

Per a tu, Gemma.

Sense tu, res seria possible.

*"El treball del mestre no consisteix tant
en ensenyar tot allò que es pot
aprendre, sinó en produir en l'alumne
amor i estima pel coneixement"*

John Locke

Prefaci

Tot i que una tesi no hauria d'estar escrita en primera persona, em permetré la llicència de fer-ho durant aquest Prefaci, doncs és la manera més natural que tinc d'expressar-me.

Segurament, la vocació de ser professor és el punt de partida i el punt final d'aquest document, i tanmateix una de les causes que aquest moment de lliurar-lo s'hagi endarrerit tant. La dedicació i la necessitat que tenia que la recerca d'aquesta tesi estigués alineada amb la meva passió docent ha fet, certament, que els temps es dilatessin. Tot i això, sens dubte, el que m'emporto en tots aquests anys de classe, em fa feliç.

Aquesta tesi partia inicialment de dues premisses que jo mateix m'exigia: en primer lloc calia trobar una relació directa entre la docència i la meva recerca i fer que el resultat d'aquesta tesi pogués tenir aplicació al món universitari; d'altra banda, aquesta tesi també havia de permetre que la tasca d'un professor fos valorada en una dimensió més global: no només permetre veure el professor com una font de coneixements, sinó com un tutor que acompanya el desenvolupament integral de l'alumne.

I això em porta a recordar què vol dir ser professor. M'agrada força la definició que fa Maria Fernanda Ramírez, professora de la Universitat de Guadalajara, a Mèxic, qui defineix el professor (Ramírez, 2009) com "una persona incansable, fervorosa i apassionada pel seu treball, a la recerca sempre de nous recursos per fer la seva tasca més eficaç i motivadora, permetent sobretot que aquesta tasca permeti obrir als nois i noies un món de possibilitats per enriquir la seva cultura. I aquest treball, per humil que sigui, és el que forja les ments i engrandeix els esperits". El professor és doncs aquell que silenciosament, amb el seu esforç i el seu exemple construeix persones amb sentit comú i capacitat d'autocrítica. Quan l'any 1998 vaig *fer* la meva primera classe, perquè certament les classes es *fan* no pas es *donen* (com les abraçades), part d'aquest text m'hagués sonat a algun idioma indesxifrabable. Amb el pas del temps, puc afirmar que aquesta tasca pren una dimensió diferent per si mateixa.

D'altra banda, també puc dir que l'evolució de la tecnologia i dels sistemes d'informació en general ha fet que la tasca del professor hagi anat acostant-se cap a sistemes més moderns i que semblen a primera vista més freds. Però no és cert. L'evolució d'Internet i la capacitat de comunicació dels alumnes i professors ha fet que tot hagi pres un nou vessant més social i alhora més enriquidor.

Ja fa anys, cap a l'any 2000, recordo quan m'incorporava al Grup de Recerca en Sistemes Intel·ligents (GRSI). Poc pensava llavors la influència que la pertinença a aquell grup em marcaria durant els següents anys de recerca i docència a la Universitat. A finals del 2004, la primera pedra d'aquest binomi docència-recerca es va forjar tot descobrint en el Diploma d'Estudis Avançats (Vernet, 2004) el món dels Sistemes Tutors Intel·ligents, sistemes consistents en petites aplicacions que guiaven l'alumne en l'aprenentatge personal. Però, amb el

naixement d'Internet, els sistemes tutors han anat evolucionant i avui en dia formen part, en més o menys mesura, de la majoria dels Sistemes d'Informació dels Centres Educatius.

Així doncs, semblava que els Sistemes Tutors Intel·ligents havien posat en mi una petita llavor per aconseguir aquest nexa que comentàvem anteriorment. A partir d'aquí, només calia seguir treballant i fer créixer les arrels d'aquesta tesi. Ser membre del GRSI em va formar durant molts anys en tècniques d'intel·ligència artificial i, més concretament, en tècniques d'aprenentatge automàtic. L'aplicació d'aquestes tècniques em va permetre descobrir com millorar aquests Sistemes Tutors i com proposar un model d'arquitectura nou enfocat a la futura Web 4.0.

I és que els Sistemes Tutors normalment sempre han funcionat de manera individualitzada, fent recomanacions als alumnes de manera particular. Jo volia anar una mica més enllà i fer que aquesta tutorització¹ estigués lligada a un entorn d'aprenentatge col·laboratiu. La relació estreta amb els membres del Grup de Recerca en Technology Enhanced Learning (GRETEL) em va donar el coneixement i les oportunitats per trobar aquest escenari.

Finalment, la desaparició del GRSI i el naixement del Grup de Recerca en Internet Technologies and Storage (GRITS), del qual en sóc membre des del 2013, em va acompanyar en la darrera part de la tesi, tot proposant la introducció de noves tècniques pensades per Entorns Virtuals d'Aprenentatge i fent ús de nova tecnologia i noves aproximacions al món d'Internet.

I aquest és un petit resum del camí que m'ha portat fins aquí. No sé si això tanca un cercle o n'obre de nous però sí que crec que les premisses que comentava a l'inici s'estan complint. I això em dóna la satisfacció de poder tancar aquesta etapa i enorgullir-me un cop més de la meua vocació.

¹ El terme “tutorització” va ser aprovat fa pocs anys pel Consell Supervisor del TERMCAT atès que aquest substantiu derivat s'ha estès àmpliament, deixant de banda les formes oficials “tutoració” i “tutoratge”.

Resum

L'evolució de la tecnologia ha produït canvis profunds en els paradigmes de l'ensenyament i, particularment, en l'aplicació d'aquests a l'aprenentatge en línia (*e-learning*). De fet va ser la pròpia revolució tecnològica la que va fer néixer aquest nou model d'aprenentatge virtual i, actualment, poques són les institucions que no compten amb alguna aplicació de l'*e-learning*, ja sigui com a alternativa al model educatiu tradicional o com a complement (*blended learning*).

La introducció de l'*e-learning*, i en general de les Tecnologies de la Informació i Comunicació (TIC), al món educatiu ha fet que la teoria *instructivista* de l'educació tradicional s'hagi desplaçat cap a un paradigma *constructivista*, generant un model molt més centrat en l'alumne. Les eines educatives han anat evolucionant cap aquest nou paradigma, on la personalització i l'adaptació són fils conductors, i els Sistemes Tutors Intel·ligents (STI) en són un bon exemple. Tanmateix, l'arribada de la Web 2.0 ha desencadenat un moment social que ha acabat marcant de nou el món educatiu. El desplegament de la teoria *connectivista*, sorgida de l'aplicació de la Web Social en l'àmbit educatiu, i la implantació de múltiples iniciatives d'*e-learning* han afavorit la proliferació d'Entorns Virtuals d'Aprenentatge (EVA) i de diferents tecnologies educatives basades en Web.

Atès que la tecnologia associada a Internet està en constant evolució, però, tot fa pensar que els entorns d'aprenentatge hauran d'evolucionar en els propers anys de manera paral·lela a com ho està fent la pròpia Web. Així, és probable que les següents generacions d'*e-learning* implementin característiques pròpies de la Web 3.0 (semàntica) i de la Web 4.0 (simbiòtica) i esdevinguin entorns on els agents intel·ligents hi tinguin un paper significatiu.

En aquesta tesi s'analitza en primer lloc quina ha estat la trajectòria que ha seguit l'educació al llarg de la història i quina influència ha tingut en la implantació dels sistemes d'aprenentatge en línia, des dels més senzills i poc adaptatius, fins als més moderns i pensats per millorar l'experiència en l'aprenentatge. A més, en vistes de la trajectòria tecnològica que es divisa, es proposa una nova arquitectura que permeti incloure, d'una banda, les capacitats dels entorns ja existents d'aprenentatge en línia, i, de l'altra, els agents intel·ligents que convertiran l'experiència de l'ensenyament a distància en una experiència adaptativa i *social*, on el concepte de *grup* tindrà cabdal importància.

Els sistemes educatius intel·ligents futurs, per tant, hauran de disposar d'una part complexa de computació avançada, aspecte abordat des del camp de la Intel·ligència Artificial, que permeti reconèixer quina és l'evolució de l'alumne en el seu aprenentatge i com aquest està interactuant i rendint amb els companys de la seva classe virtual. A més, la quantitat d'interaccions produïdes en aquests entorns generarà un gran volum de dades educatives, la *Big Learning Data*, amb informació vital que caldrà processar per millorar i adaptar el sistema a l'alumne a mesura que el curs avança, i per recollir informació valuosa per a la seva tutorització. Així, la darrera part d'aquesta tesi mostra les contribucions realitzades en Intel·ligència Artificial i els resultats de la seva implementació per crear la part intel·ligent d'aquesta arquitectura, podent

extreure d'aquesta manera el màxim rendiment d'aquests nous entorns d'aprenentatge col·laboratiu que seran realitat d'aquí a pocs anys.

Resumen

La evolución de la tecnología ha producido cambios profundos en los paradigmas de la enseñanza y, particularmente, en la aplicación de éstos en el aprendizaje en línea (*e-learning*). De hecho fue la propia revolución tecnológica la que hizo nacer este nuevo modelo de aprendizaje virtual y, actualmente, pocas son las instituciones que no cuentan con alguna aplicación del *e-learning*, ya sea como alternativa al modelo educativo tradicional o como complemento (*blended learning*).

La introducción del *e-learning*, y en general de las Tecnologías de la Información y Comunicación (TIC), en el mundo educativo ha hecho que la teoría *instructivista* de la educación tradicional se haya desplazado hacia un paradigma *constructivista*, generando un modelo mucho más centrado en el alumno. Las herramientas educativas han ido evolucionando hacia este nuevo paradigma, donde la personalización y la adaptación son hilos conductores, y los Sistemas Tutores Inteligentes (STI) son un buen ejemplo. Sin embargo, la llegada de la Web 2.0 ha desencadenado un momento social que ha marcado de nuevo el mundo educativo. El despliegue de la teoría *conectivista*, surgida de la aplicación de la Web Social en el ámbito educativo, y la implantación de varias iniciativas de *e-learning* han favorecido la proliferación de entornos virtuales de aprendizaje y de diferentes tecnologías educativas basadas en Web.

Dado que la tecnología asociada a Internet está en constante evolución, todo hace pensar que los entornos de aprendizaje deberán evolucionar en los próximos años de manera paralela a como lo está haciendo la propia Web. Así, es probable que las siguientes generaciones de *e-learning* implementen características propias de la Web 3.0 (semántica) y de la Web 4.0 (simbiótica) y se conviertan en entornos donde los agentes inteligentes tengan un papel significativo.

En esta tesis se analiza en primer lugar cuál ha sido la trayectoria que ha seguido la educación a lo largo de la historia y qué influencia ha tenido en la implantación del *e-learning*, desde los más sencillos y poco adaptativos, hasta los más modernos y pensados para mejorar la experiencia en el aprendizaje. Además, en vistas de la trayectoria tecnológica que se divisa, se propone una nueva arquitectura que permita incluir, por un lado, las capacidades de los entornos ya existentes de aprendizaje en línea, y, por otro, los agentes inteligentes que convertirán la experiencia de la enseñanza a distancia en una experiencia adaptativa y *social*, donde el concepto de *grupo* tendrá capital importancia.

Los sistemas educativos inteligentes futuros, por tanto, deberán disponer de una parte compleja de computación avanzada, aspecto abordado desde el campo de la Inteligencia Artificial, que permita reconocer cuál es la evolución del alumno en su aprendizaje y como éste está interactuando y rindiendo con los compañeros de su clase virtual. Además, la cantidad de interacciones producidas en estos entornos generará un gran volumen de datos educativos, la *Big Learning Data*, con información vital que habrá que procesar para mejorar y adaptar el sistema al alumno a medida que el curso avanza, y para recoger información valiosa para su

tutorización. Así, la última parte de esta tesis muestra las contribuciones realizadas en Inteligencia Artificial y los resultados de su implementación para crear la parte inteligente de esta arquitectura, pudiendo extraer de este modo el máximo rendimiento de estos nuevos entornos de aprendizaje colaborativo que serán realidad dentro de pocos años.

Summary

The evolution of technology has produced profound changes in the paradigms of teaching and, particularly, in their application in online learning (e-learning). In fact it was the technological revolution itself that gave birth to this new model of virtual learning and there are currently few institutions that do not have an e-learning application, either as an alternative to traditional methods or to complement them (blended learning).

The introduction of e-learning, and in general of the Information Technology and Communication (ICT) in the educational world has made the instructivist traditional education theory move to a constructivist paradigm, creating a more focused learning model. Educational tools have evolved towards this new paradigm, where customization and adaptation are the backbone of the model. Intelligent Tutoring Systems (ITS) provide a good example of this new methodology. However, the advent of Web 2.0 has created a social era which has rebranded the educational world. The deployment of the connectionist theory, arising from the implementation of the Social Web in education, and the implementation of various e-learning initiatives have led to the proliferation of virtual learning environments and different educational Web-based technologies.

Since the technology associated with the Internet is constantly evolving, everything suggests that learning environments should evolve in the coming years in parallel with the Web itself. Thus it is likely that the next generation of e-learning implements own Web 3.0 (semantic) and Web 4.0 (symbiotic) characteristics and create environments where intelligent agents have a significant role.

In this thesis we first analyze the path of education throughout history and discuss the influence it has had on the implementation of e-learning, from the simplest and less adaptive measures to the most modern, designed methods to enhance the learning experience. Furthermore, in view of the visible technological background, we propose a new architecture to include, on the one hand, the capabilities of existing online learning environments, and secondly, intelligent agents which can convert the experiences acquired in distance learning into an adaptive and social experience, where the group concept is of paramount importance.

Future intelligent educational systems must therefore have an intricate part of advanced computing, an aspect from the field of Artificial Intelligence, which recognizes the evolution of students in their learning and how they interact and perform with their virtual class mates. In addition, the number of interactions produced in these environments will generate a large volume of educational data, the Big Data Learning with vital information that must be processed to improve and adapt the system to the student as the course progresses, and to collect valuable information for tutorship. So, the last part of this thesis shows the contributions made in Artificial Intelligence and the results of their implementation to create the intelligent part of this architecture. The benefits of these new collaborative learning environments will enable us to optimize performance in coming years.

Contingut

PREFACI.....	VII
RESUM.....	IX
RESUMEN	XI
SUMMARY	XIII
CAPÍTOL 1 INTRODUCCIÓ I OBJECTIUS	1
1.1 Motivació.....	2
1.2 Punts de partida.....	5
1.2.1 Els Sistemes Tutors Intel·ligents.....	5
1.2.2 El Raonament Basat en Casos.....	7
1.3 Objectius de la tesi	8
1.4 Metodologia de recerca	11
1.5 Publicacions obtingudes.....	12
1.6 Organització de la memòria	14
CAPÍTOL 2 L'EDUCACIÓ A LA SOCIETAT DE LA INFORMACIÓ	17
2.1 Antecedents: Una història marcada per la tecnologia.....	18
2.2 El canvi de paradigma del món educatiu.....	19
2.2.1 Un nou model pedagògic	20
2.3 L'e-learning: aprendre a la societat de la informació.....	23
2.3.1 Modalitats d'e-learning	23
2.3.2 Entorns virtuals d'aprenentatge	24
2.4 Evolució i tendència de l'e-learning: des de la generació 1.0 a la 4.0.....	26
2.4.1 La creació d'un nou escenari educatiu: generacions 1.0, 1.3 i 2.0	26
2.4.2 El futur de l'e-learning: generacions 3.0 i 4.0.....	28
2.5 Resum i conclusions.....	32
CAPÍTOL 3 ANALÍTICA DE L'APRENENTATGE: EL TRACTAMENT DE LA <i>BIG LEARNING DATA</i>...35	35
3.1 <i>Big Data</i> en l'educació	36
3.1.1 Big Learning Data.....	38
3.1.2 Analítica en l'educació	40
3.2 L'analítica de l'aprenentatge	42
3.2.1 El procés d'anàlisi de l'aprenentatge.....	42
3.2.2 Model de referència de l'analítica de l'aprenentatge.....	44
3.2.2.1 Dades educatives (el què)	45
3.2.2.2 Interessats (el qui)	45
3.2.2.3 Objectius de la LA (el per què).....	46
3.2.2.4 Mètodes d'anàlisi (el com).....	47

3.3	Resum i conclusions.....	50
CAPÍTOL 4 SISTEMES EDUCATIUS INTEL·LIGENTS		51
4.1	Antecedents i evolució dels sistemes educatius intel·ligents	52
4.2	Sistemes tutors intel·ligents	54
4.2.1	Estructura d'un Sistema Tutor	55
4.2.1.1	Model de l'estudiant.....	56
4.2.1.2	Model pedagògic	57
4.2.1.3	Model de comunicació	58
4.2.1.4	Coneixement del domini.....	58
4.2.1.5	Model expert	59
4.2.2	Funcionament d'un STI.....	59
4.3	Sistemes educatius intel·ligents i adaptatius basats en web.....	60
4.3.1	Tecnologies clàssiques aplicades als AIWBES.....	61
4.3.2	Tecnologies modernes aplicades als AIWBES	62
4.4	Resum i conclusions.....	64
CAPÍTOL 5 TICVA: TUTORITZACIÓ INTEL·LIGENT DE COMUNITATS VIRTUALS D'APRENTATGE		65
5.1	Conceptualització del sistema TICVA	66
5.2	Un model que respon a preguntes	68
5.2.1	Què? El coneixement i les dades.....	69
5.2.2	Per què? La motivació i els objectius.....	70
5.2.3	Com? Processament i anàlisi	71
5.2.4	Qui? Els actors.....	72
5.3	Estructura del sistema TICVA	73
5.3.1	Mòduls del sistema	73
5.3.1.1	Mòdul de Comunicació o interfície del sistema.....	75
5.3.1.2	Coneixement del Domini.....	75
5.3.1.3	Mòdul expert	78
5.3.1.4	Filtered Reporting.....	79
5.3.1.5	Mòdul de l'Estudiant o Grup	80
5.3.1.6	Mòdul pedagògic.....	82
5.4	Un apunt de la tecnologia web	83
5.5	Resum i conclusions.....	85
CAPÍTOL 6 CONTRIBUCIONS EN ALGORISMES INTEL·LIGENTS		87
6.1	L'aprenentatge automàtic	88
6.2	Els problemes de classificació.....	89
6.2.1	Classificació mitjançant arbres de decisió.....	90
6.2.2	Classificació mitjançant regles.....	91
6.3	Metodologia de proves i dominis utilitzats.....	91
6.4	El Raonament Basat en Casos	94
6.4.1	La Memòria de Casos	95
6.4.2	Fase de recuperació.....	96

6.4.3	Fase d'adaptació	99
6.4.4	Fase de revisió	99
6.4.5	Fase d'emmagatzematge	100
6.5	Modificacions i aportacions al cicle del CBR	100
6.5.1	Diferents organitzacions de la Memòria de Casos	101
6.5.2	Predicció de l'evolució individual d'un alumne.....	102
6.5.2.1	Tècniques utilitzades per a la predicció	102
6.5.2.2	Resultats obtinguts en la predicció	105
6.5.3	Nova organització dels casos	108
6.5.3.1	Algorisme K-Means	110
6.5.3.2	Algorisme X-Means	112
6.5.3.3	Algorisme WITT	112
6.5.3.4	Algorisme NormWITT	117
6.5.3.5	Aplicació dels algorismes d'agrupament per crear grups.....	117
6.5.4	Nous mètodes de recuperació	120
6.5.4.1	Recuperació en una memòria clusteritzada.....	121
6.5.4.2	Ponderació d'atributs.....	123
6.5.4.3	Algorismes genètics	124
6.5.5	Aportacions a la fase d'emmagatzematge	125
6.5.6	Resultats obtinguts del sistema complet.....	127
6.6	Resum i conclusions.....	132
CAPÍTOL 7 CONCLUSIONS I LÍNIES DE FUTUR		135
7.1	El naixement de l' <i>e-learning</i>	136
7.2	De l'Analítica de l'Aprenentatge a la <i>Big Learning Data</i>	136
7.3	Els Sistemes Tutors Intel·ligents	137
7.4	El sistema TICVA.....	138
7.5	La predicció i la tutorització.....	140
7.6	Objectius de publicacions i resum d'aportacions	142
REFERÈNCIES		145
APÈNDIX A: ÍNDEX DE FIGURES		159
APÈNDIX B: ÍNDEX DE TAULES		163
APÈNDIX C: ACRÒNIMS UTILITZATS.....		165
APÈNDIX D: DETALL DE L'ARQUITECTURA TICVA.....		167
ANNEX: PUBLICACIONS MÉS RELLEVANTS.....		169

Capítol 1

Introducció i Objectius

“The secret of getting ahead is getting started. The secret of getting started is breaking your complex overwhelming tasks into small manageable tasks, and starting on the first one.”

Mark Twain

Segurament l'evolució d'Internet i les Comunicacions en general, tant pel que es refereix a les instal·lacions físiques com a la velocitat de servei, han permès que la manera de plantejar moltes coses a la vida hagi canviat. D'una manera paral·lela, la docència i l'aprenentatge no han quedat al marge d'aquesta revolució i han viscut de manera molt pronunciada les novetats tecnològiques d'aquests darrers temps. I és que difícilment avui en dia es pot pensar en un entorn d'aprenentatge on la tecnologia no hi intervingui d'alguna manera, ja sigui com a complement o desenvolupant-hi un paper fonamental.

El progrés tecnològic ha portat en si mateix una evolució també en les metodologies docents. Si volguéssim veure quina ha estat la trajectòria de les maneres d'ensenyar i aprendre, ens podríem remuntar a principis del segle XIX, quan arran de la Revolució Francesa van néixer els primers sistemes educatius estatals (Ministerio de Educación Cultura y Deporte, 2004). Aquells sistemes primerencs carregaven la gestió del coneixement sobre el professorat, que instruïa els estudiants impartint extenses lliçons magistrals. D'ençà d'aleshores moltes coses han canviat. La tecnologia ha anat entrant paulatinament a les aules i, amb ella, el model pedagògic s'ha anat desplaçant des de la visió instructivista del passat cap a un model molt més centrat en l'alumne.

Tot i que la tecnologia fa anys que participa d'alguna manera dels processos educatius, no és fins al preàmbul de la LOMCE², apareguda el 2006, quan se cita per primera vegada la intenció de donar especial importància a les Tecnologies de la Informació i la Comunicació com a element dins de l'ensenyament i l'educació en general. Això pot donar una idea al lector de la dificultat evident que ha tingut l'ensenyament de disposar d'eines educatives modernes per a la seva evolució. Tanmateix, la prova fefaent que la tecnologia ha entrat definitivament al món educatiu és l'aparició de sistemes d'ensenyament virtual o *e-learning*, els quals traslladen el context d'aprenentatge tradicional al món digital.

L'aparició de la primera universitat estatal - i del món - d'ensenyament virtual va arribar el 1995³ i des de llavors la majoria de centres educatius han apostat per introduir paulatinament sistemes d'aprenentatge virtuals i campus en línia. Certament, aquesta nova manera de treballar, ha aportat molts avantatges evidents, com poden ser la flexibilitat en horaris i en la localització de l'estudiant, però també ha introduït nous reptes, com ara la manera d'avaluar o la manera de treure el màxim rendiment d'un alumne, el qual, sovint, encara troba fronteres i solitud quan s'enfronta a aquests tipus d'estudi.

En aquesta tesi es presenta un nou model intel·ligent per a un entorn d'*e-learning*, el qual busca desenvolupar les eines necessàries per tal de superar els reptes que es plantegen actualment en aquests tipus d'entorns. El treball comença amb aquest capítol introductori, el qual s'estructura de la següent manera: inicialment s'exposa la motivació que ha conduït l'autor al desenvolupament d'aquesta investigació; tot seguit s'introdueixen quins són els escenaris que han servit de punts de partida per aquest treball; a continuació s'enumeren els objectius que es volen assolir durant el seu desenvolupament i la metodologia que s'ha seguit per a fer-ho; finalment, es llisten les publicacions que han permès arribar a aquest desenvolupament i es presenta l'organització de la memòria.

1.1 Motivació

Com totes les revolucions que han marcat la història de la humanitat, la revolució tecnològica que ha desencadenat l'entrada a l'era de la informació ha comportat la transformació profunda de la societat. Aquesta evolució s'ha produït especialment al llarg de la darrera dècada, quan molts dels processos que tenen lloc dins la societat s'han vist redefinits: la manera com ens relacionem, les metodologies que fem servir per adquirir coneixement, de quina manera consumim oci... Absolutament tots els àmbits socials s'han vist influïts, en major o menor mesura, per causa de la interacció amb la tecnologia i el món educatiu no s'ha vist exempt d'aquesta influència.

² Llei Orgànica per a la Millora de la Qualitat Educativa

³ La Universitat Oberta de Catalunya va ser la primera universitat del món en oferir un ensenyament virtual (<http://www.uoc.edu/portal/es/sobre-la-uoc/index.html>)

Davant del nou escenari social que està prenent forma a partir de la tecnologia, a l'educació se li presenten dos reptes importants: d'una banda ha de saber adaptar-se a aquesta nova manera de veure el món, reformulant les seves metodologies i velles concepcions; de l'altra, ha de ser prou permeable com per dur a terme una transició ràpida i adoptar la tecnologia com un element intrínsec al procés educatiu i no tractar-la com un mer complement. I és que sembla clar que el futur de l'educació passa per l'adopció de models formatius amb un fort component digital.

El primer repte l'educació l'ha afrontat mitjançant la transició entre diversos paradigmes educatius que, d'una manera formal i metodològica, reproduïen els models socials de funcionament que la societat ha anat adoptant. Així, la successiva adaptació a les diferents maneres de funcionar que es donen a la pròpia societat s'han vist representades per tres grans tendències educatives: *l'instruccionisme*, el *constructivisme* i el *connectivisme*. Paradigmes que, al seu torn, s'han vist plasmats en les eines educatives que paral·lelament han anat sorgint i que han acompanyat l'experiència educativa al llarg dels anys.

El segon repte que encara l'educació, referent a l'adopció de la tecnologia, va iniciar el seu recorregut amb les primeres iniciatives d'ensenyament assistit per ordinador, que van tenir lloc els anys 50. Des de llavors, el desenvolupament d'eines educatives ha estat sostingut. Dins la trajectòria experimentada per les eines educatives es poden destacar dues grans transicions pel que fa a les capacitats dels sistemes. La primera va tenir lloc amb l'aplicació de tècniques procedents del camp de la Intel·ligència Artificial i va donar lloc a l'aparició dels Sistemes Tutors Intel·ligents. La segona, i més recent, ha estat la que s'ha donat amb l'arribada d'Internet i l'aplicació del concepte *educació basada en Web*.

Tot el que és aprenentatge basat en Web troba el seu domini en l'*e-Learning*. I una bon exemple de l'explosió de tecnologies aplicades a l'educació que s'han succeït en aquest àmbit el representen els Entorns Virtuals d'Aprenentatge (EVA). Un EVA no és tan sols és una eina educativa en si mateixa, sinó que n'inclou moltes d'altres: repositoris d'informació, *wikis*, fòrums de debat...que fan que l'experiència d'aprenentatge sigui molt més rica. La seva aparició, com a evolució de les primitives aules virtuals, ha fet ressorgir el concepte de comunitat d'aprenentatge, ara en la seva forma digital. Així, en un EVA l'aprenentatge no tan sols és concebut com una experiència individual sinó que es busca també la dimensió col·lectiva.

Un EVA, com a comunitat d'aprenentatge que és, conté múltiples processos d'aprenentatge individuals que progressen simultàniament i que interaccionen entre si i amb el propi sistema. Com a resultat d'aquestes interaccions es genera un gran volum de dades conegut amb el nom de *Big Learning Data*. Aquest conjunt és la representació de les peces d'informació abstracta que circulen per l'entorn i que són susceptibles de ser recollides per tal d'atorgar-hi valor. No obstant, la informació en format digital dona la oportunitat de generar un coneixement de gran vàlua mitjançant el processament amb tècniques de computació avançada. Com es veurà en aquesta tesi, l'anàlisi de la *Big Learning Data* amb la finalitat d'obtenir coneixement que afavoreixi l'acció sobre el procés d'aprenentatge és una àrea que investiga l'Analítica de l'Aprenentatge (de l'anglès, *Learning Analytics*, LA).

Si bé l'*e-learning* va evolucionar a partir del que es coneixia com *educació a distància*, el fet que el seu suport sigui Internet n'atorga unes característiques diferencials. La Web és el vehicle d'aquest tipus d'aprenentatge i alhora és qui n'esculpeix, amb la seva evolució, els trets distintius. És per això que la progressió experimentada per la *World Wide Web* es veu reflectida, tot i que amb un cert desviament temporal, en la trajectòria de l'*e-learning*. La causa d'aquest desfasament respecte de la societat és a causa en certa manera als dos reptes pendents que s'esmentaven a l'inici de la secció. I és que la resistència al canvi de metodologies pedagògiques i la impermeabilitat a la introducció massiva de tecnologia fan que el món educatiu desenvolupi amb cert retard els potencials de la tecnologia que ja experimenta la Web. Així, actualment, l'*e-learning* es troba implantant característiques pròpies de la *Web social* (2.0), mentre que la Web està presenciant la intensa explotació de la seva versió semàntica (3.0) i es troba en vistes d'imaginar i dissenyar la versió simbiòtica (4.0) d'un futur proper. En aquesta direcció sembla que es desenvoluparan les properes generacions d'*e-learning* i, malgrat encara faltin uns quants anys per veure-ho com una realitat, tot fa preveure que el tret singular de les noves plataformes serà l'ús d'agents intel·ligents.

Com s'ha esmentat anteriorment, l'ús d'Intel·ligència Artificial (IA) en el món de l'aprenentatge no és nou. Cap els anys 80 van introduir-se les primeres tècniques d'IA a l'hora d'implementar eines educatives i com a resultat en van sorgir un bon nombre de Sistemes Tutors Intel·ligents (STI). Els STI es caracteritzen per la capacitat que tenen de personalitzar el procés d'aprenentatge d'un individu adaptant-lo a les seves necessitats específiques i representen la *teoria constructivista* en tota la seva essència. Darrerament aquest tipus de tecnologia educativa ha trobat un catalitzador en Internet, ja que gràcies a aquest suport ha pogut desenvolupar-se en noves àrees de recerca com és, per exemple, l'aprenentatge col·laboratiu intel·ligent. Aquest camp és relativament jove i permet explorar diversos escenaris presumibles des d'un vessant teòric. L'aprenentatge col·laboratiu intel·ligent és l'evolució que es preveu en els EVA actuals en el moment que incorporin tècniques d'IA.

Algunes de les tècniques que implementarà la IA als EVA sembla que seran: el seguiment intel·ligent, el suport adaptatiu per a la col·laboració, la tutorització intel·ligent... Aquesta darrera tècnica es pren com a base de la tesi.

En entorns de tutorització col·lectiva intel·ligent, com a evolucions tecnològiques dels sistemes tutors intel·ligents que seran, s'hauria d'administrar i recollir informació de l'alumne i del seu grup de treball per tal d'aconseguir una formació personalitzada segons cada cas. A més, a diferència dels STI tradicionals, que tenen coneixement del context de l'alumne perquè són entorns tancats i de dimensions reduïdes, aquests nous entorns han de gestionar grans volums d'informació (*Big Learning Data*) per poder saber tant el context d'un alumne com d'un col·lectiu. Aquesta informació, per tant és de vital importància i recuperar-la i tractar-la poden ser claus no només per millorar el procés d'aprenentatge sinó també per la tutorització i el seguiment continu de l'alumne i la comunitat al llarg del seu procés educatiu.

Tenint en compte tot aquest escenari, en aquesta tesi es fa una proposta d'arquitectura com a base per construir sistemes que englobin tot aquest procés. Es definirà un nou model anomenat

TICVA (Tutorització Intel·ligent de Comunitats Virtuals d'Aprenentatge) capaç de suportar les implicacions que esdevenen de la transició cap a la quarta generació de l'*e-learning*: el tractament de la *Big Learning Data*, la intel·ligència artificial i la ineludible evolució cap a la web simbiòtica.

1.2 Punts de partida

Com s'ha conclòs a la secció anterior, el propòsit principal d'aquesta memòria és proposar l'arquitectura d'un sistema que permeti abordar la tutorització intel·ligent de grups virtuals. En aquest sentit, serà imprescindible emprar com a base alguns conceptes propis de la Intel·ligència Artificial i/o de la seva aplicació.

Si s'haguessin de triar dos punts de partida d'aquesta tesi aquests serien sense dubte els Sistemes Tutors Intel·ligents i el Raonament Basat en Casos. Totes dues temàtiques han format part sempre de la recerca realitzada per l'autor i, com es veurà, han constituït el *leitmotiv* d'aquesta tesi. Tot seguit es comentarà amb més detall en què consisteixen cadascuna d'aquestes temàtiques i per què serveixen com a base i enfocament en el plantejament d'aquesta tesi.

1.2.1 Els Sistemes Tutors Intel·ligents

Tal com es comentava en el Prefaci d'aquesta tesi, un dels punts de motivació ha estat el món dels Sistemes Tutors Intel·ligents (STI). Els STI van néixer amb la finalitat d'ajudar l'alumne en el repte de l'ensenyament a distància. Segurament van ser l'evolució des d'un punt de vista tecnològic del que seria el tradicional professor particular per a l'alumne: un programari capaç de guiar l'alumne en la seva progressió i ajudar a millorar el seu rendiment en alguna matèria concreta.

Si es recorda l'estructura d'un STI clàssic es pot veure que està format per cinc mòduls ben diferenciats (Self, 1990), que es mostren a la Figura 1.

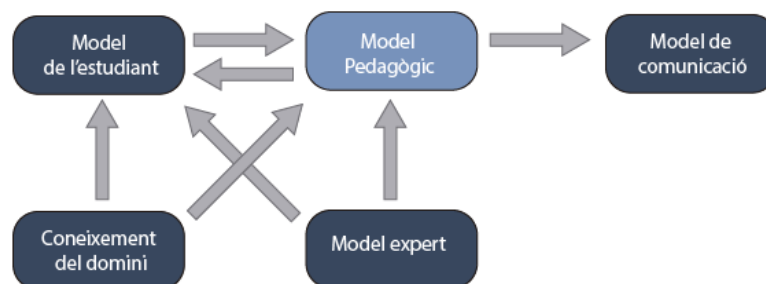


Figura 1. Mòduls que formen un Sistema Tutor Intel·ligent clàssic.

Cadascun d'aquests mòduls desenvolupa tasques rellevants i diferenciades dins del sistema, que es descriuen breument a continuació (es deixa el seu aprofundiment per al capítol 4 d'aquesta tesi).

- *Model de l'estudiant.* És el mòdul encarregat de recollir i gestionar dades històriques sobre els resultats i les característiques de l'alumne. És en aquest bloc on s'emmagatzema tota la informació històrica que s'obté dels estudiants. La manera com s'organitzi aquesta informació és vital per a trobar les dades més rellevants per al bon funcionament del sistema.
- *Model pedagògic.* És el mòdul capaç d'entendre quina és l'evolució de l'alumne i pronosticar la seva evolució. Els sistemes intel·ligents són part essencial d'aquest mòdul doncs han de ser els capaços d'entendre en quina situació es troba l'alumne i quina ha de ser la seva fita següent.
- *Model de comunicació.* És l'encarregat d'establir la interfície entre el sistema i l'estudiant o el professor. Sovint, dissenyar un bon sistema que sigui usable i accessible és una de les tasques més difícils a l'hora de desenvolupar un producte.
- *Coneixement del domini.* Aquest mòdul recull tota la informació sobre l'assignatura o matèria que s'està ensenyant, que pot incloure el propi temari, exercicis preparats, exàmens o qualsevol altre material docent.
- *Model expert.* És el sistema existent per validar les respostes i les aportacions de l'alumne. L'alumne hauria de ser autosuficient gràcies a aquest sistema i no hauria de necessitar la intervenció del professorat per tal de conèixer si les seves respostes són vàlides. Tanmateix, això no exclou que el sistema pugui incloure respostes preestablertes o solucions aportades pel professor a tall d'exemple.

Com el lector podrà veure, en el capítol 4 s'estudiaran a fons els Sistemes Tutors Intel·ligents, doncs seran la base de les posteriors aportacions que es realitzin sobre aquest model. A partir d'aquí, de manera resumida, es pot considerar que les actuacions que es duran a terme sobre un sistema d'aquestes característiques són les següents:

- Presentar una arquitectura basada en els STI clàssics que s'adapti al context col·lectiu de l'aprenentatge col·laboratiu.
- Implementar tècniques específiques que desenvolupin els aspectes intel·ligents del sistema.

1.2.2 El Raonament Basat en Casos

La segona temàtica de la que es nodreix aquesta tesi és la referent a una àrea específica de la Intel·ligència Artificial, concretament l'Aprenentatge Automàtic. En l'apartat anterior s'ha descrit que una de les parts més importants d'un Sistema Tutor Intel·ligent és justament la capacitat *intel·ligent* del Mòdul Pedagògic per tal predir quina serà l'evolució de l'alumne en funció del seu coneixement i resultats assolits. Com es veurà posteriorment són diverses les possibilitats algorísmiques que existeixen a l'hora d'afrontar aquest repte i, clarament, una de les més destacades és la tècnica anomenada Raonament Basat en Casos.

El Raonament Basat en Casos (CBR, de l'acrònim en anglès) és una metodologia de resolució de problemes complexos que intenta trobar solucions a partir d'experiències o entorns similars del passat. Mitjançant una analogia de les característiques d'un problema anterior, i a partir de la seva resolució i èxit obtingut, es proposa una solució que combina totalment o parcialment una bona solució del passat. Aamodt i Plaza, l'any 1994 (Aamodt & Plaza, 1994) van definir el que es coneix com a cicle de vida del CBR i que es representa a la Figura 2.

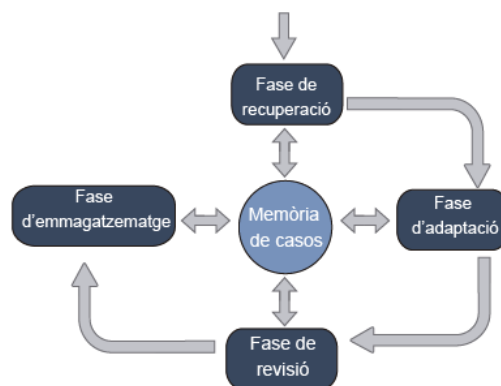


Figura 2. Cicle de vida del Raonament Basat en Casos.

Aquest cicle està compost de 4 fases diferenciades, anomenades “les 4 R”, per les inicials dels seus noms assignats. Totes les fases giren entorn de l'anomenada Memòria de Casos, entitat que recull l'experiència del sistema i emmagatzema les situacions rellevants trobades anteriorment i les solucions que es van prendre en el seu moment. L'objectiu de cadascuna d'aquestes fases és el següent:

- 1) *Recordar o Recuperar*. Quan es presenta un problema nou a resoldre cal recuperar o recordar quin o quins casos del passat tenen una similitud major al cas que s'ha presentat. Certament, la manera d'aconseguir aquest objectiu és múltiple i variada i serà en el sisè capítol on s'aprofundirà en les diferents tècniques que s'han experimentat en aquest sentit. Aquesta és la fase més important de tot el cicle i de la que dependrà l'èxit de la solució proposada.

