27%**.
- **Portal foro.** El promedio es del **0,7%**. Se concentra la semana de los comicios.
- **Web Social.** El promedio es residual del **0,08%**. Aparece solo tres jornadas pivotando el día de los comicios.

3.4.3. Elecciones Generales – 20/11/2011

>Elecciones Generales > Estrato superior

El promedio de este estrato fue del 32,47%. La **serie** de datos de visibilidad de conglomerados durante las elecciones y los datos de visibilidad de sus conglomerados fueron:



1.A.Mass_Media_Prensa.1Nacional

- **Prensa Estatal.** El promedio es del **2,5%**.
- **Prensa Regional.** El promedio es del **9,61%**. Se concentra la semana de los comicios.
- **Prensa Económica.** El promedio es del **0,29%**.
- **Blog de Prensa.** El promedio es del **1,62%**. Su visibilidad se concentra los días centrales de captura alcanzando un 2,5% el día 12.
- **Televisión.** El promedio es del **2,8%**. Su visibilidad se concentra los días centrales de captura alcanzando un **4,1%** el día 12.
- **Radio.** El promedio es del **1,73%**. Destaca el día anterior a los comicios con un **2,7%**.
- **Prensa Digital (Cibermedios).** El promedio es del **11,5%**. Se concentra a lo largo de todo el periodo y alcanza máximos el 11 y 12-nov con un 15,23% y un 14,32% la semana de los comicios.

1.A.Prensa.Internacional. El promedio es del **1,21%**.

1.B.Internet.Terra. El promedio es del **1,13%**. Se concentra los días centrales de los comicios.

>Elecciones Generales > Estrato Medio

El promedio de este estrato fue del **31,65%** y la **serie** de datos y conglomerados que alcanzaron visibilidad en el Top10 fue la siguiente:

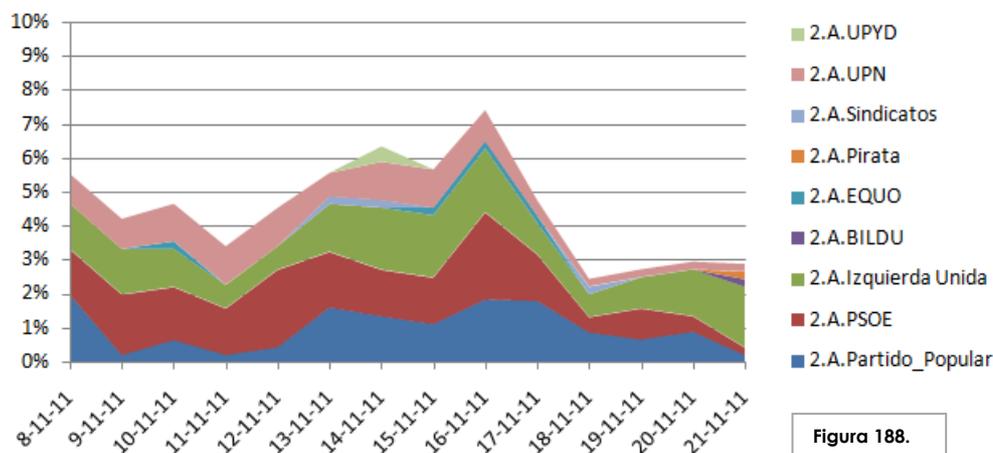


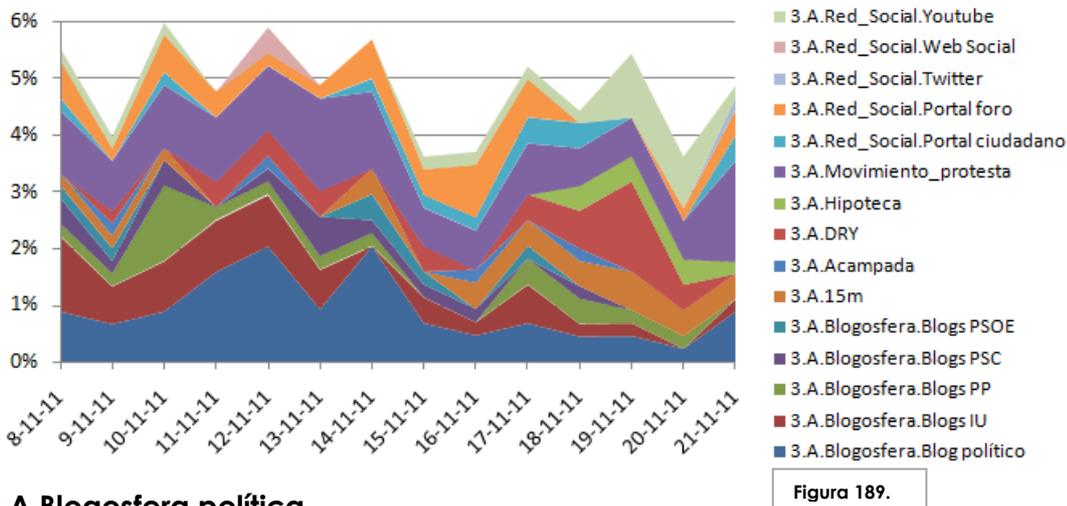
Figura 188.

2.A.Esfera política.

- **Partido_Popular.** El promedio es del **1%**. Su visibilidad se concentra los días centrales de los comicios. El día 16 logra un 1,86% pero el día de las elecciones cae hasta el 0,22%.
- **PSOE.** El promedio es del **1,32%**. El conglomerado se comporta exactamente igual al PP. Su máximo es el día 16 con un 2,565.
- **Izquierda Unida.** El promedio es del **1,26%**. Se comporta igual que los partidos popular y socialista pero el día de las elecciones alcanza un segundo máximo con un 1,78%.
- **UPN.** El promedio es del **0,74%**. Su visibilidad se concentra la primera semana de captura.
- **Resto de partidos.** Las demás agrupaciones políticas logran una visibilidad residual: BILDU un 0,02%, EQUO. el 0,06%, Partido Pirata un 0,02%, Sindicatos un 0,05% y UPYD el 0,03%.

>Elecciones Generales > Estrato Micro

El promedio de este estrato fue del **35,88%** y los conglomerados objeto de estudio visibles en el Top10 fueron:



3.A.Blogosfera política.

- **Blog político.** El promedio es del **0,92%**. Destaca el día 14 con un 2,05%.
- **Blogs IU.** Su visibilidad media es del **0,53%**. Destaca la primera semana de datos.
- **Blogs PP.** El promedio es del **0,29%**. Concentra su visibilidad el inicio del periodo y el día de las elecciones desaparece del Top10.
- **Blogs PSC.** El promedio es del **0,21%**. Aparece El día de las elecciones desaparece del Top10.
- **Blogs PSOE.** El promedio es residual del **0,1%**. Solo aparece 5 días del periodo. El día de las elecciones desaparece del Top10.

3.A.Movimiento_protesta.

- **15m.** El promedio es del **0,29%**.
- **Acampada.** El promedio es del **0,06%**. Aparece diez días del periodo de captura y sus máximos se concentran en torno al día de las elecciones.
- **DRY.** El promedio es del **0,37%**. El día 19 alcanza un 1,6%.
- **Hipoteca.** El promedio es residual del **0,11%**. Su visibilidad se concentra en torno al día de las elecciones.
- **Movimiento_protesta.** El promedio es del **1,03%**. Es más visible la primera semana de captura y el día de las elecciones, cuando alcanza el **1,78%**.

3.A.Red_Social.

- **Portal ciudadano.** El promedio es del **0,18%**. Aparece ocho de los días.
- **Portal foro.** El promedio es del **0,42%**.
- **Twitter.** Los resultados de micro-blogging suponen un promedio del **0,02%**.
- **Youtube.** Los resultados del portal de vídeo fueron un promedio del **0,27%** del total de UGC.
- **Web Social.** El promedio es residual del **0,03%**.

3.4.4. Junta de Andalucía – 12/5/2012

>Elecciones Andalucía 2012> Estrato superior

El promedio de este estrato fue del 25%. Destaca la no aparición en SERP del conglomerado 1.A.Prensa.3Económica. La **serie** de datos de conglomerados y su visibilidad durante las elecciones es la siguiente:

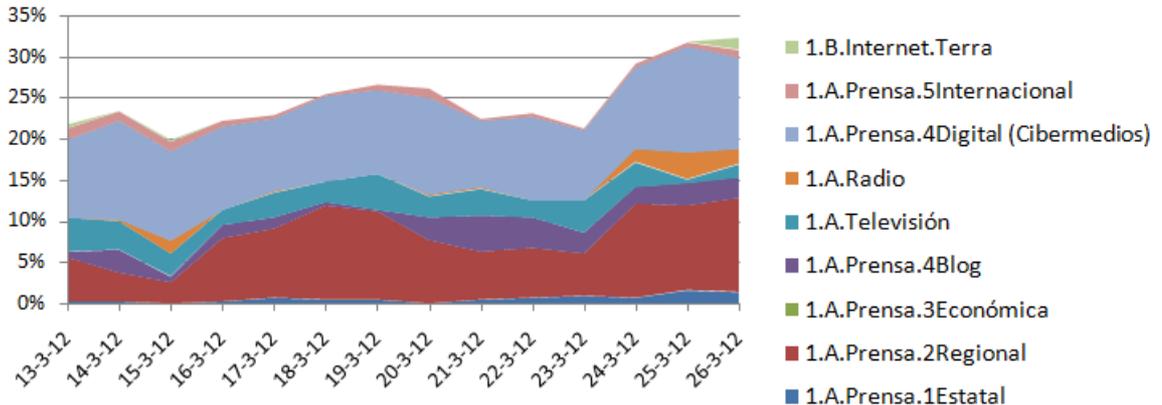


Figura 190.

1.A.Mass_Media_Prensa.1Nacional

- **Prensa Estatal.** El promedio es del **0,58%**, la segunda peor cifra en la serie. Concentra su visibilidad el día de las elecciones y el anterior y alcanza un 1,63%.
- **Prensa Regional.** El promedio es del **7,66%**. Concentra su visibilidad los días centrales del periodo y alcanza un 11,4% el día 18.
- **Blog de Prensa.** El promedio es del **2,06%**. Concentra su visibilidad la segunda semana del periodo.
- **Televisión.** El promedio es del **2,81%**. Pierde visibilidad los días cercanos a los comicios.
- **Radio.** El promedio es del **0,66%**. Su mejor día es el anterior a las elecciones con un 3,26%.
- **Digital (Cibermedios).** El promedio es del **10,38%**. Sus mejores fechas son los días 25 y 26 de marzo con un 13% y 11%.

1.A.Prensa.Internacional. El promedio es del **0,7%**. Es más visible la primera semana de captura

1.B.Internet.Terra. El promedio es del **0,15%**. Es más visible el día de las elecciones.

>Elecciones Andalucía 2012> Estrato Medio

El promedio de este estrato fue del **35,6%** y los conglomerados objeto de estudio que alcanzaron visibilidad en el Top10 fueron:

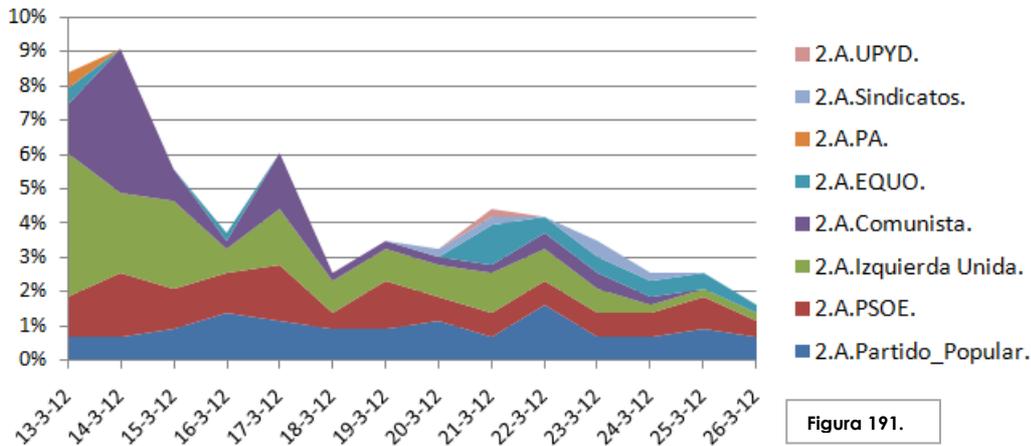


Figura 191.

2.A.Esfera política.

- **Partido_Popular.** El promedio es del **0,95%**. Logra su máximo el día 16 con un 1,4%. Su mínimo es el día de las elecciones.
- **PSOE.** El promedio es del **0,98%**. Es más visible la primera semana de captura. Su mínimo es el día de las elecciones.
- **Izquierda Unida.** El promedio es del **1,26%**. Concentra su visibilidad la primera semana de captura. Su mínimo es el día de las elecciones.
- **Comunista.** El promedio es del **0,75%**. Logra un 4,2% el día 14.
- **EQUO.** El promedio es del **0,28%**. Su máximo es el día 23 con un 0,5%.
- **Resto de partidos.** Los datos de las demás agrupaciones políticas son: Partido Andalucista un 0,03%, Sindicatos un 0,08% y UPYD con un promedio del 0,02%.