- 2) *Reutilitzar o Adaptar.* Quan es treballa en problemes no relacionats amb la classificació de conceptes, sovint esdevé indispensable l'adaptació de la solució recuperada al nou cas que arriba. Si ens trobem en un problema de classificació estàndard molt sovint és possible la reutilització de la solució recuperada per al nou cas. En canvi, en altres tipus de problemes, on els dominis i l'entorn del nou cas és diferent del que hi ha emmagatzemat, cal fer alguna variació en la solució recuperada per tal que sigui vàlida en el nou entorn.
- 3) *Revisar.* Un cop proposada la solució, adaptada o no, al nou problema, cal verificar i revisar si aquesta solució és vàlida i soluciona el problema. De vegades pot succeir que el procés d'adaptació de la solució generi una solució inconsistent o no vàlida pel domini del problema nou. Si això passa cal revisar la reutilització del cas recuperat i, si s'escau, recuperar de la Memòria de Casos altres casos alternatius.
- 4) *Retenir o Emmagatzemar.* Aquesta fase permetrà emmagatzemar a la Memòria de Casos la informació necessària per tal que la solució definida per aquest nou cas sigui recordada pel sistema. No sempre és necessari incloure nova informació, ja que si el nou cas es pot arribar a resoldre amb la informació existent, el fet d'introduir noves dades a la Memòria de Casos comportaria la repetició i redundància de les dades recordades. Per tal d'emmagatzemar aquesta informació es poden definir diverses polítiques que, més endavant en aquesta tesi, seran exposades.

En el capítol 6, per tant, es veurà com s'utilitza aquesta tècnica com a base en els algorismes intel·ligents dissenyats per implementar bona part del mòdul més complex de l'arquitectura proposada, el *mòdul pedagògic*.

1.3 Objectius de la tesi

Arribats en aquest punt, és hora de recapitular quines són les idees més importants que s'han descrit i quines són les hipòtesis d'aquesta tesi, les quals es volen validar tot assolint cinc objectius principals que es descriuran ben aviat.

D'entrada s'ha vist que els Sistemes Tutors Intel·ligents permeten poder avaluar de manera individual el rendiment d'un alumne però estan limitats pel que fa al treball en grup i l'aprenentatge col·laboratiu. A més, com a punt crític, s'ha vist que el funcionament d'aquests sistemes depèn en gran mesura de la seva part intel·ligent. Per aquest motiu, el bon disseny d'algorismes dedicats a cobrir aquest objectiu és vital per a l'èxit del sistema.

Un altre aspecte destacable és el creixent interès que hi ha a la comunitat educativa per integrar iniciatives d'aprenentatge col·laboratiu. I és que tradicionalment, tot i tractar-se d'una gran eina d'aprenentatge, la seva aplicació ha estat sovint descartada en entorns d'ensenyament no presencial justament per la limitació dels STI i dels propis sistemes i campus en línia.

L'adequació, per tant, d'un sistema que s'adapti a aquest tipus d'aprenentatge hauria de permetre poder dirigir l'alumne més eficientment en el seu procés d'aprenentatge, no només de manera individual, sinó també com un membre integrat en una comunitat virtual.

Arran de tots aquests punts de partida es poden formular ja les hipòtesis d'aquesta tesi:

Hipòtesi 1:

"És possible dissenyar un model d'arquitectura d'aprenentatge en línia que millori els models actuals tot permetent que els alumnes puguin arribar a ser supervisats de manera individual i grupal per un STI"

Hipòtesi 2:

"És possible definir, dissenyar i implementar mitjançant tècniques d'Intel·ligència Artificial la predicció de l'evolució de l'alumne tant de manera individual com col·lectiva al Mòdul Pedagògic d'un STI"

Per tal de comprovar aquestes hipòtesis cal preguntar-se doncs si és possible dissenyar una arquitectura que es recolzi en l'esperit dels STI i estableixi una metodologia per introduir un sistema de tutorització intel·ligent en línia que permeti optimitzar el rendiment i l'aprenentatge de l'aula. Per arribar a donar una resposta a aquesta qüestió es defineixen els següents 5 grans objectius científico-tècnics que també serveixen com a fites de la tesi:

PRIMER OBJECTIU: Estudi de la situació de l'aprenentatge en línia

En primer lloc, cal estudiar quin és el grau de maduresa de l'aprenentatge en línia per tal de poder albirar quina direcció s'intueix que prendrà en els propers anys. En aquest sentit, es vol estudiar com la tecnologia ha marcat la història de l'*e-learning* amb la introducció de les TIC i el profund canvi obligat en el model pedagògic. Paral·lelament, es vol analitzar les tendències de futur d'aquest sistema d'aprenentatge.

SEGON OBJECTIU: Anàlisi de la informació docent recollida

Cal determinar la millor manera d'analitzar les dades que es recullen al llarg del procés d'aprenentatge de l'alumne. Això es durà a terme mitjançant l'estudi de les metodologies existents per recollir dades abstractes i transformar-les amb informació útil que permeti prendre mesures per millorar l'aprenentatge de l'alumne. S'analitzarà què s'entén per *Big Learning Data* i s'introduirà el model de referència de l'Anàlisi de l'Aprenentatge.

TERCER OBJECTIU: Els Sistemes Tutors Intel·ligents

El punt de partida d'aquest objectiu és conèixer a fons l'estructura d'un Sistema Tutor Intel·ligent. Com ja s'ha citat, i s'aprofundirà posteriorment, un Sistema Tutor està format per diversos mòduls que tenen la seva importància dins del procés d'aprenentatge de l'alumne. Cal estudiar a fons com es relacionen aquests mòduls per identificar en quins punts d'aquest sistema es poden introduir aportacions, tals com noves tècniques d'Intel·ligència Artificial que permetin millorar el rendiment de l'alumne o noves maneres de tractar la informació recollida dels alumnes.

D'altra banda, per poder fer noves aportacions, és molt important conèixer quina ha estat l'evolució d'aquests sistemes i de l'aprenentatge en línia en general. Efectivament, l'estudi de la seva evolució al llarg dels anys i l'observació de les característiques que han estat potenciades en nous models permet veure quins són els trets més remarcables que cal estudiar per a la implementació d'un nou sistema final.

QUART OBJECTIU: Nova arquitectura intel·ligent d'aprenentatge en línia

El quart objectiu consisteix en, partint dels models de referència estudiats, fer una proposta d'arquitectura intel·ligent per a sistemes d'aprenentatge en línia que permeti introduir l'aprenentatge col·laboratiu i el concretí, tot donant rellevància a la figura *grup* d'un comunitat virtual d'aprenentatge.

CINQUÈ OBJECTIU: Contribucions en la part intel·ligent d'aquesta arquitectura, sobretot en la part de predicció i tutoria.

El cinquè i darrer objectiu té com a fita realitzar el disseny i implementació del mòdul de predicció de l'arquitectura proposada, tot fent aportacions en tècniques d'aprenentatge automàtic. A la vegada es volen estudiar els resultats obtinguts per veure quina implicació tenen en la tutorització virtual de l'alumne i en la seva progressió dins el procés d'aprenentatge.

Com a evidències del compliment de tots aquests objectius es pretén publicar les conclusions i els resultats que se n'obtinguin en revistes internacionals indexades i en publicacions nacionals. Pel que fa a les revistes indexades, en el moment d'elaborar els objectius d'aquesta tesi es van triar les següents:

- *JRPIT: Journal of Research and Practice in Information Technology*. Revista especialitzada en articles de recerca que sorgeixin de l'aplicació de les TIC al món universitari o la indústria.
- *CHB: Computers in Human Behavior*. Revista especialitzada en l'aplicació de la informàtica des del punt de vista psicològic i en l'impacte que aquesta té en els individus, grups i societat en general.

- *IJDSN: International Journal of Distributed Sensor Networks*. Revista especialitzada en xarxes i telemàtica on es publicaran els resultats d'aplicar tècniques d'IA en entorns telemàtics.

Altres publicacions i revistes no indexades però més especialitzades serviran per compartir els resultats obtinguts en les contribucions concretes realitzades en tècniques d'Intel·ligència Artificial.

1.4 Metodologia de recerca

Per assolir els objectius que s'han plantejat en aquesta tesi s'ha seguit una metodologia de recerca centrada en cinc fases diferenciades aplicades a cadascun d'ells:

1. *Identificació de la temàtica o problema a abordar*. S'ha analitzat quin domini i àrees estan involucrats en l'objectiu que s'estava tractant. Paral·lelament es llistaven els possibles *topics* i paraules clau relacionades per tal d'aconseguir tota la visió necessària de l'àrea a analitzar.
2. *Recerca bibliogràfica a fons de l'àmbit de l'objectiu*. S'ha realitzat una recerca exhaustiva i sistemàtica de publicacions relacionades amb l'àmbit d'estudi. Mitjançant l'accés a publicacions electròniques a través de la Universitat i amb el suport del servei de Biblioteca s'han pogut recollir molts resultats. Aquests resultats han estat complementats mitjançant navegació i consulta de pàgines web relacionades amb cadascun dels objectius.
3. *Interpretació i anàlisi del material recollit*. S'ha llegit i treballat cadascun dels materials aconseguits amb la finalitat de trobar propostes de solucions d'altres autors i idees que permetin suggerir possibles plantejaments als problemes o temàtiques que es volen abordar.
4. *Extracció de conclusions i actuació posterior, si s'escau*. En els objectius més teòrics l'autor s'ha limitat a extreure conclusions i models que serveixin per dur a terme accions posteriors en la recerca. Altres objectius han portat una acció concreta darrera, com pot ser la implementació d'una solució o el disseny d'una contraproposta alternativa a la identificada.
5. *Avaluació i reflexió dels resultats*. Finalment, cada objectiu s'ha tancat analitzant els resultats obtinguts i extraient-ne conclusions. Amb el director de la tesi s'han debatut els aspectes més importants amb la finalitat de dirigir i afitar els propers passos d'investigació i la publicació d'aquests, si s'escau.

Amb aquestes fases s'ha volgut assolir cadascun dels objectius de manera satisfactòria, promovent una recerca estructurada i metòdica.

En aquesta mateixa línia metodològica, la investigadora Eileen Ferrance presenta en el seu treball *Action Research* (Ferrance, 2000) una visió profunda de com s'han de realitzar el que ella anomena "passos de la recerca" i ho esquematitza en la Figura 3 que es presenta a continuació.

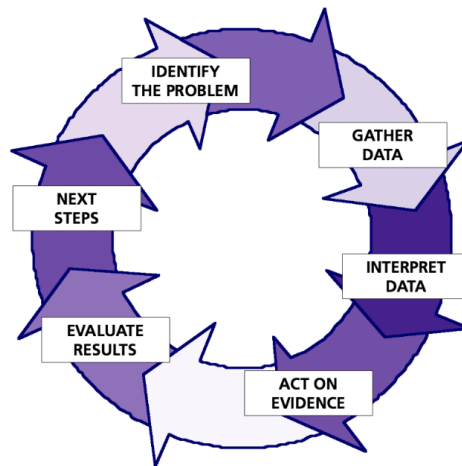


Figura 3. Cicle de metodologia de recerca proposat per Ferrance.

Com es pot comprovar, els passos proposats per aquest treball tenen força punts de coincidència amb la metodologia de recerca proposada per Ferrance. Altres autors, com per exemple Blaxter (Blaxter, 2001) proposen metodologies similars però que es poden acabar reduint o ampliant a l'exposada anteriorment. Citant a Thomas Henry Huxley, "*el mètode d'investigació científica no és res més que l'expressió necessària de la modalitat de treball de la ment humana*".

1.5 Publicacions obtingudes

Durant el camí recorregut en la consecució d'aquests objectius, s'han obtingut diverses publicacions. En la taula següent, s'indiquen les publicacions relacionades amb la tesi, per ordre cronològic invers, separades per tipus de publicació (revistes indexades a ISI, no indexades a ISI i altres publicacions) i relacionades amb els diferents objectius plantejats. Atès que moltes publicacions estan associades a diferents objectius, s'han agrupat els objectius 1, 2 i 3, en una sola columna per fer la taula més entenedora.

Taula 1. Publicacions relacionades amb la tesi classificades segons el tipus de publicació i l'objectiu associat

Publicació	O1- O2-O3	O4	O5
Revistes indexades a ISI			
IJDSN, 2015 (Vernet, Zaballos, Martin de Pozuelo, & Caballero, 2015)		X	
CHB, 2014 (Canaleta, Vernet, Vicent, & Montero, 2014)	X		
IJDSN, 2013 (Zaballos, Vernet, & Selga, 2013)			X
Revistes no indexades a ISI			
**JRPIT, 2013 (Vernet, Canaleta, Pallàs, Navarro, & Zaballos, 2013)		X	
IJAIT, 2009 (Vallejo, Zaballos, Vernet, Orriols, & Dalmau, 2009)			X
LNCS, 2006 (Fornells, Golobardes, Vernet, & Corral, 2006)			X
EU, 2003 (Vernet & Golobardes, 2003)	X		X
ACIA, 2001 (Golobardes, et al., 2001)			X
ACIA, 2000 (Golobardes, Llorà, Garrell, Vernet, & Bacardit, 2000)			X
Altres publicacions			
CLOSER, 2014 (Navarro, et al., 2014)		X	
CISTI, 2014 (Sancho-Asensio, et al., 2014)	X		
CISTI, 2014b (Solé, et al., 2014)	X		
JENUI, 2014 (Canaleta, Navarro, Solé, & López, 2014)	X		
JENUI, 2014b (Navarro, et al., 2014)	X		
JENUI, 2013 (Canaleta, Vernet, & Navarro, 2013)	X		
ISCAS, 2013 (Albó, Vernet, Canaleta, & Vilasís-Cardona, 2013)	X		
CIDU, 2012 (Canaleta, Zaballos, & Vernet, 2012)	X		
JENUI, 2012 (Vernet, Canaleta & Zaballos, 2012)	X		
SIIE, 2012 (Vernet, Canaleta, & Pallàs, 2012)		X	
SIIE, 2012b (Canaleta, Vernet, Vicent, & Montero, 2012)	X		
JENUI, 2011 (Canaleta, Vernet, & Zaballos, 2011)	X		
JENUI, 2011b (Vernet & Canaleta, 2011)	X		
CIUMM, 2011 (Vernet & Canaleta, 2011)	X		
CIUMM, 2011b (Canaleta & Vernet, 2011)	X		
FIE, 2010 (Vernet, Nicolas, Golobardes, Fornells, & Garcia-Piquer, 2010)	X		X
ICCBR, 2009 (Nicolas, et al., 2009)			X
ICITAA, 2009 (Nicolas, et al., 2009)			X
CCIA, 2008 (Vernet, et al., 2008)			X
ICSNC, 2008 (Vallejo, Zaballos, Vernet, Cutiller, & Dalmau, 2008)			X
CCIA, 2008b (Canaleta, Ros, Vallejo, Vernet, & Zaballos, 2008)		X	
JENUI, 2004 (Canaleta & Vernet, 2004)	X		
URSI, 2004 (Canaleta, Vernet, Alsina, & Montero, 2004)	X		
JIS, 2004b (Vernet & Canaleta, 2004)	X		

JENUI, 2004b (Vernet & Canaleta, 2004)	X		
CIDUI, 2004 (Alsina, Canaleta, Montero, & Vernet, 2004)	X		
JIS, 2004 (Canaleta & Vernet, TIC versus mTIC, 2004)	X		
CCIA, 2003 (Vernet & Golobardes, 2003)			X
JENUI, 2003 (Canaleta & Vernet, 2003)	X		
MAEB, 2003 (Camps, Garrell, Golobardes, & Vernet, 2003)			X
IBERAMIA, 2002 (Garrell, et al., 2002)			X
UKWCBR, 2002 (Vernet & Golobardes, An Unsupervised Learning Approach for Case-Based Classifier Systems, 2002)	X		X
CCIA, 2001 (Vernet, et al., 2001)	X		X
CCIA, 2001b (Golobardes, et al., 2001)			X
CCIA, 2000 (Golobardes, Llorà, Garrell, Vernet, & Bacardit, 2000)			X
LEARNING, 2000 (Salamó, Golobardes, Vernet, & Nieto, 2000)			X
LEARNING, 2000b (Golobardes, Vernet, & Salamó, 2000)	X		X

** Fins l'any 2012 aquesta revista ha estat indexada a ISI.

L'objectiu d'incloure aquesta taula a la introducció de la tesi és que el lector pugui anar veient com es van relacionant les diferents publicacions obtingudes per l'autor amb els diferents objectius que s'aniran treballant al llarg del document. D'aquesta manera, es vol facilitar-ne la lectura posterior i la seva correlació amb la recerca publicada per l'autor.

1.6 Organització de la memòria

La memòria d'aquesta tesi s'organitza en set capítols. En el primer capítol, el present, s'ha analitzat la motivació de la temàtica i s'ha centrat el domini d'aplicació. D'aquesta manera s'ha pretès contextualitzar al lector analitzant el marc de treball suggerit. També en aquest capítol s'han definit els objectius de la tesi que pretenen ser fites per donar resposta a les hipòtesis plantejades. Finalment s'explica com s'organitza el document per facilitar-ne la seva lectura.

El capítol segon pretén abastar el primer objectiu de la tesi: analitzar com ha evolucionat l'aprenentatge al llarg dels temps, com ha nascut el concepte d'*e-learning*, i la seva implicació en el canvi de paradigma docent. S'estudien els diferents tipus de sistemes d'aprenentatge existents i es treuen conclusions de cadascun d'ells per poder-los introduir posteriorment en l'estudi dels altres objectius.

En el següent capítol es pretén profunditzar en el concepte de la *Big Learning Data*. Són moltes les dades que avui en dia es poden recollir de l'alumne i cal determinar quina és la millor manera de fer-ho i quina metodologia cal seguir per a la seva anàlisi. Conèixer la importància

d'aquestes dades i la seva implicació en l'evolució del procés d'aprenentatge és cabdal per poder proposar un nou model millorat.

En el quart capítol, que cobreix el tercer objectiu, s'explora el món dels Sistemes Tutors Intel·ligents tot analitzant les parts que el formen i estudiant-ne el seu comportament. Es pretén entendre en quins punts d'aquests sistemes es pot intervenir, amb la finalitat d'enllaçar-ho amb els dos darrers objectius de la tesi. També es fa recerca sobre la seva evolució i quins models han estat els més estàndards a l'hora de definir aquests tipus de sistemes.

El cinquè capítol, en línia amb el quart objectiu, proposa un nou model d'aprenentatge en línia amb la finalitat de millorar i optimitzar el rendiment d'una comunitat d'aprenentatge. De la mateixa manera també es detalla de manera més tècnica quina és la seva arquitectura i com s'integrarien les darreres tècniques de desenvolupament web per tal d'aconseguir un sistema ràpid, eficient i que garanteixi la integració de tots els serveis.

El sisè capítol persegueix cobrir el darrer objectiu de la tesi. És el capítol més tècnic i s'analitzen diferents algorismes intel·ligents existents com a possibles candidats per intervenir en la creació del Mòdul Pedagògic del sistema dissenyat. Com a resultat de la recerca, es proposen nous algorismes i tècniques que han estat provats i validats en diferents dominis d'experimentació. Aquest dominis es detallen també en aquest capítol i se'n contrasten els resultats obtinguts.

Finalment, al darrer capítol s'exposen les conclusions d'aquesta tesi i s'introdueixen algunes línies de treball futur.

Capítol 2

L'educació a la societat de la informació

“Student is not a container you have to fill but a torch you have to light up.”

Albert Einstein

A llarg de les darreres dècades del segle XX, la intrusió de les Tecnologies de la Informació i la Comunicació (TIC) a tots els nivells, en bona mesura afavorida pel fenomen de la globalització, ha contribuït a redefinir molts dels àmbits organitzatius que fonamenten la societat, entre ells el món educatiu. L'arribada de les noves tecnologies als centres educatius, però, no ha implicat únicament la digitalització de les aules, sinó que ha produït un profund canvi metodològic i de model. I és que durant les darreres tres dècades, el desenvolupament de recursos en suport digital i l'aplicació de tecnologia en el procés d'aprenentatge han impulsat la transformació del sistema formatiu tradicional, dirigint-lo cap a un model molt menys instructiu i basat eminentment en l'activitat de l'estudiant (Sangrà & Stephenson, 2012). D'altra banda, els canvis en els hàbits de comportament que ha portat la societat globalitzada, ha propiciat una major generació i consum de coneixement i, alhora, una nova forma d'interaccionar-hi. L'entorn natural és "la xarxa", d'on hom pren el que necessita i ofereix el que té, formant així vincles entre integrants que sovint es materialitzen en comunitats virtuals (Fernández, 2009). Aquesta manera global de relacionar-se s'ha vist reflectida, al seu temps, en el camp de la formació: l'*e-learning* o aprenentatge en línia sorgit com a conseqüència de l'entrada de les TIC al món educatiu ha anat encaminant-se cap a un model molt més global o universal, on prima la interacció entre persones i institucions a nivell mundial i on l'aprenentatge esdevé un fet fortament influenciat pel context.

Com es veurà a continuació, en aquest capítol s'exposa breument com la tecnologia ha marcat la història de la civilització mitjançant revolucions tecnològiques. A més, s'analitza l'impacte que ha tingut en l'educació el darrer salt tecnològic que ha experimentat la societat a través de les TIC, i que ha transformat profundament el model pedagògic. Per últim es descriu l'evolució que ha sofert l'*e-learning* des de la seva aparició fins l'actualitat i s'analitzen quines sembla que seran les tendències que tindrà aquest sistema d'aprenentatge en el futur.

2.1 Antecedents: Una història marcada per la tecnologia

Des de fa més de 10.000 anys, la tecnologia ha estat el gran catalitzador de l'evolució de la civilització. El descobriment del foc, la invenció de la roda, la fosa del metall o l'aparició de la impremta, entre d'altres, han acabat determinant el curs de la humanitat. Tanmateix, tot i que l'ascensió tecnològica ha estat gairebé sempre contínua, el cert és que al llarg de la història s'han donat certes revolucions o salts tecnològics que han produït una veritable ruptura, provocant profunds canvis en la civilització del moment (Camarena, 2015).

Ja en la darrera part de l'Edat de Pedra, amb el que es coneix com a Revolució Neolítica o Agrària, la humanitat va experimentar el seu primer gran salt de desenvolupament marcat per la tecnologia. Es va passar d'un model de vida nòmada, basat en la caça i la recol·lecció, a un de sedentari, sustentat principalment per l'economia productora que oferien l'agricultura i la ramaderia. I és que sens dubte la invenció dels utensilis de pedra polida (destal, aixada...), i per tant la incursió d'una nova tecnologia, va comportar una forma inèdita d'interacció amb l'entorn, fent possible la tala d'arbres o el treball de la terra.

En una cronologia més propera, va ser la Revolució Industrial la que va marcar un abans i un després en el curs de la civilització. Els avenços tècnics, com la màquina de vapor i més tard els motors de combustió o l'energia elèctrica, i l'explosió demogràfica de l'època, que subministrava mà d'obra abundant, van afavorir el procés de creixement industrial. Aquest procés va comportar, al seu temps, una profunda transformació de la societat: D'una banda es va passar de l'economia productora agrícola i ramadera a una d'industrialitzada (economia de mercat); de l'altra, va comportar una accentuada reestructuració social (polaritzant la societat entre burgesia i proletariat).

Com no pot ser d'altra manera, la societat del coneixement en què l'ésser humà es troba immers avui dia, és fruit novament d'una revolució tecnològica. Després de passar per una època post-industrial, la humanitat ha experimentat un canvi de paradigma gràcies al desenvolupament de les TIC. Com tot salt tecnològic, però, l'assoliment del nou estadi ha comportat un profund canvi en la societat, que ha vist transformats els seus models d'organització (globalització) i interacció (televirtualitat).

Així és que, tant si es tracta de la indústria lítica de la prehistòria com de la societat de la informació actual, la història de la humanitat és una història de canvi i d'adaptació. Una narració

en què la redefinició de processos, de formes d'organització, d'estructures i d'interacció apareixen constantment com a fils conductors de l'evolució.

2.2 El canvi de paradigma del món educatiu

L'educació, per la seva banda, no s'ha vist exempta de les transformacions que la societat del coneixement o que la informació ha produït a la civilització actual. Moltes de les seves estructures i processos han estat reformulats, d'altres han desaparegut i, com en tota emergència tecnològica, s'han creat metodologies i estructures inèdites que busquen donar resposta a les necessitats originades pel nou context.

Tot i que el món educatiu ha evolucionat com ho feia la resta de la societat, el cert és que no ho ha fet al seu mateix ritme. A mitjans dels anys noranta, es va donar el període de transició entre la societat industrial i la societat de la informació, però no ha estat fins ben entrat el nou mil·lenni que el sistema educatiu ha abandonat la estanquitat amb què romania i ha començat a explotar intensivament la tecnologia en el seu model i processos (Sangrà & Stephenson, 2012).

Des de llavors són molts els països que es troben en període de revisió del seu sistema educatiu, però la lentitud amb què molts d'altres veuen els canvis que la societat de la informació produirà en el seu model, fa que existeixi el perill real que els sistemes educatius futurs no donin la resposta que la societat reclama. Per consegüent (Tiffin & Rajasingham, 1997):

"Seran més valuosos aquells models i sistemes que siguin capaços de canviar al ritme de les noves realitats i aconseguixin oferir els seus serveis a l'ampli mercat del futur del demà."

I és que, en essència, el sistema educatiu tradicional està dissenyat encara ara per donar resposta a una societat industrial on els estudiants són formats per esdevenir integrants de la societat, emulant les fàbriques i oficines: es desplacen al lloc d'estudi, comencen i acaben la jornada en un horari molt determinat, treballen en aules que són una representació primigènica de les oficines, divideixen les matèries per al seu estudi... En definitiva, es tracta d'una anticipació a la vida de després de l'escola (Tiffin & Rajasingham, 1997). El món laboral, però, està desplaçant-se ràpidament cap a un model fonamentat molt més en les comunicacions (societat de la informació) que no pas en la indústria i el transport (com ho faria la societat industrial). Així és que, si el sistema educatiu ha de preparar individus per a la societat del futur, aquest haurà d'evolucionar prou com perquè els seus alumnes acabin tenint les competències que aquest futur els requerirà. I així ho reitera Kenneth Cukier (Cukier, 2013):

"Hem de proposar als nostres fills un altre sistema educatiu, ja que aquest va ser concebut en una època diferent, a l'era industrial, mecanicista, basada en una línia de muntatge. Ara es pot adaptar a les recomanacions d'Amazon i Google que s'ajusten exactament als nostres interessos. Necessitem una educació que s'adapti a les nostres necessitats i aquesta és la millor manera d'aprendre."

2.2.1 Un nou model pedagògic

Dins de la Pedagogia actual es poden distingir tres corrents predominants pel que fa a la teoria de l'aprenentatge: l'instructivisme, el constructivisme i el connectivisme. Els dos primers paradigmes fan referència al grau d'implicació del docent durant el procés educatiu i analitzen en quina mesura l'alumne disposa de flexibilitat per a explotar els seus propis recursos amb la finalitat d'assolir un determinat objectiu formatiu (Vernet, Canaleta, & Pallàs, 2012). En canvi el tercer evoluciona a partir del concepte de xarxa social, és a dir, deixa d'estar centrat en l'individu i tracta el procés d'aprenentatge com una experiència que té lloc en comunitat (Anderson & Dron, 2011).

L'instructivisme, d'una banda, es basa fonamentalment en la figura del formador: la determinació d'objectius i metodologia es troba exclusivament a mans del docent o, per extensió, de l'entitat educativa. És l'agent formador qui determina quins són els objectius que l'alumne ha d'assolir i planifica tant la cadència com la metodologia que aquest seguirà. L'aplicació d'aquest paradigma, però, comporta l'establiment d'un patró únic per al conjunt d'alumnes ometent les característiques particulars de cada individu, particularitats que poden arribar a influir significativament en el desenvolupament del seu procés d'aprenentatge.

De l'altra banda, la teoria constructivista té el seu eix en el context particular de cada alumne i estableix el marc conceptual propi de l'individu com a punt de partida de l'aprenentatge. El docent ja no determina en exclusivitat i de manera rígida la progressió que aquest ha de seguir, sinó que orienta i subministra les eines necessàries per tal que aquest pugui assolir els objectius plantejats emprant la seva pròpia experiència personal i les capacitats que ha anat desenvolupant fins al moment. Així doncs, en aquest nou paradigma, basat en la potenciació de capacitats personals, l'educador esdevé més aviat un facilitador (Sangrà & Stephenson, 2012), que no pas un instructor.

En els darrers anys, però, la incursió de la tecnologia al medi educatiu ha començat a desplaçar les teories de l'aprenentatge cap a l'era digital. I això ha implicat la reformulació de conceptes i mètodes pedagògics (Canaleta, Vernet, & Navarro, 2013) per tal de fer-los més coherents amb el moment tecnològic i social que està vivint la humanitat. Així, segons la nova teoria connectivista ja no és possible experimentar i aprendre de manera individual, sinó que ara l'adquisició de competències es realitza a través de l'establiment de connexions formant una xarxa. L'aprenentatge, doncs, és un procés que no està sota control exclusiu de l'individu, sinó que es produeix de manera social. D'aquesta manera el coneixement va actualitzant-se i enriquint-se permanentment (Siemens & Downs, 2004).

En el context actual de l'ensenyament, el sistema instructiu es troba àmpliament implantat (sistema educatiu tradicional) i són comptades les organitzacions que aposten plenament pel mètode constructivista, i menys encara pel connectivista. I és que malgrat moltes institucions

inclouen tecnologies educatives en les seves metodologies, el cert és que la modalitat d'ús d'aquestes reproduceix més aviat el model instructiu que no pas un de basat en l'activitat de l'estudiant o fins i tot en el grup. A l'esquema que es mostra a la Figura 4 (Simon, 2007), es representen les diferents modalitats d'ús de les TIC que s'estan emprant actualment en el món educatiu classificant-les en funció de la intencionalitat metodològica amb què s'apliquen:

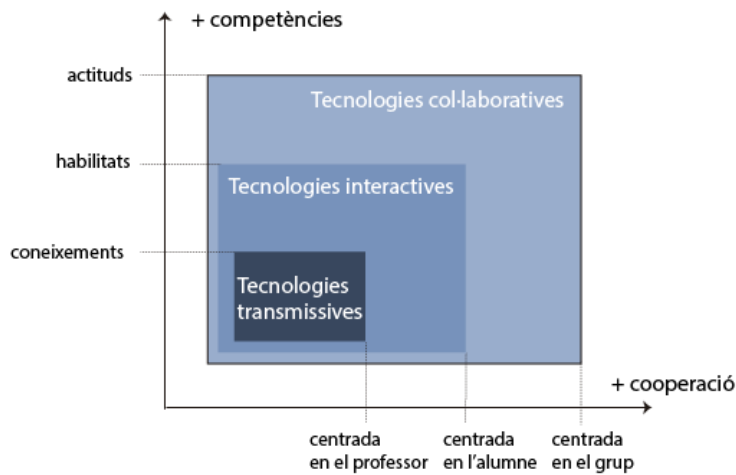


Figura 4. Tipologia de les tecnologies en la docència.

Les diferents metodologies pel que fa a l'aplicació de tecnologia educativa estan caracteritzades pels següents atributs (Simon, 2007) (Salomé & Suñé, 2011):

- **Tecnologies transmissives:** Són aquelles que exposen informació a l'alumnat a mode de presentació. Un exemple en serien les presentacions multimèdia o les pissarres digitals, en què, tot i tractar-se d'elements multimèdia, l'alumne continua essent un subjecte passiu, ja que l'instrument pedagògic continua centrat en el professor. La metodologia, per tant, malgrat que introdueix suport tecnològic, continua reproduint el sistema tradicional d'ensenyament (instructivisme). Aquesta idea es representa a la Figura 5.

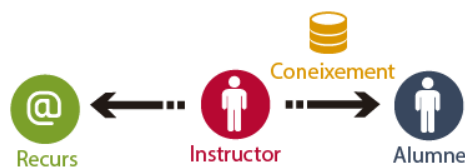


Figura 5. Tecnologia transmissiva. El model pedagògic està centrat en el professor: és aquest qui manipula la tecnologia educativa i transmet el coneixement a l'alumne com si fos un instructor.

- Tecnologies interactives: Reprodueixen un model d'aprenentatge basat en el *learning by doing* (aprendre mitjançant la pràctica). Un exemple d'aquest tipus de tecnologia són els programes d'ensenyament assistit per ordinador (EAO) o els dossiers electrònics. L'alumne disposa d'una major autonomia en el seu procés d'aprenentatge i el docent actua com a orientador del procés (constructivisme)(Figura 6).

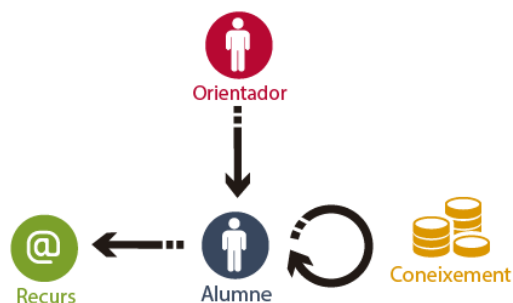


Figura 6. Metodologia interactiva. L'activitat està centrada en l'alumne: Mitjançant la interacció amb la tecnologia educativa i amb l'orientació del docent, l'alumne crea el seu propi coneixement.

- Tecnologies col·laboratives: Internet ha permès introduir a les aules entorns que permeten basar l'aprenentatge en la interacció i l'intercanvi d'idees i de recursos. En aquesta metodologia, la interconnexió no es dona exclusivament entre docent i alumne, sinó que s'estén a la resta del grup donant lloc així a una comunitat d'aprenentatge (connectivisme)(Figura 7). La base sobre la qual es fonamenta una comunitat d'aprenentatge és la del treball col·laboratiu i els campus virtuals en són un bon exemple.

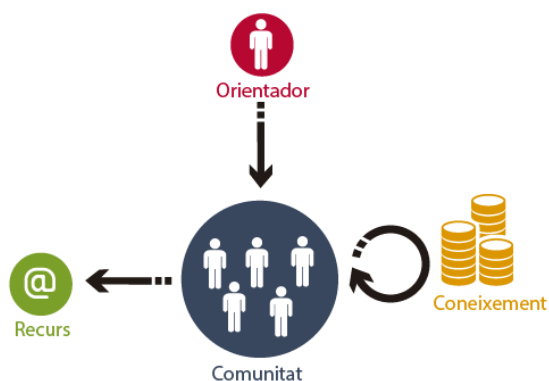


Figura 7. Tecnologia col·laborativa. L'activitat d'aprenentatge se centra en el grup d'alumnes o comunitat. Aquests, de manera col·laborativa interaccionen amb la tecnologia educativa i construeixen un coneixement grupal. El docent actua com a orientador o facilitador.

L'espectre d'aplicació de les TIC en el món educatiu abraça, per tant, des de models que reproduïxen el sistema d'ensenyament tradicional fins a models molt més orientats a donar resposta al moment social i tecnològic que la societat està travessant. Si bé les diferents metodologies en l'aplicació de tecnologies educatives han d'entendre's com a complementàries, més que no pas excloents, el cert és que el futur passa necessàriament per un procés de canvi metodològic. Sembla clar que caldrà que el sistema educatiu evolucioni cap a un model fonamentat en un canvi de visió curricular, de centrada en els continguts a centrada en les competències, i en un canvi de rols entre alumne i docent, en què l'alumne prengui un paper actiu en el seu aprenentatge i el docent esdevingui un orientador més que no pas un instructor (Salomé & Suñé, 2011).

2.3 L'e-learning: aprendre a la societat de la informació

Sovint l'aprenentatge en línia o *e-learning* s'associa equivocadament a models d'educació a distància, i és que malgrat comparteixen certes característiques, l'*e-learning* va molt més enllà de la simple introducció de la virtualitat a un sistema tradicional d'ensenyament. L'educació a distància, pot o no utilitzar tecnologia, però es basa fonamentalment en el principi de l'autoaprenentatge: l'important és garantir que l'alumne pugui realitzar el seu procés d'aprenentatge sense la necessitat de la intervenció contínua del docent. En canvi l'*e-learning*, tot i compartir la no-presencialitat del model, posa l'accent en la utilització de les TIC (especialment d'Internet) com a mitjà d'accés als continguts (Gros, 2011) i, en definitiva, facilita l'adopció de models o metodologies menys basades en l'ensenyament, per apropar-se més a aquells que se centren en l'aprenentatge (Sangrà & Stephenson, 2012).

L'*e-learning* és, per tant, un conjunt de processos d'aprenentatge que es duen a terme a través d'Internet i que reben el suport de les diferents tecnologies TIC, però en què, a diferència de l'educació a distància, predomina la comunicació entre els diferents rols del sistema donant lloc a una experiència educativa personalitzada i alhora social. Perquè, d'una banda, la tecnologia permet arribar a una gran personalització de l'activitat educativa, facilitant així a l'alumne l'autogestió del seu procés d'aprenentatge. I de l'altra, s'estableix una interacció didàctica continuada amb el docent, que actua com a orientador, i col·laborativa amb els companys, que fa que l'estudiant concebi el seu aprenentatge com un procés comunitari.

2.3.1 Modalitats d'e-learning

Existeixen moltes classificacions sobre l'*e-learning*. Algunes fan referència al dispositiu que es fa servir per accedir-hi, donant lloc a denominacions com *m-learning* (formació feta a través de dispositius mòbils) o *u-learning* (formació ubíqua o accessible des de qualsevol lloc i amb qualsevol mena de dispositiu). D'altres fan referència al mercat a què va dirigit, distingint si es tracta de l'acadèmic (institucions educatives) o bé de l'empresarial (per exemple, per a la formació d'empleats). Però una de les classificacions més importants és la que es fa des del punt

de vista de la presencialitat. I és que sovint, quan es parla d'*e-learning*, hom imagina que es tracta d'una modalitat d'aprenentatge que es duu a terme integralment de manera virtual, però no té perquè ser així. Es poden distingir tres modalitats d'*e-learning* (Seoane & García, 2010):

- **Direct e-learning o formació presencial.** En aquesta modalitat únicament s'usa Internet com a suport o complement dels cursos presencials tradicionals. Moltes de les activitats es duen a terme de manera col·laborativa com a complement de la part lectiva (per exemple realitzant activitats en forma de fòrums de discussió o *wikis*) o bé es limiten a l'execució de tasques de caràcter administratiu (per exemple, al dipòsit de treballs en bústies de lliurament).
- **Blended learning (b-learning) o formació mixta.** Es tracta d'un model de formació híbrid que busca recollir el millor de l'aprenentatge virtual i el millor del presencial. Esdevé una evolució de l'*e-learning* tradicional que busca superar les limitacions d'autoaprenentatge i sensació de solitud del primer (Fernández, 2009). En aquest tipus de formació es distribueix la càrrega lectiva entre virtual i presencial per tal d'optimitzar el procés d'aprenentatge i fer l'alumne més autònom.
- **Online Learning o formació virtual.** En aquesta modalitat el procés d'aprenentatge es duu a terme 100% a través d'Internet. Es tracta d'una modalitat completament asíncrona, doncs alumne i docent no han de perquè coincidir en lloc i moment. La seva aplicació requereix l'ús integral d'una plataforma o entorn virtual d'aprenentatge.