>Elecciones Andalucía 2012> Estrato Micro

El promedio de este estrato fue del **39,4%** y los conglomerados objeto de estudio visibles en el Top10 fueron:

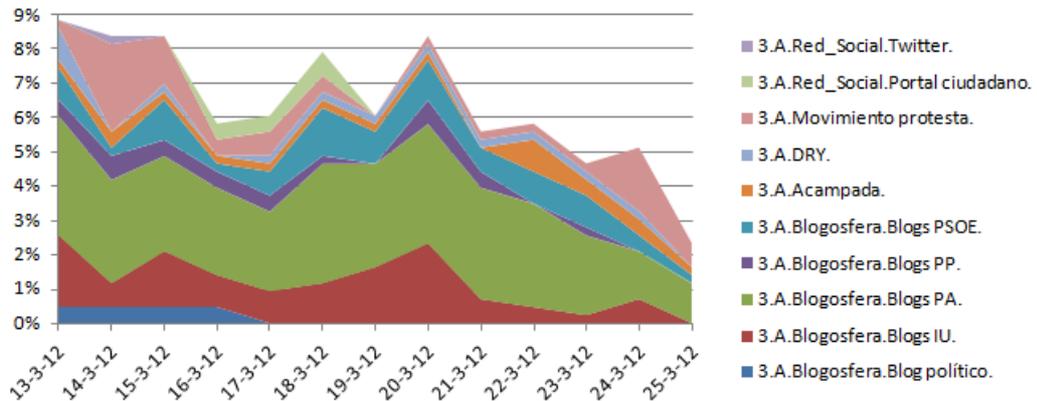


Figura 192.

3.A.Blogosfera política.

- **Blog político.** El promedio es del **0,13%**. Concentra su visibilidad al inicio del periodo y desaparece hasta el día posterior a las elecciones.
- **Blogs IU.** El promedio es del **0,96%**. Concentra su visibilidad la primera semana. El día de las elecciones desaparece del Top10.
- **Blogs PA.** El promedio es del **2,59%**. Su visibilidad se concentra la primera semana de captura, alcanzando cotas del 3,5%.
- **Blogs PP.** El promedio es del **0,30%**. Su visibilidad se concentra la primera semana de captura.
- **Blogs PSOE.** El promedio es del **0,75%**. Destaca el día 18 con un 1,4%.

3.A.Movimiento_protesta.

- **Acampada.** El promedio es del **0,33%**. Destaca el día 22 con cerca del 1%.
- **DRY.** El promedio es del **0,22%**. Aparece los días centrales de captura.
- **Movimiento protesta.** El promedio es del **0,71%**. Concentra su visibilidad el inicio del periodo y los días en torno a los comicios.

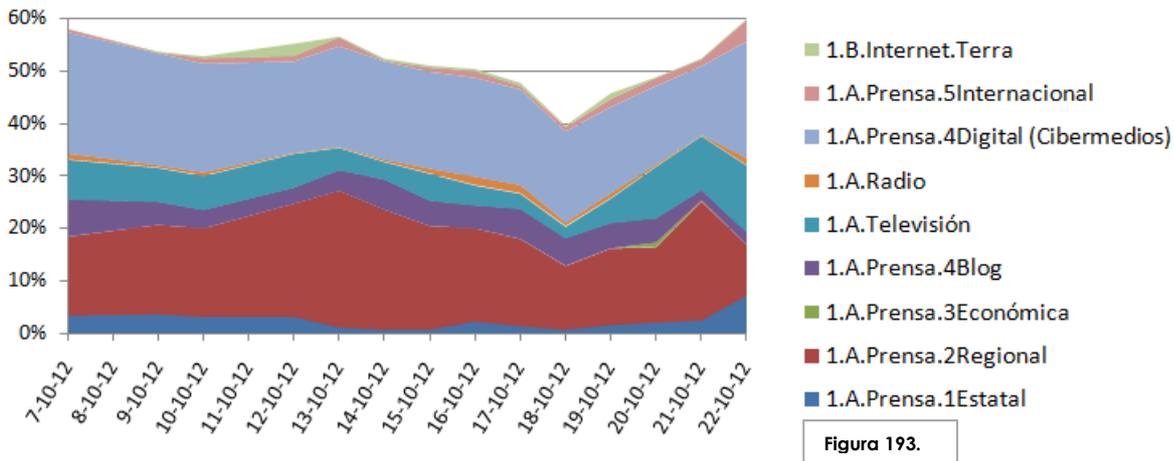
3.A.Red_Social.

- **Portal ciudadano.** El promedio es del **0,12%**. Concentra su visibilidad los días de los comicios.
- **Twitter.** El promedio es residual del **0,02%**.

3.4.5. Xunta de Galicia – 21/10/2012

>Elecciones Xunta 2012> Estrato superior

El promedio de este estrato fue del 51,8%, el mayor registrado en las series temporales de elecciones. A continuación la **serie** de datos de visibilidad y los conglomerados que alcanzaron visibilidad en la primera SERP:



1.A.Mass_Media_Prensa.1Nacional

- **Prensa Estatal.** El promedio es del **2,57%**. Concentra su visibilidad al inicio del periodo.
- **Prensa Regional.** Es el segundo conglomerado y su promedio es del **17,46%**.
- **Prensa Económica.** El promedio de este conglomerado es del 0,08%. Concentra su visibilidad días previos a las elecciones de forma similar a Web Search.
- **Blog de Prensa.** El promedio es del **4,46%**. Concentra su visibilidad la primera semana del periodo con un 7,11% al inicio de la serie.
- **Televisión.** El promedio es del **6,13%**. Concentra su visibilidad al inicio y final de la serie. Su máximo es el día posterior a las elecciones con un 12,67%, cifra superior a la registrada en Web Search.
- **Radio.** El promedio es del **0,76%**. Tiene sus máximos los días centrales de captura y el día de los comicios con un 1,33%.
- **Prensa.4Digital (Cibermedios).** El principal conglomerado y su promedio es del 18,56%. Su cota más alta es el primer día de captura del periodo con un 23,11%. El día posterior a las elecciones alcanza el 22,2%.

1.A.Prensa.5Internacional. El promedio es del 1,25%. Su máximo es el día posterior a las elecciones y logra un 4,22%.

1.B.Internet.Terra. El promedio es del 0,52%. Su máximo es el cuarto día de captura con un 2,44%.

>Elecciones Xunta 2012> Estrato Medio

El promedio de este estrato fue del **21,33%** y los conglomerados objeto de estudio que alcanzaron visibilidad en el Top10 fueron:

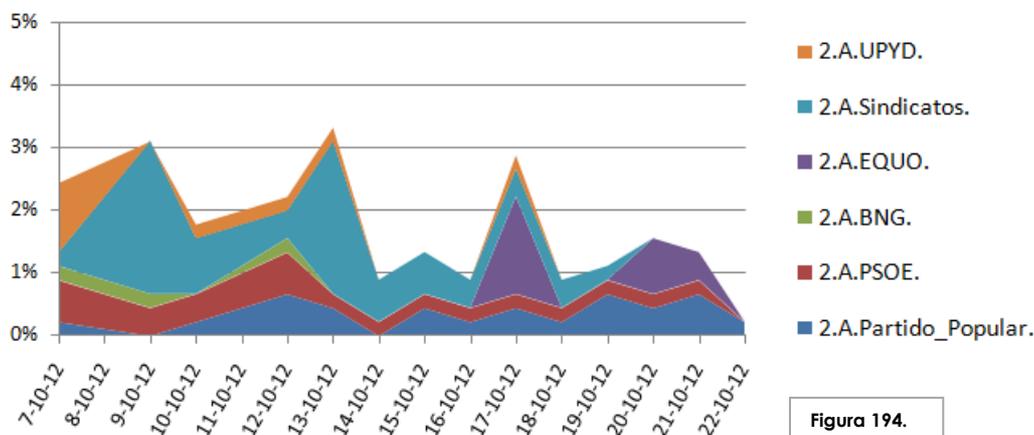


Figura 194.

2.A.Esfera política.

- **Partido_Popular.** El promedio es del 0,35%. Los días de los comicios alcanza el 0,67%.
- **PSOE.** El promedio es del 0,30%. Es más visible los primeros días de captura.
- **BNG.** El promedio es residual del 0,05%. Apenas es visible los primeros días de la serie.
- **EQUO.** El promedio es del 0,21%. Aparece los días centrales y finales de captura.
- **Sindicatos.** El promedio es del 0,67%. Destaca el día 9 con un 2,44%.
- **UPYD.** El promedio es del 0,14%. Su mejor día es el primer día de la serie.

>Elecciones Xunta 2012> Estrato Micro

El peso del Estrato Micro en las elecciones a la Xunta fue el menor de los registrados en toda la serie temporal de elecciones. El promedio de este estrato fue del **26,87%** y los conglomerados objeto de estudio apenas alcanzaron el 0,72%:

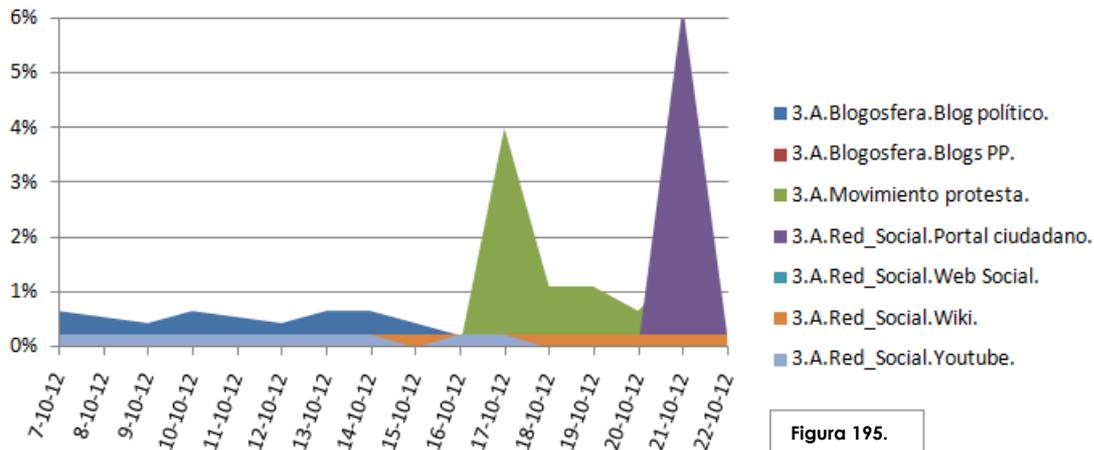


Figura 195.

3.A.Blogosfera política.

- **Blog político.** El promedio es del 0,57%. Su mejor dato es del día 18 con un 1,1%.
- **Blogs PP.** El promedio es residual del 0,03%
- **Movimiento protesta.** El promedio es del 0,67%. Su máximo es del 4% el día 17.

3.A.Red_Social.

- **Portal ciudadano.** El promedio es del 0,46%. Su mejor posición es el día de las elecciones con un 6,22%.

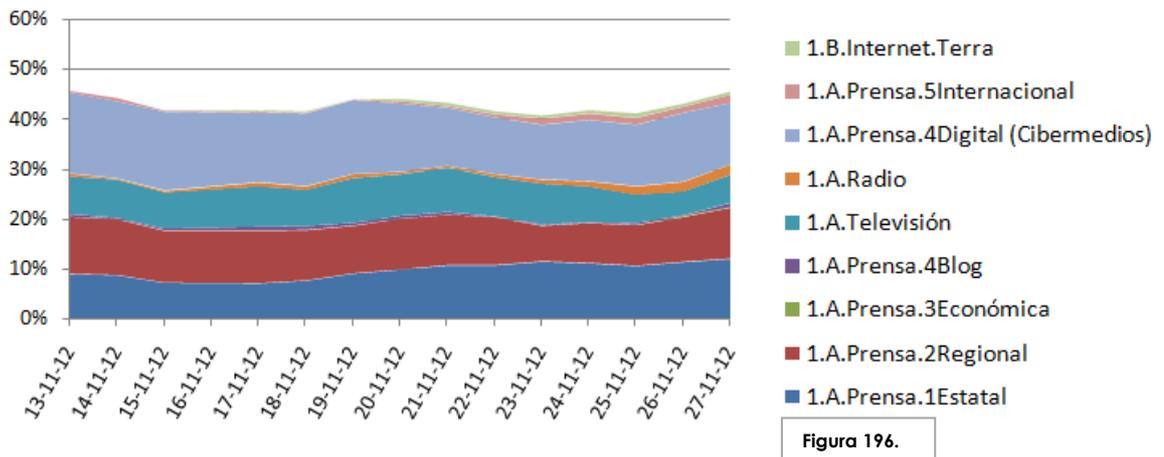
3.A.Red_Social.

- **Web Social.** El promedio es residual del 0,13%
- **Wiki.** El promedio es del 0,21%, estable todo el periodo.
- **Youtube.** El promedio es residual del 0,13%.

3.4.6. Generalitat de Catalunya – 25/11/2012

>Elecciones Generalitat 2012> Estrato superior

El promedio de este estrato fue del 43%. Los datos de visibilidad de sus conglomerados durante las elecciones fueron:



1.A.Mass_Media_Prensa.1Nacional

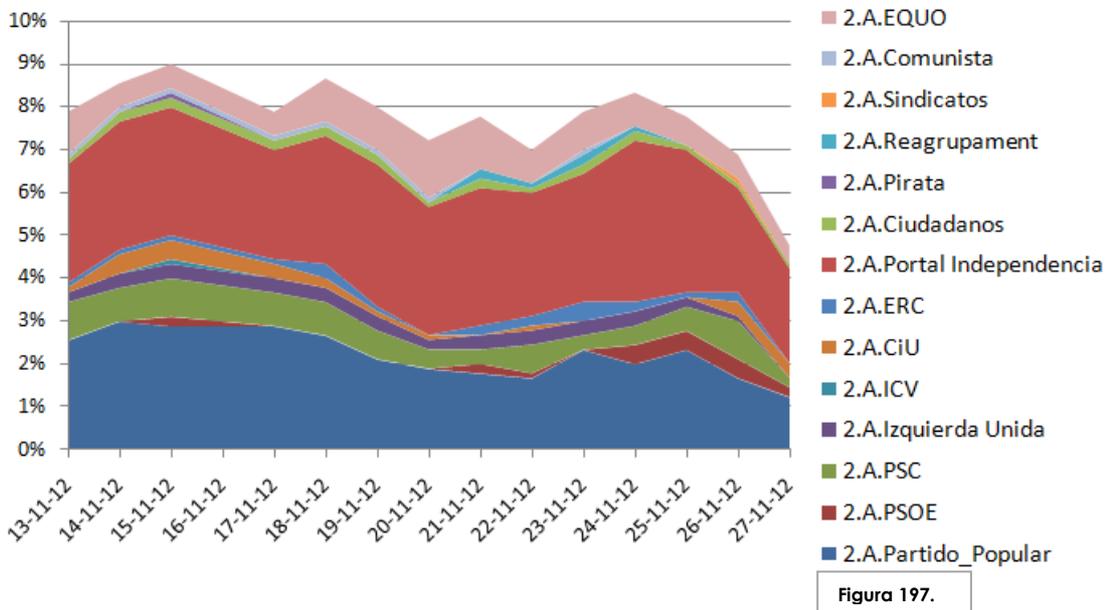
- **Prensa Estatal.** El promedio es del **9,81%**, el más elevado de la serie de elecciones. Sus mejores cifras son la segunda semana con datos del 11,5%.
- **Prensa Regional.** El promedio es del **9,70%**. Concentra mayor visibilidad la primera semana.
- **Prensa Económica.** El promedio es residual del **0,03%**.
- **Blog de Prensa.** El promedio es del **0,44%**.
- **Televisión.** El promedio es del **7,51%**. Concentra mayor visibilidad los días centrales de la muestra.
- **Radio.** El promedio es del **0,86%**. El día de las elecciones alcanza un 2%.
- **Prensa Digital (Cibermedios).** El promedio es del **13,59%**. Concentra mayor visibilidad la primera semana de captura.