2.3.2 Entorns virtuals d'aprenentatge

Els Entorns Virtuals d'Aprenentatge (EVA) reben noms diversos: plataformes, campus virtuals, LMS...però en definitiva es tracta d'eines *software* evolucionades a partir de les primitives aules virtuals (veure secció 2.4.1) que busquen donar suport al procés d'aprenentatge en l'*e-learning*. Es defineix un EVA (Salinas, 2005) com aquell espai o comunitat organitzats amb el propòsit d'assolir l'aprenentatge i que requereix certs components per tal que aquest tingui lloc: una funció pedagògica (materials i activitats d'aprenentatge), la tecnologia apropiada a aquesta (eines coherents amb el model pedagògic) i els aspectes organitzatius pertinents per a dur-ho a terme (organització de l'espai, gestió de la comunitat...).

Un EVA per tant disposa, no tant sols de la base de coneixement per a assolir l'aprenentatge en forma de material docent, sinó que també posseeix les eines necessàries que permeten l'adquisició d'aquest coneixement i la gestió de les relacions dels diferents actors del sistema (alumnes, docents, tutors...). En general es pot dir que es tracta d'un entorn que ofereix el suport tecnològic adequat per tal que es puguin dur a terme correctament les diferents fases del procés d'ensenyament i aprenentatge: la planificació, la implementació, el desenvolupament i l'avaluació (Silva, 2011).

Sovint s'utilitza diversa taxonomia a l'hora d'anomenar els Entorns Virtuals d'Aprenentatge. I si bé a nivell europeu està àmpliament acceptat el terme d'EVA (de l'anglès, *Virtual Learning Environment*, VLE), a Estats Units els termes de CMS o LMS són molt més comuns. I tot i que en moltes ocasions s'utilitzen indistintament per definir una plataforma d'*e-learning*, cadascun presenta subtils diferències:

- **Entorn Virtual d'Aprenentatge** o EVA. Descriu de manera genèrica una plataforma amb funció pedagògica que ofereix eines tecnològiques que permeten als usuaris gestionar recursos educatius per tal de dur a terme un procés d'aprenentatge. En essència es pot dir que té un plantejament constructivista: l'alumne no és "ensenyat" sinó que desenvolupa el seu propi procés d'aprenentatge. Exemples: *Moodle*, *Blackboard*...
- **Sistema de gestió de continguts** o CMS (de l'anglès, *Content Management System*). Es tracta d'un programari o plataforma que permet crear i gestionar continguts d'una web de manera automàtica. Les accions que permet fer un CMS sobre el contingut són, per exemple: editar, esborrar, publicar, atorgar permisos d'accés... Exemples: *Joomla*, *WordPress*...
- **Sistema de gestió de l'aprenentatge** o LMS (de l'anglès, *Learning Management System*). Es tracta d'un conjunt d'aplicacions integrades que permeten automatitzar l'administració d'accions de formació/aprenentatge: registre d'usuaris, estructura i gestió de cursos, seguiment de l'aprenentatge... Sovint és concebut com un sistema per a la instrucció (instructivisme), més que no pas per a l'aprenentatge. Exemples: *Moodle*⁴, *Dokeos*...
- **Sistemes de gestió de continguts d'aprenentatge** o LCMS (de l'anglès, *Learning Content Management Systems*). És un tipus de programari que combina les funcionalitats de gestió d'un LMS amb les de creació de contingut d'un CMS. Exemples: *Claroline*, *e-doceo solutions*...

En qualsevol dels casos, en els darrers anys hi ha hagut un notable creixement de les diferents plataformes digitals associades a l'educació. Moltes institucions han començat a desenvolupar intensivament la modalitat de *blended learning* per tal de complementar, amb les possibilitats que ofereix Internet, l'educació tradicional que ja desenvolupaven. Així que, el lector podrà copsar que si a les organitzacions que ja operaven plenament *on-line* s'hi afegeix aquesta nova fornada de plataformes educatives, el volum de dades educatives (interaccions, resultats...) que s'acaba generant és colossal. Aquesta informació massiva té un potencial enorme però, com és veurà al capítol següent, el seu tractament (seguiment, recol·lecció, processament...) només és abastable des del camp de la Big Data.

⁴ La principal diferència entre un LMS i un EVA és l'aplicació que se'n fa, de manera que una mateixa eina pot trobar-se en dues classes diferents segons quina sigui l'estratègia pedagògica que estableixi la institució (instrucció o aprenentatge constructiu).

2.4 Evolució i tendència de l'e-learning: des de la generació

1.0 a la 4.0

Hi ha un clar paral·lelisme entre l'evolució d'Internet i el desenvolupament que ha experimentat l'e-learning. Tant és així que la denominació que s'utilitza per a descriure les diferents etapes evolutives de l'e-learning, té una clara correspondència amb la que s'utilitza per a classificar les diferents fases de transformació que ha patit la Web en el transcurs del temps. Karrer (Karrer, 2007) distingeix tres etapes en la història de l'e-learning, que denomina: *e-learning* 1.0, *e-learning* 1.3, i *e-learning* 2.0 però a partir d'aquí molts són els autors que utilitzen aquesta denominació per parlar de generacions futures. En qualsevol dels casos cada generació representa una evolució de l'enfocament amb què s'utilitza la xarxa com a suport per a l'aprenentatge i alhora recull el conjunt d'atributs i tecnologies particulars que intervenen en aquest procés.

2.4.1 La creació d'un nou escenari educatiu: generacions 1.0, 1.3 i 2.0

En general, les primeres generacions de l'e-learning introduïen les TIC com a complement de l'educació més que no pas com a base d'una nova concepció de l'aprenentatge. Les tecnologies educatives entraven a l'aula per fomentar activitats accessòries al procés d'ensenyament, ja que la majoria d'elles reproduïen exactament formes i continguts de suports i metodologies implantades ja en l'escola tradicional. Més endavant, però, les TIC van anar evolucionant prou com per modificar substancialment el model educatiu, no tan sols adequant el material al suport (per fer-lo especialment per al seu ús en línia), sinó que també concebent una nova manera d'entendre el procés d'aprenentatge: l'alumne és actor protagonista del procés i tant materials com docents actuen com a suport i guia. A la Taula 2 es pot veure les diferents generacions d'e-learning des dels seus inicis i fins l'actualitat.

La primera etapa o generació 1.0 de l'e-learning va centrar els seus esforços en la implementació d'infraestructures tecnològiques. Hi havia la creença que la sofisticació d'un entorn tecnològic proporcionaria automàticament la tan anhelada qualitat del procés d'aprenentatge (Cabero, 2006), però el fet de deixar-se portar per la fascinació tecnològica i no revisar el model pedagògic va donar com a resultat la creació d'eines que reproduïen exactament metodologies del sistema educatiu tradicional. Així, la majoria de programari desenvolupat a l'època era de tipus instruccional i la digitalització de continguts seguia el format del llibre tradicional (Gros, 2011). Moltes de les eines educatives que van resultar d'aquesta etapa inicial, trobaven el seu suport en discs flexibles, CD ROM, DVD que estructuraven el contingut en blocs amb accés seqüencial, de manera que el procés d'aprenentatge, tot i que era individualitzat, seguia una trajectòria lineal. L'e-learning d'aquell temps es pot considerar més com a Ensenyament Assistit per Ordinador que no pas aprenentatge electrònic. I és que la inexistència de xarxes en el món educatiu durant aquesta etapa feia que els esforços se

centressin essencialment en el desenvolupament de paquets autònoms de contingut educatiu (Gómez, 2015).

Taula 2. Evolució de l'*e-learning* des dels seus orígens fins a l'actualitat. Taula desenvolupada a partir de les aportacions de (Gros, 2011) i (Seoane & García, 2010).

MODELS D'E-LEARNING	CARACTERÍSTIQUES PRINCIPALS
GENERACIÓ 1.0 Model centrat en la tecnologia	<ul style="list-style-type: none"> ▪ Implementació d'infraestructures tecnològiques ▪ Digitalització de continguts reproduint el format del llibre tradicional (es trasllada el context formatiu convencional al suport digital mantenint-ne l'estructura) ▪ Programari de tipus instruccional (desinterès pels models d'interacció i el factor humà de la formació). Prima la fascinació tecnològica.
GENERACIÓ 1.3 Model centrat en els continguts	<ul style="list-style-type: none"> ▪ Entorn virtual d'aprenentatge (model aula) ▪ Cerca d'un model de tutoria <i>on-line</i> ▪ Inici de la interactivitat ▪ Contingut generat especialment per al seu ús <i>on-line</i>
GENERACIÓ 2.0 (PRESENT) Model centrat en la metodologia	<ul style="list-style-type: none"> ▪ Té un marcat caràcter social ▪ Comunitats virtuals d'aprenentatge ▪ Posa émfasi en les eines col·laboratives

La segona generació (*e-Learning* 1.3) va posar el focus en la importància del contingut. Després de l'aparent estancament de l'enfocament excessivament tecnològic de la primera generació, els continguts prenen un paper protagonista i són desenvolupats especialment per al seu ús *on-line*. En part això va ser possible gràcies al desenvolupament massiu d'ordinadors i suports amb major capacitat d'emmagatzematge, fent factible així la incorporació d'elements multimèdia que permetien una major interactivitat i enriqueixen el contingut. Al mateix temps van sorgir les primeres representacions virtuals del concepte d'aula, molt allunyades però dels entorns socials actuals: l'usuari realitzava el procés d'aprenentatge completament fora de línia però disposava d'entorns virtuals en què es recreaven els tipus de relacions que s'estableixen a la formació presencial tradicional. En aquesta generació el model educatiu no pren excessiva rellevància i té una forta dependència de la plataforma desenvolupada.

L'any 2004, amb l'arribada de la web 2.0 es produeix una veritable revolució social i tecnològica. L'arribada de la Web Social marca un abans i un després en la manera com interaccionen els humans i aquesta mateixa transició acaba traslladant-se al model educatiu. El procés d'aprenentatge pren un patró de xarxa (Gros, 2011) on l'alumne gestiona i produeix coneixement en col·laboració amb d'altres estudiants i es concep l'educació com una activitat social. Al mateix temps s'accentua la implicació activa de l'alumne en el seu propi procés d'aprenentatge i aquest s'adapta a les seves necessitats personals (personalització). A la tercera generació d'*e-learning* (*e-Learning* 2.0), per tant, es fa especial incís en la metodologia del procés d'aprenentatge més que no pas en les eines tecnològiques que el possibiliten. Tot i així,

és una etapa molt prolífica pel que fa a tecnologia: els Entorns Virtuals d'Aprenentatge arriben a la seva màxima esplendor i es materialitzen en plataformes i campus virtuals que constitueixen veritables societats del coneixement en si mateixes.

2.4.2 El futur de l'e-learning: generacions 3.0 i 4.0

Tal i com s'ha vist a la secció anterior, sembla evident que la maduresa de l'*e-learning* està fortament lligada al component tecnològic. Els salts generacionals de l'*e-learning* troben la seva correspondència en les diferents eres del món web. Però, de la mateixa manera que el món educatiu ha estat menys permeable que la societat a l'hora d'introduir tecnologia en el seu dia a dia, retardant unes dècades la seva implantació, l'*e-learning* manté aquest desfasament respecte de la transformació que està experimentant la Xarxa. Així, si a nivell de la Web ja s'està en ple desenvolupament de la generació 3.0 (web semàntica) i elucubrant la quarta generació (web simbiòtica), l'*e-learning* tot just es troba explotant la tecnologia associada a la web 2.0 o web social. Malgrat això, el fet que l'*e-learning* reproduïx la trajectòria evolutiva de la Xarxa (Figura 8) permet que se'n pugui predir el progrés esdevenidor. I llavors, quin és el futur dels entorns d'aprenentatge?

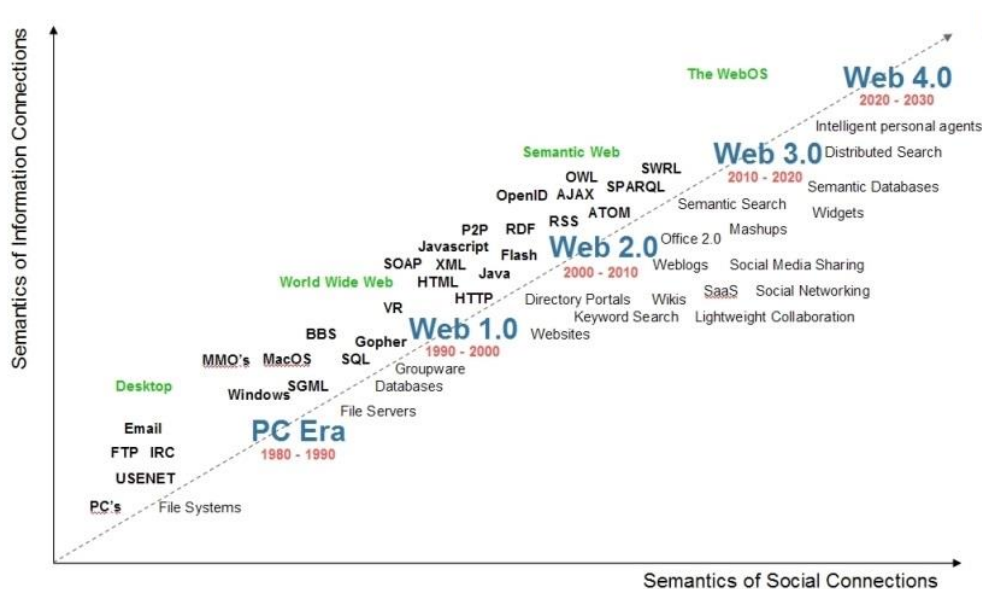


Figura 8. Representació de l'aparició dels diferents llenguatges i plataformes de programació, noves tecnologies i el moment associat d'evolució de la Web⁵.

⁵ Font: Radar Networks & Nova Spivack 2007. www.radarnetworks.com

Alguns autors ja comencen a definir les característiques de les properes generacions d'*e-learning* i qualsevol dels escenaris plausibles apunten a l'aparició de nous modes d'interacció i noves maneres de concebre l'educació (molt més oberta i universal). El tret més rellevant que tindrà el context, però, és el paper primordial que exercirà la Intel·ligència Artificial. Per tal d'entendre de quina manera els sistemes i agents intel·ligents acabaran formant part essencial del futur marc educatiu, resulta interessant posar la mirada en els diferents canvis que s'han introduït a la Xarxa des de l'aparició, anys enrere, de la coneguda com a web semàntica.

L'any 2006, el periodista John Markoff (Markoff, 2006) va encunyar per primera vegada el terme de web 3.0 (també coneguda com a web semàntica). Per aquell temps s'apuntava a la que seria una nova era d'Internet caracteritzada per l'addició d'una capa de significat a la Xarxa, concebut-la com una gran base de dades. Però el cert és que la web semàntica va molt més enllà d'aquella primitiva concepció estructural: és la unió de les tecnologies semàntiques (capaces d'interpretar el coneixement⁶) i de la pròpia representació del coneixement; atributs que en consonància donen lloc a entorns socials de computació (Aghaei, Ali, & Khosravi, 2012). Els entorns socials de computació, doncs, no són més que escenaris on humans i màquines cooperen per tal de descobrir, automatitzar i gestionar el coneixement d'una manera molt més eficient que l'actual.

Aquesta concepció de la cooperació humà - màquina és el que esdevé punt de partida per tal d'abordar el proper salt generacional de la Xarxa. Tot i que encara es tracta d'una idea en curs, la web 4.0⁷ sembla que es comença a dibuixar com un sistema simbiòtic. La idea principal és que humans i màquines interaccionin entre si sense fer distincions sobre la naturalesa d'un i d'altre. I és que si bé la web 3.0 promou les interaccions d'aquests dos actors, el cert és que ho fa en les formes: humà - humà, màquina - màquina; distingint-ne per tant la naturalesa i actuant com a sistemes separats però complementaris (cooperació). En canvi, el nou paradigma de la web 4.0 comporta una extensió d'aquest model d'interacció: busca unir intel·ligències per tal que persones i màquines es comuniquin entre si a un mateix nivell per generar una presa de decisions molt més eficient. El seu model d'interacció adopta, per tant, les formes: humà - humà, màquina - màquina, humà - màquina, màquina - humà. Es podria entendre doncs, que en aquest futur escenari humans i agents intel·ligents, més que cooperar, col·laboren.

Tot i que es pot percebre aquest escenari com una realitat remota, un entorn amb aquestes característiques no es preveu tan allunyat. Raymond Kurzweil (Kurzweil, 2009), reconegut tecnòleg i especialista en Intel·ligència Artificial, afirma que cap a l'any 2029 no hi haurà distinció entre humans i màquines, ja que la intel·ligència artificial haurà assolit en aquest


⁶ A diferència de la web actual, concebuda com una web de documents que només pot ser llegida per humans, la web semàntica es construeix amb un llenguatge específic que permet a les màquines poder-ne interpretar el contingut. Per a més detalls consulteu: Berners-Lee, T (1998) "Semantic web road map". [en línia] <http://www.w3.org/DesignIssues/Semantic.html>

⁷ La web 4.0 també és coneguda com WebOS en al·lusió al seu funcionament, que emula un sistema operatiu global.

moment el mateix nivell de desenvolupament que la humana. Segons Kurzweil les màquines intel·ligents seran capaces de combinar les habilitats en què ara els humans excel·leixen (essencialment el reconeixement de patrons) amb característiques pròpies en què ja són superiors, com la capacitat de recordar d'una manera precisa una gran quantitat de fets, cercar ràpidament en immenses bases de dades o descarregar coneixement.

En definitiva, sembla que el futur de la Web estarà fortament marcat per la interacció entre agents humans i no humans. I tot fa preveure que, tenint en compte el paral·lelisme amb què Internet i *e-learning* evolucionen, els futurs entorns d'aprenentatge incorporaran nous rols que s'hauran de repartir entre alumnes, docents i agents intel·ligents (Costa, Alves, & Rodrigues, 2013). I tot i que les properes generacions d'*e-learning* encara no estan desenvolupades i són models teòrics és interessant avançar-se a la seva arribada i observar-ne els trets distintius (Taula 3).

Taula 3. Evolució de l'*e-learning*. Taula construïda a partir de (Gros, 2011) (Seoane & García, 2010) (Moravec, 2009) i (Teten, 2007).

MODELS D'E-LEARNING	CARACTERÍSTIQUES PRINCIPALS
GENERACIÓ 1.0 Model centrat en la tecnologia	<ul style="list-style-type: none"> ▪ Implementació d'infraestructures tecnològiques ▪ Digitalització de continguts reproduint el format del llibre tradicional (es trasllada el context formatiu convencional al suport en línia mantenint-ne l'estructura) ▪ Programari de tipus instruccional (desinterès pels models d'interacció i el factor humà de la formació). Prima la fascinació tecnològica.
GENERACIÓ 1.3 Model centrat en els continguts	<ul style="list-style-type: none"> ▪ Entorn virtual d'aprenentatge (model aula) ▪ Cerca d'un model de tutoria <i>on-line</i>. ▪ Inici de la interactivitat ▪ Contingut generat especialment per al seu ús <i>on-line</i>
GENERACIÓ 2.0 (PRESENT) Model centrat en la metodologia	<ul style="list-style-type: none"> ▪ Marcat caràcter social ▪ Comunitats virtuals d'aprenentatge ▪ Posa èmfasi en les eines col·laboratives
 FUTUR	
GENERACIÓ 3.0	<ul style="list-style-type: none"> ▪ <i>Open knowledge</i> ▪ Educació construïda socialment i contextualment reinventada ▪ Ubiquïtat ▪ Interacció amb agents intel·ligents (cooperació)
GENERACIÓ 4.0	<ul style="list-style-type: none"> ▪ Interacció simbiòtica amb agents intel·ligents (col·laboració) ▪ Existència d'agents intel·ligents que orienten el procés d'aprenentatge d'una manera personalitzada (Tutors intel·ligents) ▪ Ús de raonament semàntic i heurístic per crear escenaris d'aprenentatge a temps real

Sembla que la generació 3.0 d'*e-learning* implementarà moltes de les característiques i atributs de la Web semàntica. Els entorns d'aprenentatge proporcionaran estructures d'informació contextual a base d'objectes etiquetats que podran ser de diversa tipologia (Teten, 2007). Així existirà la possibilitat que agents intel·ligents explorin informació suportada en altres formats, per exemple: imatges o veu. De la mateixa manera, els agents intel·ligents podran proporcionar recomanacions respecte els continguts educatius basant-se en les preferències de l'alumne, fent que el procés d'aprenentatge esdevingui una experiència altament personalitzada (Moravec, 2009). I és que principalment, a la generació 3.0, l'ús de tecnologies estarà enfocada a l'assistència d'alumnes i docents.

Les tecnologies educatives 4.0, per la seva banda, sembla que contindran Intel·ligència Artificial en qualsevol que sigui la seva aplicació. Actualment hi ha diverses visions sobre el que implicarà el nou escenari però aquest és un recull de les més destacades:

- Internet serà ubic: farà possible la connexió entre la realitat i la virtualitat en qualsevol moment, des de qualsevol lloc i amb qualsevol tipus de dispositiu, unint experiències entre persones i creant noves idees i tendències. L'aprenentatge serà una experiència immersiva (Moravec, 2009).
- La intel·ligència humana treballarà conjuntament amb formes avançades d'intel·ligència artificial i virtual, creant un nivell superior o *meta-nivell* de processament del coneixement (Tow, 2010).
- La comunicació entre agents humans i no humans donarà lloc a un continu intercanvi de patrons de comportament i habilitats (Costa, Alves, & Rodrigues, 2013).

Com passava en les darreres etapes evolutives d'Internet, una de les diferències principals que distingiran l'estadi semàntic del simbiòtic, dins l'evolució de l'*e-learning*, és el tipus de relació que s'establirà entre agents humans i no humans. El rol de cada actor del sistema, per tant, es veurà condicionat pel grau d'evolució d'aquestes interaccions (Taula 4):

- A la generació 3.0 s'establiran interaccions de tipus: docent - alumne, docent - agent intel·ligent, alumne - alumne, alumne - docent i alumne - agent intel·ligent. La interacció amb agents intel·ligents serà per tant unidireccional, d'agent humà a no humà. És a dir, la tecnologia intel·ligent treballarà complementàriament amb els processos humans però sense produir-hi interferències (cooperació).
- A la generació 4.0 el conjunt d'interaccions anterior es veurà ampliat amb les de tipus: agent intel·ligent - alumne, agent intel·ligent - docent i agent intel·ligent - agent intel·ligent. En aquest cas, l'entorn educatiu esdevindrà un espai de relacions simbiòtiques on tecnologia intel·ligent i processos humans col·laboren estretament al mateix nivell.

Taula 4. Generacions d'*e-learning* vs exemples tecnològics (Costa, Alves, & Rodrigues, 2013)

	1.0	2.0	3.0	4.0*
Meaning is	Dictated	Socially constructed	Socially constructed and contextually reinvented	Socially constructed, contextually reinvented and experiential
Technology is	Selected	Cautiously adopted	Everywhere	Everywhere (systemic)
Lecturing is	L/S	L/S; S/S	L/S; L/AA; S/S; S/L; S/AA	L/S; L/AA; S/S; S/L; S/AA; AA/L; AA/AA; AA/S
Classrooms location	Building	Building/online	Everywhere (infused society)	Everywhere (multidimensional society)
Lecturers are	Licensed experts	Licensed experts	Everybody	Everybody (including AA)
Hardware and software	Proprietary and costly	Open source with low cost	Low cost and used purposively	Low cost, used purposively through individual demand
Companies view of learners	Industrial workers	Like 1.0, despite a knowledge economy	Co-workers or entrepreneurs	Co-entrepreneurs

Legend: L- Lecturer| S- Student| AA- Artificial Agent

2.5 Resum i conclusions

En aquest capítol s'ha conclòs com el darrer salt tecnològic que ha experimentat la societat a través de les TIC ha transformat profundament el sistema educatiu pel que fa a la pedagogia. S'ha passat de la teoria instructivista de l'educació tradicional, a la constructivista, basada en un model molt més centrat en l'alumne. Així, la transmissió de coneixement a través de la instrucció ha donat lloc a la construcció del propi coneixement mitjançant l'adequació dels recursos formatius a les particularitats de cada individu. Si bé aquesta primera transició ha variat l'actor sobre el qual recau el procés de creació de coneixement, no ha estat fins ben entrada l'era digital que s'ha reformulat el concepte d'aprenentatge. I és que el moment social que ha desencadenat la *Web 2.0* ha descrit una nova manera de concebre aquest procés: ara es tracta d'una experiència social que té lloc en comunitat (connectivisme), en detriment de la concepció individualista del passat.

El desplegament de la teoria connectivista i la implantació de múltiples iniciatives d'*e-learning*, que troben el seu suport en les diferents tecnologies TIC, ha afavorit la proliferació d'Entorns Virtuals d'Aprenentatge. Això ha donat lloc a una experiència educativa molt més personalitzada i social. Actualment moltes institucions estan adoptant sistemes d'*e-learning* per a complementar la seva activitat docent presencial, per això tot fa pensar que l'aprenentatge evolucionarà cap a models mixtos (*blended learning*) en els que es combinarà activitat en línia i activitat presencial a parts iguals.

Per últim, com s'ha pogut copsar en les darreres seccions del capítol, l'evolució dels paradigmes educatius s'ha vist reflectida en els canvis que ha sofert l'*e-learning* des de la seva aparició fins l'actualitat. Veient doncs el paral·lelisme que s'ha donat amb l'evolució de la Web, tot fa pensar que el futur de l'*e-learning* reproduirà els models de la *Web 3.0* i de la *Web 4.0*, en què les màquines o agents intel·ligents seran actors fonamentals. És per això que, com es veurà al capítol 4, la Intel·ligència Artificial aplicada a l'educació serà un camp d'interès creixent en els propers anys.

Capítol 3

Analítica de l'aprenentatge: el tractament de la *Big Learning Data*

*“Information is the oil of the
21st century, and analytics is
the combustion engine.”*

Peter Sondergaard,
Gartner Research

Fins fa pocs anys era impensable poder arribar a recopilar, analitzar i emmagatzemar els grans volums de dades que circulen per Internet. Segons IBM, l'any 2012 s'estaven generant cada dia més de 2,5 quintilions de bytes d'informació i més del 90% d'aquesta informació havia estat produïda únicament durant els darrers dos anys (Barranco, 2012). Bona part d'aquesta descomunal quantitat de dades procedeix de múltiples dispositius i plataformes que formen part de la vida quotidiana dels humans (*smartphones*, sistemes GPS, xarxes socials ...) però molta altra és generada automàticament amb el que es coneix com comunicació màquina - màquina (per exemple usant sensors digitals). L'evolució de les tecnologies de la informació i comunicació (TIC) ha obert la porta al processament d'aquesta quantitat ingent de dades amb la finalitat de generar un coneixement que afavoreixi la predicció d'esdeveniments o la presa de decisions. Si bé durant molts anys aquesta analítica de dades massives ha restat en mans del sector empresarial amb el que es coneix com a Intel·ligència Empresarial (de l'anglès, *Business Intelligence*), poc a poc comença a fer-se lloc al món educatiu i això és, en bona mesura, gràcies a la proliferació d'Entorns Virtuals d'Aprenentatge (EVA).

Com es veurà en aquest capítol, el volum massiu de dades que es generen en els EVA, ha estès el concepte de *Big Data* per tal d'aplicar-se al món educatiu donant com a resultat la *Big Learning Data*. A la primera part del capítol es presenten les dues disciplines a què ha donat lloc l'anàlisi i processament d'aquestes dades massives: l'Anàlisi Acadèmica (AA) i l'Anàlisi de l'Aprenentatge (LA). Per últim, s'introdueix el model de referència de la LA, el qual recull les quatre dimensions crítiques que presenta l'anàlisi de l'aprenentatge i exposa diferents mètodes per abordar-lo amb la finalitat d'arribar a generar coneixement útil.

3.1 *Big Data* en l'educació

Les dades massives continuen creixent a un ritme sense precedents i s'estima que cap al 2020 cada individu generarà 1,7 megabytes d'informació nova per segon (Marr, 2015). Així doncs ens trobem davant d'una revolució que tot just està començant, però que certament tindrà un impacte colossal en les organitzacions i la societat dels propers anys. Sense anar més lluny, avui en dia són moltes les empreses que inverteixen considerables sumes de diners per tal d'implementar eines que els permetin recol·lectar i emmagatzemar dades de clients: transaccions, comunicacions, dades personals... Però quin sentit té recopilar i acumular tota aquesta informació? Quin és, per tant, el valor de la *Big Data*? Les respostes a aquestes preguntes variaran en funció de quin sigui el camp on s'apliqui, però en essència, la finalitat de la *Big Data* és la de millorar el procés de presa de decisions.

Moltes organitzacions estan basant el seu model de negoci en els següents beneficis derivats de l'ús de la *Big Data* a l'hora de prendre decisions (IBM Institute for Business Value, 2012):

- *Decisions més intel·ligents.* Aprofiten les noves fonts de dades per incrementar la qualitat de la presa de decisions.
- *Decisions més ràpides.* La captura i anàlisi en temps real de les dades permet donar suport a la presa de decisions en el moment en què realment poden tenir un impacte significatiu.
- *Decisions que marquen la diferència.* Centren els esforços de la *Big Data* cap a àrees que permeten una veritable diferenciació.

Un exemple molt paradigmàtic del potencial de la *Big Data*, és el de la coneguda cadena de supermercats americana *Wal-Mart* (Hays, 2004). Combinant dades del servei meteorològic amb les xifres de vendes de la cadena van descobrir que quan hi havia una alerta d'huracà augmentava la demanda de galetes de maduixa *Pop-Tarts* un 600%, ja que els consumidors emmagatzemaven provisions. En base a aquests resultats, la cadena va decidir començar a posicionar de manera destacada les galetes cada vegada que els serveis meteorològics alertaven d'un temporal, augmentant encara més les vendes. Així, l'extracció de coneixement assolida gràcies al creuament de dades massives procedents de dos sectors tan diferents com poden ser les vendes i la meteorologia, va permetre la detecció d'*insights* claus per definir l'estratègia comercial de la companyia.

A diferència del que pot semblar a partir de l'exemple anterior, la *Big Data* no tan sols permet prendre decisions que afecten als beneficis de grans corporacions. El sector sanitari, per exemple, en pot ser un gran beneficiari. Haig A. Peter, consultor d'*IBM Research*, afirma que actualment els facultatius diagnostiquen utilitzant únicament un 10% de la informació que hi ha disponible sobre un pacient (Montes, 2015). El lector podrà copsar, llavors, la millora que suposaria en la diagnosi el fet de poder treballar amb el 100% de les dades, magnitud únicament abastable des del camp de la *Big Data*.

El sector educatiu, per la seva banda, tampoc es queda impassible davant la revolució de les dades massives. Amb l'ús de la *Big Data* en l'educació, el procés d'aprenentatge deixarà de ser un ens monolític i immutable i donarà lloc al que es coneix com a aprenentatge adaptatiu. L'aprenentatge adaptatiu és un mètode d'ensenyament que utilitza un sistema computacional per crear una experiència personalitzada de l'aprenentatge. En aquest tipus d'aprenentatge el context educatiu s'adapta a les necessitats específiques de cada alumne, emprant el potencial de la *Big Data* per tal de fer que l'experiència de l'aprenentatge sigui totalment individualitzada. Darrerament, algunes organitzacions estan incorporant en els seus sistemes lectius eines que permeten treballar de manera adaptativa usant el potencial de la *Big Data*. Aquest és el cas de la Universitat Estatal d'Arizona, amb el seu programa de matemàtiques assistides per ordinador: Un *software* recull tota la informació i dades de cada estudiant: les notes, les habilitats i dificultats (inclús registra les seves vacil·lacions a l'hora d'usar el ratolí de l'ordinador). En acabar, uns algorismes predictius comparen les estadístiques amb les dades recollides entre desenes de milers d'estudiants i personalitzen el material en funció de les necessitats de cada individu (Sánchez, 2015).

L'anàlisi de les dades massives que es generen en els processos educatius mitjançant tècniques d'intel·ligència artificial, permetrà un coneixement valuós sobre l'estudiant i, per extensió, sobre la comunitat educativa. I és que les dades recopilades no tan sols descriuen la progressió de cada alumne, sinó que també ho fan del procés d'aprenentatge de tot el grup. Així, l'extracció de coneixement a partir de l'anàlisi de patrons sobre aquest volum de dades pot beneficiar la manera de fer front a futures casuístiques del procés d'aprenentatge. En aquest sentit, Kenneth Cukier (Cukier, 2013) descriu de manera exemplificant com aquesta informació extreta d'una experiència d'educació massiva pot ser emprada per millorar el procés d'aprenentatge de futures promocions :

"En donar una classe a Coursera⁸ sobre l'aprenentatge automàtic, el professor d'Stanford Andrew Ng va observar que al voltant de dos mil estudiants van contestar erròniament a una pregunta d'una activitat determinada; però tots presentaven exactament la mateixa resposta incorrecta. Clarament, tots estaven cometent el mateix error. Ara bé, quin era aquest? Investigant una mica, va descobrir que tots estaven invertint dues equacions algebraïques en un algorisme.

⁸ Plataforma d'educació virtual gratuïta desenvolupada la Universitat d'Stanford amb la finalitat d'oferir educació massiva a la població.

Així que ara, quan altres estudiants cometen el mateix error, el sistema no tan sols els comunica simplement que estan equivocats, sinó que els dóna una pista per tal que verifiquin els seus càlculs."

Així doncs, la *Big Data* té el potencial d'esdevenir un factor diferencial en l'educació del futur. Tot fa pensar que en els propers anys el processament i ús de dades massives educatives s'anirà traduint en un ensenyament més personalitzat i centrat en la millora permanent dels mètodes d'aprenentatge. Tot i això, aquest futur encara resta una mica llunyà i els experts pronostiquen que no serà fins d'aquí a 5 - 10 anys que la tecnologia de la *Big Data* estarà plenament implantada en els diferents sectors de la societat, incloent l'àmbit educatiu (Figura 9).

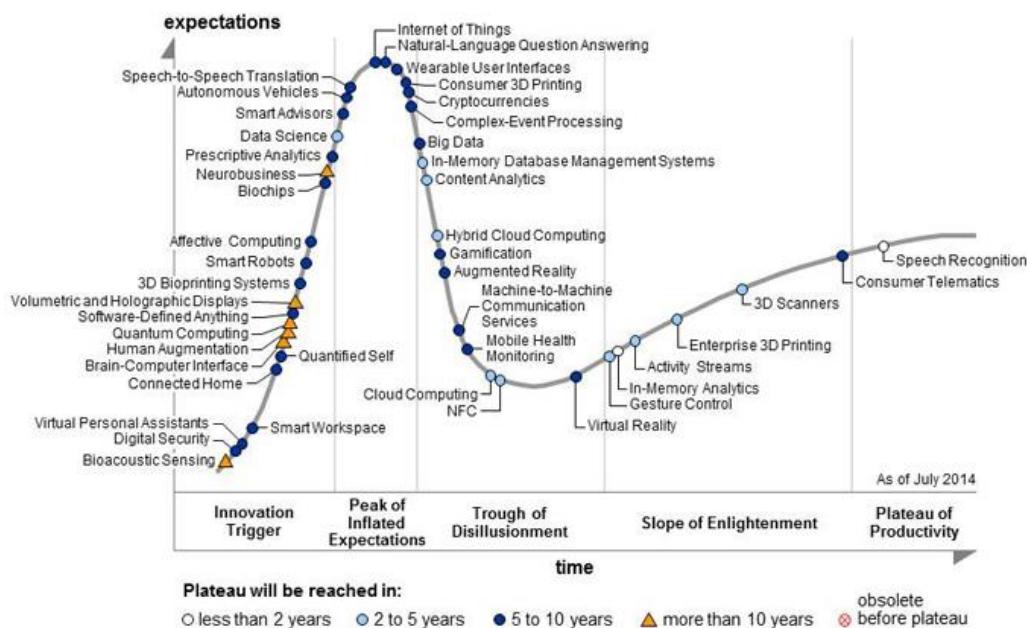


Figura 9. Cicle per a les tecnologies emergents (Gartner, 2014)

3.1.1 Big Learning Data

Quan un alumne interacciona amb el contingut d'un curs o una activitat dins d'un EVA, les seves accions van deixant un rastre digital imperceptible que, en certa manera, descriu el procés d'aprenentatge d'aquest individu: quines pàgines ha consultat, quins camins ha seguit per resoldre una activitat, quants intents ha necessitat per solucionar un problema, quanta estona ha estat aturat davant d'un repte, etc. El conjunt d'informació que es va generant a partir de les diferents interaccions, individuals i col·lectives, que tenen lloc dins d'un EVA s'anomena *Big Learning Data*. La *Big Learning Data* és, per tant, l'aplicació de la *Big Data* al món educatiu.

El terme *Big Learning Data* descriu tres dimensions fonamentals (Figura 10) sobre la informació (Masie, 2014):

- *Volum*. Representa la quantitat de dades i es refereix a les massives quotes d'informació susceptibles de ser recol·lectades. La *Big Learning Data*, per exemple, permet recopilar informació de milers d'estudiants que estan realitzant un mateix curs o activitat formativa o recollir múltiples dades sobre un mateix estudiant. Així, és possible generar una gran quantitat de coneixement que descriu com és el procés d'aprenentatge tant individual com col·lectiu.
- *Velocitat*. La velocitat amb què la informació és generada, processada i analitzada s'incrementa dia a dia. I tot i que actualment aquest procés no pot realitzar-se en temps real, es busca que en un futur la captura, emmagatzematge i anàlisi siguin instantanis. I que permeti, per exemple, orientar un alumne en cas que aquest estigui desenvolupant erròniament una activitat basant-se en les estratègies d'èxit de milers d'estudiants que s'hi hagin enfrontat prèviament.
- *Varietat*. La *Big Learning Data* permet vincular un gran nombre de variables de fonts i tipologia diferents: es tracten múltiples dades de diferent naturalesa (estructurades, no estructurades o semi-estructurades) procedents tant de fonts tradicionals com de no tradicionals. En el procés d'aprenentatge podrien ser, per exemple, dades sobre la progressió en l'aprenentatge, sobre les diferents tipologies de procediments o de comportaments, informació dels resultats que s'obtenen al llarg del procés d'aprenentatge, del nombre i tipologia d'interaccions...

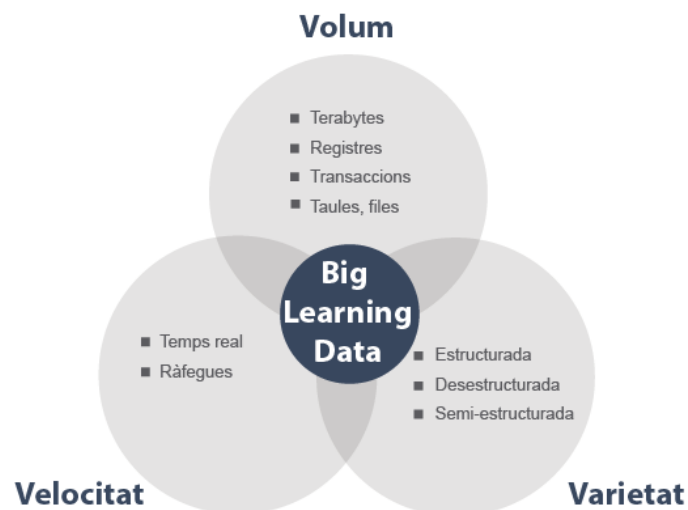


Figura 10. Dimensions de la *Big Learning Data*.

La *Big Learning Data* té un potencial enormement valuós per a les institucions educatives. Les dades han de ser vistes i tractades com si fossin l'essència de les organitzacions i no mers complements. I això només és possible si es crea una cadena de valor separada dels altres elements que formen la institució. Com es pot veure a la Figura 11 (i s'estudiarà més extensament a l'apartat 3.2.1), aquesta cadena consta de tres passos: la recollida de dades, la seva anàlisi i accions de postprocessament. Així, a l'entrada de la cadena hi haurà les dades brutes procedents de diferents fonts (EVA, STI, *social media*...), les quals passaran per un extens processament (eines de *Big Data* i anàlisi), per tal de donar-hi valor. Per tant, a mesura que aquestes dades es tractin aniran adquirint valor aconseguint que a la sortida de la cadena s'obtinguin valuosos *insights* que permetin millorar els processos educatius (Niemeijer, 2014).

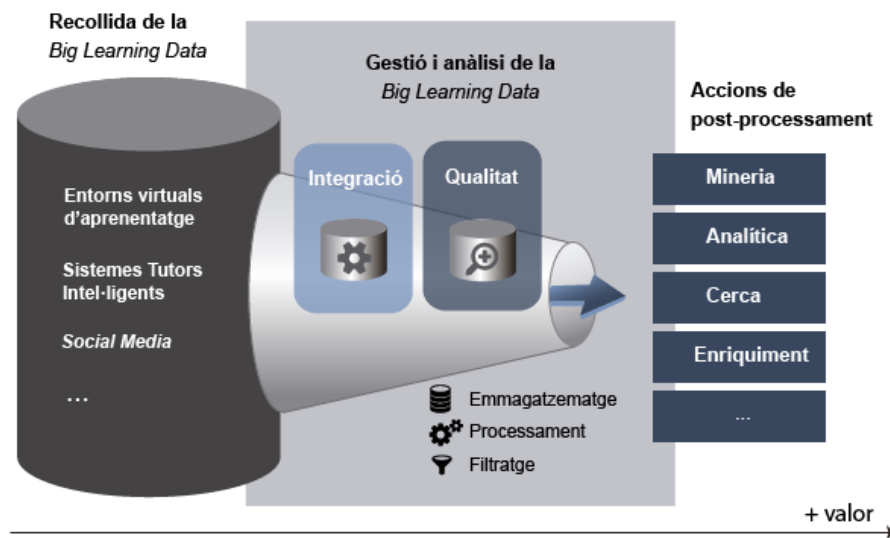


Figura 11. Cadena de valor de la *Big Learning Data*.

3.1.2 Analítica en l'educació

Tradicionalment s'ha parlat d'intel·ligència empresarial (de l'anglès, *business intelligence*) per descriure el conjunt d'estratègies, aplicacions i tecnologies enfocades a la creació de coneixement sobre l'organització mitjançant l'anàlisi de dades relacionades amb el negoci. Mitjançant la recopilació, emmagatzemament i anàlisi de dades massives es pretén millorar la presa decisions estratègiques. Quan aquest tipus d'analítica s'aplica a l'educació, es parla de dos grans sectors: analítica acadèmica i l'analítica de l'aprenentatge (Siemens, et al., 2001) (Taula 5).

Taula 5. Anàlisi acadèmica i anàlisi de l'aprenentatge. Adaptat de (Siemens, et al., 2001).

TIPUS D'ANALÍTICA	NIVELL O OBJECTE D'ANÀLISI	BENEFICIARI
Anàlisi de l'aprenentatge	Nivell personal: anàlisi del rendiment personal en relació als objectius d'aprenentatge, recursos i hàbits d'estudi de la resta de companys	Alumnes, educadors i personal docent
	Nivell curs: xarxes socials, desenvolupament conceptual, anàlisi del discurs, "currículum intel·ligent"	
	Departamental: modelat predictiu, patrons d'èxit/fracàs	Alumnes i educadors
Anàlisi Acadèmica	Institucional: perfils d'alumnes, rendiment acadèmic	Administradors, <i>marketing</i>
	Regional: comparació entre sistemes, qualitat i estàndards	Administradors
	Nacional i internacional	Govern nacional, OCDE, UNESCO

L'Anàlisi acadèmica o *Academic Analytics* (AA) és la millora dels processos d'organització, fluxos de treball i distribució de recursos. Analitza dades referents als usos que fa l'estudiant del sistema i té com a finalitat millorar l'eficàcia de l'organització (Siemens, 2010). Per exemple, les primeres iniciatives d'AA, desenvolupades als anys noranta, aplicaven mètodes d'anàlisi per detectar les possibles millores de les institucions educatives: posaven l'accent en la gestió de les matriculacions o en la predicció de l'èxit acadèmic dels estudiants (taxa de graduats) (Chatti, Dyckhoff, Schroeder, & Thüs, 2012). Per tant, aquest tipus d'anàlisi no treballa estrictament sobre el que s'entén com a procés d'aprenentatge, ja que tot i que pren dades sobre els processos associats a l'estudiant, s'enfoca a què es compleixin els objectius de les institucions educatives i no pas de l'alumne.

L'Anàlisi de l'aprenentatge o *Learning Analytics* (LA), per la seva banda, es basa en el mesurament, la recopilació, l'anàlisi i la presentació d'informació sobre els alumnes i els seus contextos, i té com a objectiu entendre i millorar els entorns d'aprenentatge en què es produeix (Gasevic, 2012). És a dir, l'anàlisi té posat el seu focus en millorar el rendiment de l'estudiant per tal que aquest assoleixi un major èxit en el seu procés d'aprenentatge. En definitiva, la LA és molt més específica que l'AA: el focus de la primera està exclusivament en el procés d'aprenentatge.

Ambdues disciplines tindran gran rellevància en el futur del panorama educatiu. I és que la idea és simple però alhora transformadora: l'anàlisi aplicada a l'educació proporcionarà nous models que permetran millorar la docència, l'aprenentatge, l'eficiència organitzativa i la presa de decisions. En aquest sentit Long i Siemens (Long & Siemens, 2011) apunten:

"Els intents d'imaginar el futur de l'educació solen posar l'accent en les noves tecnologies —dispositius ubics, dissenys d'aules flexibles, innovadores eines de

visualització...— Però el factor més crucial que donarà forma al futur de l'educació superior és una cosa que en realitat no podem tocar o veure: la Big Data i l'analítica. Basar les decisions en dades i evidències sembla increïblement obvi, i de fet, la recerca en aquest camp indica que la presa de decisions basada en dades millora els processos de l'organització i n'augmenta la productivitat."

3.2 L'analítica de l'aprenentatge

La proliferació d'Entorns Virtuals d'Aprenentatge (EVA), sistemes de gestió de l'aprenentatge (LMS), Sistemes Tutors Intel·ligents (STI) i Entorns d'Aprenentatge Personals (PLE) produeix una extensa quantitat de dades susceptibles de ser rastrejades. Però malgrat aquests entorns d'*e-learning* actualment ja emmagatzemen dades d'usuari de manera automàtica, la seva explotació per a millorar el procés d'aprenentatge és molt limitada. Aquí precisament resideix la importància de l'analítica de l'aprenentatge (LA) i és que aquests conjunts de dades educatives, ben processades, proporcionen noves oportunitats per avaluar teories d'aprenentatge, donar suport a l'alumne en el seu procés d'aprenentatge o crear sistemes educatius més òptims, oportunitats que només poden ser explotades amb les eines que proporciona l'analítica de dades massives (Greller & Drachsler, 2012).

Com genèricament ha estat definida, l'analítica és el processament de dades mitjançant anàlisi estadística i models explicatius i predictius que busca detectar *insights* que permetin prendre decisions i actuar sobre temes complexos (Díaz & Brown, 2012). Com a cas particular de l'analítica, doncs, l'analítica de l'aprenentatge (LA) consisteix en la recollida i anàlisi de dades sobre alumnes. Però de la mateixa manera que l'analítica té com a objectiu trobar *insights*, la LA no constitueix en si mateixa un objectiu final, sinó que es tracta d'un mitjà que proporciona als integrants d'un entorn d'aprenentatge (alumnes, docents, administradors...) un millor i més profund coneixement del procés d'aprenentatge per tal que hi puguin actuar. Així, la LA comprèn tant l'anàlisi del procés d'aprenentatge (a través dels clics, mapes de calor, anàlisi social de l'aprenentatge, sistemes de recomanació...) com l'acció sobre aquest (reelaboració de currículum, personalització i adaptació, predicció, intervenció o determinació de competències) (Siemens, 2010).

3.2.1 El procés d'anàlisi de l'aprenentatge

Les dades educatives són la base sobre la qual s'estructura el procés d'anàlisi de l'aprenentatge. Aquest procés s'inicia amb la captura de dades brutes, sense sentit, que després de ser pertinentment tractades es transformen en informació de gran valor. De manera genèrica, el procés de l'analítica de l'aprenentatge es basa en tres etapes (Chatti, Dyckhoff, Schroeder, & Thüs, 2012) que sovint es van succeint de manera iterativa (Figura 12):

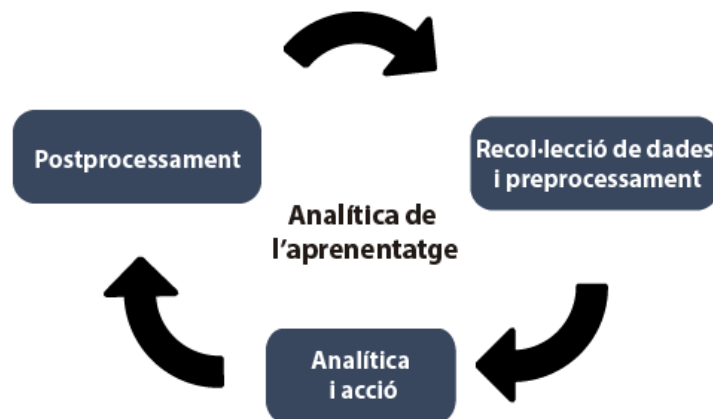


Figura 12. Procés de l'analítica de l'aprenentatge.

- *Recol·lecció de dades i preprocessament.* La primera etapa dins la LA és la recol·lecció de dades procedents de múltiples sistemes i entorns educatius. Aquest procés permetrà en fases posteriors extreure coneixement descobrint patrons útils, motiu pel qual resulta crucial seleccionar correctament les dades en el seu origen. A més, el volum de dades inicial pot ser excessivament gran o contenir atributs irrelevantes per al tipus de coneixement que se'n vol extreure, per tant, és necessària una fase de preprocessament abans de procedir a l'anàlisi. La fase de preprocessament o preparació de dades no tan sols permetrà seleccionar-les de manera més acurada sinó que alhora les adequarà al format específic que emprà cada mètode de LA. Per fer-ho possible, en aquesta fase se solen aplicar tècniques pròpies del camp de la mineria de dades: *data cleaning*, *data transformation*, *data reduction*, *data integration*... (Han & Kamber, 2006).
- *Anàlisi i acció.* Prenent com a base les dades preprocessades i en funció de l'objectiu que es pretén assolir amb la seva anàlisi (quin tipus de patrons es vol obtenir), s'apliquen unes tècniques o altres. En les tècniques d'anàlisi no tan sols s'inclouen l'anàlisi estadística o la visualització, sinó també les accions que es poden dur a terme amb aquesta informació, com són: el seguiment, la predicció, la intervenció, la recomanació... (Bienkowski, Feng, & Means, 2012).
- *Postprocessament.* Per aconseguir una millora de l'exercici de l'analítica cal post-processar les dades. Això inclou recopilar dades de fonts addicionals, determinar nous atributs, utilitzar noves mètriques, refinar el conjunt de dades inicial o variar la tècnica d'anàlisi (Chatti, Dyckhoff, Schroeder, & Thüs, 2012).

3.2.2 Model de referència de l'anàlisi de l'aprenentatge

Els models de referència o *frameworks* són abstraccions que defineixen un conjunt estandaritzat de conceptes i pràctiques amb l'objectiu de resoldre una problemàtica determinada. En el cas de l'anàlisi de l'aprenentatge, Chatti i altres (Chatti, Dyckhoff, Schroeder, & Thüs, 2012) proposen un model de referència que té com a propòsit esdevenir una guia que orienti el desenvolupament de nous mètodes i serveis de LA (Figura 13).

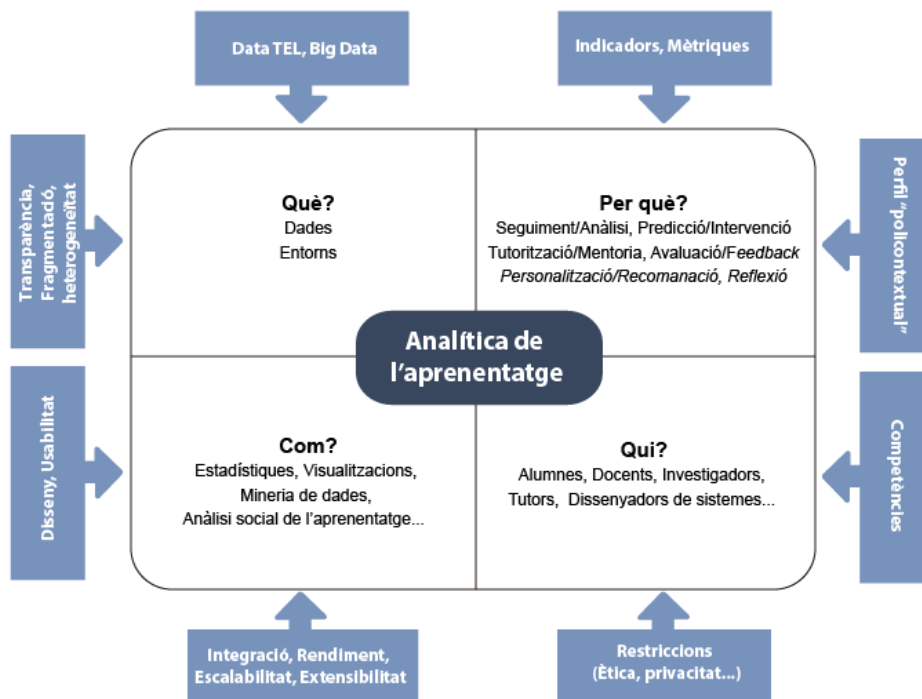


Figura 13. Model de referència de l'anàlisi de l'aprenentatge. Adaptat de (Chatti i altres, 2012).

El model identifica les quatre⁹ dimensions crítiques que han de ser contemplades a l'hora d'enfrontar-se al disseny d'aplicacions o mètodes per tal d'assegurar una correcta explotació del LA en benefici educatiu. Aquestes dimensions són:

- *Què?*: Quin tipus de dades recopila, analitza i usa el sistema per a l'anàlisi?

⁹ Greller i Drachsler (Greller & Drachsler, 2012) estenen aquest model de referència fins a les 6 dimensions, afegint-hi explícitament les àrees de *limitacions internes* (referint-se al nivell de competències que han de tenir els actors del sistema per poder avaluar correctament els resultats) i *restriccions externes* (factors que limiten l'aplicació de l'anàlisi de l'aprenentatge: legalitat, ètica, privacitat etc.).

- *Qui?:* A qui està dirigit l'anàlisi? Quins rols són els que participen del procés?
- *Per què?:* Amb quina finalitat s'analitzen les dades recopilades?
- *Com?:* De quina manera s'implementa l'anàlisi de les dades recopilades?

3.2.2.1 Dades educatives (el què)

La base de la LA és la informació, concretament les dades que es generen en un entorn educatiu. Aquest entorn pot ser de qualsevol mena: Entorn Virtual d'Aprenentatge (EVA) o *Learning Management System* (LMS), sistemes d'aprenentatge adaptatiu intel·ligent (incloent sistemes tutors intel·ligents (STI) o sistemes adaptatius hipermèdia (AHS)), cursos en suport web, *social media*, Entorns d'Aprenentatge Personal (PLE), etc. El repte de la LA és, per tant, trobar la manera d'agregar i integrar conjunts de dades brutes des de múltiples i heterogènies fonts, moltes vegades en formats diferents, per tal de crear coneixement educatiu útil.

Aquests conjunts de dades educatives, fonamentals per a l'anàlisi de l'aprenentatge, es poden classificar en diferents tipologies, entre les quals, les més significatives són (Santamaría, 2012):

- *Dades personals.* Informació sobre l'individu com, per exemple, el nom, l'edat,...
- *Dades d'interacció.* Dades sobre el comportament de l'usuari dins del sistema (missatges que deixa, rutes de navegació que segueix, accions que fa...).
- *Dades relacionals.* Connexions entre usuaris (seguidors, amics,...).
- *Dades de context.* Dades que mostren el context on es troba l'usuari (ubicació, sensors de moviment...).
- *Dades textuals.* Els continguts textuals creats pels alumnes, per exemple, poden ser avaluats amb tècniques de mineria de dades.

3.2.2.2 Interessats (el qui)

L'aplicació de la LA pot implicar a diferents actors: alumnes, docents, tutors o agents intel·ligents, administradors d'entorns d'aprenentatge, investigadors, etc. De manera general es poden distingir dos tipus d'interessats (Greller & Drachsler, 2012):

- *Els titulars de les dades.* Són els actors del sistema dels quals se n'està analitzant les accions. En cas que es tractés d'Analítica Acadèmica (AA) aquests podrien ser qualsevol persona o grup de la institució, però en la LA només es tracten alumnes i grups d'alumnes (classe).

- *Els clients de les dades.* Són els actors del sistema que es beneficien de l'anàlisi. En aquest cas pot tractar-se de qualsevol rol que té relació amb el procés d'aprenentatge. Per exemple, un alumne podria estar interessat en conèixer la seva progressió per tal d'avaluar si compleix o no amb les seves expectatives, un professor podria estar interessat en conèixer l'evolució dels diferents estudiants d'una aula, etc.

3.2.2.3 Objectius de la LA (el per què)

La LA no tan sols persegueix extreure coneixement a partir de dades educatives, sinó que també busca que hi hagi una acció posterior als resultats, és a dir, una aplicació d'aquest coneixement. Des del punt de vista dels diferents rols o participants de la LA existeixen diferents objectius o beneficis com poden ser (Siemens, et al., 2001) (Chatti, Dyckhoff, Schroeder, & Thüs, 2012) (Briggs, 2014):

- *Personalització.* La *Big Learning Data* canviarà la manera com es dissenyen als components (cursos, activitats...) dels EVA. L'alumne deixarà de ser un element genèric i uniforme i es procedirà a la personalització de components en funció de les seves necessitats específiques.
- *Seguiment i anàlisi.* El tractament de les dades massives que es generen en els processos educatius pot proporcionar molta informació sobre l'experiència d'aprenentatge dels alumnes. En ser possible rastrejar la cadena d'accions digitals que deixa un alumne quan interacciona amb els components d'un EVA, es pot arribar a traçar quin ha estat el seu procés d'aprenentatge. Així, a partir de l'anàlisi d'aquest procés per al conjunt d'estudiants, el docent pot detectar patrons que l'ajudin, per exemple, a implementar activitats o a corregir metodologies.
- *Predicció.* La predicció té com a finalitat desenvolupar un model que, a partir de les activitats realitzades per l'alumne, permeti pronosticar quin serà el rendiment en el procés d'aprenentatge futur. Aquest model pot ser emprat, per exemple, per subministrar ajuda addicional als alumnes que ho necessitin per tal d'assolir les expectatives educatives.
- *Tutorització o mentoria.* El fet de disposar de dades sobre el procés d'aprenentatge de l'alumne fa que sigui possible recomanar-li activitats dins d'un curs o fer un seguiment de les seves fites. Des d'una perspectiva més global, també permet orientar-lo al llarg del seva vida educativa ja sigui traçant itineraris o ajudant-lo a preparar nous reptes.
- *Avaluació i feedback.* El coneixement de la progressió de l'aprenentatge dels diferents alumnes que formen o han format part d'un EVA pot beneficiar no tant sols al propi alumne, sinó també a l'experiència educativa de futurs estudiants. Així, gràcies al coneixement generat per part d'altres individus es pot orientar un alumne en les seves accions: guiant-lo en la cerca de solucions, recomanant-li itineraris d'activitats....

- *Eficiència.* El tractament de dades massives educatives en permetrà l'extracció de coneixement, que podrà ser comunicat als actors dels EVA per tal que puguin prendre decisions i actuar tenint una major perspectiva sobre el context. Per exemple: si un alumne vol realitzar una activitat A, el sistema podria aconsellar-lo de realitzar abans una activitat B per tal de tenir un major èxit en la consecució de la primera. I és que la *Big Learning Data* pot avaluar el percentatge d'èxit en la metodologia d'assoliment d'aquesta activitat A tractant la informació recollida de l'experiència d'altres alumnes: quants han superat A sense fer B, quants fent B abans que A,... i orientar l'alumne en qüestió per tal millori l'eficiència del seu procés d'aprenentatge.
- *Motivació.* El fet que un alumne pugui percebre com la *Big Learning Data* pot ajudar-lo en el seu procés educatiu fa que mantingui un bon nivell de motivació. L'individu no s'arriba a trobar mai en escenaris dels quals no sap sortir i que acabarien generant-li frustració, ja que el sistema l'orienta per tal que no tinguin lloc aquest tipus de situacions. És a dir, es redueix el desgast que provocaria una situació així a través de la detecció precoç d'estudiants en risc i la generació d' alertes (tant per a alumnes com per a docents).
- *Coneixement del procés d'aprenentatge.* Gràcies a la *Big Learning Data* es podrà tenir un major coneixement dels diferents contextos educatius que es donen en un determinat entorn. Permetrà saber, per exemple, si una activitat era massa complicada o excessivament senzilla, quines parts d'un examen han costat més de superar, quins són els continguts que els estudiants consulten amb més freqüència, quines són les seccions més recomanades entre companys...

3.2.2.4 Mètodes d'anàlisi (el com)

El repte de la LA és dissenyar i desenvolupar eines estadístiques, de visualització, de filtratge i de mineria que siguin útils i usables per tal que els interessats puguin assolir els seus objectius analítics. En aquest sentit, els investigadors estan adaptant tècniques procedents de diferents camps com són: la mineria de dades educativa (Koedinger, D'Mello, & MCLAughlin, 2015), la gamificació, el marcatge automàtic, els sistemes de recomanació, la tutorització intel·ligent, la visualització de la informació, la computació lingüística o l'analítica social de l'aprenentatge, entre d'altres (UNESCO Institute for Information Technologies, 2012).

L'ús d'una o altra tècnica dependrà, en bona mesura, dels objectius que es pretenguin aconseguir. En qualsevol dels casos sembla que hi ha algunes tendències clares, pel que fa a metodologies. A (Chatti, Dyckhoff, Schroeder, & Thüs, 2012) es destaquen quatre grans àrees tècniques descobertes a partir de l'anàlisi de publicacions científiques on es manifestava l'ús d'alguna tècnica específica per a fer el procés d'anàlisi de l'aprenentatge i el de mineria de dades (*Data Mining*, de l'anglès) (Figura 14): l'aprenentatge automàtic, l'anàlisi estadística, la visualització d'informació i l'anàlisi social de l'aprenentatge.

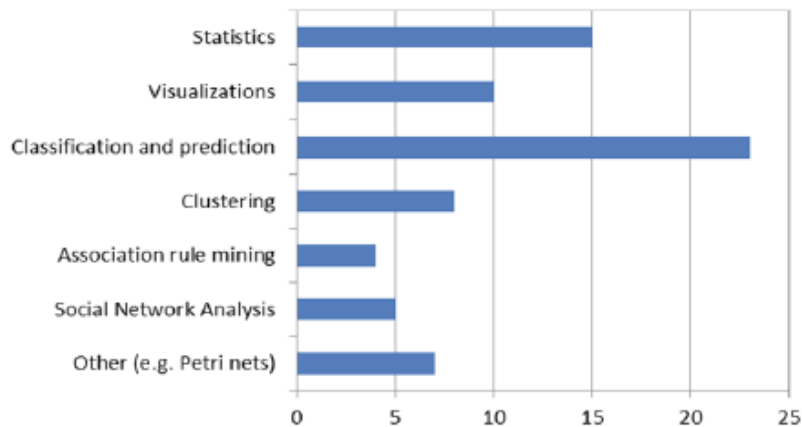


Figura 14. Mètodes aplicats a la LA en funció del nombre de publicacions científiques consultades.

- *Aprenentatge automàtic*: Durant el procés de Minería de Dades s'usen diferents tècniques d'aprenentatge com poden ser la classificació i predicció, el *clustering* o l'ús de regles d'associació (Romero & Ventura, 2010).
 - *Classificació i predicció*. És el procés de trobar un model o funció que descriu classes de dades o conceptes, i que tingui com a objectiu poder predir la classe d'altres objectes, la classe dels quals és desconeguda. El model derivat es basa en l'anàlisi d'un conjunt de dades conegudes (*training set*), com poden ser les classes prèviament conegudes.
 - *Clustering*. A diferència de la classificació i predicció, que partia de classes conegudes, el *clustering* treballa amb objectes dels quals no se'n coneix la classe. És un de les tècniques principals de l'*aprenentatge no supervisat*, com es veurà més endavant en aquesta tesi. El *clustering* és, per tant, l'organització de les dades en grups de manera que els integrants d'un grup presentin unes característiques similars entre si, però diferents dels integrants d'altres grups.
 - *Regles d'associació*. S'empren per a descobrir relacions entre variables.
- *Anàlisi estadística*. Moltes de les eines de gestió de l'aprenentatge actuals ja incorporen estadístiques bàsiques sobre la interacció que fa l'estudiant amb el sistema: el nombre de visites, el temps de connexió, el nombre de missatges penjats per estudiant...
- *Visualització d'informació*. Permet la representació de dades abstractes d'una manera visual i amigable per tal d'afavorir-ne la interpretació. Un exemple és el de la Figura 15.

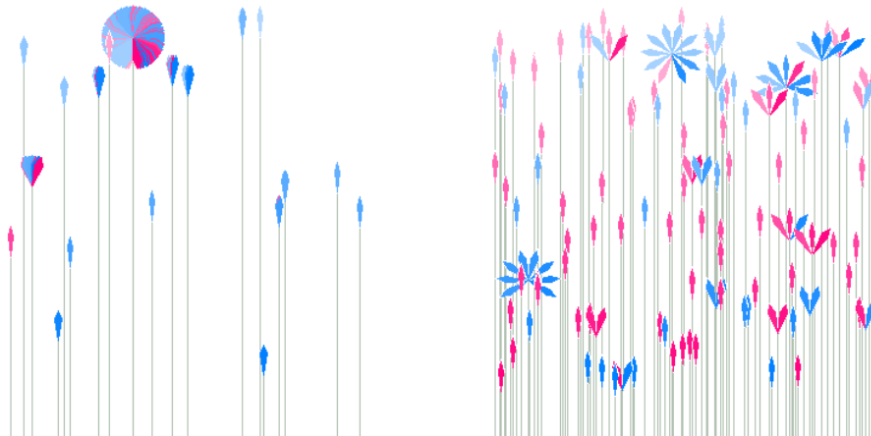


Figura 15. Representació visual de la participació en un fòrum (Xiong & Donath, 1999). A l'esquerra s'observa un grup amb una veu dominant, amb un gran nombre de respostes en blau. A la dreta un grup més democràtic, amb un gran nombre de posts inicials en magenta. L'alçada de les flors denota quant de temps ha estat l'usuari al tauler.

- *Anàlisi social de l'aprenentatge* (de l'anglès *Social Network Analysis*, SNA). Es tracta d'una disciplina basada en la teoria de xarxes o de grafs que permet visualitzar i analitzar les relacions que s'estableixen en les xarxes socials (Figura 16).

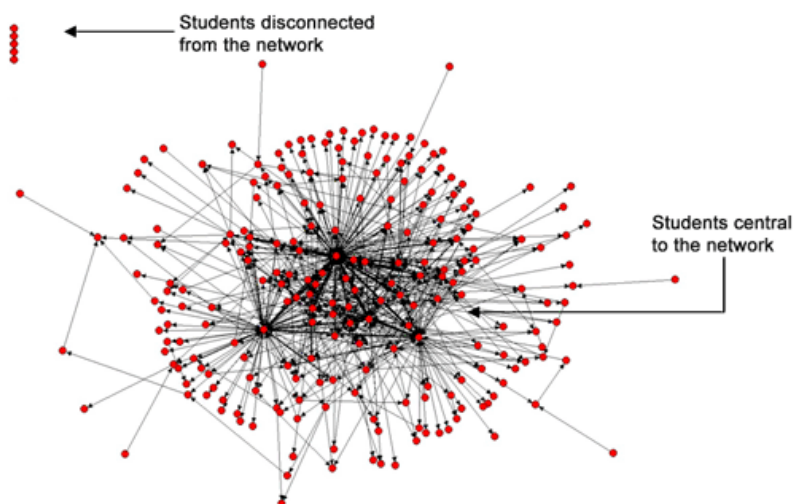


Figura 16. Diagrama SNA que mostra les connexions entre alumnes dins d'un entorn virtual (Dawson, 2010).

3.3 Resum i conclusions

S'ha constatat que en els darrers anys hi ha hagut un interès creixent en l'anàlisi automàtica de les dades que es generen durant el procés d'aprenentatge, és a dir, de la *Big Learning Data*. I és que la gran quantitat d'informació que es produeix a través de la interacció amb les tecnologies educatives pot ser de gran valor, sempre i quan es pugui processar de manera correcta per tal d'extreure'n coneixement.

Com s'ha vist en aquest capítol, l'Analítica de l'Aprenentatge (LA) implementa una metodologia de tres passos per a fer aquest tractament: la recollida de dades i preprocessament, l'anàlisi i acció sobre les dades i, per últim, el postprocessament. En la transició d'un pas a l'altre, les dades massives van adquirint valor fins que al final del procés esdevenen coneixement útil.

Com el lector ha pogut veure, es pot fer front a l'anàlisi de la *Big Learning Data* des de diferents punts de vista i l'ús d'una tècnica d'anàlisi o altra depèn, en bona mesura, del tipus de coneixement que es vol obtenir al final. Malgrat això, sembla que hi ha certes tecnologies que despunten respecte d'altres pel que fa al tractament de les dades massives, entre les quals hi ha: la mineria de dades, l'anàlisi estadística, la visualització d'informació i l'anàlisi social de l'aprenentatge.

Capítol 4

Sistemes educatius intel·ligents

“Children have to be educated, but they have also to be left to educate themselves”

Ernest Dimnet

L'aplicació d'Intel·ligència Artificial (IA) al món educatiu ha esdevingut en els darrers anys un camp d'interès creixent. No en va l'aplicació de tècniques específiques del camp de la IA en diversos contextos docents ha demostrat que pot afavorir la millora dels processos d'aprenentatge. Tant és així que al llarg dels últims anys hi ha hagut una notable proliferació de sistemes educatius intel·ligents (SEI). Els SEI són eines amb finalitat didàctica que sorgeixen dels esforços de la IA que, aplicada a l'educació, busca desenvolupar sistemes que s'adaptin a les particularitats de l'individu i tinguin un fort component de computació avançada. Exemples de SEI els trobem en tecnologies com són l'Aprenentatge Col·laboratiu Suportat per Ordinador (de l'Anglès, *Computer Supported Collaborative Learning*, CSCL), els sistemes de co-aprenentatge (de l'anglès, *Learning Companion System*, LCS), els sistemes d'avaluació automàtica o l'aprenentatge basat en jocs. Però si ha hagut una tecnologia rellevant en aquest àmbit durant els darrers vint anys ha estat la dels Sistemes Tutors Intel·ligents (STI).

Els STI han esdevingut importants eines pedagògiques que han demostrat una destacable efectivitat a l'hora d'incrementar el nivell d'aprenentatge i motivació dels alumnes (Anderson, Corbert, Koedinger, & Pelletier, 1995). Mitjançant la personalització i adaptació de continguts i estratègies pedagògiques, els STI actuen com a orientadors i tutors de l'alumne en el seu procés d'aprenentatge.

Al llarg de la seva existència s'han implementat múltiples iniciatives basades en la tutorització intel·ligent. Tanmateix, l'arribada de la *World Wide Web* ha suposat un important salt evolutiu en aquests sistemes. A partir de l'arribada d'Internet i la seva corresponent aplicació al món educatiu, han sorgit diverses tecnologies que implementen el model d'*educació basada en Web*. Així, els STI han anat evolucionat cap al que es coneix com a sistemes educatius intel·ligents i adaptatius basats en Web.

En aquest capítol es presenta l'evolució dels sistemes d'aprenentatge assistits per ordinador des de la instrucció assistida per ordinador (de l'anglès, *Computer-Assisted Instruction*, CAI) fins a la instrucció intel·ligent assistida per ordinador (de l'anglès, *Intelligent Computer-Assisted Instruction*, ICAI), que va suposar la primera interacció entre el món educatiu i el camp de la intel·ligència artificial. A partir d'aquí s'estudien amb més profunditat els Sistemes Tutors Intel·ligents (STI) i la seva evolució cap als sistemes educatius intel·ligents i adaptatius basats en Web (de l'anglès *Adaptive and Intelligent Web-Based Educational Systems*, AIWBES).

4.1 Antecedents i evolució dels sistemes educatius intel·ligents

La idea d'introduir eines informàtiques a l'educació es remunta als anys 50, quan les grans universitats americanes van començar a posar en marxa les primeres iniciatives d'Instrucció Assistida per Ordinador (de l'anglès, *Computer-Assisted Instruction*, CAI). Els sistemes CAI inicials eren coneguts amb el nom de *programes lineals*. Aquests sistemes mostraven el contingut de manera seqüencial, seguint un ordre preestablert que no podia ser alterat de cap manera, i l'alumne havia d'anar superant seccions per poder anar avançant per material. La metodologia de funcionament dels primers sistemes CAI doncs posava l'èmfasi en el material i no en l'alumne. Cap a finals de la dècada aquests sistemes van anar evolucionant cap als *programes ramificats*, més sensibles a la interacció de l'alumne. Aquests, igual que els programes lineals, tenien un nombre fix de temes però, a diferència dels primers, permetien certa flexibilitat d'accés als continguts. És a dir, l'accés a una o altra secció es duia a terme en funció de les respostes de l'alumne i no seguint un ordre fix preestablert.

L'arribada dels microordinadors, a finals dels anys 70, va provocar una notable expansió dels sistemes CAI a les universitats. Això va posar de manifest els beneficis que podien aportar les eines informàtiques al procés d'aprenentatge i en va sorgir una nova filosofia educativa: l'alumne aprenia d'una manera més eficient si s'enfrontava a problemes que tinguessin una dificultat adequada per a ell, més que no pas seguint explicacions sistemàtiques (Urretavizcaya, 2001). A partir d'aquesta filosofia els sistemes CAI van anar derivant cap a programes més adaptatius, que adequaven el contingut de les activitats al coneixement inicial de l'alumne.

Tot i les diverses evolucions experimentades pels sistemes CAI, no va ser fins els anys 80 que l'ensenyament assistit per ordinador va experimentar un gran salt qualitatiu. Per aquells temps, el camp de la Intel·ligència Artificial havia desenvolupat les primeres iniciatives que intentaven emular la lògica humana i la introducció d'aquests avenços al món educatiu va impulsar una

nova generació d'eines educatives intel·ligents formant una nova àrea anomenada ICAI (de l'anglès, *Intelligent Computer-Assisted Instruction*). Aquestes noves eines educatives derivades de l'aplicació de la IA als sistemes CAI rebien el nom de Sistemes Tutors Intel·ligents.

La missió dels STI és bàsicament la d'acompanyar a l'estudiant en el seu procés d'aprenentatge (Nesbit, Adesope, & Liu, 2014). I és que, a diferència dels sistemes CAI, en què l'ordinador és simplement un vehicle per a la instrucció, els STI permeten transformar la tecnologia (ordinador) en un instructor dinàmic i no pas un simple vehicle estàtic d'informació. Com es veurà més extensament a l'apartat 4.2, la base dels STI és tenir la capacitat d'afitar correctament el coneixement del domini (el que es vol ensenyar) i posseir el suficient coneixement sobre l'estudiant per tal de dirigir i adaptar el seu procés d'aprenentatge.

Taula 6. Paradigmes principals d'Intel·ligència Artificial aplicada a l'educació (Brusilovsky & Peylo, 2003).

	AI-CAI Paradigm	ITS Paradigm	AIWBES Paradigm
Time span	1970-ies	1980-1990-ies	End of 1990-ies – 2000-ies
Goal	Replace primitive CAI in transferring knowledge	Support problem solving	Comprehensive support
Context	Classroom without teachers	Classroom with a facilitator or self-study	Impendent self-study
Learning material	All learning material inside the system, most often presentations, but also exercises and problems.	No presentation material inside the system, but problems are often included.	Rich learning material on-line: presentations, examples, problems.
Technologies	Curriculum sequencing and intelligent solution analysis are the core technologies.	No course sequencing or adaptive hypermedia. Interactive problem solving support is the core technology.	Extensive use of adaptive hypermedia. Curriculum sequencing and intelligent solution analysis become widespread again. A range of Web-inspired technologies appears.
System completeness	All systems focus on single intelligent technology.	Most systems focus on single intelligent technology.	Most systems focus on several intelligent technologies.
Platform	Mainframes and mini-computers.	Personal computers	WWW

Els STI han estat i encara són eines didàctiques fonamentals en l'ensenyament suportat per ordinador i especialment pel que fa a l'aplicació d'Intel·ligència Artificial en educació. Tanmateix, l'arribada d'Internet ha suposat un nou ressorgiment de les tecnologies educatives. La interacció dels processos d'aprenentatge amb la *World Wide Web* ha donat lloc a noves maneres d'interaccionar amb el coneixement i noves formes de relació. I aquesta nova visió s'ha

vist reflectida, en bona mesura, en les noves fornades d'eines educatives intel·ligents. A la Taula 6 es pot observar l'evolució dels sistemes d'aprenentatge assistit per ordinador: des de l'aplicació inicial de la IA als CAI, fins els darrers sistemes d'aprenentatge intel·ligents i adaptatius basats en web (de l'anglès *Adaptive and intelligent Web-based educational systems*, AIWBES), que s'estudiaran més detalladament a l'apartat 4.3.

4.2 Sistemes tutors intel·ligents

Com s'ha introduït a l'apartat anterior, la incursió de la Intel·ligència Artificial al camp de l'educació ha donat lloc a una veritable revolució del model d'aprenentatge. Mitjançant la introducció de tècniques d'IA s'ha buscat potenciar el desenvolupament dels sistemes d'aprenentatge assistits per ordinador fins a assolir sistemes intel·ligents, altrament coneguts com a Sistemes Tutors Intel·ligents, capaços d'interaccionar amb l'usuari d'una manera personalitzada i participar activament en el seu procés educatiu.