1.A.Prensa.5Internacional. El promedio es del **0,69%**. La segunda semana logra un promedio del 1,23%.

1.B.Internet.Terra. El promedio es del **0,39%**. Es visible la segunda parte del periodo.

>Elecciones Generalitat 2012> Estrato Medio

El promedio de este estrato fue del **27,74%** y los conglomerados visibles en el Top10 fueron:

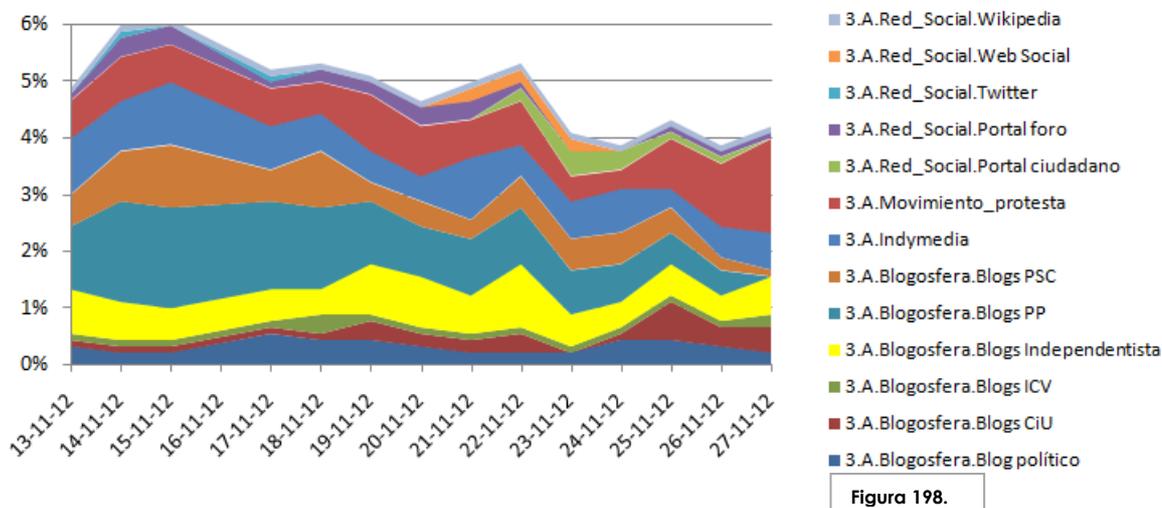


2.A.Esfera política.

- **Partido_Popular.** El promedio es del **2,21%**. Su visibilidad es ligeramente superior la primera semana de datos.
- **PSOE.** El promedio es residual del **0,15%**. Es más visible el día anterior a los comicios.
- **PSC.** El promedio es del **0,62%**. Es más visible la primera semana de datos.
- **Izquierda Unida.** El promedio es del **0,27%**. Desaparece el día de las elecciones.
- **Portal Independencia.** Es el principal conglomerado y su promedio es del **2,97%**. Su visibilidad es constante todo el periodo pero cae el día de los comicios hasta el 2,2%.
- **Resto de partidos.** Las demás agrupaciones políticas destacables son **CiU** con un 0,18%, **ERC** alcanza un 0,17%, **ICV** el 0,01%, **Ciudadans** el 0,17% y **EQUO** el 0,81%. UPYD desaparece del Top10 respecto al año 2010.

>Elecciones Generalitat 2012> Estrato Micro

El promedio de este estrato fue del **29,25%** y los conglomerados que alcanzaron visibilidad en el Top10 fueron:



3.A. Blogosfera.

- **Blog político.** El promedio es del 0,33%. A finales de la primera semana alcanza su máximo en torno al 0,5%.
- **Blogs CiU.** El promedio es del 0,23%. Máximo un 0,67% el día 25.
- **Blogs ICV.** El promedio es residual del 0,13%.
- **Blogs Independentista.** El promedio es del 0,66%. Concentra su visibilidad los días centrales.
- **Blogs PP.** El promedio es del 1,01%. La primera semana logra el 1,78% de visibilidad. Desaparece el día de los comicios.
- **Blogs PSC.** El promedio es del 0,55%. El más visible la primera semana de comicios.

3.A. Movimiento_protesta.

- **Indymedia.** El promedio es del 0,72%. Primera semana
- **Movimiento_protesta.** El promedio es del 0,79%. Día de los comicios con un 1,67%.

3.A. Red_Social.

- **Portal ciudadano.** El promedio es del 0,09%
- **Portal foro.** El promedio es del 0,17%
- **Twitter.** El promedio es residual del 0,02%
- **Web Social.** El promedio es del 0,05%
- **Wikipedia.** El sitio web Wikipedia tuvo resultados en español y catalán que acumularon un promedio es del 0,11%.

3.4.7. Tendencias Google Blog Search

Estrato Superior. La evolución de este estrato dentro Blog Search muestra el incremento de resultados ajenos a la opción de búsqueda propia del Micro Estrato. El estrato de los medios convencionales aumenta su visibilidad en un **32%** durante el periodo de estudio de 2010 a 2012, destacando en las dos elecciones a la Generalitat el incremento del 16% en al 43%, desplazando los contenidos blog y reafirmando el dominio del Top10 por parte de los grupos mediáticos. Por tanto podemos ver que la democratización de visibilidad posibilitada por la búsqueda universal de Google (que facilita la inclusión de sitios web de otras esferas) ha sido acosada por los grandes medios cuya estrategias de posicionamiento y creación de contenidos han logrado contraprogramar los contenidos blog.

Estrato Medio. El estrato mantiene su visibilidad en torno a un 32% durante el periodo de estudio de 2010 a 2012 y su evolución es similar a la detectada en la opción de búsqueda Web Search.

Estrato Micro. Este estrato se ve reducido en más de un 90% durante el periodo de estudio de 2010 a 2012. La evolución general del estrato en el hipotético canal de la blogosfera viene marcada por las dos elecciones de referencia a la Generalitat que muestran una evidente reducción de visibilidad. El conglomerado que más visibilidad pierde es el de 'Red de Blogs', que pasa del 41,6% al 23,86% del total.

Generalitat Catalunya – 28 Noviembre 2010.

Estrato Superior. La visibilidad de los medios convencionales se mantiene estable todo el periodo en torno a un **16%**. El día anterior a las elecciones cae hasta el **11,59%** y remonta el día post-electoral hasta el **17,73%** siguiendo un patrón visto en las otras opciones de búsqueda de Google.

Estrato Medio. Se mantiene estable todo el periodo en torno al 28% pero cae el día posterior a las elecciones a un 20%.

Estrato Micro. Con un promedio del 56%, es más visible los días de la segunda semana de captura y alcanza su máximo a final de la serie con un 63,2%.

Elecciones Municipales-22 Mayo 2011.

Estrato Superior. La visibilidad se mantiene estable todo el periodo en torno a un **27,5%**. El día de las elecciones logra su máximo con un **33,72%**. A diferencia de otras elecciones, los medios de prensa no incrementan su visibilidad el día post-electoral.

Estrato Medio. La visibilidad promedio es del **32,6%**. Pierde visibilidad conforme avanza la serie y logra su mínimo el día de las elecciones con un **27,9%**.

Estrato Micro. La visibilidad se incrementa todo el periodo desde el **32%** a inicio de la serie hasta el **46,5%** el día posterior a las elecciones, cayendo el día de los comicios hasta el 38,3%. Dentro de este estrato destaca el conglomerados **3.C.Red_Blogspot.** con una media del **22,25%** y **3.C.Red_Wordpress.** con un **5,2%**.

Elecciones Generales- 20 Noviembre 2011.

Estrato convencional. El promedio de visibilidad es de un **31,9%**. Su visibilidad máxima la alcanza el séptimo día de captura con un **35,5%** y el día anterior a las elecciones es de **33,33%**. Alcanza mínimos a inicios de la segunda semana de la serie con cifras en torno al 28%.

Estrato Medio. La visibilidad promedio es del **31,1%** y cae al final de la primera semana de la serie. Destaca la visibilidad del conglomerado 'Universidad' con un promedio del 1,6%. El conglomerado 'Administración' tiene una visibilidad residual pero aparece en SERP el día previo a las elecciones.

Estrato Micro. La visibilidad se mantiene estable todo el periodo sobre un **35,25%**. Alcanza sus máximos el primer y último día de la serie con un 38,7% y un 26,9% respectivamente. Destaca el conglomerado **3.C.Red_Blogspot.** con una media del **16%** y **3.C.Red_Wordpress.** con un **7,25%**.

Junta de Andalucía - Marzo 2011.

Estrato convencional. El promedio de visibilidad es de un **25%**. El día de menor que menor visibilidad logra es el tercer día de la serie con un 20% y a partir de ahí su tendencia es creciente hasta el **32,33%** el día posterior a las elecciones.

Estrato Medio. El promedio del estrato es de un **35,6%**. Concentra su mayor visibilidad los primeros días de la serie con un **41,4%** y va disminuyendo en visibilidad progresivamente hasta el final de la serie con un **33,02%**.

Estrato Micro. La visibilidad promedio es del **39,4%** y concentra sus días de más visibilidad al inicio de la segunda semana pero cae los días posteriores a las elecciones a mínimos del 34,65%. Destaca el conglomerado **3.C.Red_Blogspot.** con una media del **16,5%** y **3.C.Red_Wordpress.** con un **8,4%**.

Xunta Galicia – 21 Octubre 2012.

Estrato convencional. El promedio de visibilidad es el mayor registrado en toda la serie de elecciones con un **51,8%**. El día posterior a las elecciones es del **60%**, cifra más alta registrada en Blog Search

Estrato Medio. La visibilidad se mantiene en un promedio del **20,6%**. El día 18 de octubre se alza hasta el **30,4%** pero cae hasta el **12,44%** el día posterior a las elecciones.

Estrato Micro. La visibilidad se tiene un promedio del **27,5%**. Alcanza el máximo el día anterior a las elecciones con un **31,8%**. Destacan los conglomerados **3.C.Red_Blogspot.** con una media del **15,17%** y **3.C.Red_Wordpress.** con un **3,84%**.

Generalitat Catalunya - 25 Noviembre 2012.

Estrato convencional. El promedio de visibilidad es de un **43%**. El día posterior a las elecciones es de **45,8%**.

Estrato Medio. El promedio del estrato es de un **27,74%**. Concentra su mayor visibilidad al final de la primera semana de la serie con cifras en torno al **29,4%** y va disminuyendo en visibilidad progresivamente hasta el final de la serie con un **23,44%**.

Estrato Micro. La visibilidad se mantiene estable todo el periodo sobre un **29,25%**. Alcanza mayores cotas de visibilidad la segunda semana de captura. Destaca el conglomerado **3.C.Red_Blogspot.** con una media del **10,75%** y **3.C.Red_Wordpress.** con un **11,34%**.

4. Conclusiones

4.1 Hipótesis de partida

Para esta sección, revisaremos primero la hipótesis inicial o conjetura de partida y después haremos lo mismo con las preguntas de investigación.

El medio Internet, y más concretamente los resultados obtenidos a través de buscadores, podría suponer un aumento de la visibilidad para los medios digitales considerados alternativos, trayendo esto como consecuencia un aumento de la diversidad en las fuentes de información disponibles para los ciudadanos.

Actualmente, ya nadie pone en duda que Internet se ha convertido en la fuente de información global y los medios impresos están siendo reemplazados por una nueva generación de medios, algunos basados en la Web 2.0 que nació en 2004 y en la revolución social que propició esa nueva tecnología.

En este nuevo contexto, de forma paralela al incremento del uso de Internet como principal medio de comunicación, encontramos que la propia red ha moldeado los hábitos de consumo de información de la generación Google, un grupo de usuarios compuesto por las franjas de edad más jóvenes de la sociedad del conocimiento que, siendo consumidora de los contenidos posicionados en el Top10 de buscadores, está altamente expuesta a los sitios que mayor visibilidad alcanzan en motores de búsqueda, particularmente en Google.

Esta generación pronto se convertirá en población adulta, y cuando eso ocurra en una escala temporal próxima, se verán interactuando con una Internet y una blogosfera que ya han adquirido un rol predominante entre los ciudadanos de la Unión Europea, exponiendo a la sociedad a unas ventajas (en el mejor de los casos) o a unas amenazas que bien podrían dejar de ser un modelo teórico como el expuesto en el estado de la cuestión para convertirse en una realidad de carácter más o menos virulento.