A diferència dels sistemes tradicionals d'ensenyament, que conceben el procés d'aprenentatge com una experiència basada en la presentació i memorització de conceptes, els STI es basen en el concepte de *learning by doing* o aprendre a través de la pràctica. En els sistemes tradicionals un docent exposa una gran quantitat d'informació als alumnes, aquests la memoritzen i després són sotmesos a una sèrie d'avaluacions que miren de verificar la correcta adquisició del coneixement. I si bé és cert que la memorització pot resultar una eina vàlida per adquirir coneixement, aquesta no garanteix que els conceptes apresos se sàpiguen posar en pràctica correctament. Els STI miren de suplir aquesta carència creant entorns de d'aprenentatge on es requereix que l'alumne apliqui els seus coneixements i habilitats mitjançant el desenvolupament d'activitats específiques amb la finalitat, no tant sols d'avaluar l'adquisició de coneixement, sinó també que aquest es construeixi formant una estructura sòlida. Per tant, els STI permeten a l'alumne retenir coneixements i aplicar-los d'una manera més efectiva que els sistemes clàssics d'ensenyament (Amela, 2010).

Els STI, en definitiva, neixen amb la finalitat d'ajudar l'alumne en alguns dels seus reptes d'aprenentatge. I és que segurament van ser concebuts com l'evolució tecnològica del que seria el tradicional professor particular per a l'alumne: una eina capaç de guiar l'alumne en la seva progressió per tal que aquest millori el seu nivell en alguna matèria concreta. Un tutor intel·ligent, és llavors, " un sistema *software* que utilitza tècniques d'IA per tal de representar coneixement i que interactua amb els estudiants per tal d'ensenyar-los-el" (Vanlehn, 1988). En aquest sentit la seva missió és la de dirigir l'alumne al llarg d'un domini del coneixement, resolent durant el procés tasques com són l'elaboració d'una estratègia de tutorització, la generació d'exercicis a mida, la resolució pedagògica d'aquests exercicis i l'explicació de la solució (González, 2004).

Es considera que els primers Sistemes Tutors Intel·ligents van sorgir els anys 70 amb l'aparició dels sistemes *Scholar* (Carbonell, 1970), *Steamer* i *Why* (Stevens & Collins, 1977), i *Buggy*

(Brown & Burton, 1978). Posteriorment van anar evolucionant i en van sorgir altres més avançats com *Sophie* (Brown & Burton, 1982) o *Guidon* (Clancey, 1991), sempre amb la missió d'ensenyar una matèria concreta. En els darrers anys destaquen els nascuts dins d'universitats com ara l'*Andes* (Gertner & Vanlehn, 2000) a la Universitat de Carnegie Mellon o el *CircSim* (Shah, 1997) a l'Ilinois Institute of Technology. La Universitat de Canterbury també ha desenvolupat entre els anys 2003 i 2006 diferents STI per l'ensenyament de conceptes informàtics.

Més recentment han aparegut sistemes que modelen estudiants per tal de descobrir emocions en ells (Jaques, Conati, & Harley, 2014) o per tal que, per exemple, puguin aprendre una llengua estrangera (Aslan, Galip, Ozlem, & Mustafa, 2014) (Reddy & Sasikumar, 2014), física (Myneni, Narayanan, & Rebello, 2013) o robòtica (Belghith, Nkambou, & Kabanza, 2012). Aquests sistemes s'han provat també en entorns no universitaris, integrant-los a les aules físiques per millorar el procés d'aprenentatge dels alumnes (Dolenc & Abersek, 2015).

4.2.1 Estructura d'un Sistema Tutor

L'arquitectura que presenten els STI (Figura 17) i que ja s'ha apuntat a la introducció d'aquesta tesi, és molt variada i depèn en gran mesura de l'enfocament que s'hagi pres a l'hora de definir-ne el disseny conceptual. Tot i així es poden distingir certes característiques comunes a tots ells (Urretavizcaya, 2001):

- Presenten separadament la matèria que s'ensenyava (*coneixement del domini*) i les estratègies per ensenyar-la (*model pedagògic*).
- Caracteritzen l'alumne i recullen els seus resultats (*model de l'estudiant*) per a procurar-li un ensenyament individualitzat.
- Disposen d'una interfície que facilita la comunicació alumne-tutor (*model de comunicació*).

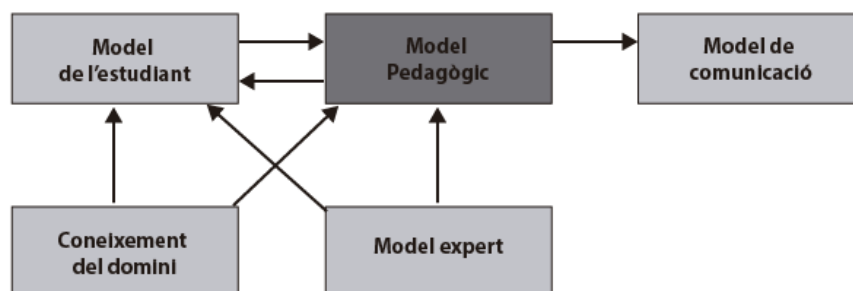


Figura 17. Mòduls que formen un Sistema Tutor Intel·ligent i les interaccions que s'hi donen.

En general tot i que reben diversa nomenclatura, els STI disposen d'aquests quatre mòduls: coneixement del domini, model pedagògic, model de l'estudiant i model de comunicació. Tanmateix, alguns autors (Self, 1990) (Beck, Stern, & Haugsjaa, 1996) hi identifiquen un cinquè component: el *model expert*. Sovint aquest model s'inclou com a part del coneixement del domini però aquí s'ha considerat que té prou entitat com per esdevenir un mòdul independent. Així, l'arquitectura d'un STI clàssic seria la que es mostra a la Figura 17.

4.2.1.1 Model de l'estudiant

El modelat de l'alumne és una característica fonamental que distingeix els CAI tradicionals dels STI, per la seva capacitat d'adaptació a les necessitats específiques de l'alumne (Cataldi & Lage, 2009). És a dir, el sistema és capaç de determinar quin és el coneixement del domini que posseeix l'alumne per tal de poder adequar l'estratègia d'aprenentatge i el tipus i nivell de dificultat de les activitats.

El mòdul de l'estudiant conté, per tant, el coneixement que caracteritza l'alumne i representa aspectes tan diversos com són el coneixement que té aquest sobre el domini o les habilitats mínimes que ha de posseir per tal de poder realitzar una tasca. Aquest mòdul s'encarrega també de recollir el comportament evolutiu de l'alumne durant diferents sessions de treball i generar un històric. Així és capaç de crear un model que descriu "què és el que sap i el que no sap" l'estudiant, per tal de poder adaptar posteriorment el sistema a les seves particularitats (González, 2004).

Existeixen diversos mètodes per a representar la informació sobre l'estudiant (Smith, Min, & Mott, 2015), però les tècniques més comunes són la *superposició de models* i les *xarxes bayesianes*. La *superposició de models* és la tècnica estàndard (Carr & Goldstein, 2012) i representa el coneixement de l'estudiant com un subconjunt del coneixement de l'expert (Figura 18).

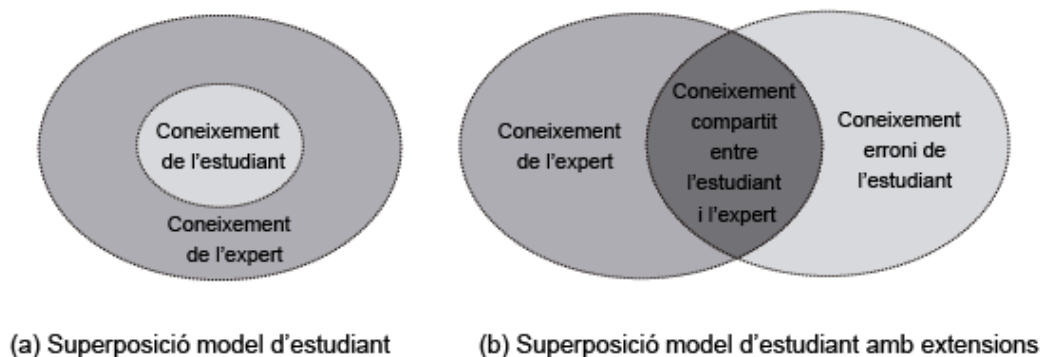


Figura 18. Superposicions del model d'estudiant

En aquest tipus de representació, el STI presenta material a l'alumne amb l'objectiu que el seu coneixement acabi assolint el nivell expert. En d'altres aproximacions se suggereix que el coneixement de l'alumne no ha de perquè estar inscrit dins del de l'expert, ja que el primer pot tenir conceptes erronis sobre un domini. En aquest cas el model representa aquest desviament amb una extensió i d'aquesta manera el STI és capaç d'actuar pertinentment per corregir el coneixement erroni de l'alumne (Beck, Stern, & Haugsjaa, 1996).

L'altra tècnica que s'utilitza per a determinar el coneixement de l'estudiant són les *xarxes bayesianes* (Martin & VanLehn, 1995) (Figura 19). En aquest cas, el coneixement s'implementa com un model probabilístic en un graf que relaciona l'estat del coneixement de l'alumne amb les interaccions amb el STI. Així, cada node del sistema esdevé una petita porció de coneixement a assolir per part de l'alumne.

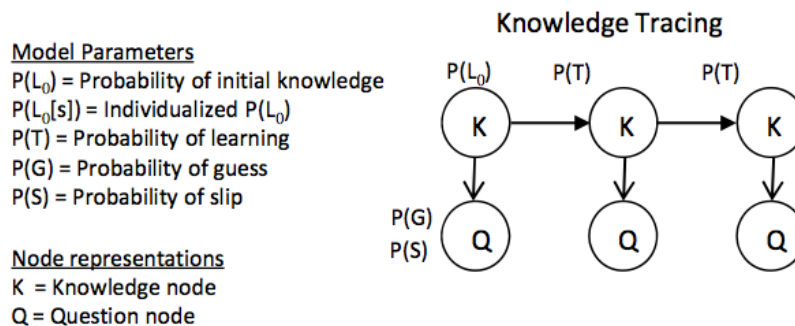


Figura 19. Model bàsic de xarxa bayesiana que permet fer un seguiment de l'adquisició de coneixement.

4.2.1.2 Model pedagògic

El model pedagògic conté la representació del coneixement expert en els àmbits relatius a processos d'avaluació, aprenentatge i metodologia d'ensenyament. És, per tant, el component que determina els plans instruccionals interpretant el model de l'estudiant respecte del currículum que reflecteix el mòdul del coneixement del domini (Urretavizcaya, 2001). En poques paraules, es pot dir que modela el procés d'ensenyament: és capaç d'entendre quina és l'evolució de l'alumne i pronosticar la seva evolució per tal de determinar quina ha de ser la seva fita següent.

La predicció de l'evolució de l'alumne i la presa de decisions són parts essencials d'aquest mòdul. Com es veurà més extensament al capítol 6, on s'efectuaran aportacions quant a tècniques i alternatives, els algorismes intel·ligents donen resposta a aquestes qüestions. Però de manera genèrica es pot afirmar que l'STI ha de prendre decisions sobre tres àmbits (Beck, Stern, & Haugsjaa, 1996):

- *Selecció del tema.* L'STI examina el model de l'estudiant per tal de determinar sobre quins temes cal que l'alumne treballi. Per exemple, si l'estratègia pedagògica indica que

cal fer incís en la revisió, l'STI selecciona un tema que l'alumne ja hagi vist. En canvi si indica que cal presentar nova informació, el sistema selecciona un tema que l'alumne no hagi treballat prèviament.

- *Generació del problema.* Un cop s'ha determinat el tema a tractar, el STI genera un problema per tal que l'alumne el resolgui. La dificultat d'aquest la determinaran tant el coneixement del domini com el model de l'estudiant.
- *Feedback.* Es tracta d'un nivell de decisió que afecta al moment en què l'alumne es troba amb alguna dificultat i no sap com continuar. En aquests casos l'STI decideix quanta ajuda ha de proveir a l'estudiant.

4.2.1.3 Model de comunicació

El model de comunicació s'encarrega d'establir la interfície entre el sistema i l'estudiant. És a dir, és el mòdul que permet a l'alumne comunicar-se amb l'STI. Aquest component controla les interaccions amb l'alumne i conté els mecanismes de representació (imatge, so, animacions, llenguatge oral i escrit...) necessaris que fan possibles la presentació d'informació d'una manera adequada per tal que aquest pugui realitzar amb èxit el seu procés d'aprenentatge.

Tot i que es tracta d'un mòdul més aviat perifèric, no per això resulta menys rellevant dins de l'STI. I és que l'èxit d'un programa educatiu no rau exclusivament en el coneixement que contingui, sinó que depèn en gran mesura de la riquesa comunicadora de què disposi (González, 2004). En aquest sentit, com demostren algunes investigacions en psicologia cognitiva (Mayer & Moreno, 2000), és fonamental tenir en compte alguns principis de disseny multimèdia per aconseguir potenciar el procés d'aprenentatge.

4.2.1.4 Coneixement del domini

Aquest mòdul recull tota la informació sobre l'assignatura o matèria que s'està ensenyant i pot incloure el propi temari, exercicis preparats, exàmens o qualsevol altre material docent. Entre els seus submòduls destaquen (Cataldi & Lage, 2009):

- *Paràmetres bàsics del sistema.*
- *Coneixements.* Són els continguts que s'han de carregar al sistema a través dels conceptes, les preguntes i els exercicis.
- *Elements didàctics.* Són les imatges, vídeos, sons i material multimèdia en general que es proporciona a l'alumne per tal que li sigui més fàcil adquirir el coneixement durant la sessió pedagògica.

Amb la introducció del model de Bolonya la manera com s'ha anat emmagatzemant la informació dins del coneixement del domini ha anat variant i ara els conceptes de competència i resultat d'aprenentatge són indispensables que s'hi vegin representats. A (Vernet, Nicolas, Golobardes, Fornells, & Garcia-Piquer, 2010) es presenta una metodologia alternativa per tal d'incloure aquesta informació que permet incorporar aquestes característiques essencials de l'educació com a submòduls del coneixement del domini.

4.2.1.5 Model expert

El model expert d'un STI presenta similituds amb el mòdul de coneixement del domini pel que fa a l'emmagatzematge de la informació que ha de ser ensenyada a l'alumne. Però, a diferència del primer, va molt més enllà de ser una simple representació del coneixement: es tracta d'un model sobre com algú amb habilitats sobre un domini representa el coneixement (Beck, Stern, & Haugsjaa, 1996). És a dir, descriu quin és el coneixement desitjable sobre una matèria i quines són les habilitats que cal tenir per tal assolir-lo.

El coneixement expert emmagatzemat en aquest mòdul permet a l'STI analitzar les accions de l'alumne i comparar-les amb les de l'expert amb la finalitat d'avaluar què és el que sap i el que no sap. Per tant la funció essencial d'aquest component es pot dir que és la de validar les respostes i aportacions de l'alumne.

4.2.2 Funcionament d'un STI

El cicle de funcionament de l'STI s'inicia quan el sistema identifica l'usuari. En aquest moment, el *model pedagògic* prepara la sessió de treball adaptant els continguts i la metodologia a les necessitats específiques de l'usuari. Per tal de poder realitzar aquesta personalització, el model pedagògic elabora un pla de sessió mitjançant la consulta del *model de l'estudiant*, el *model expert* i el *coneixement del domini*.

A partir del *coneixement del domini*, l'STI obté els conceptes que s'han d'ensenyar i del *model de l'estudiant* recupera l'històric d'activitats i les seves preferències. A partir d'aquests paràmetres i basant-se en la fita a assolir, gràcies a les especificacions del *model expert*, el *model pedagògic* determina quin és el coneixement previsible que té l'alumne i elabora el pla de treball d'acord amb aquest. Aquest pla de treball serà dirigit cap al *mòdul de comunicació* per tal que l'alumne pugui posar en pràctica les activitats planificades.

A mesura que es vagi desenvolupant la sessió, el *model pedagògic* anirà fent el seguiment de la progressió de l'alumne i n'avaluarà els resultats. Per a fer-ho emprerà el *model expert* sobre el domini i l'alumne en qüestió. Un cop realitzada l'avaluació, si el *model pedagògic* identifica un baix rendiment de l'alumne o que aquest necessita reforç en alguna àrea, replanificarà el pla de treball per adaptar-lo a les noves necessitats que presenta l'estudiant; en cas contrari continuarà desenvolupant la planificació preestablerta.

4.3 Sistemes educatius intel·ligents i adaptatius basats en web

Malgrat que els Sistemes Tutors Intel·ligents han ocupat, i encara ocupen, un lloc destacat dins l'ampli sector de les eines educatives intel·ligents, l'arribada d'Internet ha donat lloc a l'aparició de noves tecnologies educatives intel·ligents. Com es veurà a continuació, molts d'aquestes sistemes es conceptualitzen sobre els fonaments dels propis STI i implementen tècniques que ja s'utilitzaven en ells. De fet, els propis STI han anat evolucionant gràcies a la Web, revisant les seves característiques intrínseques per adaptar-se a la nova plataforma. En aquest nou context, les velles tecnologies s'han actualitzat i fins i tot ha sorgit una nova metodologia pedagògica, com és l'*educació basada en Web*¹⁰.

La investigació i el desenvolupament de sistemes d'aprenentatge a través d'Internet, o el que és el mateix, de sistemes educatius basats en Web, és una àrea que ha experimentat un notable creixement en els últims anys. I és que el benefici que suposa no tenir dependència d'una aula física ni d'una plataforma específica fa que aquesta tecnologia presenti un gran potencial per a l'educació del present i del futur. Tanmateix, bona part dels recursos educatius que existeixen actualment a la Web no són més que pàgines estàtiques que no ofereixen cap mena d'interactivitat ni adaptabilitat. Com declaren Brusilovsky i Peylo (Brusilovsky & Peylo, 2003), es tracta de recursos educatius basats en Web que presenten un enfocament "simplement-posat-a-la-Web". Els sistemes educatius intel·ligents i adaptatius basats en Web pretenen ser l'alternativa a aquest model de recursos educatius "estàtics".

Els sistemes educatius intel·ligents i adaptatius basats en Web (de l'anglès *Adaptive and intelligent Web-based educational systems*, AIWBES) busquen proveir d'una major adaptació construint un model d'objectius, preferències i coneixement de cada alumne per tal de poder-se adaptar a cada moment a les necessitats específiques de l'individu. Al mateix temps pretenen esdevenir "eines intel·ligents" capaces de realitzar tasques tradicionalment desenvolupades per agents humans, com poden ser la tutorització o la detecció de desviaments en l'aprenentatge.

Com es pot veure, la definició dels sistemes AIWBES no dista gaire de la que s'ha vist en les anteriors seccions respecte de les primeres tecnologies adaptatives i els Sistemes Tutors Intel·ligents. En part això és així perquè els AIWBES desenvolupen aquesta mateixa tecnologia però en una nova plataforma com és Internet. D'aquí que, com es veurà a l'apartat 4.3.1, es parli de tecnologies clàssiques aplicades als nous sistemes basats en Web. Els AIWBES, però, no només tracten amb actualitzacions de velles tecnologies sinó que, com es veurà a l'apartat 4.3.2, el propi entorn ha fet sorgir noves tècniques (tecnologies modernes) derivades de camps de recerca vinculats estretament amb la *World Wide Web*.

¹⁰ Alguns autors utilitzen el terme d'*e-learning* com a sinònim de l'*educació basada en Web*, però de fet, el primer es tracta d'una aplicació pràctica dels conceptes que defineix aquesta metodologia. L'*e-learning* cobreix un ampli ventall d'aplicacions i processos com són, per exemple, l'aprenentatge basat en web, l'aprenentatge suportat per ordinador i la col·laboració digital (Devedzic, 2006).

4.3.1 Tecnologies clàssiques aplicades als AIWBES

Donada la gran varietat de les tecnologies associades als AIWBES, aquests sistemes sovint es classifiquen segons el camp d'origen de les tècniques que empren. Així, de manera general es pot dir que els AIWBES tenen les seves arrels en dues tecnologies clàssiques (Figura 20):

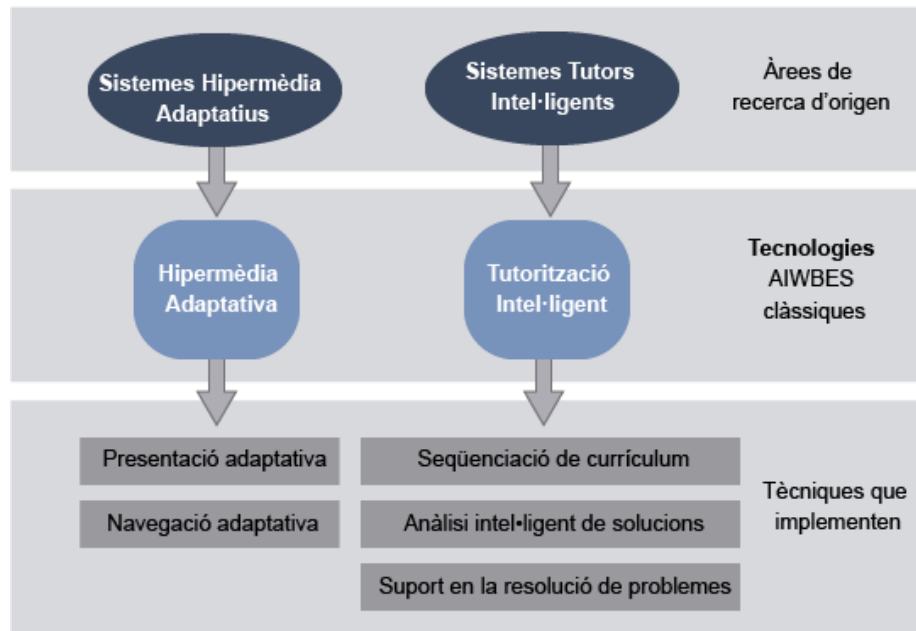


Figura 20. Tècniques d'adaptació emprades en els AIWBES procedents de tecnologies clàssiques (STI i HA). Diagrama realitzat en base a (Brusilovsky & Peylo, 2003) i (González, 2004)

- **Sistemes tutors intel·ligents.** Com s'ha vist anteriorment, els STI tenen com a finalitat personalitzar el procés d'aprenentatge de l'alumne, orientar-lo en la seva progressió a través de l'experiència de l'aprenentatge i proveir-lo de recursos per tal que pugui superar amb èxit les dificultats amb què es trobi en el procés. En aquest sentit, les tecnologies més representatives que aporten els STI als AIWBES són:
 - *Seqüenciació de currículum.* El seu objectiu és proveir l'alumne de la seqüència més personalitzada possible d'activitats, continguts i eines de suport per tal que faci el seu procés d'aprenentatge de la manera òptima.
 - *Anàlisi intel·ligent de solucions.* Té com a finalitat avaluar la correctesa de les respostes de l'alumne, indicant no tan sols la seva validesa (correcta/incorrecta), sinó també si es tracta d'una resposta incompleta. En aquest darrer cas, el

sistema és capaç de detectar quines són les peces incorrectes de la resposta i donar un *feedback* dels errors a l'alumne.

- *Suport en la resolució de problemes.* El seu objectiu és proveir l'alumne amb ajuda intel·ligent per tal que sigui capaç de enfrontar-se amb èxit a un problema.
- **Sistemes Hipermedia Adaptatius (HA).** Els HA tenen com a objectiu incrementar la funcionalitat de la hipermedia clàssica mitjançant l'accés personalitzat a la informació (González, 2004). A través d'un model de domini (que estructura els continguts) i un model d'usuari (amb els objectius, preferències i coneixement de l'individu), la HA construeix un model d'adaptació per tal d'adaptar la presentació d'informació segons les característiques de l'alumne i de les seves necessitats. Les diferents maneres d'adaptar un sistema hipermedia a l'usuari són:
 - *Adaptació de contingut.* La seva finalitat és adaptar el contingut que es presenta a cada node hipermedia (pàgina) a les característiques del model d'estudiant emmagatzemat al sistema (que conté els objectius d'aprenentatge, les preferències i coneixements previs de l'alumne).
 - *Adaptació de la navegació.* Té com a objectiu facilitar la navegació de l'alumne pels continguts canviant l'aparença dels enllaços per tal de crear un "camí òptim" per a l'aprenentatge.

4.3.2 Tecnologies modernes aplicades als AIWBES

En una primera revisió de la classificació original, vista a l'apartat anterior, (Brusilovsky, 1999) va identificar un grup complementari amb tecnologies inspirades en Internet i que no tenien relació directa amb els sistemes educatius pre-Internet (com són les tecnologies clàssiques exposades anteriorment). Tanmateix, en posteriors actualitzacions (Brusilovsky & Peylo, 2003) aquest grup ha estat dividit en tres subcategories que són les que s'indiquen a continuació:

- **Filtratge adaptatiu de la informació.** Es tracta d'una tecnologia procedent del camp de la Cerca i la Recuperació d'Informació (de l'anglès, *Information Search and Retrieval*, ISR) i consisteix en trobar un conjunt d'ítems rellevants per als interessos de l'usuari dins d'un conjunt de documents basats en text.
- **Aprenentatge col·laboratiu intel·ligent.** Consisteix en un conjunt de tecnologies sorgides a partir de la combinació de tècniques procedents dels STI i de l'aprenentatge col·laboratiu suportat per ordinador (CSCL). Un exemple en seria la formació adaptativa de grups de treball, el suport adaptatiu per a la col·laboració o els estudiants virtuals. Nombrosos autors tenen en compte aquesta característica (Olsen, Aleven, & Rummel, 2015).

- **Seguiment intel·ligent de classe.** Són sistemes que permeten rastrejar les diferents accions de l'alumne dins l'entorn d'aprenentatge.

Els cinc grups resultants de tecnologies AIWBES i els seus camps d'origen poden veure's a la Figura 21. En un primer nivell es poden observar les àrees on s'han originat aquestes tecnologies. D'una banda, es poden veure les àrees dels Sistemes Hipermedia Adaptatius i dels Sistemes Tutors Intel·ligents, l'origen de les quals està vinculat a l'aparició dels ordinadors personals. De l'altra banda, es troben les àrees vinculades a l'aparició d'Internet. Aquestes busquen donar resposta tant a la gestió de les dades massives que es generen a la Xarxa com al desenvolupament del seu màxim potencial. Així, en aquest grup es troben àrees com la Cerca i Recuperació d'Informació, l'Aprenentatge Col·laboratiu suportat per ordinador o l'Aprenentatge Automàtic i la Minería de Dades. En un segon nivell es poden apreciar les tecnologies que deriven de l'aplicació d'aquestes àrees al món Web i que esdevenen disciplines de recerca. Finalment, en un tercer nivell, s'exemplifiquen algunes de les tècniques que implementen aquestes tecnologies o àmbits de recerca per tal d'assolir el seu propòsit.

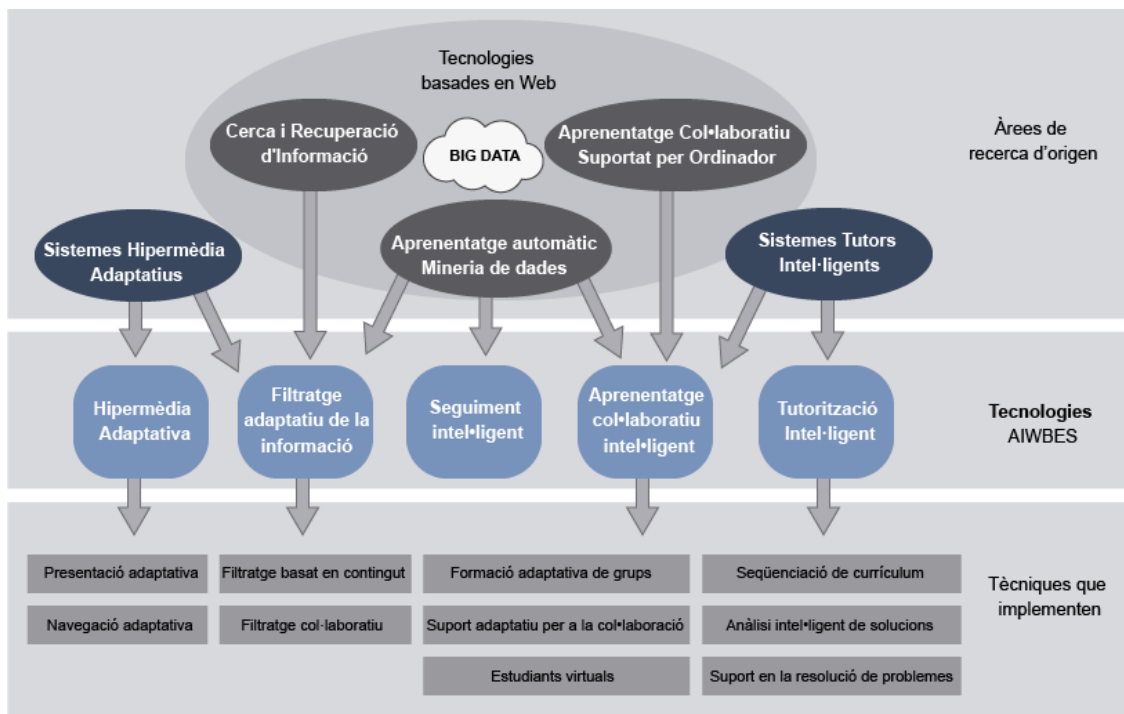


Figura 21. Els cinc grups de tecnologies AIWBES modernes i les seves tècniques principals. Diagrama elaborat a partir de (Brusilovsky & Peylo, 2003).

4.4 Resum i conclusions

L'aplicació de tècniques específiques de la IA en el món educatiu ha donat lloc a l'aparició de nombrosos sistemes educatius intel·ligents (SEI), el màxim exponent dels quals ha estat la tecnologia dels STI. Com s'ha vist en aquest capítol, els STI van ser creats amb la finalitat de guiar l'alumne al llarg d'un domini del coneixement. Tanmateix els STI són concebuts com eines individuals, personalitzades, on la capacitat de treballar en grup és inexistent. Serveixen tant a l'alumne com al professor per fer-ne el seguiment, però no estan pensats per valorar el rendiment d'un grup o d'una aula virtual.

L'arribada d'Internet, però, va suposar un important progrés en els SEI i fruit d'això han sorgit diverses tecnologies d'*educació basada en Web*. Com a conseqüència d'aquest nou escenari, l'arquitectura dels STI s'ha anat actualitzant i ha donat lloc al que es coneix com a *sistemes educatius intel·ligents i adaptatius basats en Web*. A més, s'han incorporat noves tècniques procedents de tecnologies basades en Web, que d'alguna manera permeten treballar l'aprenentatge intel·ligent des d'un punt de vista grupal o col·laboratiu.

Tot i que es tracta d'una tecnologia emergent, i en molts casos només ha estat desenvolupada de manera conceptual en certes àrees de recerca, el cert és que d'aquí a pocs anys l'*e-learning* tindrà la maduresa suficient per incorporar aquest tipus d'intel·ligència en els seus processos. És per això que a l'apartat que segueix es presenta un model que vol donar resposta a la tutorització dels alumnes i grups que formen una comunitat virtual d'aprenentatge.

Capítol 5

TICVA: Tutorització Intel·ligent de Comunitats Virtuals d'Aprenentatge

“There are two fundamental equalizers in life – the Internet and education.”

John Chambers,
CEO of Cisco Systems

En capítols anteriors s'ha vist com l'educació ha evolucionat molt en les darreres dècades. L'aplicació de la tecnologia al món educatiu ha produït nous entorns d'aprenentatge que han permès situar l'alumne, no com un simple espectador del procés educatiu, sinó com un actor més dins de la xarxa que forma l'experiència de l'aprenentatge. En definitiva, a través dels capítols precedents, s'ha pogut copsar com l'evolució de la tecnologia ha portat al redisseny i reenfortament dels programes formatius tradicionals, donant com a resultat models i eines fortament fonamentats en l'adquisició de competències i la interacció entre individus.

Internet ha estat el màxim exponent d'aquesta revolució tecnològica i precisament la Web ha estat la plataforma de desenvolupament per excel·lència de moltes de les tecnologies educatives que han sorgit durant la darrera dècada. Amb l'educació basada en Web i els Entorns Virtuals d'Aprenentatge, nascuts d'aquesta filosofia, s'obre tot un ventall de possibilitats a l'hora de traçar estratègies pedagògiques que millorin el rendiment del procés d'aprenentatge.

Sembla evident, doncs, que els EVA presenten un gran potencial per al desenvolupament de processos d'aprenentatge ja que, si més no, aquests sistemes compten amb tecnologies que permeten tant la personalització de l'aprenentatge com la socialització en el mateix. De tota manera el grau de maduresa que ha assolit aquesta tecnologia en l'actualitat encara resulta insuficient per desplegar tot el seu potencial i, en aquest sentit, es pot afirmar que la investigació en aquesta àrea resultarà de gran interès a l'hora de dissenyar els escenaris d'aprenentatge futurs.

En aquest capítol es presenta un escenari d'aprenentatge que mira de combinar les millors característiques dels EVA amb les millors dels STI. Com a resultat s'obté un model que l'autor ha anomenat TICVA (Tutorització Intel·ligent de Comunitats Virtuals d'Aprenentatge) que busca encarar el procés d'aprenentatge virtual cap a la següent generació d'*e-learning* (*e-learning 4.0*). Al primer apartat es mostra la conceptualització del sistema TICVA partint del model de referència de l'analítica de l'aprenentatge. Tot seguit es presenta l'arquitectura del sistema desenvolupada a partir de l'estructura d'un STI clàssic i, finalment, es tracten breument les tecnologies actualment disponibles per tal d'encarar la futura implementació del sistema.

5.1 Conceptualització del sistema TICVA

Com s'ha vist al capítol anterior, la transició dels sistemes CAI als STI va comportar una reformulació de metodologia didàctica. Si bé els sistemes CAI, com el seu nom indica, es basaven en la instrucció de coneixements, la integració de la Intel·ligència Artificial va donar pas a una metodologia eminentment constructiva, que és la que implementen els STI.

Els STI tenen com a objectiu guiar l'alumne al llarg d'un domini de coneixement mitjançant l'adaptació d'activitats a les seves necessitats específiques. És a dir, un STI té com a característica fonamental la personalització de l'aprenentatge. Tanmateix, precisament per aquest motiu, els STI són concebuts com eines individuals on la capacitat de treballar en grup és inexistent. És cert que serveixen tant a l'alumne com al professor per fer-ne el seguiment, però no estan pensats per valorar el rendiment d'un grup o d'una aula virtual.

Tal com s'ha descobert en capítols anteriors, avui en dia, és impensable l'aprenentatge en línia sense el concepte de xarxa i treball en grup. És evident que la personalització de l'aprenentatge és un tret diferencial respecte models generalistes del passat, però la concepció de l'aprenentatge com un procés social no pot ser menystinguda. A partir d'aquestes premisses, es proposa un sistema en què aquests factors fonamentals convergeixen i permeten crear un entorn en què sigui possible produir un procés d'aprenentatge personalitzat i alhora social.

Com s'ha dit, el sistema TICVA mira de combinar el millor dels EVA amb el millor dels Sistemes Tutors Intel·ligents: del primer en pren la capacitat que ofereix el sistema de realitzar el procés d'aprenentatge en comunitat, com una experiència grupal; del segon, n'adopta la capacitat intel·ligent de predicció i adaptació a les necessitats de l'individu. Així en resulta un sistema capaç, no tant sols d'orientar intel·ligentment a un individu en el seu procés d'aprenentatge, sinó també de fer-ho a una comunitat.

En treballs anteriors (Vernet, Canaleta, Pallàs, Navarro, & Zaballos, 2013) (Vernet, Canaleta, & Pallàs, 2012) es va fer una primera aproximació del que hauria d'arribar a ser un sistema d'aprenentatge en línia amb aquestes capacitats. Així, l'estructura elemental i la conceptualització del sistema TICVA és la que mostra el diagrama de la Figura 22.

Sistema Tutor Intel·ligent

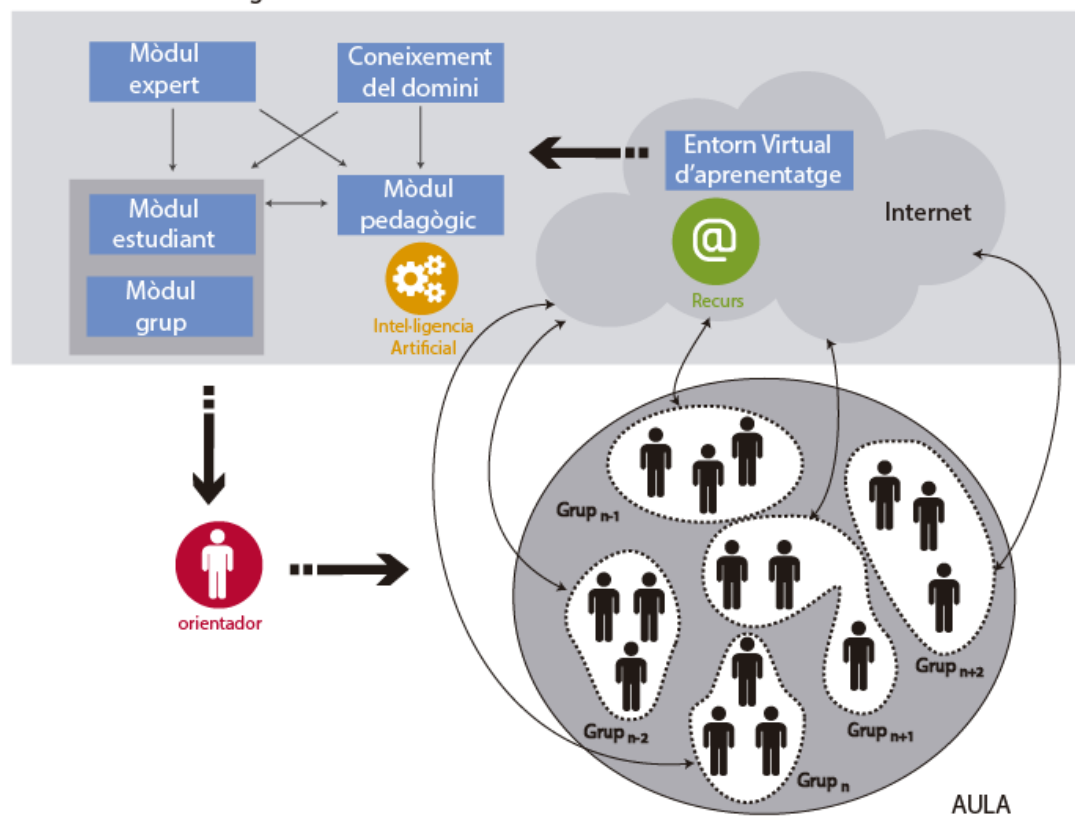


Figura 22. Conceptualització de l'arquitectura TICVA.

Com es pot observar, tot i tractar-se d'un plantejament a gran escala ja es poden identificar els trets distintius del que suposa introduir la tutorització intel·ligent als EVA. D'una banda, la relació dels alumnes amb l'EVA i la xarxa d'Internet en general i, de l'altra, el control de l'evolució d'aquests alumnes mitjançant el mòdul pedagògic del STI, el qual rep un *feedback* continu del seu procés d'aprenentatge. En paral·lel el professor/orientador, a partir de les dades enregistrades pel sistema, interacciona amb l'STI per actuar sobre la comunitat d'aprenentatge.