Bajo estas circunstancias, en esta tesis hemos intentado analizar el contexto en el que se han desarrollado los hábitos informativos de esta generación Google, para lo cual, por un lado hemos desarrollado y presentado un sistema de análisis que nos permite

Capítulo 4. Conclusiones

comparar los resultados de buscadores en forma de una propuesta conceptual derivada del modelo teórico de Infoesfera, y por otro lado, bajo esta perspectiva común hemos mostrado los resultados de los seis comicios electorales objeto de estudio.

En paralelo al periodo de investigación, en estos últimos años hemos asistido a la expansión de una nueva forma de acceder a la información web a través de unas redes sociales que han ido creciendo gracias a la implantación de la web móvil y que están en los aledaños de convertirse en el principal medio de acceso a la web en detrimento de los resultados de buscadores, hasta ahora gatekeepers por excelencia del ecosistema Internet.

Siendo la hipótesis de partida de nuestra investigación la posible democratización de la visibilidad gracias al buscador Google, a lo largo de la exposición de resultados de la serie temporal completa de elecciones, hemos podido constatar que durante el periodo de estudio la blogosfera ha perdido peso en las SERPs de buscadores, atenuando, por un lado la fuente principal del peligro apuntado en el estado de la cuestión, pero atenuando a la vez el papel diversificador de la web en cuanto a la oferta informativa a los ciudadanos; un papel que hubiera podido actuar, eventualmente, como una nueva fuerza democratizadora.

4.2 Preguntas de investigación

Hasta aquí el (re) examen de la conjetura inicial. A continuación, presentamos las principales conclusiones siguiendo esta vez el orden de las preguntas de investigación (PI) presentadas en la introducción:

PI 1: ¿Es posible desarrollar el modelo de la Infoesfera enunciado por J. Peretti (2001) y W. L. Bennet (2003), pasando del planteamiento teórico a una plasmación del modelo con datos reales en un contexto electoral?

La respuesta a esta primera pregunta de investigación la encontramos desarrollada en el capítulo dos correspondiente al apartado de sistema de análisis.

A pesar de que el modelo teórico fue apenas esbozado a partir del análisis empírico de un caso particular como fue 'Nike fábrica de explotación', y que en el ámbito académico el modelo no ha sido trabajado por otros investigadores, sí ha sido posible el desarrollo y plasmación del modelo con datos reales en un contexto electoral.

La tarea de desarrollar el modelo fue el primer y gran problema metodológico a la hora de plantearnos cómo trabajar con los datos recuperados de las SERPs. El planteamiento teórico inicial basado en un esquema básico, cuando no superficial, de la composición de Internet del año 2001 fue el punto de partida para emprender nuestro particular camino en la elaboración de un sistema de análisis evolucionado que permitiese englobar toda la web actual. Por suerte el esquema básico inicial basado en el reduccionismo a tres niveles se mantiene a día de hoy, pudiendo reconocer los tres estratos en parte gracias a una definición amplia. Este esquema, por lo tanto, ha permitido dar cabida a las nuevas tipologías web que no existían en el momento de su elaboración y que han podido ser categorizadas según el esquema de Infoesfera.

En el caso electoral, el tamaño finito del principal grupo de páginas objeto de estudio (páginas electorales y de partidos políticos) supuso un estímulo a la hora de empezar a vislumbrar una posible categorización de los sitios web en una ontología con aspiraciones a convertirse en un marco global.

En un estadio inicial del proceso se contó con datos de SERPs en bruto, es decir, sin ningún tipo de categorización, por lo que el primer paso fue desarrollar dicha ontología junto a una metodología de análisis basada en la categorización esbozada en el modelo teórico de Infoesfera. Después de las primeras elecciones a la Generalitat de 2010 contamos ya con una primera versión de la ontología web. El análisis de la primera base de datos de resultados permitió establecer un primer panel de categorías sobre la que trabajar y aplicar la metodología de análisis. Las subsiguientes bases de datos correspondientes a los siguientes procesos electorales ayudaron a ampliar y mejorar la ontología, puliendo de forma paralela el sistema de análisis.

Tras dos años de capturas de datos en el contexto electoral, el cúmulo de URLs y sitios web permitió no solo establecer un mapa preciso de los principales actores del nicho de búsqueda, sino también de todos los resultados colaterales, es decir, de todas las páginas web que en un momento u otro aparecieron reflejadas en SERPs.

El primero de los resultados conseguidos ha sido el desarrollo de este novedoso sistema para analizar los resultados de búsqueda en contexto electoral que permite llevar a cabo las comparaciones longitudinales y transversales entre los diferentes paquetes de datos capturados, por lo que podemos concluir que el desarrollo del modelo de Infoesfera ha sido llevado a cabo con éxito.

PI 2: En tal caso, ¿sería posible aplicar y poner a prueba el modelo evolucionado de la Infoesfera a través de los resultados de los buscadores?

La respuesta a esta segunda pregunta de investigación ha sido plasmada en el capítulo tres correspondiente a los resultados de investigación.

El modelo evolucionado de Infoesfera ha sido llevado a la práctica a través de los resultados de buscadores de los seis comicios electorales seleccionados, permitiendo apreciar diferencias significativas en su evolución a lo largo del tiempo.

Sin embargo, la continua necesidad de mantenimiento y actualización de sitios web debido a la mutabilidad/caducidad de la web hace que la base de datos de sitios web que componen la ontología deba ser actualizada regularmente, dificultando el reaprovechamiento de datos de análisis transcurridos periodos de tiempo largos entre

capturas. Los grandes actores de la Infoesfera es probable que permanezcan inmutables durante largos periodos de tiempo, sin embargo la obsolescencia de los sitios menores y los cambios de composición resultante de actualizaciones de los algoritmos de búsqueda hace que muchos de los jugadores secundarios sean renovados en poco tiempo, requiriendo por tanto una actualización casi constante de los nuevos sitios web detectados.

A pesar de este impedimento temporal, la aplicación del modelo de análisis resultante es posible ejecutarlo a diferentes niveles, partiendo del más global con el análisis de proporciones y evolución de los tres estratos principales, y pudiendo llegar a un análisis pormenorizado de aspectos muy concretos y específicos, tanto sobre la composición de los estratos como de cualquiera de los conglomerados en los que se han categorizado cada uno de los sitios web.

Una vez desarrollado y testeado el modelo ha sido posible aplicarlo a las diferentes opciones de consulta que ofrecen los motores de búsqueda, permitiendo analizar el funcionamiento y composición diferenciado de Web, News y Blog Search. De igual forma, el modelo ha permitido comparar los resultados de los tres principales buscadores comerciales, Google, Bing y Yahoo, en cualquiera de los niveles de análisis derivados de la composición de la Infoesfera.

Una vez vistos los resultados de la investigación, consideramos por tanto que la puesta a prueba del modelo evolucionado a través de los resultados (SERPs) de buscadores ha resultado exitoso, permitiendo visualizar cuál ha sido la composición y evolución de los resultados de buscadores durante el periodo 2010-2012 y medir la Infoesfera en el ámbito temático de los resultados de búsquedas políticas durante unas elecciones.

PI 3: Una vez actualizado el modelo, ¿sería posible extrapolarlo a otros ámbitos temáticos?

Una vez desarrollado el modelo y aplicado en una primera instancia a los resultados electorales podemos constatar que es posible su extrapolación a otros ámbitos de estudio de la web además de los resultados de buscadores.

Aunque los diferentes niveles de categorización han venido marcados por la esfera política como principal objeto de estudio, la ontología desarrollada engloba la

totalidad de sitios web aparecidos en resultados y es posible replicarla siguiendo el mismo esquema en cualquier otra esfera temática.

La categorización del Estrato Superior incluye todos los medios españoles convencionales y digitales, pudiendo ser reaprovechada en cualquier ámbito temático circunscrito dentro del ámbito de la web española. La expansión a otros ámbitos territoriales requiere únicamente de una base de datos de medios actualizada, por lo que es fácilmente extrapolable al estudio de la esfera web de otros países.

La posibilidad de expandir la ontología correspondiente a los estratos Medio y Micro es también factible, permitiendo aplicar el esquema general de coronas y conglomerados para analizar cualquier paquete de datos de sitios web que gravite en torno a cualquier tema central de análisis seleccionado. Para la elaboración de esta tesis se han llevado a cabo análisis en diferentes campos temáticos como han sido los de identidad digital de empresas, búsquedas informacionales relacionadas con el mundo de la abogacía o resultados de novedades editoriales. Estos experimentos preliminares han tenido como consecuencia la categorización de nichos temáticos que aparecen ubicados en las coronas secundarias de la Infoesfera y que han sido agrupados en colectivos. Cualquiera de estos temas es proclive a ser analizado como campo temático central de otros estudios, siendo únicamente necesaria la actualización de sitios web categorizados. La expansión de la ontología para dar cabida a nuevos temas es el único inconveniente, ya que cualquier nuevo estrato medio temático requiere de una nueva base de datos actualizada relativa a dicho ámbito temático, cuyo desarrollo puede requerir de una inversión importante en tiempo y medios dependiendo de la cantidad de sitios relacionados.

La categorización en detalle del estrato Micro muestra las mismas posibilidades e inconvenientes que el estrato Medio, en especial en el caso de la aplicación del modelo general a resultados de la blogosfera dada la cantidad ingente de resultados potencialmente adscritos a un ámbito temático.

Por otro lado, además de la aplicación del modelo general a otros campos temáticos dentro de los resultados de buscadores, existe la posibilidad de utilizar el mismo esquema de análisis en cualquier otro ámbito. La expansión natural del modelo a otros ámbitos es el de las redes sociales. Dada la limitación de recursos para el desarrollo de

esta tesis, el esquema aquí planteado agrupa todos los resultados sociales en conglomerados homogéneos, siendo la tipología de resultados UGC uno grupo más a analizar dentro del comportamiento general de buscadores. En nueva fase de desarrollo metodológico sería posible no solo avanzar en el estudio de los diferentes medios sociales adscribiéndolos a afiliaciones específicas, sino también aplicar la ontología a los diferentes sitios web citados dentro de cada red social.

PI 4: ¿El modelo puede ayudar a identificar factores críticos de éxito en reputación e identidad digital?

Los datos estadísticos permiten identificar un conjunto de páginas web y conglomerados que pueden ayudar a definir la identidad digital de diferentes tipos de búsquedas. La base de datos de resultados ha permitido identificar un amplio abanico de tipos de sitios web, cada uno de ellos asociado a una tipología específica de búsqueda.

Por tanto, si entendemos como identidad digital al conjunto de resultados mostrados por los buscadores, gracias al estudio de la relación existente entre un determinado tipo de búsquedas y los diferentes sitios mostrados como respuesta por los buscadores, en una fase posterior podemos ayudar a construir o modificar la identidad digital de búsquedas de la misma categoría. De esta forma, si por ejemplo nos centramos en la identidad digital de búsquedas navegacionales de tipo personal o corporativo, el modelo de Infoesfera puede ser aplicado para analizar la composición exacta de los conglomerados asociados a este nicho temático, pudiendo construir una identidad digital análoga dentro del mismo nicho inyectando contenidos en los conglomerados detectados con una probabilidad alta de replicación en resultados. En el caso de la identidad digital personal, el conocimiento previo de los sitios web vinculados a este tipo de búsquedas puede servir tanto para replicar una nueva identidad como para modificar resultados de SERPs ya existentes desplazando resultados críticos mediante la creación e inyección de contenidos afines en los conglomerados que definen el nicho.

Con el objetivo de estructurar las estrategias de posicionamiento web a través de los contenidos ya posicionados y optimizar contenidos de nueva creación, en el análisis del estado de la cuestión hemos presentado una lista sistematizada de los 205 factores de posicionamiento conocidos a día de hoy. Esta lista permitirá analizar los sitios web

Capítulo 4. Conclusiones

aparecidos en SERPs vinculadas a cualquier nicho temático, permitiendo estudios comparativos que facilitarán la optimización de nuevos contenidos.

Por tanto concluimos que el modelo de análisis es de gran utilidad en cualquier estrategia SEO ya que puede servir en la creación de nuevas identidades digitales corporativas o personales publicando contenidos en los conglomerados y/o sitios web que hayan alcanzado una visibilidad más alta en capturas previas. Estudios estadísticos de composición de identidad digital pueden servir de referencia en estrategias de posicionamiento planificado para cualquier nicho del *long tail* permitiendo, por ejemplo, adelantarse a necesidades informativas o al comportamiento de usuarios posicionando con antelación contenidos de búsquedas estacionales de cualquier acontecimiento.

4.3 Consideraciones finales y futuros desarrollos

En un futuro próximo nos planteamos la aplicación del modelo de Infoesfera y de la metodología desarrollada en cuatro escenarios:

El primero es una extensión natural del trabajo realizado hasta el momento, reaprovechando la ontología web desarrollada en el ámbito político y aplicándola a la base de datos correspondiente a las elecciones europeas de mayo de 2014. Con este nuevo proceso electoral podremos añadir una nueva serie temporal a las seis ya analizadas, ofreciéndonos una perspectiva que abarque desde el 2010 hasta la actualidad.

Un segundo escenario es aplicar el modelo a un nuevo contexto, en especial al de las redes sociales. La extracción de enlaces citados en medios sociales como Facebook o Twitter permitirá ver la correspondencia de los medios que están en boca de los usuarios en redes sociales con los resultados de Google. La aplicación en el ámbito de redes sociales y comunicación política nos permitirá analizar el discurso mediático de los partidos a través de los contenidos volcados en redes sociales, conocer las fuentes en las que se respaldan y establecer comparativas con otros partidos.

Un tercer escenario donde desarrollar la metodología es el de la identidad digital corporativa, en especial la relacionada con la composición de resultados en buscadores y la evolución de un valor bursátil. Con el objetivo de identificar casos de guerra de información centrada en nichos de búsqueda económicos se aplicará el modelo de análisis a resultados en el Top10 con el objetivo de detectar noticias económicas que puedan estar asociadas a variaciones en el precio de valores bursátiles.