A més, es percep la importància del grup, i per extensió de la comunitat, en el procés d'aprenentatge de l'individu. Si observem la figura, el mòdul pedagògic es dota d'intel·ligència artificial per tal de controlar la comunitat virtual d'aprenentatge mentre que aquesta comunitat d'alumnes es reflecteix a la figura mitjançant els grups d'alumnes representats. Tot i que el lector pot tenir la primera impressió que la introducció d'algorismes intel·ligents en l'aprenentatge busca desposseir els professors d'un paper rellevant en l'orientació dels estudiants, el cert és que aquests pretenen ser més un complement que no pas un substitut, ja que, si bé la tasca de seguiment i orientació és relativament senzilla en un grup d'estudiants d'una aula física, el cert és que la translació d'aquesta al context digital en dificulta el procés. I és que, a diferència del que passaria en una aula física, en el món virtual és molt difícil que el docent pugui arribar a

saber a temps quines mancances tenen els alumnes o si funcionen bé els grups de treball. En aquest sentit la part intel·ligent del sistema pretén esdevenir una versió digital del docent, una mena d'avatar del professor, que l'informi dels fets que passen a l'entorn virtual i l'ajudi a la presa de decisions.

5.2 Un model que respon a preguntes

Modelar el sistema TICVA pot resultar complicat d'abordar si únicament es contempla la dimensió pedagògica. Perquè si hi ha un fet rellevant dels EVA és la immensa quantitat d'informació educativa que s'hi genera. En essència les dades que circulen per un EVA (interaccions, resultats, usuaris...) representen múltiples processos d'aprenentatge d'una forma confusa i abstracta. Però si d'alguna manera es pogués capturar aquesta informació i processar-la, el coneixement extret donaria una visió clara del que està passant a l'EVA.

Com s'ha vist anteriorment, un Sistema Tutor Intel·ligent és capaç d'assolir aquest nivell de coneixement en un entorn molt reduït, d'un sol individu. Però, què passaria si això es pogués ampliar a tota una comunitat? En tant que el sistema TICVA pretén emular el comportament d'un STI però per a més d'un individu, és necessari que pugui arribar a tractar la *Big Learning Data* que es genera en un EVA i, d'aquesta manera, poder tenir una idea clara del que està succeint a l'entorn. Per això cal enfocar el model tenint en compte també la dimensió analítica.

És per aquest motiu que s'ha decidit iniciar el modelat del TICVA a partir de l'Analítica de l'Aprenentatge i, per tant, del tractament de la *Big Learning Data*. Com s'ha vist al capítol 3, el procés de tractament de les dades massives comporta contemplar diferents dimensions de les mateixes per tal de poder determinar les tècniques més adequades per a dur a terme el procés d'extracció del coneixement. Per aquesta raó s'ha agafat com a punt de partida el model proposat per Chatti (Chatti, Dyckhoff, Schroeder, & Thüs, 2012) a l'hora de definir les bases del que hauria de contemplar l'arquitectura que es vol dissenyar.

Aquest model defineix les quatre dimensions crítiques (“Què?”, “Per què?”, “Com?” i “Qui?”) que cal tenir en compte en el moment d'abordar un disseny que pretengui analitzar processos d'aprenentatge.

A la Figura 23 es mostra el model TICVA desenvolupat a partir del model de referència de l'Analítica de l'Aprenentatge, en què es dona resposta a aquestes qüestions. I a continuació es descriuen breument cadascuna d'aquestes dimensions.

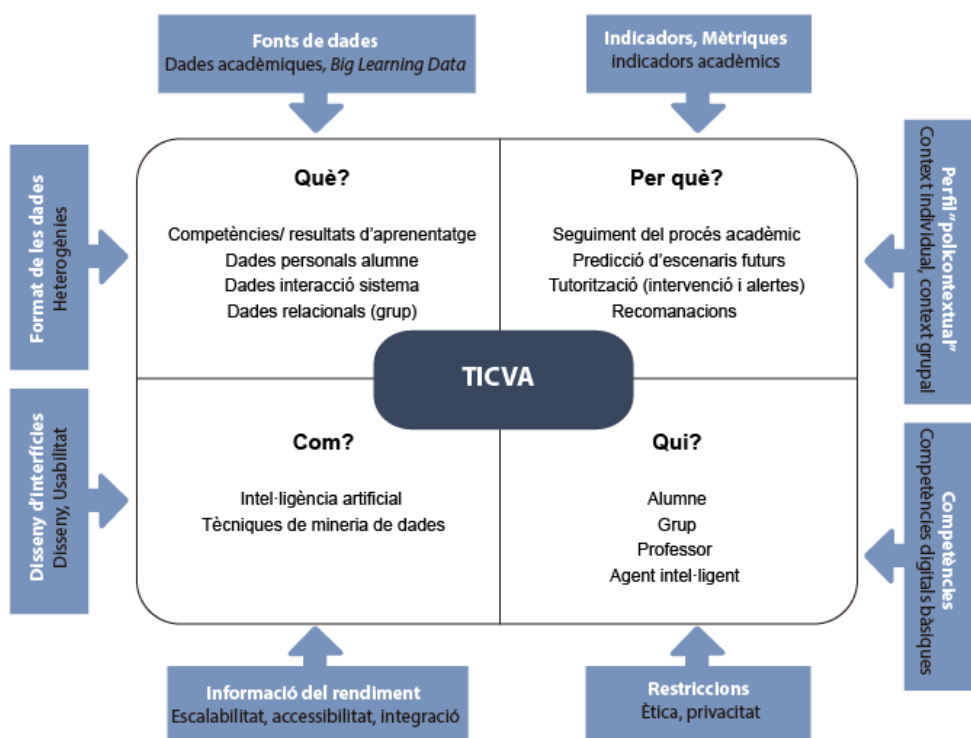


Figura 23. Model conceptual del TICVA que respon a les quatre dimensions crítiques plantejades per Chatti.

5.2.1 Què? El coneixement i les dades

Com s'ha vist al capítol 3, la base de l'anàlisi de l'aprenentatge és la informació. La seva finalitat és crear coneixement educatiu útil a partir de dades brutes procedents de múltiples fonts. En el cas del TICVA les dades originals susceptibles de ser processades provenen de diferents fonts i són de diversa tipologia. Així, el repte del TICVA és trobar la manera d'agregar i integrar aquests conjunts de dades heterogènies per tal que, mitjançant algunes tècniques d'anàlisi i processament que es veuran més exhaustivament al capítol següent, se'n pugui extreure coneixement profitós.

De manera genèrica es pot dir que hi ha quatre àmbits diferents dels quals es vol recol·lectar dades dins del sistema TICVA: dades sobre les competències i els resultats d'aprenentatge, dades personals de l'alumne, dades d'interacció amb el sistema i, finalment, dades relacionals.

Com és sobradament conegut, tot sistema d'aprenentatge en línia incorpora una sèrie de coneixements que volen ser transmesos, transformats sovint en conceptes més abstractes com ara sessions, temes o assignatures. Si bé aquesta informació com a tal no té una implicació directa amb l'anàlisi del procés d'aprenentatge, els canvis produïts pel Procés de Bolonya han fet que d'alguna manera el resultat de la interacció amb aquesta informació sigui de crucial

interès en l'experiència educativa. Així, la forma d'organitzar i adquirir els coneixements ha evolucionat i s'ha anat transformant en un conjunt d'activitats, associades a la vegada a un conjunt de competències i resultats d'aprenentatge. És per tant imprescindible, en primer terme, que l'arquitectura del TICVA s'adapti a aquest nou escenari (Vernet, Nicolas, Golobardes, Fornells, & Garcia-Piquer, 2010) i sigui capaç internament d'organitzar la informació associada als conceptes que es volen transmetre seguint els models actuals educatius basats en competències. A més, com en qualsevol sistema basat en la personalització, la caracterització de l'individu és un element fonamental a l'hora de crear un entorn altament adaptatiu. És per això que la gestió de dades personals i preferències de l'usuari també és una variable que ha de ser contemplada pel model.

Un altre tipus de dades que entra en joc en el TICVA és la que es recull de la interacció de l'alumne o el grup d'alumnes amb el sistema. En efecte, els sistemes en línia han de ser dissenyats amb la capacitat d'adquirir qualsevol informació útil provinent de l'alumnat per tal de poder-la utilitzar en un futur per a una millor personalització, experiència d'usuari i tasques de tutoria posterior. De la mateixa manera, en entorns on l'alumne necessita dispositius físics per poder avançar en l'aprenentatge (Vernet, Canaleta & Zaballos, 2012), caldrà que aquests dispositius estiguin preparats per poder-se configurar per tal de recollir aquestes dades i transmetre-les remotament al sistema central (Albó, Vernet, Canaleta, & Vilasís-Cardona, 2013).

Finalment, des del punt de vista relacional, el model TICVA ha de permetre contemplar el procés d'aprenentatge com a resultat de la interacció, no tan sols d'un individu amb el sistema, sinó també amb d'altres actors del mateix. Així, d'alguna manera s'han de poder gestionar dades procedents de la interacció de comunitats d'usuaris o grups amb el propi sistema i dades que provinguin de la interacció d'individus dins del propi grup.

5.2.2 Per què? La motivació i els objectius

La motivació principal del modelat d'un sistema com aquest és la de donar resposta a la necessitat d'orientar d'una manera més intel·ligent i efectiva els processos d'aprenentatge que es donen en un entorn virtual. En aquest sentit, es pot dir que els objectius que persegueix el sistema TICVA són diversos i es podrien classificar en quatre categories: el seguiment del procés acadèmic, la predicció d'escenaris futurs, la tutorització mitjançant intervencions i alertes i, per últim, la recomanació.

Per tal d'aconseguir fer un bon seguiment del procés acadèmic de l'alumne o el grup d'alumnes és indispensable que el sistema reculli informació, la processi i la mostri al professor o tutor de l'aula virtual per tal que aquest, mitjançant la seva anàlisi, pugui prendre decisions al respecte. Cal doncs que l'arquitectura contempli l'existència d'un mòdul especialitzat en la generació d'informes configurables i valoracions dels alumnes a partir de la informació recopilada.

La predicció de l'evolució de l'alumne és un aspecte clau a l'hora de possibilitar la intervenció, en cas de necessitat, en el seu procés d'aprenentatge. Tanmateix, és segurament la part més complicada de l'arquitectura, doncs les dades recollides plantegen exclusivament informació sobre el context present i l'anticipació que es busca a l'hora d'intervenir en un procés d'aprenentatge requereix fer una extrapolació al context futur. Per aconseguir-ho, caldrà incorporar algorismes d'intel·ligència artificial que permetin predir un context futur. D'aquesta manera, el professorat disposarà tant d'informació relativa a com funcionen els grups de treball establerts a l'aula virtual (trajectòria col·lectiva), com de la situació individualitzada de cada alumne (trajectòria individual). En el proper capítol s'analitzen diferents possibilitats per dur a terme aquest mòdul del sistema i es mostren els resultats obtinguts després d'una primera implementació que valida l'eficàcia dels algorismes dissenyats.

Com s'ha esmentat, la tutorització de l'alumne és un altre objectiu clau del sistema. Amb la tutorització es pretén orientar l'alumne en el seu procés d'aprenentatge i corregir-ne els possibles desviaments mitjançant una intervenció directa o indirecta o l'emissió d'alertes als interessats. Tal com es planteja el sistema TICVA aquesta tutorització pot venir des de fonts diferents:

- *Tutorització a partir de les dades recollides pel sistema.* El professor pot utilitzar el mòdul d'analítica i d'informes per extreure conclusions de l'estat d'un alumne i pot intervenir en la tutoria individualitzada o grupal de l'aula virtual (Vernet, Canaleta, & Pallàs, 2012).
- *Tutorització intel·ligent a partir de la proposta del sistema de predicció.* El sistema de predicció individual o grupal informa el professor d'aquells alumnes dels quals s'està observant una situació anòmala. El propi sistema de tutoria automatitzada pot fer propostes de canvis de grup d'alumnes, en cas que el problema provingués del funcionament d'un grup en qüestió. D'altra banda, quan el sistema intel·ligent detecta que la predicció de l'evolució de l'alumne produeix un resultat negatiu, pot realitzar dues intervencions: iniciar una acció personalitzada de recomanació o bé avisar el professor perquè aquest intercedeixi en el procés d'aprenentatge.

Finalment la recomanació del sistema va lligada a la personalització i a l'avaluació de l'alumne. Per exemple, si un alumne no resol correctament un exercici, el sistema expert integrat en l'arquitectura pot proveir aquest alumne d'altres exercicis complementaris que li permetin entendre els conceptes que ha d'aprendre. Des d'un altre punt de vista, la personalització de la resposta o de la recomanació és un factor cabdal per atraure l'atenció de l'alumne. El mode de navegació que utilitza l'alumne en el sistema, així com la manera com es relaciona amb els companys són dades que es recullen a l'arquitectura i que permeten donar una resposta acurada.

5.2.3 Com? Processament i anàlisi

L'aplicació d'una o altra tècnica de tractament i anàlisi de dades dependrà en certa mesura de l'objectiu que es vulgui aconseguir. És a dir, la selecció d'una o altra metodologia es farà en

funció del tipus de coneixement útil que es vol obtenir al final. En el cas del model que es presenta, els algorismes intel·ligents han estat la tècnica principal en el processament de la informació.

Els algorismes d'aprenentatge automàtic permetran analitzar la informació recollida de l'alumne al llarg del curs per tal de poder analitzar quina és la seva evolució en el procés d'aprenentatge en un instant de curs i poder guiar-lo i tutoritzar-lo per millorar la seva situació. Aquests algorismes barrejaran, com es veurà al proper capítol, tècniques d'aprenentatge supervisat com és el Raonament Basat en Casos, amb d'altres de no supervisades com poden ser els algorismes d'agrupament o *clustering*.

D'altra banda, l'anàlisi de les dades recollides implicarà processos de Minería de Dades. En efecte, no només s'analitzaran les dades per tal de detectar evolucions en la trajectòria de l'alumne, sinó que caldrà analitzar les dades que es van recollint per automatitzar processos i generar els informes necessaris pel professor que li facilitin una posterior tutoria. A més, l'alumne ha de disposar de diferents eines de visualització de la informació que es va generant que li permetin tenir perspectiva tant del seu aprenentatge, com de l'aprenentatge aconseguit a nivell de grup.

5.2.4 Qui? Els actors

Tractant-se d'un sistema concebut per a l'aprenentatge col·laboratiu supervisat per agents humans i agents intel·ligents, el model TICVA esdevé un entorn on conviuen quatre diferents tipus d'actors o interessats:

1. *L'alumne*. Com en tot sistema d'aprenentatge en línia l'alumne és la peça central del sistema. La seva interacció correcta amb el sistema marcarà en bona mesura el seu progrés en el procés d'aprenentatge. El seu estil d'aprenentatge també serà molt clau per identificar-ne el seu perfil (Lathan, Crockett, & Mclean, 2013).
2. *El grup*. El conjunt d'alumnes han de veure també el sistema com un espai de treball en grup, on aquest reacciona no només a nivell individual sinó també a nivell col·lectiu. La resposta del sistema pot anar dirigida, per tant, als propis membres del grup o al professor, en cas que siguin necessàries accions posteriors de tutorització grupal.
3. *El professor o docent*. Naturalment el professor ha de poder fer el seguiment de tot el que està succeint i determinar com els seus alumnes progressen en l'assimilació dels continguts. A més, aquest ha de disposar d'un cert control sobre el funcionament dels grups i fer-ne un seguiment acurat, fent ús del *feedback* que li proporcioni el sistema.
4. *L'agent intel·ligent*. Com a sistema que ha de pertànyer a la generació simbiòtica d'entorns d'aprenentatge web, l'agent intel·ligent ha de ser un actor més que es relacioni, en cas necessari, amb altres agents i a la vegada amb el professor i l'alumne.

5.3 Estructura del sistema TICVA

Després d'haver definit quines són les dimensions crítiques del sistema TICVA, en aquest apartat es procedeix a representar-ne l'arquitectura. En primer lloc es presenta una primera aproximació amb una arquitectura molt bàsica (Figura 24), inspirada en el model tradicional d'un sistema tutor intel·ligent. A continuació es presenta un model més evolucionat, que permet suportar les quatre dimensions plantejades a la secció anterior.

5.3.1 Mòduls del sistema

El sistema TICVA es recolza en la nomenclatura tradicional dels sistemes tutors intel·ligents per tal que el lector pugui fer un paral·lelisme entre els mòduls tradicionals d'un sistema tutor i els mòduls proposats en aquesta arquitectura. Tot i així, com es veurà ben aviat, aquests mòduls es componen de diverses peces de *software* que, tot i que no es representen de manera explícita en un sistema tutor intel·ligent tradicional, han estat incloses en la mesura del possible dins del mapa estàndard del que seria un sistema tutor per tal de facilitar-ne la ubicació conceptual. Malgrat tot, hi ha hagut aportacions que no s'han pogut incloure en cap dels mòduls tradicionals així que s'han afegit com a capes externes o mòduls complementaris als ja coneguts.

Com es pot veure a la Figura 24 l'estructura general de l'arquitectura TICVA intenta partir dels mòduls tradicionals d'un STI. A diferència del model tradicional apareix una nova entitat corresponent al mòdul de grup, el qual inclou la informació corresponent al procés d'aprenentatge d'un conjunt d'alumnes que es relacionen i col·laboren entre ells. Així, el mòdul de l'estudiant clàssic s'inclou dins d'aquest mòdul superior.

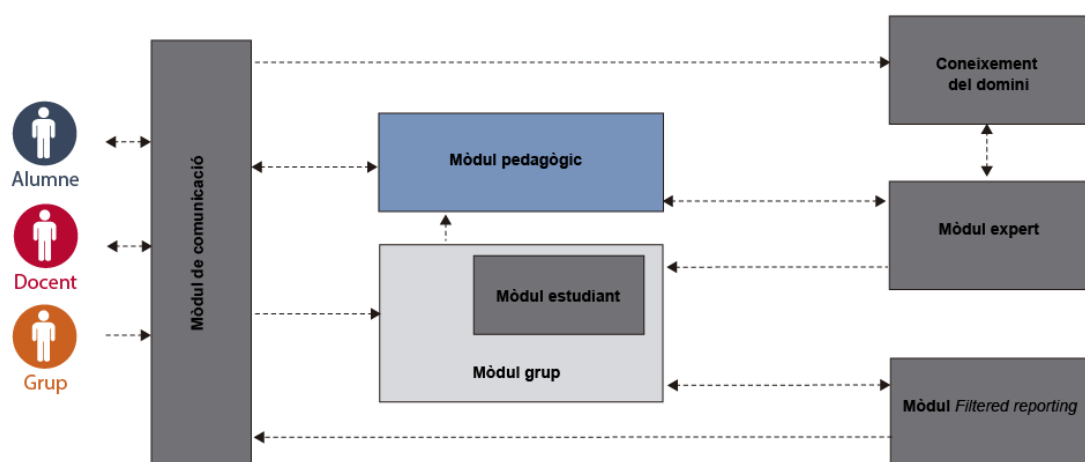


Figura 24. Arquitectura bàsica del sistema TICVA dissenyat.

En la Figura 25 d'aquesta pàgina es mostra una visió global de l'arquitectura dissenyada, amb tots els mòduls connectats entre ells. Aquesta figura es pot veure ampliada, per poder observar-la amb més deteniment i claredat a l'Apèndix D d'aquesta tesi. A continuació s'analitzaran cadascun d'aquests mòduls indicant quines són les característiques rellevants de tots ells.

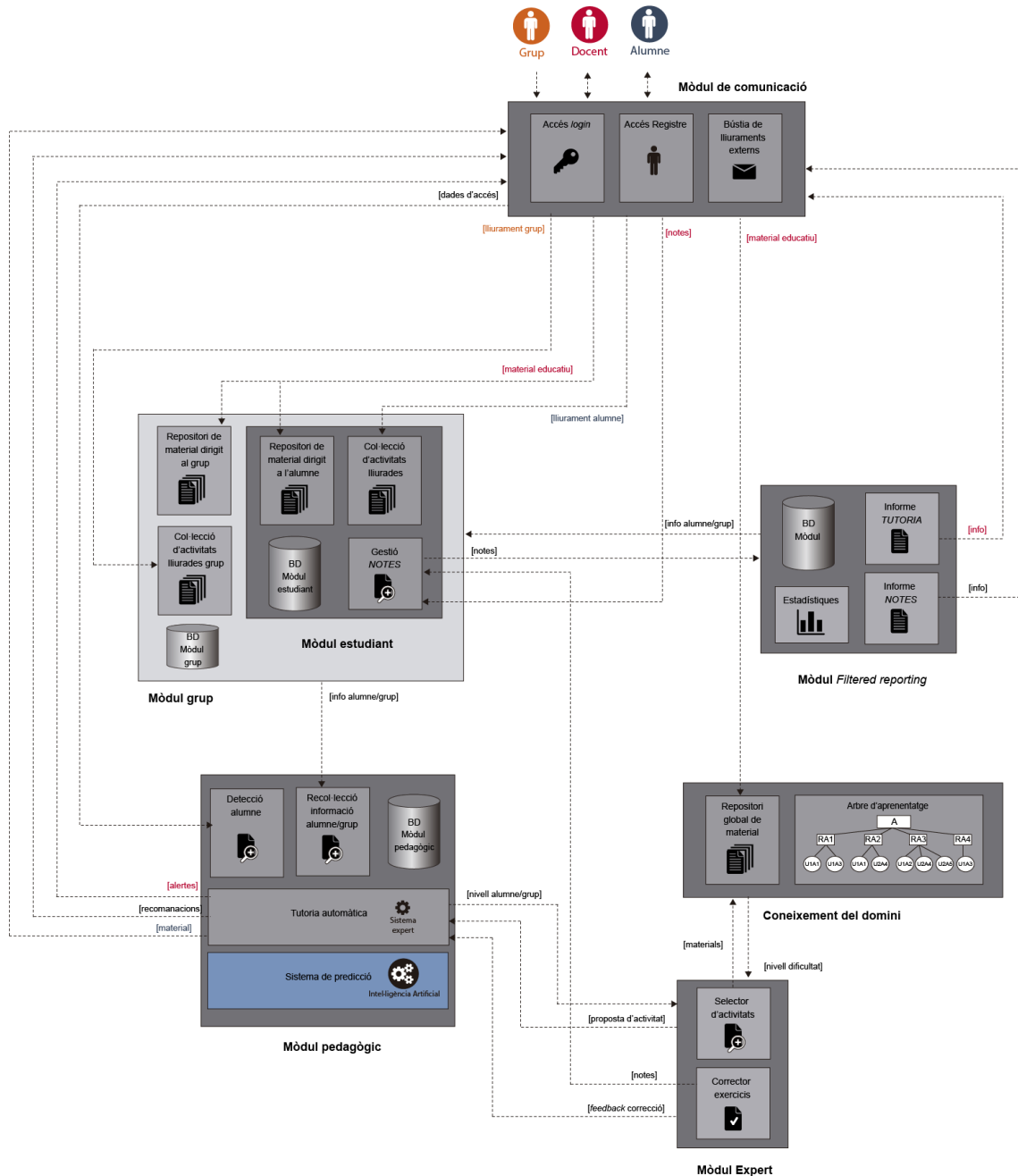


Figura 25. Arquitectura completa del sistema TICVA dissenyat.

5.3.1.1 Mòdul de Comunicació o interfície del sistema

El primer punt de contacte dels alumnes amb el sistema TICVA és el Mòdul de Comunicació. Aquest mòdul inclou bona part del *frontend* de l'aplicació, la gestió del *login* i el registre d'usuari. La gestió del *login* i del registre, tot i que es tracta de processos de *backend*, es representen en aquest mòdul indicant que l'accés es fa a través de la interfície del sistema. El submòdul de *login* autenticarà l'usuari i enviarà l'identificador d'aquest al Mòdul Pedagògic, perquè es pugui recuperar la informació necessària i s'habilitin els perfils d'alumne, grup o professor, segons s'escaigui. Cal dir que tot el programari desenvolupat sota aquesta arquitectura hauria de fer ús d'un entorn *responsive* i que faciliti l'ús de diferents dispositius mòbils per establir la connexió (Canaleta & Vernet, TIC versus mTIC, 2004).

Aquest mòdul també ha d'incorporar la passarel·la necessària perquè els materials i lliuraments externs arribin al sistema. Així, si un alumne vol lliurar un document aliè al sistema, el submòdul *Bústia de lliuraments externs* s'encarregarà de recollir-lo i enviar-lo al Mòdul de l'Estudiant/Grup perquè quedi arxivat i sincronitzat amb la resta de materials interns.

En general, el Mòdul de Comunicació hauria d'estar implementat, tal i com s'apunta al final del capítol, amb una tecnologia que permeti crear entorns usables, lleugers i que renderitzin el més ràpid possible.

La Figura 26 representa l'estructura interna d'aquest mòdul i les dependències que té amb els altres mòduls del sistema.

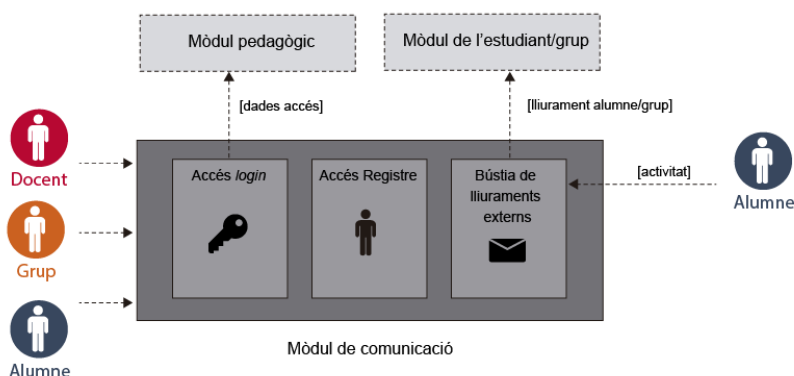


Figura 26. Arquitectura detallada del Mòdul de Comunicació dissenyat.

5.3.1.2 Coneixement del Domini

El mòdul de Coneixement del Domini és l'encarregat de gestionar tots els continguts relacionats amb el temari de l'assignatura o matèria que l'alumne està aprenent. Tal com s'ha comentat

anteriorment, la introducció del procés de Bolonya (Bologna Declaration, 1999) en els processos d'aprenentatge ha fet que la manera de dissenyar i emmagatzemar aquesta informació hagi canviat substancialment (Canaleta & Vernet, 2004). Així, si antigament els continguts de les assignatures seguien la sèrie "Assignatura-Tema-Activitat" i no es tenien en compte ni els Resultats d'Aprenentatge ni les Competències associades (Figura 27), avui en dia els nous entorns d'aprenentatge en línia han de ser plantejats per contenir aquesta informació formant una estructura inversa, on el que és més important és el Resultat d'Aprenentatge o la Competència que es vol adquirir. Les activitats, per tant, han de ser plantejades de la millor manera per tal que es compleixi que mitjançant la seva realització es pugui assolir un resultat d'aprenentatge o competència.

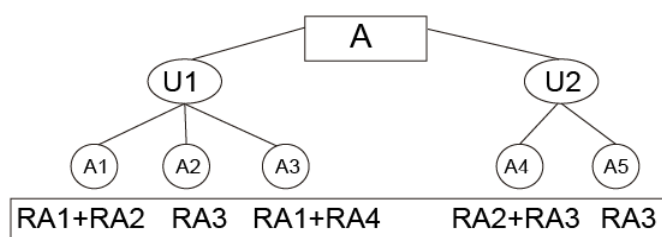


Figura 27. Arbre de coneixement clàssic, pre-Bolonya (A = Assignatura, U=Unitat o Tema, A=Activitat, RA = Resultat d'Aprenentatge).

A la publicació (Vernet, Nicolas, Golobardes, Fornells, & Garcia-Piquer, 2010) es va presentar una nova manera de reorganitzar aquesta informació per tal que es representi el coneixement relatiu al procés d'aprenentatge d'acord amb els principis establerts per Bolonya. Aquesta estructuració es resumeix breument a la Figura 28. Derivada d'aquesta mateixa qüestió en resulta l'avaluació de l'alumne per competències. En aquest sentit, la possibilitat que l'alumne pugui ser avaluat seguint aquest model competencial es va veure en profunditat a (Canaleta, Vernet, & Zaballos, 2011).

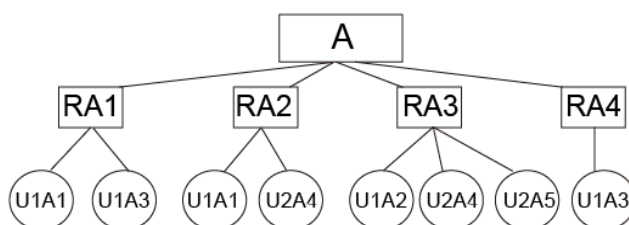


Figura 28. Nova representació del coneixement després dels canvis introduïts pel model de Bolonya.

Com es pot veure a la Figura 28, l'assignatura divideix el coneixement segons les competències i els resultats d'aprenentatge que l'alumne ha d'assolir. Per simplicitat, a la figura només es mostra l'assignatura estructurada segons els resultats d'aprenentatge. En aquesta nova estructuració, per tal que l'alumne assoleixi cadascun dels resultats d'aprenentatge, es dissenyen un conjunt d'activitats a mida que permeten que l'alumne adquireixi els coneixements i les competències projectades. La notació U_xA_y indica que el node representa l'activitat y de la Unitat o Tema x . A diferència de la versió pre-Bolonya, en què el Coneixement del Domini estava estructurat de la manera clàssica i les competències s'assignaven a posteriori (a partir de les activitats realitzades), la nova estructura permet emmagatzemar les activitats associant-les als resultats d'aprenentatge.

Així doncs, en el sistema TICVA tota aquesta informació està representada per un arbre de coneixement, on les activitats es creen i es classifiquen com a material destinat a aconseguir un determinat Resultat d'Aprenentatge. A (Canaleta, Vernet, Vicent, & Montero, 2014) s'analitza també en profunditat la importància cabdal de dissenyar els continguts a aprendre a partir de les competències que l'alumne necessita adquirir. A nivell de blocs dins del mòdul, per tant, el Coneixement del Domini estarà format per tota la representació d'enllaç arborescent entre les diferents activitats i materials de les assignatures, més el repositori digital de tota aquesta informació (Figura 29).

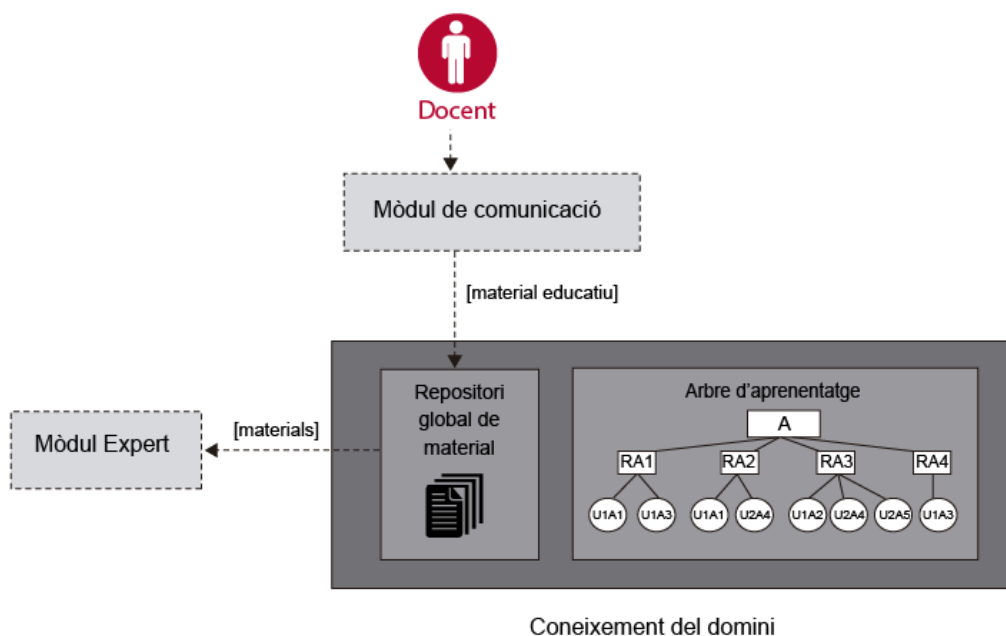


Figura 29. Arquitectura detallada del Mòdul del Coneixement del Domini dissenyat.

5.3.1.3 Mòdul expert

El mòdul expert del sistema TICVA tindrà dues tasques principals que es mapejaran en els corresponents submòduls: poder avaluar els diferents exercicis que lliuren els alumnes amb el submòdul *Corrector exercicis* i ser capaç de seleccionar els exercicis o materials necessaris que sol·liciti el Mòdul Pedagògic segons un cert criteri que rebí (*Selector d'activitats*). La Figura 30 representa gràficament l'estructura d'aquesta part del sistema.

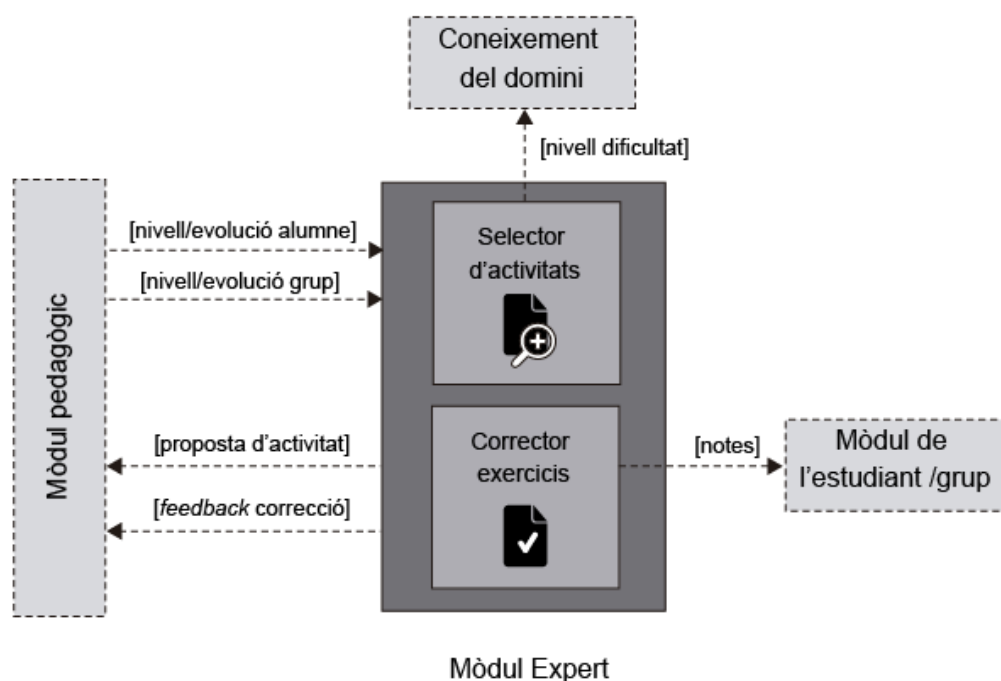


Figura 30. Arquitectura detallada del Mòdul Expert dissenyat.

Els exercicis que lliuren els alumnes poden ser individuals o grupals, així que el submòdul de correcció haurà de ser capaç d'avaluar els alumnes tenint en compte aquests dos punts de vista. El resultat de l'avaluació haurà de ser enviat al Mòdul de l'Estudiant o Grup que s'encarregarà d'arxivar-lo i actualitzar el perfil de cada alumne, el qual, com ja s'ha tractat en altres treballs (Canaleta, Zaballo, & Vernet, 2012), ha de conèixer com ha estat avaluat i rebre les retroaccions corresponents.

El Mòdul Pedagògic demanarà al Mòdul Expert un conjunt de material a oferir a l'alumne. Per fer-ho, el Mòdul Expert rebrà una valoració del nivell de l'alumne calculada pel Mòdul Pedagògic i un "tipus de material sol·licitat" que el permeti, o bé millorar el coneixement, reforçar-lo o sincronitzar-lo amb la resta de companys del seu grup, segons el cas. Per recuperar

aquesta informació haurà de connectar amb el Coneixement del Domini, on es troba tot el material de les assignatures.

5.3.1.4 Filtered Reporting

La generació d'informes i estadístiques estarà localitzada en el mòdul de *Filtered Reporting*. D'una banda, l'alumne ha de poder consultar les seves notes individuals i/o grupals sempre que ho necessiti. De l'altra, el professor ha de recuperar aquesta informació tant per realitzar l'avaluació contínua de l'alumne com per obtenir informació per a les tutories personalitzades. Algunes primeres aproximacions d'implementació ja van ser apuntades a (Vernet & Canaleta, 2004).

Així aquest mòdul tindrà contacte directe amb el Mòdul de l'Estudiant o Grup que li subministrarà informació dels alumnes per generar les estadístiques i també amb el Mòdul de Comunicació ja que haurà de subministrar aquesta informació cap a l'exterior del sistema. La Figura 31 mostra les relacions d'aquest mòdul amb la resta de mòduls, així com les parts que el formen.

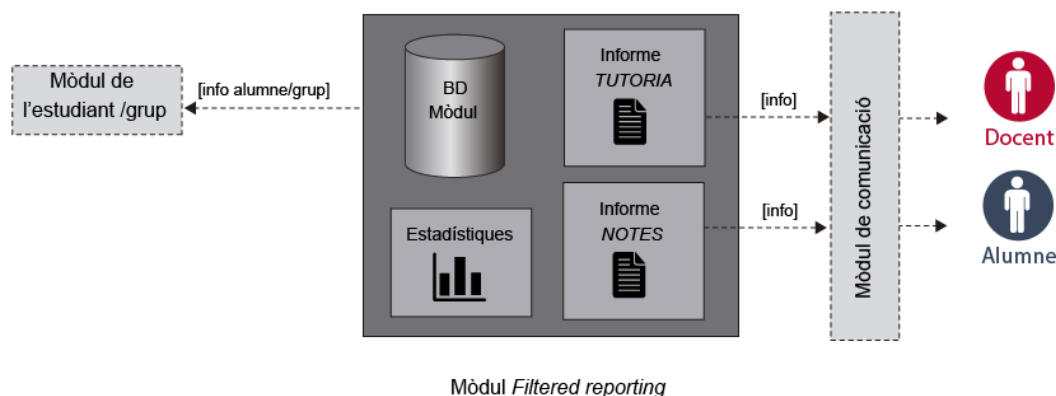


Figura 31. Arquitectura detallada del Mòdul *Filtered Reporting* dissenyat.

Com es pot veure, aquest mòdul disposarà d'una part de la base de dades distribuïda del sistema. Aquí s'aniran emmagatzemant les estadístiques i els càlculs de notes que es necessitin tenir accessibles per generar els informes cap al professor i l'alumne. Els submòduls d'informes connectaran amb aquesta base de dades i, amb el suport del submòdul d'estadístiques, prepararan la informació que se sol·liciti des de l'exterior.

Així doncs, el disseny d'aquest mòdul ha de ser prou net i entenedor per tal de permetre al professor consultar tota la informació de manera ordenada. Un disseny de pantalla de *backend* com el *Katniss Premium Admin Template*¹¹ és una bona solució, doncs amb un panell de control integral es pot gestionar i consultar totes les dades referents als alumnes. De totes maneres, són moltes les solucions finals que es poden aplicar i que dependran del moment d'implementació d'aquest mòdul.

5.3.1.5 Mòdul de l'Estudiant o Grup

El mòdul de l'estudiant bàsic que s'havia vist en el Sistemes Tutors Intel·ligents tradicionals ha de ser ampliat perquè l'entitat *grup* hi tingui cabuda. La Figura 32 descriu com s'hauria d'ampliar aquest mòdul per tenir en compte aquest aspecte.

De cada alumne s'haurà de tenir arxivada diferent informació:

- *Col·lecció d'activitats lliurades.* Tots els lliuraments de l'alumne han de quedar emmagatzemats en aquest mòdul. Estaran disponibles per a tots els actors del sistema excepte pel grup. Aquesta zona haurà de tenir accés directe amb el Mòdul de Comunicació amb una interfície amigable i de fàcil ús.
- *Repositori de material dirigit a l'alumne.* Els materials i activitats que necessiti un alumne en un determinat moment hauran d'estar igualment disponibles per a la seva consulta.
- *Emmagatzematge i consulta de notes.* L'estudiant ha de tenir disponible sempre les qualificacions i retroaccions de les seves activitats. La informació que introdueixi el mòdul expert provinent de la correcció automàtica o la correcció directa del professor quedarà registrada en la base de dades d'aquest mòdul. En cas que calgui un informe complet s'utilitzarà el submòdul corresponent de *Filtered Reporting*.

Pel que es refereix a l'entitat *grup*, el mòdul hauria de tenir també un punt d'entrada diferent que es representa a la Figura 32. De la mateixa manera que un alumne pot accedir a les seves activitats, enunciats o lliuraments, també ha de poder accedir a l'espai grupal, on trobarà també activitats especials per ser resoltes en grup i els espais necessaris per poder fer el dipòsit de les activitats realitzades amb altres companys de l'aula virtual.

¹¹ <http://themeforest.net/item/katniss-premium-admin-template/3878281>

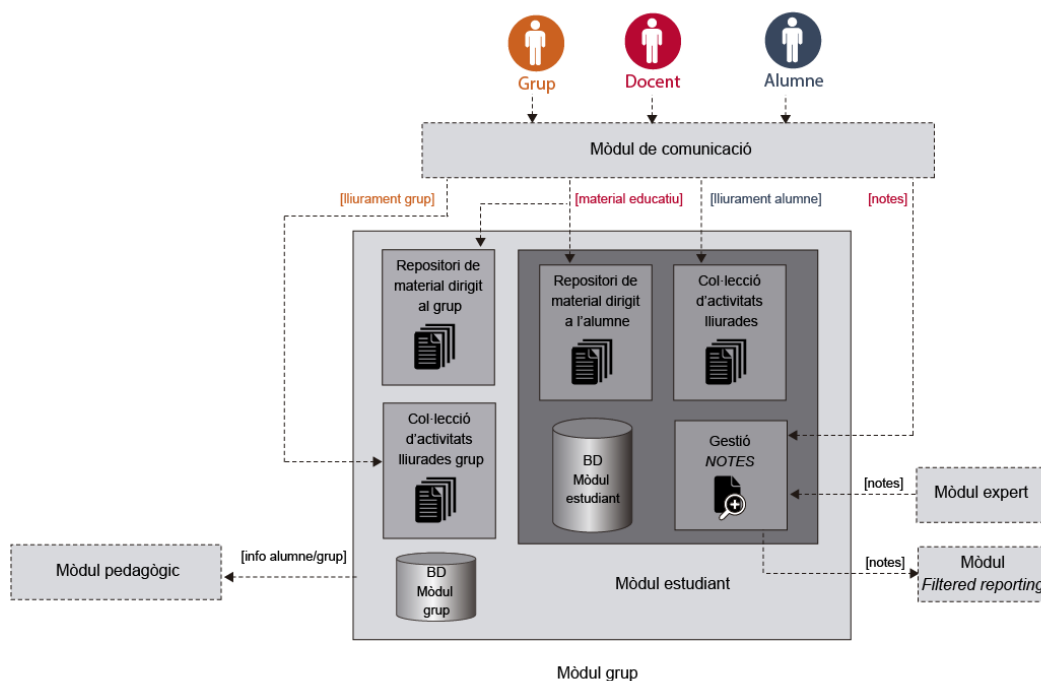


Figura 32. Arquitectura detallada del Mòdul de l'Estudiant/Grup.

Si es fa un resum de totes les dades necessàries per al Mòdul Pedagògic que hauria de recollir el sistema TICVA i que modelen a l'estudiant/grup, es podrien caracteritzar de la manera següent:

- *Qualificacions de l'alumne*: cal emmagatzemar totes les qualificacions obtingudes per l'alumne al llarg del període de curs avaluat.
- *Informació del grup o aula assignada*: s'ha de tenir la informació referent a l'aula virtual assignada a l'estudiant.
- *Informació del grup de treball*: cal saber amb quins altres estudiants es relacionarà l'alumne (quins grups formarà).
- *Altra informació acadèmica*: informació no lligada a les qualificacions, com ara el número de repeticions de l'assignatura.
- *Informació de comportament*: seria interessant disposar de dades com la velocitat en la resolució dels exercicis o quin període de resposta té l'alumne a l'hora de començar la resolució d'una pràctica o exercici.

Com es veurà en el capítol següent, s'ha realitzat la implementació del submòdul de predicció de l'alumne disposant únicament de les quatre primeres classes de dades.

5.3.1.6 Mòdul pedagògic

El mòdul més important del sistema és el Mòdul Pedagògic. És l'encarregat de fer el seguiment de l'alumne, avaluant en tot moment quina és la seva evolució i progrés i avisant, ja sigui amb recomanacions o alertes, dels possibles problemes d'un alumne concret o del funcionament d'un grup virtual. També es pot encarregar de l'estratègia pedagògica i de com es presenten els materials a l'alumne per millorar el seu procés d'aprenentatge (Burns & Capps, 1988).

El Mòdul Pedagògic es compon dels diferents paquets de programari que es representen a la Figura 33 i que es descriuen a continuació:

1. *Detecció de l'alumne.* Aquesta part del mòdul s'encarrega d'identificar l'alumne dins la classe virtual a partir de les dades d'accés que li proporciona el Mòdul de Comunicació.
2. *Recol·lecció d'informació de l'alumne o del grup.* Un cop identificat l'alumne cal recuperar tota la informació existent de l'alumne (o del grup, si s'escau) per poder executar posteriorment el sistema de predicció amb tota la informació registrada al sistema. Informació que pot anar des de les pròpies qualificacions de l'alumne fins al seu temps de resposta (Jarusek & Pelanek, 2012).
3. *Sistema de predicció.* El sistema de predicció és la part *intel·ligent* del Mòdul Pedagògic i és la responsable d'avaluar l'estat d'un alumne o un grup per tal detectar possibles problemes en el procés d'aprenentatge. En el proper capítol s'estudiarà a fons aquest mòdul i s'analitzaran els resultats d'una implementació realitzada utilitzant Intel·ligència Artificial.
4. *Tutoria automàtica.* Les recomanacions automàtiques a l'alumne o les alertes dirigides al personal docent es generen mitjançant aquest paquet de programari. En definitiva, es tractaria d'un petit sistema expert capaç d'interpretar la resposta donada pel sistema de predicció i llançar el missatge de sortida cap al professorat. El disseny final de les interaccions concretes que se succeeixin pot ser clau perquè el sistema tingui un bon comportament tutorial (Bourdeau & Grandbastien, 2010).

Una representació completa d'aquest mòdul és la que es pot veure a la Figura 33 de la pàgina següent. Es destaca de color especial el sistema de predicció, al qual s'hi farà especial incís al capítol 6.

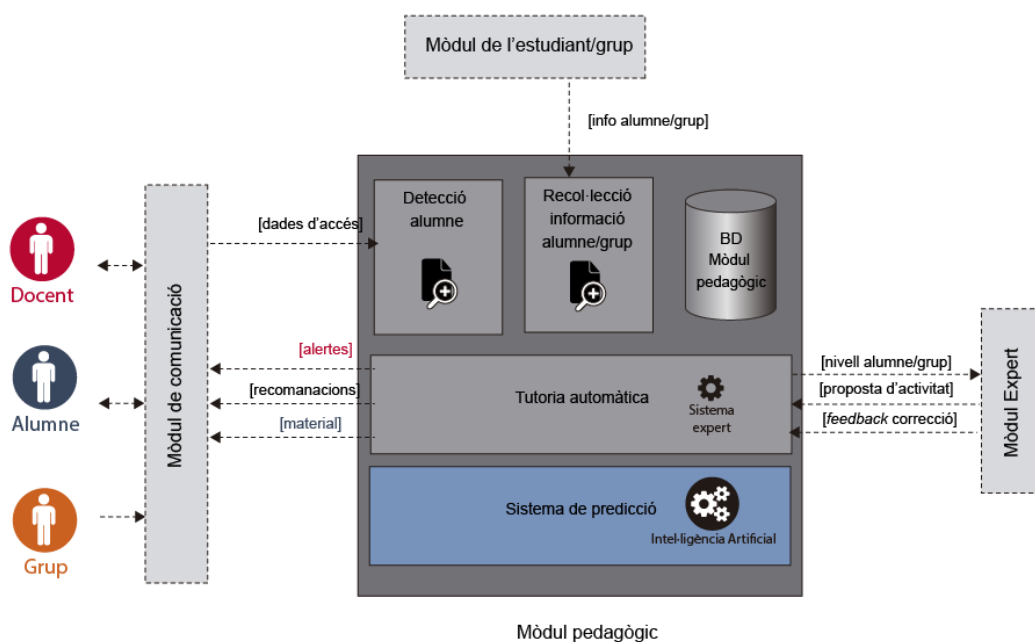


Figura 33. Arquitectura detallada del Mòdul Pedagògic.

5.4 Un apunt de la tecnologia web

Tot i que les tecnologies web estan contínuament debatent-se, canviant i avui en dia s'estan desenvolupant grans avenços en la programació d'entorns web, sí que es poden apuntar certes possibilitats pel que a la implementació de la plataforma basada en aquesta arquitectura. Certament el desenvolupament de bona part del codi hauria de respectar els principis SOLID (Oloruntoba, 2015). Això no treu que altres característiques tecnològiques són igualment importants de tractar i que podem resumir en els següents titulars destacats:

1. *Base de dades distribuïdes.* Com s'ha vist en els diferents representacions dels mòduls del sistema, la base de dades està distribuïda al llarg d'aquests i no segueix un model relacional estàndard. Actualment existeixen tecnologies com MongoDB¹², que ofereixen la distribució d'informació orientada a document, que encaixa perfectament amb la filosofia d'emmagatzematge introduïda a TICVA.
2. *Un mòdul de comunicació àgil i efectiu.* Les darreres tecnologies de *frontend* permeten una renderització de les pantalles i una interactivitat que cal imposar en tot sistema

¹² <https://www.mongodb.com/>

d'aprenentatge en línia. Així, una tecnologia com React¹³, creada i mantinguda per empreses importants del sector web com Facebook o Instagram, és potser la millor solució actualment disponible.

3. *Un llenguatge de servidor ràpid i orientat a objectes.* Moltes són les alternatives en aquest aspecte i que dependran del moment de la implementació. Actualment llenguatges com PHP, Python, Java o Go¹⁴, creat per Google, són llenguatges que permeten desenvolupaments eficients en entorns web.
4. *Un maquinari a l'alçada de l'ús del sistema.* El maquinari necessari per desenvolupar un sistema així anirà directament relacionat amb el tràfic de dades generat. Així, inicialment tot el programari podria estar allotjat en una mateixa màquina dotada d'un bon processador i bona memòria RAM, però, certament, si fos necessari caldria configurar un o més servidors d'arxius estàtics, replicar els servidors de bases de dades i els servidors web, i afegir servidors de *cache* de manera distribuïda.

En definitiva, totes les tecnologies que depenguin de l'entorn web dependran també del moment d'implementació. I si bé no es pot afitar perfectament aquest àmbit, sí que se'n poden distingir uns trets fonamentals: la tecnologia triada ha de ser perfectament escalable i preparada per ser dimensionada segons la necessitat o demanda; ha de permetre que l'alumne pugui interactuar amb el sistema des de qualsevol dispositiu mòbil i que l'entorn s'adapti a les particularitats d'aquest; i a més cal tenir en compte que l'arquitectura que s'està proposant ha de facilitar que la seva implementació atorgui a l'alumne una experiència d'usuari completa.

Paral·lelament, la gestió i la monitorització dels elements controlats des d'una plataforma web està convergint cap a una nova manera de gestionar els recursos: és l'anomenada Web de les Coses (*Web of Things*, de l'anglès). L'autor presenta a (Vernet, Zaballos, Martín de Pozuelo, & Caballero, 2015) una possible arquitectura en aquest sentit i una aplicació pràctica en un dels dominis (no docents) utilitzats durant aquesta tesi. Sens dubte, la seva aplicació a l'entorn docent podria fer-se també de manera satisfactòria seguint aquestes pautes apuntades.

Si bé la implementació de l'arquitectura del TICVA està subjecta a la maduració de certes tecnologies, no passa el mateix amb la part més intel·ligent del sistema, la qual actualment ja es pot desenvolupar de manera completa. En aquest cas, com s'ha plasmat en el capítol següent, en aquests moments ja és possible abordar el disseny i desenvolupament del submòdul de predicció del Mòdul Pedagògic.

¹³ <https://facebook.github.io/react/>

¹⁴ <https://golang.org/>

5.5 Resum i conclusions

En aquest capítol s'ha fet la proposta d'arquitectura TICVA, a partir de les mancances que s'han detectat en els sistemes tutors intel·ligents actuals basats en web. S'han respost les preguntes del model proposat per Chatti, que corresponen a les quatre dimensions crítiques que cal abordar a l'hora de dissenyar models que analitzin el procés d'aprenentatge, i s'han enllaçat directament amb la implicació que tenen en els diferents mòduls i actors que formen part del sistema.

Pel que fa al disseny dels diferents mòduls s'han extret diferents conclusions que se citen a continuació. En referència al Coneixement del Domini s'ha vist la importància de poder emmagatzemar i servir els materials evitant el model rígid d'ensenyaments caducs. En aquest sentit, s'ha revisat i aplicat la proposta presentada amb anterioritat en consonància amb les indicacions del Pla Bolonya.

Al mateix temps, s'ha vist que caldrà disposar d'un Mòdul Expert que avaluarà la qualitat dels exercicis lliurats pels estudiants i donarà suport al professorat en la correcció de les activitats acadèmiques.

Pel que fa al Mòdul de Comunicació, aquest cal que interactuï amb la resta de mòduls del sistema per garantir una interfície amigable i segura, com es requereix en entorns acadèmics (Canaleta & Vernet, 2003).

Els Mòduls més importants del sistema són el Mòdul de l'Estudiant o de Grup i el Mòdul Pedagògic. El primer permet tenir sempre actualitzat un perfil de l'alumne o del grup i un registre de totes les seves activitats lliurades i qualificacions obtingudes. El segon, destaca per incorporar principalment la part intel·ligent, que es comentarà amb més profunditat al següent capítol, i un sistema expert de recomanació útil per al personal docent.

En definitiva, s'ha constatat que hi ha molts entorns que recullen molta informació de l'alumne, però pocs que la tracten. En aquest sentit, s'ha dissenyat una proposta d'arquitectura que pretén incorporar tota la potència de coneixement que ofereix la *Big Learning Data* i la possibilitat de consultar-la mitjançant el mòdul de *Filtered Reporting*. Al mateix temps, el recull de les interaccions de l'alumne amb la resta de companys virtuals permet valorar l'*aprenentatge social* de cada individu.

Per últim, una altra conclusió que es pot treure és que el sistema ha de permetre la interacció de múltiples persones de les que és recull infinitat de dades del procés d'aprenentatge. Per tant, sembla bastant indiscutible que els sistemes en línia que es dissenyin han de ser sistemes web d'alt rendiment, on el tràfic d'informació que viatja pel sistema sigui suportat per la infraestructura. Aquest aspecte és clau perquè el sistema tingui èxit: una arquitectura ben dimensionada, optimitzada a nivell de rendiment, escalable i fàcilment mantenible.

Capítol 6

Contribucions en algorismes intel·ligents

*“Is there anyone so wise as to learn
by the experience of others?”*

Voltaire

Un Sistema Tutor Intel·ligent, per la seva pròpia definició necessita d'un conjunt de tècniques integrades que permetin dotar al sistema d'una capacitat de raonament. Tal com s'ha vist en el capítol anterior, el mòdul pedagògic integra la major part d'aquesta intel·ligència i és l'encarregat de donar suport al professorat en el procés d'aprenentatge de l'alumne.

Com es veurà ben aviat en aquest capítol, el Raonament Basat en Casos (*Case-Based Reasoning* de l'anglès, CBR) és la tècnica principal implicada per dur a terme el disseny d'aquest mòdul. La versatilitat que ofereix a l'hora de classificar situacions segons diferents estats del passat i les possibilitats que ofereix en problemes de classificació, el fan un candidat idoni per aquest objectiu. Tot i així, la particularitat de la seva aplicació final fa imprescindible una discussió a fons del seu funcionament tot estudiant les diferents modificacions que s'haurien d'aplicar a la tècnica estàndard per tal d'assolir la finalitat establerta.

Tenint en compte que les modificacions dels algorismes s'han hagut de provar exhaustivament, el lector veurà que els dominis de prova emprats en les validacions de les tècniques, sovint no són provinents només de resultats acadèmics de l'alumnat, sinó que provenen de jocs de dades o projectes utilitzats en el Grup de Recerca. L'avantatge d'utilitzar aquests jocs en les validacions dels algorismes és que, a diferència de la mida reduïda de conjunts de dades educatives recollides, els jocs de dades genèrics permeten provar els algorismes en diferents condicions. Tot i així, com més endavant es tractarà, els jocs de dades utilitzats respecten la tipologia de

dades que es pot esperar d'un alumne: un conjunt d'atributs variables i la possibilitat de donar una classificació final en més d'una classe. Aquests dos factors són els més importants per tal de realitzar aquesta validació.

Tanmateix el recull de les dades dels alumnes provinents d'un model basat en una quarta generació d'aprenentatge, es trasllada a una línia futura d'aquesta tesi doctoral. Són molts els factors que intervenen en la tria de les dades adients i el moment *Big Data* que es viu en l'actualitat dificulta i allarga aquesta tasca necessària.

Com es veurà tot seguit, aquest capítol s'estructura en sis blocs ben diferenciats. En primer lloc, es fa una introducció de què s'entén per aprenentatge automàtic i quines tècniques són les més destacables en aquest àmbit. A continuació es dona una visió del funcionament estàndard del Raonament Basat en Casos, tot descrivint les seves característiques principals i les fases que el componen. Els següents tres apartats se centren en cadascuna de les fases més importants d'aquesta tècnica d'aprenentatge artificial a la vegada que introdueixen les millores i contribucions que s'han definit per tal d'assolir el resultat final. Finalment, l'últim punt extreu les conclusions d'aquest capítol.

6.1 L'aprenentatge automàtic

L'aprenentatge automàtic, també anomenat aprenentatge artificial, és una branca del camp de la Intel·ligència Artificial que té per objectiu *dotar de la capacitat d'aprendre a certs dispositius*. Dins de la comunitat científica aquesta descripció no ha estat lliure de polèmiques. Potser una de les coses més difícils de definir és què s'entén per *aprendre*. Si se cerca al diccionari la definició d'aquesta paraula es pot trobar una frase similar a aquesta: “Adquirir la coneixença, la pràctica d'alguna cosa amb l'estudi o l'atenció”¹⁵. Tot i que aquestes paraules puguin tenir una repercussió més enllà del camp de treball actual, certament, una de les missions dels algorismes és assolir aquesta capacitat de coneixença, mitjançant la qual puguin ser capaços de predir o generalitzar una determinada situació o comportament.

Així, les aplicacions de l'aprenentatge artificial s'estenen més enllà de la pròpia informàtica. Molts són els sectors que utilitzen ordinadors o dispositius en general amb aquestes característiques: hospitals, cadenes de muntatge, entitats financeres, sistemes de control d'accés (Vernet & Canaleta, 2004)... En tots ells l'aprenentatge artificial ha ajudat en bona mesura a millorar la qualitat de vida o els productes finals que s'estan generant.

En el món de l'aprenentatge artificial es poden classificar de manera general les diferents tècniques existents mitjançant la següent taxonomia:

¹⁵ Definició procedent del Grup Enciclopèdia Catalana (<http://www.diccionari.cat>)

- *Aprenentatge supervisat.* Recull aquelles tècniques d'aprenentatge automàtic en les que es disposa d'un *supervisor* de les dades d'entrada, és a dir, es coneix a priori la correspondència entre les dades d'entrada i les de sortida. Sovint aquest tipus d'aprenentatge inclou problemes de classificació en els quals un conjunt de dades d'entrada ha de ser etiquetat amb una sortida a partir d'un conjunt d'entrenament definit anteriorment. De vegades pot succeir que algunes de les dades d'entrada no estiguin classificades i requereixin de la intervenció d'un expert que interactuï amb el sistema; en aquest cas ens trobem en l'anomenat *aprenentatge actiu*.
- *Aprenentatge no supervisat.* En aquest cas no es disposa d'un patró o una resposta per a les dades existents i els propis algorismes han de ser capaços de generalitzar i reconèixer nous patrons per tal de poder etiquetar noves entrades. Molts d'aquests mètodes es basen en models de mineria de dades, on el descobriment d'informació no evident es fa del tot imprescindible.
- *Aprenentatge per reforçament.* Aquest tipus d'aprenentatge s'aplica sovint quan l'entorn d'aprenentatge és variant en el temps. El resultat de l'algorisme i la sortida produïda en l'entorn serveix per intentar millorar el rendiment en la següent iteració.

Segons com es miri, existeixen altres models d'aprenentatge però que la majoria d'investigadors consideren derivats o combinacions dels aquí descrits. En aquest treball es combinen els dos primers conjunts, essent la tècnica principal emprada (CBR) pertanyent a l'aprenentatge supervisat, mentre que altres tècniques que es proposen més endavant, com són els algorismes d'agrupament (*clustering*, en anglès) o els algorismes genètics, són tècniques que pertanyen a la segona classificació o a totes dues alhora.

6.2 Els problemes de classificació

Dins de l'aprenentatge supervisat els problemes de classificació són els més comuns. Consisteix en disposar d'un conjunt de dades a les quals se li assigna una etiqueta o classe. L'objectiu és poder deduir quina relació hi ha entre les dades d'entrada i la classe establerta per tal de poder *classificar* futures instàncies en la classe correcta. Aquesta necessitat és palesa en molts camps de la vida quotidiana. Alguns exemples podrien ser, per exemple, la classificació de certs pacients mèdics en funció d'uns símptomes, la classificació dels clients d'una entitat bancària segons el seu perfil de morositat o risc, el resultat d'un partit de futbol en funció de les característiques i la situació dels equips que s'enfronten o, lligant-ho amb el cas que ens ocupa, el nivell de progressió d'un alumne en un determinat moment del curs.

Així doncs, un problema de classificació està compost per una sèrie d'instàncies $I = \{I_1, I_2, \dots, I_n\}$ on cadascuna d'elles tenen un conjunt d'atributs $A = \{A_1, A_2, \dots, A_n\}$ que la defineixen i una classe resultant que pot trobar-se entre el conjunt de classes possible $C = \{C_1, C_2, \dots, C_n\}$. Els atributs de cada instància poden ser atributs reals o nominals i la classe acostuma a estar sempre associada a una etiqueta. De manera formal, un problema de classificació consisteix en

classificar un nou conjunt d'instàncies I' en algunes de les classes contingudes en el conjunt C a partir de la informació dels atributs $A' = \{A'_1, A'_2, \dots, A'_n\}$ que les defineixen.

Històricament la classificació d'instàncies s'ha vingut efectuant de moltes maneres diferents. L'objectiu de tot procés ha estat entendre la importància de cadascun dels atributs i generalitzar la seva relació per tal de deduir-ne la seva etiqueta. Certament, el Raonament Basat en Casos és una de les tècniques més utilitzades per a la classificació d'instàncies i és l'eix central de la classificació efectuada en aquesta tesi. No obstant, altres tècniques són també destacables. A continuació es comentaran els altres dos tipus de classificació més importants que s'han esdevingut en la història: la classificació amb arbres de decisió i la classificació fent ús de regles.

6.2.1 Classificació mitjançant arbres de decisió

Sens dubte, un dels mètodes tradicionals per excel·lència a l'hora de dur a terme una classificació ha estat l'arbre de decisió. Ideat per Quinlan (Quinlan, 1986) als anys vuitanta, aquesta solució proposava classificar les instàncies tot construint el coneixement en format d'arbre segons el valor dels atributs concrets. L'algorisme de construcció d'arbres més famós ha estat l'anomenat ID3. Un exemple del resultat d'aquest tipus de classificació és el que s'observa a la Figura 34.

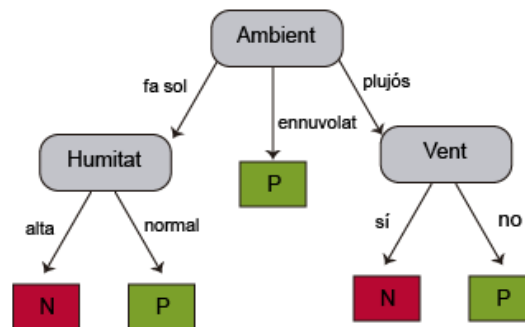


Figura 34. Exemple del resultat d'una classificació utilitzant l'algorisme ID3

L'arbre que es mostra com a exemple resol el problema següent: “en quines situacions es pot participar en una cursa de muntanya?”. Com es pot veure a les fulles de l'arbre en resulten les classes finals (P=situació positiva; N=situació negativa). Seguint amb l'exemple mostrat es pot veure que l'algorisme intenta escollir l'atribut més discriminant com a primera opció. Els valors dels atributs són excloents entre ells. Així, aquest arbre es llegeix de la següent manera pel que fa als casos positius: “si fa sol i la humitat és normal, es pot córrer una cursa de muntanya”, o bé “si l'ambient és plujós però no fa vent, també es pot participar a la cursa”, o bé “si està ennuvolat, també es pot anar a la cursa”.

Aquest algorisme té certes restriccions, com per exemple que els atributs han de ser discrets i els valors que poden prendre han de ser excloents entre ells. Millores posteriors del mateix autor van portar a d'altres algorismes que construïen arbres de decisió més eficients com l'algorisme C4.5 (Quinlan, 1993). En general els arbres de decisió resultants d'una classificació han de ser el més simples possibles que classifiquin correctament tot el conjunt de dades d'entrenament.

6.2.2 Classificació mitjançant regles

Una regla es pot definir de manera senzilla utilitzant una sentència condicional de l'estil “*SI succeeix quelcom LLAVORS conseqüència*”. Si se segueix l'exemple de l'apartat anterior el lector pot veure que les conclusions que s'han extret de l'arbre de decisió són exactament regles amb aquest format.

Així doncs, construir un classificador mitjançant regles té com a objectiu aconseguir un conjunt de regles que permetin, de la mateixa manera que en els arbres de decisió, classificar correctament els exemples d'entrada. Com es pot imaginar es pot aconseguir fàcilment convertir un arbre de decisió en un conjunt de regles equivalent.

Paral·lelament, els algorismes genètics (Goldberg, 1989) s'han utilitzat amb èxit per a la construcció de regles que classifiquin un conjunt d'instàncies. Aquests algorismes simulen el comportament de l'evolució natural de Darwin proposant solucions formades per individus resultants del creuament d'altres individus, aconseguint així una població diversa que permet trobar un conjunt de regles resultants força encertat. Hi ha dues aproximacions principals per aconseguir el conjunt de regles finals:

1. *Aproximació de Pittsburgh*. Proposada per De Jong (De Jong, 1975), en aquesta aproximació cada conjunt de regles candidat està representat per un sol individu. Així, l'evolució dels individus provocarà que el millor individu final sigui l'escollit, representant el conjunt de regles finals del classificador.
2. *Aproximació de Michigan*. Proposada per Holland (Holland, 1975), aquesta aproximació tracta cada regla com un individu i per tant el conjunt de regles final és la totalitat de població.

Més endavant, en aquest mateix capítol, es veurà com s'han utilitzat els algorismes genètics dins del Raonament Basat en Casos per assolir un objectiu diferent.

6.3 Metodologia de proves i dominis utilitzats

Com s'ha comentat, malgrat l'aplicació final dels resultats d'aquesta tesi es troba en un àmbit educatiu, no només s'ha utilitzat aquest domini de dades per testejar el funcionament dels

algorismes proposats sinó també en altres conjunts de dades que respecten una tipologia similar. Concretament, s'han utilitzat dades provinents dels següents camps:

- *Entorn educatiu.* Es disposen de dades anònimes provinents d'alumnes de diferents assignatures de Graus i Màsters. Molt sovint les dades corresponen a una *fotografia* acadèmica de l'alumne en un moment determinat del curs amb la finalitat de poder predir la seva evolució al final del curs, com es representa a la Figura 35.

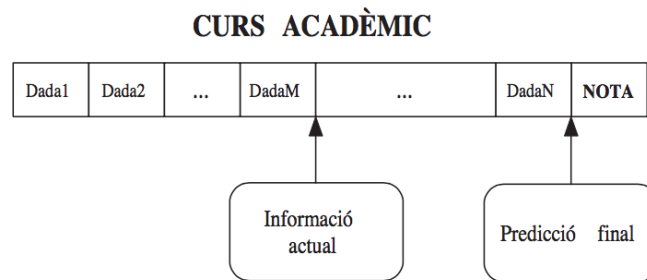


Figura 35. Representació gràfica del conjunt de dades provinents d'un entorn educatiu.

- *Entorn mèdic.* S'han utilitzat també dades anonimitzades corresponents a diagnòstics de càncer de mama provinents del projecte FIS 00/0033 (Martí, et al., 2000) i de biòpsies provinents de diferents centres hospitalaris. Aquestes dades arribaven preprocessades digitalment amb tots els atributs calculats i preparats per poder ser utilitzats en format numèric. A la Figura 36 es pot veure un exemple d'una d'aquestes instàncies i la seva segmentació posterior.

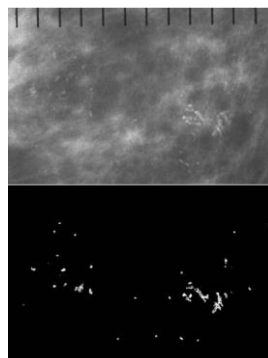


Figura 36. Imatge digital d'una mamografia i la seva segmentació posterior.

- *Entorn telemàtic.* Com es veurà posteriorment, certes dades provenen de problemes d'optimització en l'enrutament de paquets de dades (Vallejo, Zaballos, Canaleta, & Dalmau, 2008). L'objectiu en aquest conjunt de dades és optimitzar l'enrutament de paquets en una tipologia de xarxa com per exemple la que es mostra a la Figura 37.

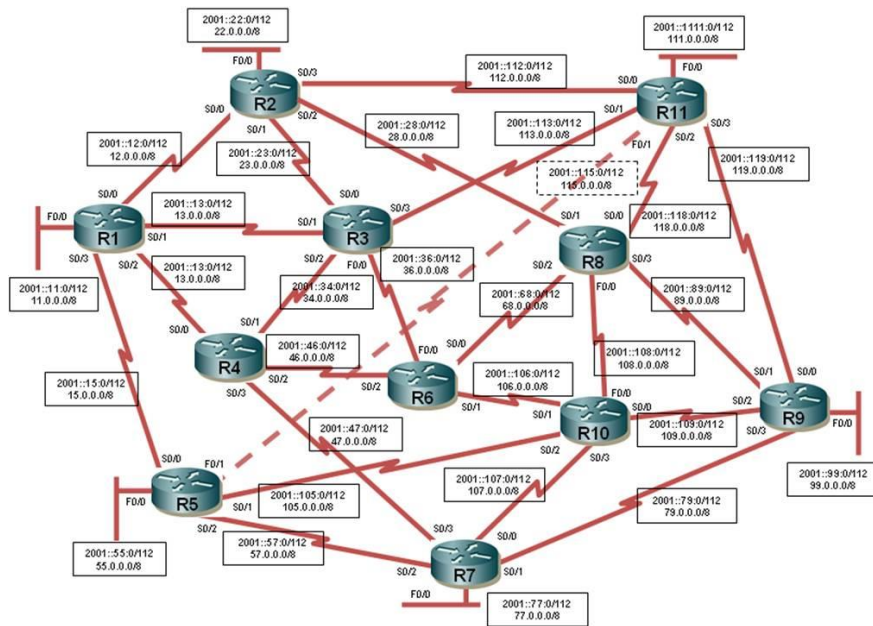


Figura 37. Tipologia de xarxa exemple utilitzada per optimitzar l'enrutament de paquets en el problema relacionat amb l'entorn telemàtic.

- *Repositori UCI*. Molts dels algorismes s'han provat en diferents conjunts de dades provinents de l'*UCI repository* (Lichman, 2013). Això permet validar els algorismes amb problemes que presenten diferents característiques en la distribució i tipus de les dades, així com en el nombre de classes finals. A la Taula 7 es mostren els conjunts de dades que s'han utilitzat més d'aquest repositori. De cadascun d'ells es detallen el nom complet, el nom abreujat, el nombre d'instàncies disponibles i la quantitat de classes possibles. El motiu de la tria d'aquests jocs concrets és la seva variabilitat en el nombre de classes i quantitat d'instàncies.

Taula 7. Conjunts de dades de l'UCI repositori més utilitzats per a la validació dels algorismes. Per cadascun d'ells, s'indica la seva etiqueta abreujada, el nombre d'instàncies, el nombre d'atributs per instància, quants atributs són nominals i quants numèrics i el nombre de classes de classificació.

Nom del joc de dades	Abrev.	Instàncies	Atributs	Nom.-Num.	Classes
Glass	GL	214	10	0-10	6
Ionosphere	IO	351	34	0-34	2
Iris	IR	150	4	0-4	3
Multiplexor 11	MX	2048	10	10-0	2
Vowel	VO	990	13	3-10	11
Sonar	SO	208	60	0-60	2
Vehicle	VE	846	18	0-18	4

- *Problemes sintètics*. S'han generat també alguns problemes sintètics per descobrir certes característiques concretes de comportament dels algorismes i també per simular situacions especials en un entorn educatiu.

En general, les validacions dels algorismes s'han efectuat usant tècniques de validació creuada i s'han comparat els resultats realitzant una prova de *t-student* amb un nivell de significació del 95%. Concretament, les tècniques de validació creuada que s'han utilitzat són:

- *Particionat simple o holdout method*. Es parteix el conjunt de dades en dues parts: una part d'entrenament i una de test. Aquest mètode proporciona un temps de resposta molt ràpid però implica una variança dels resultats molt àmplia.
- *k-fold stratified cross-validation*. El conjunt de dades es divideix en k particions i s'escull una d'elles com a conjunt d'entrenament. A continuació es prenen els restants $k-1$ paquets i s'agrupen tots junts per formar el conjunt d'entrenament, contra el qual es prova l'algorisme. Això es repeteix per a totes les combinacions de paquets possibles i es fa la mitjana del resultat obtingut. L'adjectiu *stratified* en aquest tipus de prova proporciona la certesa que en tots els conjunts de dades generats està garantida la presència de com a mínim una instància de cada classe resultant.
- *Leave one out*. Consisteix en deixar només una instància per a test i utilitzar la resta de dades com a conjunt d'entrenament. De fet, seria com portar a l'extrem el mètode anterior fent que k sigui exactament igual al nombre de dades disponibles.

El programari s'ha implementat majoritàriament fent ús dels llenguatges C i C++ i s'ha tingut en compte que els fitxers d'entrada siguin compatibles amb el format ARFF utilitzat per WEKA (Witten & Frank, 2000). Alguns dels resultats obtinguts també provenen de l'eina *Sagittarius*, implementada íntegrament en Java (Solé, et al., 2014).

6.4 El Raonament Basat en Casos

El Raonament Basat en Casos és una tècnica d'aprenentatge artificial, nascuda l'any 1977 de mans de Schank i Abelson (Schank & Abelson, 1977), que integra l'aprenentatge dins de la pròpia resolució dels problemes. S'inspira en el concepte d'analogia, tècnica que usem els humans per resoldre la majoria dels problemes de la vida quotidiana, en la qual els problemes nous que es presenten s'intenten resoldre *recordant* (Riesbeck & Schank, 1989) com s'havien resolt altres similars anteriorment. Al mateix temps, la resolució de nous casos aporta informació i coneixement per tal de servir d'ajuda per resoldre altres casos en el futur.

Aquesta tècnica difereix d'altres tècniques d'Intel·ligència Artificial bàsicament en dos aspectes. El primer és que el CBR és capaç d'usar coneixement específic de l'experiència prèvia, ja que un nou problema es resol buscant el cas ja resolt anteriorment més similar a la situació actual. En segon terme, el CBR posseeix un aprenentatge incremental, en el sentit que s'emmagatzemen, si s'escau, noves experiències a mesura que es resolten. Com a tècnica d'aprenentatge supervisat rellevant, el CBR espera que els casos anteriorment resolts o utilitzats per a l'entrenament estiguin correctament etiquetats i classificats.

El cicle de vida del Raonament Basat en Casos va ser definit formalment per (Aamodt & Plaza, 1994) i va ser resumit en un esquema on el cicle se centra en la Memòria de Casos i en les quatre fases que s'exposen a continuació i que es representen a la Figura 38, ja introduïda a la introducció d'aquesta tesi.

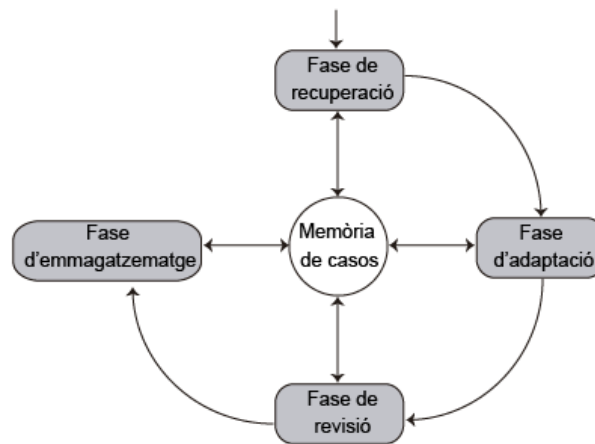


Figura 38. Cicle bàsic del Raonament Basat en Casos proposat per Aamodt i Plaza

6.4.1 La Memòria de Casos

La Memòria de Casos és la part més rellevant d'un sistema de Raonament Basat en Casos. En aquesta memòria s'emmagatzemen els casos que han de servir d'experiència passada per resoldre noves situacions. L'organització de la Memòria de Casos pot arribar a ser molt complexa, depenent del tipus de casos que es disposin. La manera més comuna és que els casos siguin instàncies per ells mateixos d'experiències basades, però existeixen moltes altres maneres d'organitzar la memòria de casos, com poden ser a partir de generalitzacions, abstraccions, índexs o Paquets d'Organització de Memòries (MOP) (Riesbeck & Schank, 1989).