Por último, y a modo de compendio de la tesis, desarrollaremos una aplicación dinámica que permita por un lado el análisis de la composición de resultados de buscadores, y por otro que analice los factores de posicionamiento de cada uno de los resultados de una SERP, cotejando cuáles de los 205 factores expuestos determinan cada posicionamiento particular.

5. Bibliografía

Bibliografía

- Abadal, E.; Codina, L. (2006). *¿Es Google una amenaza para la diversidad cultural?*. Anuario ThinkEPI, El profesional de la información, 2006. p. 94- p.96. Recuperado de <http://dialnet.unirioja.es/descarga/articulo/2555770.pdf>
- ABC . (2013). *Los motivos por los que Twitter ha sido un éxito en su salida a bolsa*. Recuperado de <http://www.abc.es/economia/20131108/abci-motivos-twitter-sido-exito-201311081131.html>
- About.com. (n.d.) *What is Google Caffeine*. Recuperado el 20 de octubre de 2014 <http://webtrends.about.com/od/webportals/a/what-is-google-caffeine.htm>
- Apostatando que es gerundio. (2006). *¿Quién controla los medios?*. Recuperado de <http://cincominutos.com/apostasia/blog/?p=48>
- Attachmedia. (2012). *Infografía - Google Penguin*. Recuperado de <http://blog.attachmedia.com/google-penguin-infografia/>
- Barry Schwartz. (2014). Searchenginewatch. *Study: Google Universal Results Show Up For 85% Of All Searches: Videos In 65% & Maps In 1%*. Recuperado de <http://searchengineland.com/study-google-universal-results-show-85-searches-videos-65-maps-1-194477>
- Barry, C.; Lardner, M. (2011). *A Study of First Click Behaviour and User Interaction on the Google SERP*. Information Systems Development, 2011 - Springer Recuperado de http://link.springer.com/chapter/10.1007/978-1-4419-9790-6_7
- Barysevich, A. (2012). Searchenginejournal. *An SEO's Ultimate Post-Penguin Checklist [INFOGRAPHIC]*. Recuperado de <http://www.searchenginejournal.com/an-seos-ultimate-post-penguin-checklist-infographic/48228>
- Becker, L. B., McCombs, M. E., & McLeod, J. M. (1975). *The Development of Political Cognitions*. En Steven H. Chafee (Ed.), *Political Communication: Issues and Strategies for Research* Beverly Hills, CA:Sage, págs. 21-63.
- Bennett, W.L. (2003). *Communicating Global Activism: Strengths and Vulnerabilities of Networked Politics*. *Information, Communication & Society*. Recuperado de http://ccce.com.washington.edu/projects/assets/working_papers/communicatingglobalactivism.pdf
- Bennett, W.L. (2004). *Political communication, citizenship and democracy: From the mass society to personal information networks*. Presentation slide show distributed by author.
- Björneborn, L. (2004). *Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach* . Ph.D. thesis from the Department of Information Studie. Royal School of Library and Information Science, Denmark. Recuperado de

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.2353&rep=rep1&type=pdf>

Blázquez, M. (2012). *Sistemas de Recuperación en Internet. Webmetría y análisis de páginas web*. Recuperado de <http://ccdoc-sistemasrecuperacioninternet.blogspot.com.es/2012/11/webmetria-y-analisis-de-paginas-web.html>

Blumler, J. G. (2001). *The third age of political communication*. *Journal of publicaffairs*, 1, (3), 201-209.

Brin, S.; Page, L. (1998). *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Stanford University. Recuperado de <http://infolab.stanford.edu/~backrub/google.htm>

Broder, A.; et al. (2000). *Graph structure in the web*. Recuperado de <http://www9.org/w9cdrom/160/160.html>

Burton, G. (2005). *Media and Society. Critical perspectives*. Open University Press. New York.

Cabra, F.; Marciales, G.P. (2009). *Nativos digitales: ¿ocultamiento de factores generadores de fracaso escolar?* Recuperado de <http://www.rieoei.org/rie50a06.pdf>

Cameron, A. (n.d.) Alister Cameron // Blogologist. Recuperado el 20 de octubre de 2014 <http://www.alistercameron.com/>

Casares, J. (2013). *Cómo detecta Bing el spam*. Recuperado de <http://www.durky.com/c-mo-detecta-bing-el-spam/>

Casares, J. (2013). *Google quality rater webspam*. Recuperado de <http://www.durky.com/google-quality-rater-webspam/>

Casares, J. (2013). *Search Ranger AntiSpam de Bing*. Recuperado de <http://www.durky.com/search-ranger-antispam-de-bing/>

Casares, J. (n.d.). *Determinar la credibilidad de un sitio web*. Recuperado el 20 de octubre de 2014 <http://www.durky.com/determinar-la-credibilidad-de-un-sitio-web/>

Chaffee, S.H.; Metzger, M.J.(2001). *The End of Mass Communication?* *Mass Communication & Society*, 2001, 4(4). 365-379.

Charlton, G. (2013). *Google's new 'Hummingbird' search algorithm: the experts' view*. Recuperado de <http://econsultancy.com/es/blog/63486-google-s-new-hummingbird-search-algorithm-the-experts-view>

Chomsky, N.; Herman, E. (1988). *Manufacturing Consent: The Political Economy of the Mass Media*. Pantheon. New York.

Chomsky, N.; Herman, E. (1995). *Los cinco filtros*. Recuperado de http://wiki.medialab-prado.es/images/0/0a/Textos_sobre_propaganda.pdf

Bibliografía

CIBER research team (2008). *Information Behaviour of the Researcher of the Future*. University College London. Recuperado de <http://www.bl.uk/news/pdf/googlegen.pdf>

Cinco Días. (2012). *Las claves del batacazo de Facebook en Bolsa*. Recuperado de http://cincodias.com/cincodias/2012/05/22/mercados/1337902006_850215.html.

Codina, Ll. (2004). *Posicionamiento Web: Conceptos y Ciclo de Vida*. Recuperado de http://www.upf.edu/hipertextnet/numero-2/posicion_web.html

Combes, B. (2007). *The Search for Information and the Net Generation*. International Association of School Librarianship. Selected Papers from the 36th Annual Conference.

Content farm. (n.d.). En Wikipedia. Recuperado el 20 de octubre de 2014 http://en.wikipedia.org/wiki/Content_farm

Crum, Chris. (2010). Webpronews.com. *Google News SEO Tips – Ranking in News Search*. Recuperado el 27 de octubre de 2014 <http://www.webpronews.com/topnews/2009/09/02/google-news-seo-tips-ranking-in-news-search>

Daniel Deceuster.(2012).Moz. *The Penguin Update & How Google Identifies Spam*. Recuperado el 20 de octubre de 2014 <http://moz.com/ugc/the-penguin-update-how-google-identifies-spam>

Danny Sullivan. (2007). Searchenginewatch. *Google Launches “Universal Search” & Blended Results*. Recuperado de <http://searchengineland.com/google-20-google-universal-search-11232>

Danny Sullivan. (2012). Searchenginewatch. *The Google Dance Is Back*. Recuperado de <http://searchengineland.com/google-dance-is-back-134125>

Das, A. (2011). *Google's Global panda update hit UK-Losers & Winners!* Recuperado de <http://anirbann.wordpress.com/2011/04/13/googles-global-panda-update-hit-uk-losers-winners/>

Datacenterknowledge. (2009). *Who Has the Most Web Servers?* Recuperado de <http://www.datacenterknowledge.com/archives/2009/05/14/whos-got-the-most-web-servers/>

Dean, B. (2014). *Google's 200 Ranking Factors: The Complete List*. Recuperado de <http://backlinko.com/google-ranking-factors>

doPDF. (2011). *A Google Panda recovery case study*. Recuperado de <http://www.dopdf.com/forum/topic/dopdf-a-google-panda-recovery-case-study>

Durán, J.; Nieto, S. (2010). *El arte de ganar: Cómo usar el ataque en campañas electorales exitosas*. Editorial Debate. Recuperado de http://books.google.es/books?id=1z82247MTtgC&hl=es&source=gbs_navlinks_s

Bibliografía

EFE. (2011). El mundo.com. *LinkedIn sube más de un 130% en el mejor estreno en Wall Street desde 2000.* Recuperado de <http://www.elmundo.es/elmundo/2011/05/19/navegante/1305789520.html>

El País.(2014). *Turquía levanta el bloqueo de Twitter después de que el Constitucional lo declarara ilegal.* Recuperado de http://internacional.elpais.com/internacional/2014/04/03/actualidad/1396534745_356724.html

Eliance. (2007). *Comparativa del modelo de búsqueda de Universal Search.* Recuperado de <http://www.eliance.com/aha/infographics/google-universal-search.aspx>

Epstein, R.; Robertson, R.E. (2013). *Democracy at Risk: Manipulating Search Rankings Can Shift Voters' Preferences Substantially Without Their Awareness.* American Institute for Behavioral Research and Technology. Recuperado de <http://aibr.org/downloads/EPSTEIN and Robertson 2013-Democracy at Risk-APS-summary-5-13.pdf>

European Parliament. (n.d.) *Jorgo CHATZIMARKAKIS – European Parliament.* Recuperado el 27 de octubre de 2014 http://www.europarl.europa.eu/meps/en/28243/JORGO_CHATZIMARKAKIS_home.html

European Parliament. (n.d.) *Marianne Mikko – European Parliament.* Recuperado el 27 de octubre de 2014 http://www.europarl.europa.eu/meps/es/28421/MARIANNE_MIKKO_home.html

Excite. (2008). *Europa regulará los blogs con un DNI para bloggers.* Recuperado de <http://zonadigital.excite.es/europa-regulara-los-blogs-con-un-dni-para-bloggers-N1651.html>

Financial Times. (2007). *Web censorship spreading globally.* Recuperado de <http://www.ft.com/cms/s/2/1dbb5faa-d268-11db-a7c0-000b5df10621.html>

Firstviewonline. (2011). *Why SEO is Important For Businesses.* Recuperado de <http://firstviewonline.com/why-seo-is-important-for-businesses/>

Garry Przyklen. (2011). Searchenginewatch. *Top 5 Google Panda Update SEO Survival Tips.* Recuperado de <http://searchenginewatch.com/article/2080032/Top-5-Google-Panda-Update-SEO-Survival-Tips>

Genbeta. (2013). *La salida a bolsa de tres grandes empresas tecnológicas: Google, LinkedIn y Facebook.* Recuperado de <http://www.genbeta.com/genbeta/la-salida-a-bolsa-de-tres-grandes-empresas-tecnologicas-google-linkedin-y-facebook>

Goldenberg, G. (2010). *Blog Search Ranking Factors – Google's BlogSearch Algorithm.* Recuperado de <http://seoroi.com/seo-faq/blog-search-ranking-factors-googles-blogsearch-algorithm/>

Bibliografía

Gonzalo, C. (2004). Hipertext.net. *La selección de palabras clave para el posicionamiento en buscadores*. Recuperado de http://www.upf.edu/hipertextnet/numero-2/palabras_clave.html

Gonzalo, C. (2006). *Tipología y análisis de enlaces web: aplicación al estudio de los enlaces fraudulentos y de las granjas de enlaces*. Recuperado de http://www2.ub.edu/bid/consulta_articulos.php?fichero=16gonza2.htm

Gonzalo, C. (2010). *Guerra de medios digitales. Google News, el ABC.es y la cobertura de la búsqueda 'estatut' durante la crisis informativa de la 'dignitat de Catalunya'*. Jornades Catalanes d'informació i documentació. Recuperado de http://www.cobdc.net/12JCD/wp-content/materials/comunicacions/GONZALO_guerra_medios_digitales_dignitat_catalunya.pdf

Google Caffeine. (n.d.). En Wikipedia. Recuperado el 20 de octubre de 2014 http://en.wikipedia.org/wiki/Google_Search#Google_Caffeine

Google Insidesearch. (2012). *Another step to reward high-quality sites*. Recuperado de <http://insidesearch.blogspot.com.es/2012/04/another-step-to-reward-high-quality.html>

Google Insidesearch. (n.d.). *Comparativa de la estructura de índice de Google Caffeine Search*. Recuperado de <http://www.google.com/insidesearch/features/instant/about.html>

Google Instant. (n.d.). En Wikipedia. Recuperado el 20 de octubre de 2014 http://en.wikipedia.org/wiki/Google_Search#Instant_Search

Google webmaster central. (2010). *Our new search index: Caffeine*. Recuperado de <http://googlewebmastercentral.blogspot.com.es/2010/06/our-new-search-index-caffeine.html>

Google webmaster. (2011). *Más recomendaciones para construir sitios de alta calidad*. Recuperado de <http://googlewebmaster-es.blogspot.com/2011/05/mas-recomendaciones-para-construir.html>

Google. (2005). *Systems and methods for improving the ranking of news articles*. US 20050060312 A1. Recuperado el 20 de octubre de 2014 <http://www.google.com/patents/US20050060312>

Google. (2008). *Interleaving Search Results*. US Patent Application 20080140647. Published June 12, 2008. Filed: December 6, 2007. Recuperado el 20 de octubre de 2014 <http://www.google.de/patents/US20080140647>

Google. (2013). *System and method for search engine result ranking*. Recuperado el 20 de octubre de 2014 <http://www.google.de/patents/US20120130814>

Google. (n.d.). *Solicitudes de reconsideración por penalizaciones*. Recuperado de <http://support.google.com/webmasters/bin/answer.py?hl=es&answer=35843>

Bibliografía

Google. (n.d.). *Comparativa del tiempo de búsqueda de Google Caffeine Search*. Recuperado el 20 de octubre de 2014 <http://www.google.com/insidesearch/features/instant/about.html>

Google. (n.d.). *Indexing and retrieval of blogs*. Recuperado el 20 de octubre de 2014 <http://www.google.com/patents/US7765209>

Google. (n.d.). *Ranking blog documents*. Recuperado el 20 de octubre de 2014 <http://www.google.de/patents/US8244720>

Google. (n.d.). *Ayuda de Editor de Google Noticias*. Recuperado de <https://support.google.com/news/publisher/?hl=es#topic=4359865>

Google. (n.d.). *Sitios incluidos en Google Noticias*. Recuperado de <https://support.google.com/news/publisher/answer/40249?hl=es&topic=2511925&ctx=topic>.

Guinsberg, E. (2005). *Control de los medios, control del hombre. Medios masivos y formación psicosocial*. Plaza y Valdés Editores.

Harry, D. (2010). *The fire horse trail. Google Blog Search; under the hood*. Recuperado de <http://www.huomah.com/Search-Engines/Algorithm-Matters/Google-Blog-Search-under-the-hood.html>

Hidalgo, F.J. (2011). *wwwwhat's new. Google vs. granja de contenidos [infografía]*. Recuperado de <http://wwwwhatsnew.com/2011/03/20/google-vs-granja-de-contenidos-infografia/>

Höchstötter, N; Lewandowski, D. (2009). *What Users See – Structures in Search Engine Results Pages*. Information Sciences 179, 1796–1812. Recuperado de <http://core.kmi.open.ac.uk/download/pdf/11888238.pdf>

Huerta, I. (2011). *Qué son y qué no son Granjas de Contenido*. Recuperado de <http://blog.ikhuerta.com/que-son-y-que-no-son-granjas-de-contenido>

Ibáñez Micó, J. (2011). *Extracción de datos web usando técnicas de screen-scraping*. Tesis: Universidad Politécnica de Valencia. Recuperado de http://riunet.upv.es/bitstream/handle/10251/10104/PFC_Javier_Iba%C3%83%C2%B1ez.pdf?sequence=1

Ingeniería social (seguridad informática). (n.d.). En Wikipedia. Recuperado el 20 de octubre de 2014 [http://es.wikipedia.org/wiki/Ingenier%C3%ADa_social_\(seguridad_inform%C3%A1tica\)](http://es.wikipedia.org/wiki/Ingenier%C3%ADa_social_(seguridad_inform%C3%A1tica))

Jansen, B.; et al. (2008). *Determining the informational, navigational, and transactional intent of Web queries*. Information Processing & management. Recuperado de http://faculty.ist.psu.edu/jjansen/academic/pubs/jansen_user_intent.pdf

Bibliografía

Joachims, T. et al (2005) *Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search*. Cornell University. ACM Transactions on Information Systems, Vol. 25, No. 2. Recuperado de http://www.cs.cornell.edu/people/tj/publications/joachims_etal_07a.pdf

Jonah Peretti. (n.d.). En Wikipedia. Recuperado el 20 de octubre de 2014 http://en.wikipedia.org/wiki/Jonah_Peretti

Jongales. (2011). *List of content farms*. Recuperado de <http://www.jongales.com/blog/2011/02/14/list-of-content-farms/>

Kinniburgh, J.B. (2006) *Blogs and Military Information Strategy*. Information Operations Sphere. Joint Special Operations University. Recuperado de http://www.au.af.mil/info-ops/iosphere/iosphere_summer06_kinniburgh.pdf

Knowledge Graph. (n.d.). En Wikipedia. Recuperado el 20 de octubre de 2014 http://es.wikipedia.org/wiki/Knowledge_Graph

Koopmans, R.; Zimmermann, A. (2003). *Internet: A New Potential for European Political Communication?*. ISSN 1612-1635, Berlin.

La mercadotecnia viral. (n.d.). En Wikipedia. Recuperado el 20 de octubre de 2014 http://es.wikipedia.org/wiki/Mercadotecnia_viral

La Vanguardia. (2013). *Hackers' chinos han accedido al armamento secreto del Pentágono*. Recuperado de <http://www.lavanguardia.com/internet/20130528/54374921591/hackers-chinos-han-accedido-al-armamento-secreto-del-pentagono.html>

Lee, K. (2005). PRWeb. *Did-it, Enquiro, and Eyetools Uncover Google's Golden Triangle*. Recuperado de <http://www.prweb.com/releases/2005/03/prweb213516.htm>

Lewandowski, D. (2008). *Search engine user behaviour: How can users be guided to quality content?*. Information Services & Use 28 (2008) 261–268. 261. DOI 10.3233/ISU-2008-0583. IOS Press.

List_of_Google_hoaxes_and_easter_eggs. (n.d.). En Wikipedia. Recuperado el 20 de octubre de 2014 http://en.wikipedia.org/wiki/List_of_Google_hoaxes_and_easter_eggs#Search_Engine

López García, G. (2006). *Comunicación digital y líneas de fractura en el paradigma de la agenda setting*. Doxa comunicación nº4.

López-Escobar, E. (1996). *Agenda-setting: investigaciones sobre el primero y el segundo nivel*. Communication & Society 9(1 y 2), 9-15. Recuperado de http://www.unav.es/fcom/communication-society/es/articulo.php?art_id=286

Maarek, P. J. (2009). *Marketing político y comunicación*. Editorial Paidós. Barcelona.

Bibliografía

Mahour, B; Tiwari, A. (2014). *A Ranking Algorithm for News Data Streams*. International Journal of Computer Applications (0975 – 8887). Volume 93, No 16. Recuperado de <http://research.ijcaonline.org/volume94/number6/pxc3895683.pdf>

Marketing Directo. (2010). *El 52,2% de los internautas españoles leen blogs*. Recuperado de <http://www.marketingdirecto.com/actualidad/digital/el-522-de-los-internautas-espanoles-leen-blogs/>

Mason, I. (2011). *Panda 2.6 Recovery Case Study #1 - SEHabitat*. Recuperado de <http://sehabitat.com/blog/panda-2-6-recovery-case-study-1/>

McCombs, M. (2005). *A Look at Agenda-setting: past, present and future*. Journalism Studies, Volume 6; Number 4; 2005, pp. 543-557.

Megan Boler (2008). *Digital media and democracy: tactics in hard times*. Massachusetts Institute of Technology. Recuperado de http://static.ow.ly/docs/9657-digital_media_and_democracy_7fM.pdf

Microsiervos. (2013). *Educación y nativos digitales: @blogoff desmontando mitos*. Recuperado de <http://www.microsiervos.com/archivo/mundoreal/educacion-y-nativos-digitales-desmontando-mitos.html>

Mikko, M. (2008). *DRAFT REPORT on concentration and pluralism in the media in the European Union*. Committee on Culture and Education. Recuperado de <http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A6-2008-0303&language=EN#title1>

Miller, M. (2012). Searchenginewatch. *10 Keys to Ranking on Google & Bing During Breaking News Events*. Recuperado de <http://searchenginewatch.com/article/2159184/10-Keys-to-Ranking-on-Google-Bing-During-Breaking-News-Events-Study>

Miranda Miller. (2012). Searchenginewatch. *Search Engines Are Winning the War on Content Farms [STUDY]*. Recuperado de <http://searchenginewatch.com/article/2135047/Search-Engines-Are-Winning-the-War-on-Content-Farms-STUDY>

Modelo de propaganda. (n.d.). En Wikipedia. Recuperado el 20 de octubre de 2014 http://es.wikipedia.org/wiki/Modelo_de_propaganda

Netcraft (2014). *June 2014 Web Server Survey*. Recuperado de <http://news.netcraft.com/archives/category/web-server-survey/>

Nicholas, D. et al. (2003). *Digital Information Consumers, Players and Purchasers: Information Seeking Behaviour in the New Digital Interactive Environment*. Aslib Proceedings, vol. 55, n.º 1/2.

Bibliografía

- Odei. (2013). *Evolución de los algoritmos de Google durante el periodo 2009-2013*. Recuperado de <http://blog.odei.es/google-hummingbird-nuevo-algoritmo-motor-busqueda/>
- Orozco, L.A.;Chavarro, D.A. (2006). *De historia y sociología de la ciencia a indicadores y redes sociales*. Observatorio Colombiano de Ciencia y Tecnología OCyT. Recuperado de <http://es.scribd.com/doc/235946073/17310117-VVAA-de-Historia-y-Sociologia-de-La-Ciencia-a-Indicadores-y-Redes-Sociales-2006-1#scribd>
- Page, J. (2014). Semrush. *How to survive an angry Panda Update*. Recuperado de <http://www.semrush.com/blog/publications/how-to-survive-an-angry-panda-update/>
- Peretti, J. (2003). *Culture Jamming, Memes, Social Networks, and the Emerging Media Ecology The "Nike Sweatshop Email" as Object-To-Think-With*. University of Washington. Recuperado de <http://depts.washington.edu/ccce/polcommcampaigns/peretti.html>
- Peter J. Meyers. (2013).Moz. *Mega-SERP: A Visual Guide to Google*. Recuperado de <http://moz.com/blog/mega-serp-a-visual-guide-to-google>
- Puro Marketing. (n.d.). *Las redes sociales se consolidan como vía de acceso a las noticias*. Recuperado de <http://www.puromarketing.com/16/22400/redes-sociales-consolidan-como-via-acceso-noticias.html>
- Quondos. (2014). *Lista con 100 factores clave que encuentras en posts virales*. Recuperado de <http://www.quondos.com/los-100-factores-para-crear-un-post-viral/>
- Quondos. (2014). *Los 200 factores que Google tiene en cuenta para posicionar tu página*. Recuperado de <http://www.quondos.com/los-200-factores-que-google-tiene-en-cuenta-para-posicionar-tu-pagina>
- Quora. (n.d.) *Information Retrieval: What is inverted index?* Recuperado el 15 de septiembre de 2014 <http://www.quora.com/Information-Retrieval/What-is-inverted-index>
- Reuters Institute. (2014). *Digital News Report 2014*. Recuperado de <http://www.digitalnewsreport.org/>
- Reuters Institute. (2014). *Reuters Institute Digital News Report*. Recuperado de <http://www.unav.edu/documents/3786985/0/Reuters-Institute-Digital-News-Report-2014-Espana.pdf>
- Ritzer, G. (1996). *La McDonalización de la sociedad. Un análisis de la racionalización en la vida cotidiana*. Editorial Ariel. Barcelona.
- Rodríguez, E. (2011). *Panda en perspectiva: qué busca Google con esta actualización - Internet Advantage*. Recuperado de <http://www.internetadvantage.es/blog/seo/panda-en-perspectiva-que-busca-google-con-esta-actualizacion/>

Bibliografía

Schwartz, W. (1995). *Information Warfare: Chaos on the Electronic Superhighway*. Thunder's Mouth Press, NY.

Search Engine Land Infographics. (2013). Searchenginewatch. *Infographic: An Anatomy Of Google's Dynamic Search Results Page*. Recuperado de <http://searchengineland.com/infographic-google-search-results-174190>

Searchmetrics. (2011). *6 tips to make things work with Google News*. Recuperado de <http://blog.searchmetrics.com/us/2011/07/20/6-tips-to-make-things-work-with-google-news/>

Searchmetrics. (2012). *Google Ranking Factors UK 2012*. Recuperado de <http://www.searchmetrics.com/en/white-paper/google-ranking-factors-uk-2012/>

SEOMoz. (2012). *Video Penguin de SEOMoz*. Recuperado de <https://plus.google.com/u/1/+SEOMoz/posts/hHaXg8Rs5Lf>

SEOMoz. (n.d.). *Google Algorithm Change History | SEOMoz*. Recuperado el 25 de octubre de 2014 <http://www.seomoz.org/google-algorithm-change>

Shachtman, N. (2009). *Wired. New U.S. Foreign Policy: Building Social Networks*. Recuperado de <http://www.wired.com/2009/12/new-us-foreign-policy-building-social-networks/>

Shaw, D. L. et al. (2001). *Agenda Setting Theory and Public Opinion Studies in a Post-Mass Media Age*. WAPOR Annual Conference. Roma.

Shey.net. (2001). *Personalize, NIKE iD. email correspondence*. Recuperado de <http://www.shey.net/niked.html>

Sky News. (2008). *Hong Kong Bank Mobbed By Savers*. Recuperado de <http://news.sky.com/story/635509/hong-kong-bank-mobbed-by-savers>

Slawski, B. (2008). *Seobythesea. HOW GOOGLE UNIVERSAL SEARCH AND BLENDED RESULTS MAY WORK*. Recuperado de <http://www.seobythesea.com/2008/06/how-google-universal-search-and-blended-results-may-work/>

Slawski, B. (2013). *Seobythesea. The Google Hummingbird Patent?*. Recuperado de <http://www.seobythesea.com/2013/09/google-hummingbird-patent/>

Slawski, B. (n.d.). *SEO By The Sea*. Recuperado el 20 de octubre de 2014 <http://www.seobythesea.com/about-seo-by-the-sea/>

Takeshita, T. (2005). *Current critical problems in Agenda-setting research*. *International Journal of Public Opinion Research*, Volume 18; Number 3.

Top-rankin.com. (2011). *Google mueve ficha: Larry Page, Panda Update y +1*. Recuperado de <http://www.top-rankin.com/google-se-mueve-larry-page-panda-update/>

Bibliografía

- Travers Scout, D. (2008). *Tempests of the Blogosphere: Presidential Campaign Stories that Failed to Ignite Mainstream Media*. in Boler, M.: *Digital Media and Democracy: Tactics in Hard Times* (271-300) MIT Press. Recuperado de <http://web.mit.edu/comm-forum/mit4/papers/scott.pdf>
- Tricot, C. (2013). ¿Qué hacen Panda Penguin y Hummingbird? Qué deben hacer las empresas. Recuperado de <http://objetivoseo.wordpress.com/2013/11/29/que-hacen-panda-penguin-y-hummingbird-que-deben-hacer-las-empresas/>
- Valdés, L. (2012). *Análisis de Internet 2.0 y la influencia de la agenda-setting*. Recuperado de <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/19881/3/lgvaldesmTFM0113memoria.pdf>
- Whitney, D. C., & Becker, L. B. (1982). *Keeping the gates for gatekeepers: the effects of wire news*. *Journalism Quarterly*, 59(1), págs. 60-65.
- Wordstream. (2010). *Web Spam: The Definitive Guide for SEOs*. Recuperado de <http://www.wordstream.com/blog/ws/2010/04/28/web-spam-guide>
- Worldwidewebsite.com (n.d.) *The size of the World Wide Web (The Internet)*. Recuperado de <http://www.worldwidewebsite.com>
- Xavier Peytibi. (2010). *El arte de la guerra (en las redes sociales)*. Recuperado de <http://www.xavierpeytibi.com/2010/01/20/el-arte-de-la-guerra-en-las-redes-sociales/>
- Yahoo (n.d.) *Yahoo Finanzas*. Recuperado el 20 de octubre de 2014 <https://es.finance.yahoo.com/>
- Young, R.D. (2011). *Searchenginewatch. Google News: Key Ranking Factors [Study]*. Rob D. Young, September 26, 2011. Recuperado de <http://searchenginewatch.com/article/2112126/Google-News-Key-Ranking-Factors-Study>
- Youtube. (2009). *News Search SEO*. Recuperado de http://www.youtube.com/watch?feature=player_embedded&v=hg8xgoULIIE

Glosario de términos

Agenda: temas abordados por gobiernos e instituciones. Las agendas objetivas de los distintos estados expresan las prioridades en función de intereses y amenazas comunes. Sin embargo, en los hechos cada país tiene su propia agenda, que es determinada por las urgencias de sus gobiernos. Por razones de popularidad electoral u otros motivos domésticos, a menudo se elevan a la cabeza del quehacer estatal asuntos de importancia secundaria.

Agenda-setting ([o teoría del establecimiento periodístico de temas](#)): La teoría del establecimiento periodístico de temas de discusión, también conocido por el anglicismo teoría de la agenda-setting postula que los medios de comunicación de masas tienen una gran influencia sobre el público al determinar qué historias poseen interés informativo y cuánto espacio e importancia se les da. El punto central de esta teoría es la capacidad de los medios de comunicación para graduar la importancia de la información que se va a difundir, dándole un orden de prioridad para obtener mayor audiencia, mayor impacto y una determinada conciencia sobre la noticia. Del mismo modo, deciden qué temas excluir de la agenda. Más claramente, la teoría del "establecimiento de la agenda" dice que la agenda mediática, conformada por las noticias que difunden los medios informativos cotidianamente y a las que confieren mayor o menor relevancia, influye en la agenda del público.

Agregador: lector de feeds. Software *online* u *offline* que recibe de forma automática los posts de los blogs previamente suscritos gracias al RSS

Android: plataforma de código abierto para dispositivos móviles que está basada en el sistema operativo Linux.

Astroturfing: término utilizado en campañas de relaciones públicas en el ámbito de la propaganda electoral y los anuncios comerciales que pretende dar una impresión de espontaneidad, fruto de un comportamiento con base social. El objetivo de una campaña de este estilo es disfrazar las acciones de una entidad política o comercial como la reacción pública espontánea e independiente frente a otra entidad, producto, servicio, etc.

Audioblog o podcast: creación de archivos de sonido, generalmente en formato mp3 y su distribución mediante un archivo de manera que permita suscribirse y usar un programa que lo descargue para que el usuario lo escuche en el momento que quiera, generalmente en un reproductor portátil.

Beers&Blogs: término que se usa para determinar un encuentro en el mundo real de autores y participantes en weblogs. Se trata de reuniones informales que sirven para poder conocer cara a cara a personas con las que comparte la afición por Internet y los blogs. Suelen convocarse en cafeterías o restaurantes y de ahí beers (cervezas en inglés), que se junta formando un término similar a rock & roll. A los encuentros suelen acudir autores, así como también personas que habitualmente comentan en las bitácoras implicadas, o simplemente lectores de algunas de las bitácoras de los autores, sin excluir amistades y relaciones de diferente tipo.

Blogólogo: autor de un discurso tecnocultural sobre los blogs, sus técnicas, su taxonomía, sus expresiones, las causas de su éxito o fracaso, sus impactos, sus relaciones con la evolución tecnológica y con las diversas formas sociales (periodismo, educación, comunicación corporativa o política, literatura, psicología, etc.).

Blogosfera: universo de los weblogs en la Red, así como la cultura que generan los blogueros.

Blogger: autor de un *weblog*. También se usa bloguero.

Blogroll: repertorio de enlaces organizados por categorías o temas. Constituyen los favoritos, o sitios amigos del propietario del blog.

Blogware: Software que se utiliza para la gestión y mantenimiento de un blog.

C3I (C4I): el C3 es la fórmula con que los militares abrevian Comando (mando en castellano), Control y Comunicaciones. La I es por la inteligencia. Hoy se habla de C4; la cuarta C agregada corresponde a Computación.

CMS: *Content Manager System*, herramienta software que permite al usuario gestionar dinámicamente y de forma sencilla, los elementos que componen un sitio web. Desde la creación de páginas, redacción, diseño, archivos a permisos.

Contrainformación:

Contrainformación alternativa: La contrainformación alternativa o simplemente contrainformación es la información transmitida por grupos independientes no vinculados, en principio, a un poder político o económico ofreciendo información independiente acerca de los problemas del mundo, en contraposición a los medios de comunicación creados bajo grupos empresariales. Este tipo de información suele hacer énfasis en las reivindicaciones de diferentes movimientos sociales o de distintas ideologías.

El término contrainformación es utilizado, principalmente, en medios alternativos de izquierdas. Sin embargo, es polémico, ya que muchos consideran que los medios de "contrainformación" son los verdaderos medios informativos, pues no manipulan sus contenidos por condicionantes externos, y deberían ser medios de comunicación convencionales los que recibieran la denominación de "Falsimedia", término que se utiliza habitualmente para referirse a ellos.

Motivación. El objetivo de este tipo de contrainformación es ofrecer un punto de vista alternativo para contrarrestar las informaciones sesgadas que se han difundido por otros medios. Su existencia se fundamenta en que la información emitida por medios convencionales es de interés para colectivos importantes (gobiernos, anunciantes, grupos de poder, grupos de presión, etc.) que, desde su punto de vista, resaltan las noticias para no generar pérdidas económicas y de confianza y así dar a la población más claridad sobre sus denuncias y notas periodísticas.

Su principal herramienta de difusión, aunque no exclusiva, es Internet. En ocasiones los grupos editores trabajan a través de un servidor autogestionado.

Conversación: fenómeno que consiste en la publicación de una serie de artículos o comentarios entre diferentes blogs, que abordan un mismo tema y que se enlazan entre sí, criticándose, haciendo referencias, etc. El conjunto de artículos así enlazados da lugar a lo que se denomina una conversación. Para los *blogger*, la blogosfera es una “gran conversación”.

Copyleft: forma de licencia y puede ser usado para modificar el derecho de autor de obras o trabajos, tales como software, documentos, música, y obras de arte. Los contenidos de un blog se suelen ofrecer gratis a la comunidad, pero con reserva de ciertos derechos, como los de cita o uso no comercial.

Creative Commons: en español, bienes comunes creativos. Organización no gubernamental sin ánimo de lucro que desarrolla planes para ayudar a reducir las barreras legales de la creatividad, por medio de nueva legislación y nuevas tecnologías. También el nombre dado a las licencias desarrolladas por esta organización.

Dailymotion: servicio de hospedaje de vídeos en Internet, con sede en París.

Delicious: sitio web en el que, previo registro, y a gracias a unos botones que se añaden al navegador, permite almacenar las páginas favoritas y tenerlas disponibles desde cualquier ordenador.

Desinformación: entrega de información falsa para desorientar a un adversario. Es una estrategia muy antigua, cuya meta práctica es que el destinatario de la desinformación no pueda distinguir la realidad de la ficción. Los servicios de inteligencia son usuarios frecuentes de las técnicas de desinformación. Existen dos tipos básicos de operaciones: las ofensivas, que buscan confundir al enemigo, y las defensivas, que se dirigen a la fuerza propia para mantener la moral.

Edublog: blog orientado al aprendizaje y a la educación que implementa recursos *online* para las comunidades de estudiantes, alumnos y profesores.

Egosurfing: práctica que consiste en buscar en la red, en bases de datos, medios escritos u otros documentos menciones del propio nombre o de los demás.

Facebook: sitio web formado por muchas redes sociales relacionadas con una escuela, universidad, trabajo, región, etc.

Feed: fuente web o canal web. medio de redifusión de contenido web. Se utiliza para suministrar información actualizada frecuentemente a sus suscriptores.

Filtro social: se aplica un filtro social, cuando se emplean herramientas en donde un grupo de usuarios vota noticias, ideas y demás propuestas por otros, para así

seleccionar las más votadas o valoradas. Dos de los sitios más conocidos son *Digg* y *Menéame*.

Flickr: sitio web que permite almacenar, ordenar, buscar y compartir fotografías y videos en línea.

Flog: blog falso o escrito con una identidad falsa. También llamado *fake blog*.

Follower: seguidor, usuario o cuenta dentro de la plataforma Twitter que, por cualquier razón, desea seguirnos.

Folksonomía: método de clasificación colaborativa de contenidos normalmente a través de *tags* o etiquetas.

Foursquare: aplicación basada en la geolocalización web aplicada a las redes sociales. La geolocalización permite localizar un dispositivo fijo o móvil en una ubicación geográfica.

Gardening: borrado manual de comentarios de un blog que no respetan las normas de netiqueta: son absurdos, obscenos, contienen insultos graves, etc.

Geolocalización: posicionamiento con el que se define la localización de un objeto espacial (representado mediante punto, vector, área, volumen) en un sistema de coordenadas y datum determinado. Este proceso es utilizado frecuentemente en los Sistemas de Información Geográfica.

Geek: Anglicismo para calificar a una persona que comparte una gran fascinación, quizás obsesiva, por la tecnología. Es más un estilo de vida y una forma de ser que una afición concreta por las nuevas tecnologías. La "cultura geek" está también íntimamente relacionada con la ciencia ficción la fantasía épica o los videojuegos, comics, etc.

Glocalización: término que nace de la mezcla entre globalización y localización, y que consiste en la globalización de la información local.

Google News: opción de búsqueda de noticias de Google. El portal de Google News España cerró a partir del 16 de diciembre de 2014.

Google Blogsearch: opción de búsqueda de blogs de Google.

Google bombing: actividad consistente en forzar el posicionamiento de una página en los resultados de Google utilizando criterios externos a la página, en concreto, utilizando el pagerank. El procedimiento básico del google bombing consiste en enlazar repetidamente un mismo sitio web usando una determinada palabra clave como anclaje del enlace (es decir, como parte de la etiqueta).

Google Buzz: servicio que permite a los usuarios el uso de herramientas como el intercambio de hipervínculos, fotos, videos y actualizaciones por medio de *Gmail*.

Granja de enlaces ([link farms](#)): sitio web que contiene miles de enlaces salientes con el objetivo de incrementar el PR de las páginas objetivo.

Granja de contenido ([content farms](#)): sitio web que emplea a escritores freelance para generar grandes cantidades de contenido de texto diseñado para satisfacer los algoritmos de los motores de búsqueda. Pueden albergar [millones de artículos](#).

Guerra asimétrica: se llama así al choque en que se miden contrincantes de pesos y medidas muy dispares, por contraste, la guerra simétrica es la que opone a fuerzas equivalentes.

Guerra psicológica: el conjunto de actividades de propaganda destinado a influir en la opinión pública tanto propia como extranjera. Las operaciones psicológicas tienen un blanco preciso: "las mentes y los corazones" de combatientes y civiles. Esta dimensión de la guerra adquirió rigurosidad científica durante la Segunda Guerra Mundial, cuando el ejército de los EE.UU. acuñó la siguiente definición: "La guerra psicológica es la coordinación y el uso de todos los medios (...) que tienden a destruir la voluntad del enemigo por alcanzar la victoria ya a dañar su capacidad económica y política de obtenerla; a privar al enemigo de apoyo, asistencia o simpatía por parte de aliados, asociados o neutrales".

H.264: nuevo estándar de codificación que permite la compresión de video a bajos *bitrates* consiguiendo la mejor calidad de imagen de video. El estándar H.264 es el formato de codificación de video digital mas innovador para la distribución de video de alta definición a través de Internet. La tecnología del H.264 cuenta permite disfrutar de la mejor calidad de video con una pequeña velocidad de bits.

Herramientas sociales: ver software social.

HTML5: (HyperText Markup Language, versión 5) es la quinta revisión importante del lenguaje básico de la World Wide Web, el HTML. Establece una serie de nuevos elementos y atributos que reflejan el uso típico de los sitios web modernos.

Inteligencia colectiva: sistema que intenta unir la experiencia de un grupo en lugar de limitarse a la de un único individuo, de forma que el conocimiento se genera a partir de la colaboración.

Java FX: familia de productos y tecnologías de Sun Microsystems, adquirida por Oracle Corporation, para la creación de Rich Internet Applications (RIAs), esto es, aplicaciones web que tienen las características y capacidades de aplicaciones de escritorio, incluyendo aplicaciones multimedia interactivas.

Lector de feeds: ver agregador.

Lifestream: agrupa el flujo de información ofrecido por una persona a través de las redes sociales en las que participa (por ejemplo, MySpace, Facebook o Twitter).

Marcadores sociales: forma de almacenar, clasificar y compartir enlaces en Internet o en una Intranet.

Marketing viral: estrategia de marketing que consiste en incentivar, de alguna forma, a la gente a que hable y difunda un producto, una empresa o una idea, de manera espontánea y adquiera en automático la validez y credibilidad que los foros publicitarios tradicionales no gozan.

Mashup: aplicación web que utiliza información de diversas fuentes relevantes para crear un nuevo servicio en base a ellas. La mayor parte de los *mashup* están basados en los mapas de Google y emplean programas conocidos como APIs para insertar más fácilmente los contenidos de otras páginas.

Medios alternativos: los [medios alternativos](#) son los medios de comunicación (periódicos, radio, televisión, revistas, películas, Internet, etc) que ofrecen información alternativa a los medios de comunicación convencionales. A inicios de la década del 2000 los medios alternativos no eran todavía accesibles a la mayoría del público, pero fueron ganando visibilidad a medida que la sociedad de la información se afianzaba en España. En la actualidad los medios alternativos cuentan con un creciente número de seguidores en su comunidad y más allá de sus límites geográficos gracias a internet.

Medio alternativo de información: Medios y redes informativas como Indymedia, Diagonal, IskraDigital, Radio Klara, Rebelión, Tercera Información, Periodismo Humano, Insurgente, Periodistas 21, Dos Manzanas, Nodo50, La Haine, Kaosenlared, Taringa, Punto de Vista y Propuesta son significativos [medios alternativos](#).

Medios sociales: medios de comunicación social donde la información y en general el contenido es creado por los propios usuarios mediante el uso de las nuevas tecnologías, que permiten un fácil uso y acceso mediante poderosas tecnologías de edición, publicación e intercambio.

Meme: unidad teórica de información cultural transmisible de un individuo a otro, o de una mente a otra, o de una generación a la siguiente. En relación con la blogosfera, son conversaciones distribuidas que se van contagiando de un blog a otro, de manera que el blog que origina el *meme* enlazado desde muchos otros blogs, recibe muchas visitas, dependiendo del éxito del mensaje. Se podría decir que los *memes* son campañas de márketing viral.

Menéame: sitio web basado en la participación comunitaria en el que los usuarios registrados envían historias que los demás usuarios del sitio (registrados o no) pueden votar, promoviendo las más votadas a la página principal.

Metadatos: Datos sobre datos. En el contexto del posicionamiento web son datos incluidos en la propia página web y que ayudan a caracterizar el sitio debido a que los motores de búsqueda las utilizan (en parte) para determinar su relevancia. Los metadatos adoptan, al menos estas cinco formas en el seno de páginas web: etiqueta, etiquetas en la sección, atributos title en etiquetas de anclaje, atributos title en etiquetas de imágenes, atributos alt en etiquetas de imágenes. Por tanto, a efectos

de posicionamiento web, también se consideran metadatos (o tienen un efecto similar) etiquetas y atributos distintos de las etiquetas tipo.

Microblogging: también conocido como *nanoblogging*, es un servicio que permite a sus usuarios enviar y publicar mensajes breves (alrededor de 140 caracteres), generalmente de sólo texto.

Microsoft Silverlight: estructura para aplicaciones web que agrega nuevas funciones multimedia como la reproducción de vídeos, gráficos vectoriales, animaciones e interactividad, en forma similar a lo que hace Adobe Flash.

Moblog: blog pensado para ser visto en un móvil.

Nanomedio: publicaciones en Internet con formato blog y temática acotada que se publican con ánimo de lucro. También denominados nanopublicaciones (del inglés *nanopublishing*).

Nativo digital: aquel que nació cuando ya existía la tecnología digital. La tecnología digital comenzó a desarrollarse con fuerza en 1978, por lo tanto, se considera que los que nacieron después de 1979 y tuvieron a su alcance en el hogar, establecimientos de estudio y de recreación computadoras o celulares pueden considerarse Nativos Digitales.

Netiqueta: palabra derivada del francés *étiquette* (buena educación) y del inglés *net* (red). Se refiere a las reglas o normas de comportamiento en Internet, cuando usamos el correo electrónico, o realizamos cualquier aportación en un sitio web (blogs, foros de discusión, etc.). En este sentido, la netiqueta adapta a las nuevas tecnologías las reglas de etiqueta que existen en el mundo real.

Operaciones psicológicas: actividades dirigidas a audiencias enemigas, amistosas o neutrales que buscan influir en las percepciones y el comportamiento, afectando el cumplimiento de objetivos políticos y militares. Tanto en tiempos de paz como de guerra, frente al enemigo y las audiencias propias los bandos buscan influir en el estado de ánimo de la población, ya sea para elevar la moral de los suyos o para minar la de los adversarios.

PageRank: algoritmo utilizado para asignar de forma numérica la relevancia de las páginas web indexados en el motor de búsqueda Google. Una página web tiene mayor pagerank cuantos más enlaces recibe por parte de páginas webs de todo tipo y de cualquier punto del ciberespacio y que, a su vez, cuanto mayor sea el Pagerank de los sitios de origen de los enlaces.

Palabra clave: término respecto al cual se persigue la optimización de una página web. Puede ser una palabra única o un conjunto de palabras.

PdR (Página de resultados): página resultante de una consulta en un motor de búsqueda compuesta por enlaces orgánicos y patrocinados. En inglés **SERP**, de las siglas Search Engine Result Pages.

Permalink: enlace permanente, *URL*, dirección de Internet propia que tiene cada entrada.

Ping: notificación automática o manual de la actualización de un blog o de la inclusión de un enlace hacia otro (ver *trackback*).

Podcasting: práctica que consiste en la elaboración de grabaciones sonoras, a la manera de programas radiofónicos, que son distribuidos como archivos de sonido digital usando el formato RSS. Su uso permite a los usuarios suscribirse y usar un programa que descarga estas grabaciones automáticamente en el ordenador, para poder transferirlas después, generalmente a un reproductor portátil. La palabra *podcasting* es un acrónimo de las palabras iPod y *broadcasting*.

Popularidad: puede significar o bien tráfico (o sea número de visitas y de páginas vistas en una web) o bien número de enlaces que recibe una página o un sitio web. La tendencia mayoritaria consiste en reservar el término popularidad para expresar el tráfico, y reservar el término visibilidad para expresar el número de enlaces que recibe una web.

Rastreador (crawler), spider (araña) o robot: es un programa que explora la Web de forma sistemática con dos objetivos principales, 1) Interactuar con los servidores de sitios web para descargar páginas web (*robots.txt*) u otros documentos, y 2) Obtener nuevas direcciones (*URL*) para añadir a su lista de enlaces pendientes de revisar.

Pagerank: algoritmo basado en popularidad de enlaces creado por Larry Page, fundador de Google

Post: entrada o anotación que se hace en un blog. Puede ser texto, imagen, etc.

Propaganda: cualquier forma e comunicación diseñada para influir en las opiniones, emociones o comportamientos de un grupo, ya sea de manera directa o indirecta.

RAW: formato de archivo digital de imágenes que contiene la totalidad de los datos de la imagen tal y como ha sido captada por el sensor digital de la cámara fotográfica.

Redes sociales: espacio de diálogo y coordinación, a través del cual se vinculan personas u organizaciones en función de un objetivo común y sobre la base de normas y valores compartidos. Las redes sociales han permitido generar relaciones de colaboración, poner en común recursos, desarrollar actividades en beneficio de los participantes, ampliar y estrechar vínculos, crear sentido de pertenencia y socializar conocimientos, experiencias y saberes, al establecer relaciones de intercambio y reciprocidad.

Relevancia: capacidad de satisfacer una necesidad de información que presenta una página un sitio web. Se dice que un recurso es muy relevante si es muy útil para solucionar una necesidad de información. El cálculo de relevancia de cada motor de búsqueda combina diversas medidas. Se sabe que Google utiliza más de 200 factores de ranking a la hora de generar los resultados de búsqueda.

Referrer logs: mecanismo mediante el cual podemos conocer de donde proceden los visitantes de un blog o una página web. **RSS:** *Real Simple Syndication*, tecnología que permite la distribución automática online para enviar las actualizaciones de un blog a los agregadores y lectores de blogs (como *Bloglines*, *Google Reader*, *Newsgator...*) y a otros blogs. Cuando los lectores se suscriben a un *RSS feed*, reciben automáticamente las actualizaciones que se hacen en el momento que se producen.

Sandbox: una sección de pruebas temporal de Google donde se analizan páginas web recién indexadas.

SEO (Search Engine Optimization):

SEO Black Hat: conjunto de técnicas de posicionamiento web fraudulentas que pueden ser penalizadas por los buscadores.

Scrapping de contenidos: copiar contenidos de un sitio web y reutilizarlos en la creación de otro nuevo. Estas copias idénticas pueden estar mejor optimizadas SEO que los artículos originales.

Smartphone: teléfono inteligente. Término comercial para denominar a un teléfono móvil que ofrece mas funciones que un teléfono común. Casi todos los teléfonos inteligentes son móviles que soportan completamente un cliente de correo electrónico con la funcionalidad completa de un organizador personal. Una característica importante de casi todos los teléfonos inteligentes es que permiten la instalación de programas para incrementar el procesamiento de datos y la conectividad.

Software social: herramientas que basan su existencia en las necesidades o fines de comunicación de las personas y que normalmente forman una comunidad con intereses comunes.

Spam: prácticas destinadas a forzar una alta posición de una página web para una o más palabras clave, sin que tal posición tenga relación con la relevancia real de la página web.

Splog: abreviatura de *spam blog*. Blog sin contenido original, que reproduce contenidos de otros blogs, creado para ganar enlaces.

Tablet PC: ordenador portátil en el que se puede interactuar a través de una pantalla táctil o Multitáctil. El usuario, puede utilizar una pluma *stylus* para trabajar con el ordenador sin necesidad de teclado físico, o ratón.

Tags: etiquetas, palabras que se asignan a un *post* para clasificarlo según la temática y que permiten identificar rápidamente su contenido. Estas etiquetas se suelen colocar conjuntamente en forma de 'nubes' en las que las más empleadas aparecen con mayor tamaño que el resto.

Tecnoinfluenciadores: grupo de personas que utiliza las tecnologías más avanzadas para abrir debates sobre temas de actualidad, crear percepciones sobre tendencias empresariales y hasta pueden transformar los comportamientos de compra de productos de algunas personas, mediante sus recomendaciones y opiniones.

Thread: serie de mensajes que han sido posteados como respuestas a otros en grupos de noticias, foros, etc. Un foro suele tener muchos threads que poseen distintos temas de conversación. Leyendo cada mensaje de un thread en orden, se puede ver cómo se va desarrollando esa discusión.

Trackback: enlace inverso que permite conocer qué enlaces apuntan hacia un determinado *post*. De ese modo avisa a otro *weblog* que se está citando uno de sus *posts*.

TrafficRank: medida del tráfico de un sitio web debido a Alexa (www.alexa.com), empresa que realiza análisis y mediciones sobre tráfico en internet y que proporciona también un directorio de sitios web. El traffic rank de un sitio indica tanto el número de usuarios que visita la web, como el número de páginas vistas en el sitio por los usuarios (en nuestro estudio aportaremos la posición de los cibermedios de prensa digital más destacados como un indicador más para la construcción del ranking).

Trending Topic: son las palabras más significativas que se están usando en Twitter en un momento determinado. Estas palabras conforman un top 10 que aparece en las búsquedas de Twitter, y es una pista sobre lo que en ese momento más se está conversando.

Trol: usuario de Internet que utiliza mecanismos de interactividad (por ejemplo, la sección de comentarios de un blog) para sabotear los debates en curso, escribiendo mensajes incoherentes (*flames*) para enfadar a los participantes.

Twitter: (del inglés gorjear, parlotear, trinar) es una red social y servicio de *microblogging* que permite a sus usuarios enviar y leer micro-entradas de texto de una longitud máxima de 140 caracteres denominados *tweets*.

Retweet: consiste en decir mediante tu usuario de Twitter lo que ha comentado otro usuario, ya que lo consideramos interesante.

URL: secuencia de caracteres, de acuerdo a un formato modélico y estándar, que se usa para nombrar recursos en Internet para su localización o identificación, como por ejemplo documentos textuales, imágenes, videos, presentaciones digitales, etc.

USER-GENERATED CONTENT (UGC): contenido generado por usuarios, término que engloba una amplia gama de contenidos en los medios de comunicación disponibles que han sido creados por diferentes usuarios. Este concepto se acuñó durante el año 2005 gracias a las facilidades y las opciones que la web 2.0 permiten en las plataformas web y que se basan en la libertad y diversidad de información. Estamos hablando pues de todas las tecnologías de medios digitales, tales como bases de datos de preguntas y respuestas, vídeo digital, blogs, foros, podcasts, redes sociales, la fotografía móvil y los wikis.

Videoblog o vlog: blog cuyas entradas son en formato vídeo.

Warlog: tipo de blog especializados en narrar los acontecimientos que se sucedieron durante la Segunda Guerra del Golfo.

Web 3.0: neologismo que se utiliza para describir la evolución del uso y la interacción en la red a través de diferentes caminos. Ello incluye, la transformación de la red en una base de datos, un movimiento hacia hacer los contenidos accesibles por múltiples aplicaciones *non-browser*, el empuje de las tecnologías de inteligencia artificial, la web semántica, la Web Geoespacial, o la Web 3D.

Web Semántica: significado añadido a los datos, en forma de metadatos, de modo que los ordenadores puedan entender mejor la información que existe en la World Wide Web.

Wiki: proviene de la lengua hawaiana y significa rápido. Es la aplicación de servidor que permite que los documentos alojados (las páginas *wiki*) sean escritos de forma colaborativa a través de un navegador, utilizando una notación sencilla para dar formato, crear enlaces, etc. Cuando alguien edita una página de este tipo, sus cambios aparecen inmediatamente en la web.

Wikipedia: enciclopedia libre multilingüe basada en la tecnología wiki. Se escribe de forma colaborativa por voluntarios, permitiendo que la mayoría de los artículos sean modificados por cualquier persona con acceso mediante un navegador web.

World Wide Web: Sistema de documentos de hipertexto enlazados y accesibles a través de Internet creado por Tim-Berners Lee y Robert Cailliau.

Youtube: sitio web en el cual los usuarios pueden subir y compartir vídeos.