Tanmateix, una altra manera d'organitzar la memòria consisteix en agrupar casos similars i convertir la memòria de casos en un conjunt d'agrupacions o *clusters* que resumeixen diferents tipologies de casos. En aquest treball, com es veurà més endavant, s'ha utilitzat també aquesta

forma d'organització. Altres maneres d'organitzar la memòria de manera no lineal, com ara representacions jeràrquiques, també són acceptades per determinats dominis.

No totes les experiències noves que es troben s'han d'emmagatzemar a la Memòria de Casos. Normalment aquells casos que no aporten informació nova no s'acostumen a registrar en la memòria de casos, doncs experiències similars es poden deduir a partir de la informació existent a la Memòria de Casos. Tot i així, existeixen diverses polítiques en aquest sentit, algunes de les quals s'han usat en aquest treball i seran comentades més endavant.

Malgrat tot, un cas enregistrat en la Memòria de Casos hauria de comptar com a mínim amb la següent informació:

- *Característiques del cas.* Això hauria d'incloure un conjunt d'atributs que defineixin el cas amb una sèrie de valors registrats per cadascun d'ells. No obstant, no sempre tots els atributs tenen valors assignats. A vegades es poden trobar el que s'anomenen *casos incomplets*, en els quals algun d'aquests atributs té un valor desconegut. Els tipus dels atributs d'un cas poden ser numèrics, discrets o qualsevol variant més complexa.
- *Solució al cas.* Cadascun dels casos de la memòria de casos hauria d'aportar informació respecte a l'etiqueta assignada per aquesta situació. Tot i així, pot ser possible topar-se amb algun cas o el valor de la classe és indefinit. Aquesta informació pot ser útil igualment per resoldre nous problemes.

Cal fer notar també, que en funció del domini de resolució o depenent de l'aplicació final que es planteja, pot ser útil emmagatzemar informació extra del cas, tal com descripcions textuales de certs atributs, respostes predeterminades per informar a l'usuari o informació rellevant addicional al cas.

De la mateixa manera, altres informacions que es poden emmagatzemar poden tenir a veure amb restriccions que s'apliquen a aquest cas o justificacions de les decisions que s'han pres per etiquetar aquest cas amb aquesta classe. En ocasions també s'inclou informació en el cas sobre quin ha estat el motiu pel qual va fallar aquest cas quan va ser escollit o, si ha estat reparat, quina estratègia es va seguir per fer-ho. Com es pot observar, un cas pot arribar a contenir molta informació i la manera de definir-lo quan es vol codificar serà clau perquè el sistema creat tingui un bon rendiment.

6.4.2 Fase de recuperació

La fase de recuperació és la primera de les quatre fases que defineixen un cicle de Raonament Basat en Casos. L'objectiu d'aquesta fase és recuperar de la Memòria de Casos el cas o casos

més similars al que s'està valorant. Per tal de procedir a aquesta recuperació cal tenir en compte els següents factors:

- *Velocitat en la recuperació.* Si la memòria de casos és molt gran és necessari definir una estratègia de recuperació abans d'executar el procés pròpiament dit. L'ús d'índexs que permetin un accés directe a les dades pot ser essencial per al bon rendiment del sistema.
- *Nombre de casos a recuperar.* Tot i que en general l'objectiu és sempre trobar el cas més general, molt sovint, en una primera iteració, es recuperen un conjunt de casos candidats. La definició del nombre idoni de casos també és un paràmetre important pel sistema i pot ser cabdal per donar una resposta encertada.

El procés de recuperació consta dels següents passos diferenciats:

1. Cerca del cas o casos més similars al nou cas
2. Ordenació dels resultats obtinguts
3. Selecció del millor cas o aplicació d'una estratègia que inclogui més d'un cas recuperat

Tradicionalment, quan els casos tenen una estructura estàndard, és a dir, un conjunt d'atributs finit i una classe definida i el domini és *pla*, on tots els casos continguts a la memòria de casos tenen el mateix nombre d'atributs, la cerca dels casos més similars s'acostuma a fer mitjançant la definició d'una funció de similitud entre dos casos. Aquesta funció de similitud té en compte les diferències entre els atributs dels dos casos. La més usada és l'anomenada *Nearest Neighbour* o *Veí Més Proper* (Aha, 1998), la qual computa la similitud de dos casos tot usant una mesura de similitud global. Una de les mesures de similitud més utilitzades en el Raonament Basat en Casos és la distància de Minkowski, definida de la següent manera:

Siguin dos casos X i Y , definits pels següents atributs:

$$X = \{x_1, x_2, \dots, x_n\} \quad Y = \{y_1, y_2, \dots, y_n\}$$

Es defineix la distància de Minkowski d com:

$$d = \sqrt[p]{\sum_{i=1}^n [S(x_i, y_i)]^p} \quad (1)$$

On S és la funció de similitud aplicada entre dos atributs qualssevol dels casos i p és la generalització de la distància, tal que per $p \geq 1$ defineix tres mètriques diferents usades habitualment com a funcions de similitud entre dos casos:

- Distància de Hamming ($p=1$)
- Distància Euclidiana ($p=2$)
- Distància Cúbica ($p=3$)

La manera més simple de definir la similitud entre dos atributs depèn de si aquests atributs són continus o discrets. En cas que els atributs siguin continus es defineix S com:

$$S(x_i, y_i) = |x_i - y_i| \quad (2)$$

En canvi, quan els atributs són discrets, la manera més simple i tradicional de definir aquesta similitud és la que es defineix a l'equació 3:

$$S(x_i, y_i) = \begin{cases} 0, & x_i \neq y_i \\ 1, & x_i = y_i \end{cases} \quad (3)$$

En un entorn educatiu, per exemple, els atributs en general són continus ja que reflecteixen les qualificacions obtingudes per un alumne o el temps de resposta en solucionar un problema. No obstant això, ens podem trobar amb atributs discrets relacionats amb aquest domini com ara la classe a la que assisteix aquest alumne, saber si és repetidor o no o les seves preferències d'usuari.

En tots els casos cal tenir en compte que certs atributs poden tenir una ponderació diferent a l'hora de comparar dos casos. Així, els atributs més importants haurien de tenir més rellevància en la comparació, mentre que aquells que no siguin tan importants no haurien d'afectar en gran mesura a aquesta similitud. Per incloure aquest aspecte, es pot introduir la influència dels pesos de cada atribut a la distància de Minkowski de la següent manera:

$$d = \sqrt[p]{\sum_{i=1}^n w_i [S(x_i, y_i)]^p} \quad (4)$$

Com es pot veure a l'equació 4, la ponderació dels atributs es duu a terme amb la introducció d'un vector de pesos $W = \{w_1, w_2, \dots, w_n\}$ per cadascun dels atributs que formen els casos continguts a la memòria de casos.

Un altre problema que cal resoldre a l'hora de trobar la similitud entre dos casos és el marge de valors que pot prendre cadascun dels atributs continguts en el cas. Així, dos atributs que tinguin un recorregut molt diferent poden influir significativament en el resultat de la comparació entre els dos casos. Per evitar això s'introdueixen sovint tècniques de *normalització* de les dades (Fayyad, Piatesky-Shapiro, Smyth, & Uthurusamy, 1996) amb la finalitat d'ajustar-ne els rangs i fer que els atributs siguin totalment comparables. Per exemple, no és mateix comparar un

atribut com l'edat d'una persona, que pot tenir una variació d'un centenar de valors, amb el nombre de fills que pot tenir, on el valor diferirà d'un cas a un altre en poques unitats.

S'ha comentat anteriorment que la recuperació més tradicional es fa utilitzant l'algorisme *Nearest Neighbour (NN)* com a funció de similitud global. Existeixen variacions d'aquest algorisme per tal de poder recuperar els k casos més similars convertint així l'algorisme en un k -*NN*. En aquest cas es poden seguir diverses polítiques de votació posteriors per decidir quina classe és l'assignada al nou cas que arriba.

S'han provat també altres funcions de similitud dissenyades mitjançant Programació Genètica i presentades a (Camps, Garrell, Golobardes, & Vernet, 2003). S'ha comprovat que aquest tipus de funcions no són aplicables a l'entorn plantejat en aquesta tesi, atès el seu gran cost de computació i la seva dependència directa de les dades existents per a cada alumne.

6.4.3 Fase d'adaptació

La fase d'adaptació és la que ha de permetre ajustar la solució proposada per tal que sigui una solució vàlida per al problema que es vol resoldre. Normalment, en problemes de classificació, la funció aplicada és la Identitat i la solució inicialment proposada serveix directament com a solució del problema.

En altres entorns hom pot trobar-se en situacions on calgui fer alguna substitució, *reinstanciació*, ajust de paràmetres, cerca especialitzada o noves cerques dins de la memòria de casos per tal de complementar la resposta proposada. Altres mètodes possibles són mètodes que introdueixin alguna mena de transformació en la solució, ja sigui perquè s'aplica una reparació al cas proposat o bé perquè s'aplica certa heurística pròpia del problema que es vol resoldre. En situacions especials es pot dividir el problema en diferents passos de resolució per intentar aprendre de cada pas i tractar-los en la seva totalitat com la resposta del problema. En aquest cas s'arriba a una solució incremental: és el que s'anomena *analogia derivacional*.

6.4.4 Fase de revisió

Un cop es té una solució candidata del problema que es vol resoldre, el Raonament Basat en Casos requereix que aquesta solució es comprovi/provi en l'entorn o domini del propi problema. En certa manera, el sistema avalua la bondat de la solució i interactua per tal de saber si la solució proposada serà correcta.

A vegades, s'afegeix una simulació paral·lela o la intervenció d'un expert del domini del problema. En cas que la prova no sigui acceptable, s'intenta identificar els motius del resultat erroni i el sistema necessita corregir el resultat atenent a l'informe que li arriba. Aquesta

correcció pot passar per una nova adaptació del cas o per proposar una solució alternativa, ja sigui la mateixa reparada o una altra de diferent.

Com s'ha comentat, en el treball present, on els problemes que es plantegen són de classificació, tant la fase d'adaptació com la de revisió seran senzilles, doncs normalment no cal cap adaptació addicional a la proposta de solució i aquesta proposta és compatible amb el domini del problema que s'està resolent.

6.4.5 Fase d'emmagatzematge

Finalment, la fase d'emmagatzematge és l'encarregada d'incorporar coneixement al sistema i realitzar l'aprenentatge automàtic. Quan un cas es resol, positivament o negativament, el sistema ha de decidir si cal emmagatzemar aquesta situació a la Memòria de Casos, per tal que en situacions futures es pugui utilitzar l'experiència assolida.

Existeixen diferents criteris i alguns de nous que s'han aplicat en aquesta tesi per tal de decidir si un cas resolt cal ser emmagatzemat. En posteriors apartats es podrà comprovar quines polítiques s'han definit en aquest sentit. Cal tenir en compte que la decisió d'emmagatzemar una solució a la memòria de casos cal que valori com a mínim dos factors importants:

- *Excepcionalitat del cas.* Si es detecta que el cas resolt és un cas especial que no pot ser resolt amb la informació existent a la Memòria de Casos, llavors pot ser bo analitzar si aquest cas cal que sigui emmagatzemat. De la mateixa manera, si el cas resolt ha estat erroni i emmagatzemar aquesta situació pot permetre evitar futurs errors, també cal considerar la possibilitat de registrar-lo al sistema.
- *Mida de la Memòria de Casos.* Sense dubte, no es pot emmagatzemar a la Memòria tots els casos que es troben, doncs la seva mida influirà tant en la seva organització com en el rendiment del sistema. Per tant, aquest factor és crucial i valdrà la pena valorar-lo a l'hora de prendre la decisió correcta.

A l'apartat 6.5.5 es mostraran les aportacions realitzades en aquest sentit i s'analitzaran les seves conseqüències a nivell de rendiment i fiabilitat del sistema.

6.5 Modificacions i aportacions al cicle del CBR

A continuació s'introduiran les aportacions realitzades dins del cicle del Raonament Basat en Casos per tal d'assolir els objectius que han de permetre construir el sistema final. Com s'ha

comentat ja en el capítol anterior, les aportacions que es faran cercaran demostrar bàsicament dues circumstàncies:

- És possible predir l'evolució de l'alumne utilitzant algorisme intel·ligents, i concretament, fent ús de Raonament Basat en Casos.
- És possible modificar el funcionament i el cicle del Raonament Basat en Casos per tal que es pugui predir l'evolució d'un grup d'alumnes virtual per millorar-ne el seu rendiment i fer-ne la seva tutorització posterior.

Per assolir el primer objectiu s'utilitzarà l'estructura d'una Memòria de Casos tradicional, mentre que per al segon objectiu caldrà fer-ne algunes variacions. En aquest capítol primer es veuran les tècniques que s'han dissenyat i analitzat per assolir la primera fita i posteriorment s'introduiran les modificacions que porten al segon objectiu.

6.5.1 Diferents organitzacions de la Memòria de Casos

S'ha vist doncs que un dels objectius finals que es persegueix en aquesta tesi és poder treballar de manera directa amb grups d'alumnes existents a l'aula virtual. La manera de representar aquesta situació dins d'un sistema de CBR no és compatible amb l'estructura normal d'una Memòria de Casos. Normalment, com també s'ha comentat anteriorment, la Memòria de Casos, en un sistema de Raonament Basat en Casos tradicional, està organitzada de tal manera que fa correspondre cada instància amb un cas.

Per tant, sembla bastant raonable pensar que caldrà una modificació al sistema per tal de poder treballar amb grups d'alumnes tot fent que, d'una banda, cada grup històric vingui representat per un cas dins de la Memòria de Casos, i d'altra banda, cadascun dels casos actius amb els que s'està treballant (l'aula pròpiament dita) també segueixin una estructura que permeti tractar amb més d'un individu a la vegada. Per aconseguir tal objectiu caldrà introduir l'opció de modificar l'organització de la Memòria de Casos i introduir algorismes d'agrupació o *clustering* en el conjunt de test, que afectaran a la inicialització del sistema, a la fase de recuperació i a la fase d'emmagatzematge del cicle de CBR.

Tanmateix, quan es vol predir l'evolució d'un alumne cal donar l'opció que aquesta predicció sigui a nivell individual o grupal. Per aquest motiu, la memòria de Casos s'ha organitzat tot donant la possibilitat d'interactuar amb ella en aquests dos escenaris.

A continuació es veuran els resultats que s'obtenen en la predicció individual de les evolucions dels alumnes, per passar posteriorment a una organització que permeti treballar amb informació grupal.

6.5.2 Predicció de l'evolució individual d'un alumne

Tot Sistema Tutor hauria de tenir la possibilitat de poder avaluar quin és el grau d'assoliment de coneixements que té un alumne en un moment determinat del curs, per tal de poder-lo orientar i intentar millorar la seva situació tan aviat com sigui possible. Aquesta predicció s'ha de basar en les dades que s'han recollit de l'alumne fins al moment i, per tant, la fiabilitat d'aquest sistema també dependrà de la quantitat i qualitat de les dades disponibles.

A continuació s'analitzaran les tècniques i aportacions introduïdes en el Raonament Basat en Casos per a realitzar la predicció de l'evolució de l'alumnat i els resultats obtinguts en aquest sentit.

6.5.2.1 Tècniques utilitzades per a la predicció

Tot i que el funcionament de les tècniques presentades a continuació s'han testejat en altres problemes de classificació (Vernet, 2004) (Camps, Garrell, Golobardes, & Vernet, 2003) (Nicolas, et al., 2009) (Golobardes, Llorà, Garrell, Vernet, & Bacardit, 2000) (Golobardes, et al., 2001), en aquest apartat es mostraran només els resultats dins de l'àmbit educatiu. Les proves s'han realitzat a partir de dades anonimitzades corresponents a ensenyament presencial recollides en diferents promocions de les assignatures de "Metodologia i Tecnologia de la Programació" (648 alumnes) i "Disseny i Programació Orientats a Objecte" (104 alumnes) de primer i segon curs de diferents Graus d'Enginyeria, així com dades procedents del Màster en Formació del Professorat (29 alumnes) (Canaleta, Vernet, Vicent, & Montero, 2014) impartit a la Universitat. Malgrat les proves s'han realitzat en diferents assignatures, els resultats que es presenten en aquesta tesi se centren en els de l'assignatura amb més alumnes matriculats, doncs en les altres assignatures els resultats no eren estadísticament significatius.

A l'hora de predir quina seria l'evolució individual d'un alumne, s'han definit dos objectius diferents:

1. Intentar predir una valoració a "alt nivell", tot indicant si l'alumne aprovarà o suspèn timerà l'assignatura que està cursant.
2. Afinar més el resultat anterior i predir en quina situació es troba exactament, tot definint els següents cinc intervals possibles de sortida:
 - a. L'alumne porta una trajectòria on sembla que tindrà temptacions de no presentar-se a l'examen final de l'assignatura.
 - b. L'alumne es presentarà a l'examen però els suspèn timerà clarament, amb una qualificació compresa entre 0 i 4 punts.

- c. L'alumne tindrà un resultat just amb possibilitats tant de suspendre com d'aprovar l'examen. Situem en aquest interval tots els alumnes dels que es prediu una qualificació final per sobre o igual a 4 punts però inferior a 7.
- d. Els alumnes notables es qualifiquen també en un grup independent, amb notes finals compreses entre 7 i 9 punts.
- e. L'últim grup correspon a alumnes excel·lents, on la seva evolució fa pensar que finalment assoliran una qualificació superior o igual a 9 punts.

D'altra banda, també s'ha perseguit un tercer objectiu, que passa per descobrir quina és la ponderació real de les diferents informacions que es posseeixen dels alumnes (Vernet, et al., 2001). És a dir, amb referència a l'assignatura que s'està predint, és interessant saber quina és la influència en la trajectòria de l'alumne de cadascun dels exàmens, exercicis o altres proves que s'hagin recollit.

Tot i que en funció de l'assignatura analitzada, el nombre de dades recollides és diferent, la Taula 8 mostra quines han estat les dades mínimes comunes de tots els conjunts de dades que s'han considerat per iniciar la predicció de l'alumne.

Taula 8. Conjunt de característiques mínimes escollides per tal de realitzar la predicció de l'evolució d'un alumne.

Característica

Classe/Grup

Indicador de repetidor

Examen 1

Examen 2

Pràctica 1

Pràctica 2

El significat de cadascuna d'aquestes dades és el següent:

- *Classe/Grup*. Indica en quina de les possibles aules ha assistit l'alumne. Tot i que els resultats haurien d'indicar que el professor que ha tingut l'alumne no té influència en el resultat final, es vol analitzar aquesta possibilitat.
- *Indicador de repetidor*. Aquest valor indica si l'alumne està repetint l'assignatura o no.
- *Examen 1*. Resultat del primer examen realitzat per l'alumne.
- *Examen 2*. Resultat del segon examen realitzat per l'alumne.
- *Pràctica 1*. Nota final de la primera pràctica realitzada per l'alumne.
- *Pràctica 2*. Nota final de la segona pràctica realitzada per l'alumne.

Per tal de millorar els resultats estàndard del Raonament Basat en Casos, s'han introduït les següents variacions:

1. Noves funcions de similitud utilitzades:

Les proves s'han realitzat utilitzant la distància de Minkowski ponderada indicada a l'equació 4 definida anteriorment, la distància de Mahalanobis (Campbell, 1980) basada en la matriu de covariàncies, la distància Equal Variance que iguala l'efecte de la variància en els atributs multiplicant cada atribut per l'invers de la seva desviació típica al quadrat i amb la distància de Clark definida de la següent manera:

$$d = \sqrt{\sum_1^n \frac{|x_i - y_i|^2}{|x_i + y_i|^2}} \quad (5)$$

2. Possible ignorància d'atributs:

S'han dissenyat diferents polítiques que permeten ignorar certs atributs en circumstàncies especials. Concretament s'han introduït les següents polítiques:

- i. *Ignore Mode Mean Std.* Aquesta política ignora aquells atributs que tenen una desviació estàndard superior a la mitjana de les desviacions dels atributs existents.
- ii. *Ignore Mode Worst Std.* En aquest cas s'ignora l'atribut que té una desviació típica més alta, proposant que aquest atribut és el que desestabilitza el sistema.

3. Ponderació d'atributs:

La ponderació dels atributs s'ha fet tenint en compte tres aproximacions diferents:

- i. Sense ponderació
- ii. Ponderació calculada a partir d'Algorisme Genètic (Vernet, et al., 2001). En l'apartat 6.5.6 s'analitza amb més profunditat les aplicacions realitzades amb aquesta tècnica d'Aprenentatge Artificial.
- iii. Ponderació introduïda pel professor de l'assignatura

6.5.2.2 Resultats obtinguts en la predicció

Els resultats obtinguts són els indicats a la Taula 9. Cal indicar que els resultats que es mostren a continuació són els obtinguts quan s'intenta predir l'evolució final de l'alumne a una de les 5 classes indicades anteriorment. S'han utilitzat les dades corresponents al curs més nombrós de l'assignatura de primer curs, pels motius que s'han explicat a l'apartat 6.5.2.1. Per avaluar la bonesa dels resultats s'ha experimentat també amb altres jocs de dades d'entorns no docents i no s'han utilitzat altres cursos de la mateixa assignatura perquè les dades no eren comparables entre elles (no es disposava del mateix tipus d'atributs) i tampoc es volia comprovar una tendència en les dades, sinó avaluar la qualitat del procés de predicció.

Taula 9. Eficàcia mitjana, màxima i mínima assolida amb les dades disponibles, en la predicció a 5 classes, tenint en compte les diferents ponderacions aplicades als atributs.

Tipus de ponderació	Eficàcia	Màxim	Mínim
Sense ponderació	61,18 %	76,56 %	43,75 %
Ponderació Genètics	67,82 %	85,94 %	60,23 %
Ponderació Professor	64,21 %	81,25 %	46,87 %

En els resultats sense ponderació dels atributs, s'ha donat la mateixa importància a tots els atributs. Com es pot veure, l'eficàcia resultant a partir del coneixement dels atributs aportat pel professor està fins i tot per sota a la proporcionada pels Algorismes Genètics, provocant que la ponderació automàtica millori els resultats de l'expert. Això pot indicar que les prediccions fetes per sistemes automàtics, com ara tutors intel·ligents, poden ser inclús més fiables que les realitzades per un professor habitual de l'assignatura.

A continuació, a la Taula 10 es pot observar quina ha estat la ponderació que ha estat donada a cadascun dels atributs després d'un entrenament realitzat amb Algorismes Genètics i la ponderació fixada per un humà expert.

Taula 10. Ponderacions resultants obtingudes usant Algorismes Genètics i informació d'un professor expert.

Característica	Ponderació GA	Ponderació professor
Classe/Grup	2,3 %	0 %
Indicador de repetidor	11,6 %	15 %
Examen 1	5,8 %	30 %
Examen 2	58,1 %	30 %
Pràctica 1	11,6 %	10 %
Pràctica 2	11,6 %	15 %

Analitzant la taula anterior és interessant veure com la ponderació automàtica efectuada pels Algorismes Genètics detecta dues coses importants:

1. La classe o el grup de l'alumne influeix, tot i que en petita mesura, en el resultat final de l'alumne. Això porta a la següent reflexió: implica doncs aquest resultat que el professor concret que tingui l'alumne en una aula presencial és important per al seu procés d'aprenentatge?
2. El segon examen és el més important en les assignatures de les que es disposa d'informació, mentre que el primer examen, on la dificultat encara és baixa, és relativament poc significatiu.

S'han dut a terme també proves de tendència de resultats eliminant per complet els dos indicadors que no corresponen a resultats directes de l'alumne. Per fer-ho s'ha utilitzat l'algorisme *Apriori* de regles d'associació de descobriment d'informació (Agrawal & Ramakrishnan, 1994) i que ha permès descobrir les següents tendències en les dades numèriques:

Núm.	Tendència	Soporte (%)	Confianza (%)
1	SI Examen2=NP y Practica1=SUSPENSO --> Practica2=SUSPENSO	13.0	96.0
2	SI Examen1=SUSPENSO y Practica1=SUSPENSO --> Practica2=SUSPENSO	13.0	92.0
3	SI Practica1=SUSPENSO --> Practica2=SUSPENSO	18.0	92.0
4	SI Examen1=NP --> Examen2=NP	11.0	91.0
5	SI Examen1=SUSPENSO y Examen2=NP --> Practica2=SUSPENSO	14.0	76.0
6	SI Examen2=NP --> Practica2=SUSPENSO	21.0	71.0
7	SI Examen1=SUSPENSO --> Examen2=SUSPENSO	41.0	61.0
8	SI Examen1=SUSPENSO --> Practica2=SUSPENSO	35.0	52.0
9	SI Examen1=SUSPENSO y Practica1=NOTABLE --> Practica2=SUSPENSO	11.0	49.0
10	SI Examen1=SUSPENSO y Examen2=SUSPENSO --> Practica2=SUSPENSO	19.0	47.0

Figura 39. Tendències detectades en les dades d'ensenyament eliminant completament aquelles dades que no depenen de l'esforç de l'alumne. Per cada regla obtinguda s'indica el seu suport i la seva confiança.

Només s'han tingut en compte les 10 regles que mostraven un percentatge de confiança més elevat i la resta s'han descartat per tenir un valor de confiança massa baix. Com es pot observar en aquest cas, la tendència amb un grau de suport major és aquella que informa de la importància de superar el primer examen. Si això no succeeix, les possibilitats de suspendre el segon examen són elevades. Això no contradiu els resultats anteriors, sinó que els complementa, doncs tot i ser en general el segon examen el més important, la tendència informa de com influeix a l'alumne el resultat previ obtingut en el primer examen. De fet, si s'analitza la correlació existent entre el resultat de l'examen 1 i l'examen 2, es veu que aquesta és força alta, com es pot observar a la Figura 40.

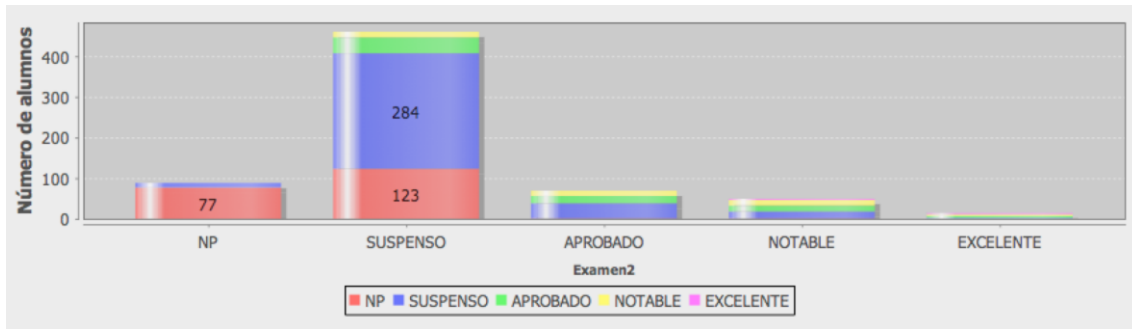


Figura 40. Correlacions existents entre les dades corresponents a l'examen 1 i a l'examen 2 del conjunt de dades d'ensenyament.

En efecte, com es pot veure a la Figura 40, la majoria d'alumnes que ha suspès el primer examen, també suspensen el segon. El gràfic s'ha construït de la següent manera: a l'eix de les X s'han repartit tots els alumnes segons la seva qualificació al primer examen. A continuació, per cada resultat obtingut en aquest eix, s'han pintat a l'eix de les Y, i de diferents colors, els resultats obtinguts pels alumnes en el segon examen.

Per finalitzar aquest apartat es vol mostrar els resultats obtinguts (Taula 11) si només es vol predir l'aprobat o el suspens de l'alumne. En aquests cas els resultats que s'indiquen són ja utilitzant la millor configuració i la ponderació proposada pels Algorismes Genètics.

Taula 11. Eficàcia mitjana, màxima i mínima assolida amb les dades disponibles, tenint en compte la ponderació proposada pels Algorismes Genètics i classificant a l'alumne en només dues classes (aprobat o suspès), emprant *10-fold crossvalidation*.

Tipus de ponderació	Eficàcia	Màxim	Mínim
Ponderació Genètics	90,66 %	98,44 %	87,04 %

Com a conclusió d'aquest apartat, es pot afirmar que la predicció de l'evolució individual d'un alumne és factible. Revisant els resultats obtinguts, es veu com s'arriba a llindars per sobre del 98% d'encert, només utilitzant la poca informació comuna que es tenia dels diferents conjunts de dades educatives disponibles. Això fa pensar que si, d'una banda, es disposen de totes les dades disponibles al llarg d'un curs normal i, de l'altra, es té informació extra intrínseca a un aprenentatge en línia, la fiabilitat del sistema final encara serà més rellevant.

Aquesta informació va ser publicada en diversos articles i va tenir repercussió mediàtica, com es pot observar a la Figura 41, on es mostra un retall d'un diari sud-americà que feia referència a la recerca que s'havia efectuat.



Figura 41. Retall del diari sud-americà Clarín on es fa ressò dels resultats obtinguts pel sistema de predicció implementat.

En treballs recents s'han presentat altres sistemes de predicció de l'evolució de l'alumne. Així, a (Villagrà-Arnedo, Gallego-Durán, Molina-Carmona, & Llorens-Largo, 2015) es dissenya un sistema de predicció progressiu, basat en les activitats realitzades pels estudiants de manera setmanal. Aquest sistema és gamificat i automàtic i permet extreure característiques rellevants de la interacció de l'alumne amb el sistema. D'altra banda, el sistema *GradeForeseer* (Salamó, et al., 2015) és un altre sistema presentat recentment que ofereix una interfície gràfica i la possibilitat de realitzar una predicció exhaustiva fent ús de diferents classificadors recollits a WEKA i una llibreria específica de Raonament Basat en Casos, escollint el millor classificador segons el cas. En aquest cas igualment es fa la predicció binària (aprovat o suspès) i la predicció en 4 classes.

Altres treballs presentats en els darrers anys intenten predir el rendiment acadèmic de l'alumne fent ús de sistemes difusos i xarxes neuronals, com a (Sevindik, 2011), o es questionen el rendiment de sistemes de predicció mitjançant models ensamblats (Baker, Pardos, & Gowda, 2011).

6.5.3 Nova organització dels casos

Normalment, com s'ha comentat en apartats anteriors, la memòria de casos en un sistema de Raonament Basat en Casos està organitzada de tal manera que fa correspondre cada instància del domini amb un cas. Aquesta situació permet fer la predicció individualitzada dels alumnes però no és la millor organització si es vol treballar amb el concepte de grup. Com que un dels objectius d'aquesta tesi és també analitzar l'evolució dels grups de treball, caldrà modificar l'estructura base de la Memòria de Casos i del conjunt de test (aula virtual) per assolir-ho.

Així doncs, la primera modificació que es vol introduir al sistema és el fet de poder treballar amb grups d'alumnes i que cada grup vingui representat per un cas dins del sistema. Per

aconseguir tal objectiu caldrà modificar l'organització dels casos tot introduint algorismes d'agrupació o *clustering* que afectaran a la inicialització del sistema, a la fase de recuperació i a la fase d'emmagatzematge. D'altra banda, com a efecte colateral, el fet de tenir un nombre inferior d'instàncies un cop clusteritzada l'aula i la Memòria de Casos, provocarà l'assoliment d'un altre objectiu: millorar el rendiment del sistema tot augmentant el temps d'execució i garantir una execució en temps real.

L'objectiu dels algorismes de *clustering* és trobar agrupaments de casos que maximitzin la cohesió dels casos d'un grup i la distància entre elements situats en altres grups. L'agrupació de casos es fa de manera no supervisada i pot arribar a ser subjectiva, essent indispensable una mesura de similitud global entre dues instàncies qualssevol.

Principalment, existeixen cinc categories d'algorismes d'agrupació:

- *Agrupació per particionat*. Es coneix a priori el nombre d'agrupacions que es volen realitzar. L'objectiu d'aquesta mena d'algorismes és optimitzar un criteri de particionat, sovint anomenat funció de similitud o distància entre agrupacions. Un exemple d'aquesta categoria seria l'algorisme *K-Means* que s'analitzarà a continuació amb més profunditat.
- *Agrupació basada en la densitat*. Es creen agrupacions d'acord a un nivell mínim de densitat. Són útils quan les formes de les agrupacions poden ser totalment arbitràries o quan hi ha regions de l'espai on la densitat dels elements és molt més gran comparada amb la resta. Els algorismes d'aquests tipus més reconeguts són DBSCAN (Density-Based Spatial Clustering of Applications with Noise) (Ester, Kriegel, & Xu, 1995), OPTICS (Ordering Points to Identify Clustering Structure) (Ankerst, Breunig, Kriegel, & Sander, 1999) i DENCLUE (DENsity-based CLUstEring) (Hinneburg & Keim, 1998).
- *Agrupació jeràrquica*. Es crea una descomposició jeràrquica d'un conjunt de dades donat, és a dir, el resultat de l'algorisme és un arbre d'agrupacions. Existeixen dues polítiques diferents pel que fa a la creació d'aquest arbre: *bottom-up* o *top-down*. En la primera s'aplica el que s'anomena *clustering* aglomeratiu, on inicialment tots els objectes estan en una agrupació diferent i iterativament es van agrupant en *clusters* més grans fins que s'esdevé alguna determinada condició de finalització. En canvi, la segona política és la que s'aplica en procediments d'agrupació per divisió, on inicialment tots els objectes formen part del mateix *cluster* i paulatinament es van dividint fins arribar també a una condició, com ara el nombre d'agrupacions finals o una distància lliardar. A (Han & Kamber, 2000) es poden veure diversos exemples d'aquest tipus d'agrupació.
- *Agrupació basada en graells (grid-based methods)*. Es divideix l'espai en un nombre finit de cel·les, les quals formen una estructura en forma de graella. L'avantatge principal d'aquest mètode és que el temps d'execució es manté constant, independentment dels objectes que es volen agrupar. En aquesta tipologia d'algorismes es poden trobar exemples com CLIQUE (Clustering High- Dimensional Space) (Agrawal, Gehrke, Gunopulos, &

Raghavan, 1998), STING (STatistical INformation Grid) (Wang, Yang, & Muntz, 1997) i WaveCluster (Sheikholeslami, Chatterjee, & Zhang, 1998).

- *Agrupació per model.* Aquest mètode defineix un model matemàtic i probabilístic per cada agrupació que es vol crear i tracta de trobar el millor ajust de les dades per tal que encaixin en el model dissenyat. Aquests mètodes sovint es basen en aproximacions estadístiques o fan ús de xarxes neuronals. Alguns exemples rellevants serien AUTOCLASS (Cheeseman & Stutz, 1996), COBWEB (Fisher, 1987) o les Support Vector Machines (SVM) (Smola & Sch, 1998).

De tots els models d'agrupació exposats, en aquesta tesi el treball s'ha centrat en algorismes que fan agrupació per particionat a causa que conceptualment són els més propers a la definició de *k-NN* (*Nearest Neighbour*) associat al Raonament Basat en Casos. A més, aquests tipus d'algorismes parteixen l'espai de cerca en una estructura que segueix la forma d'un diagrama de Voronoi que es representa a la Figura 42.

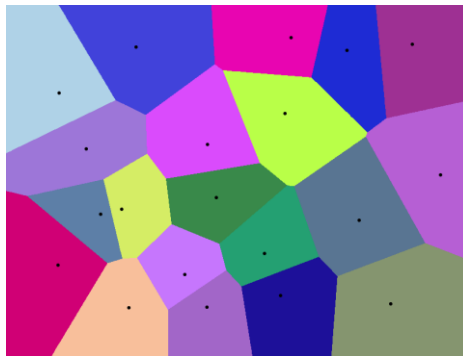


Figura 42. Exemple de diagrama de Voronoi

Aquest tipus de diagrama presenta una particularitat que interessa en els objectius d'aquesta tesi: petits canvis en la disposició dels elements, afecten en poca mesura en la distribució de la resta d'elements, és a dir, l'estructura resultant és força robusta al canvi. Si es té al cap l'objectiu final i es pensa que els elements són grups d'alumnes, si posteriorment es vol fer algun canvi de grup, aquest canvi no afectarà en gran mesura a l'estabilitat general de la classe.

A continuació s'analitza el funcionament dels algorismes utilitzats i com l'aplicació de cadascun d'ells modifica l'estructura del conjunt de casos.

6.5.3.1 Algorisme K-Means

L'algorisme *K-Means* (Hartigan, 1975) (Hartigan & Wong, 1978) és un algorisme d'agrupació o *clustering* classificat com a mètode de particionat i recol·locació. Cadascuna de les agrupacions o *clusters* tenen un centroide o punt mig calculat a partir de la mitjana o mitjana ponderada dels casos inclosos (d'aquí prové el seu nom).

La funció objectiu d'aquest algorisme es basa en els errors o distàncies entre els punts a col·locar i els centroides que es van definint dinàmicament. Això provoca que aquest algorisme sigui vàlid només si la totalitat dels atributs existents són numèrics. Un altre inconvenient d'aquest algorisme és que cal definir a priori el nombre d'agrupacions que es volen dur a terme.

A la Figura 43 es pot veure l'algorisme en funcionament i com convergeix al cap de poques iteracions.

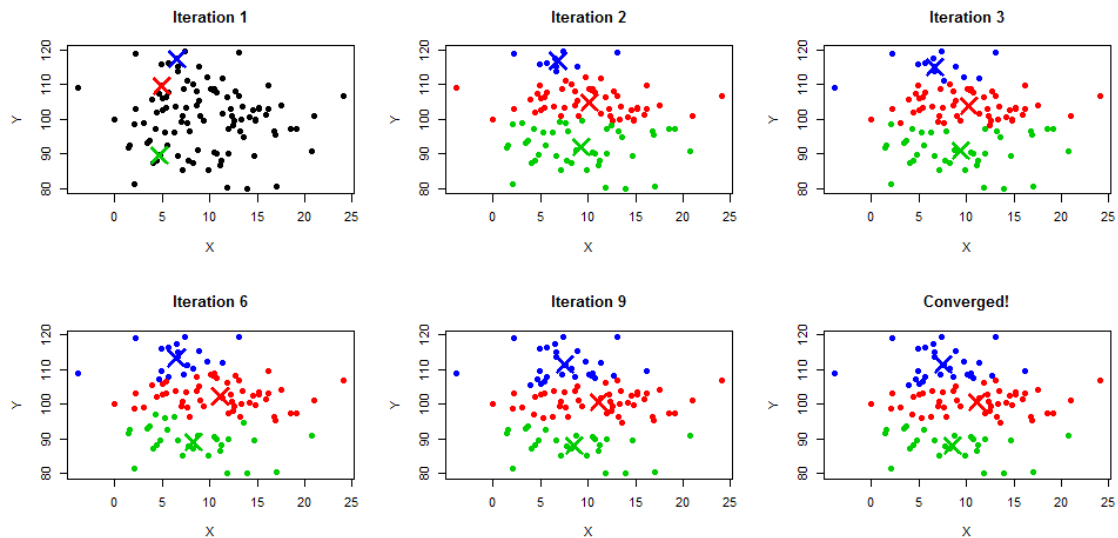


Figura 43. Exemple de funcionament de l'algorisme *K-Means*. Els punts marcats amb una X corresponen als centroides de cadascun dels *clusters*. Imatge en línia: Learn by Marketing.

L'algorisme general es podria resumir de la manera que es descriu a l'Algorisme 1.

```

escollirElementsCentroides()
assignarElementsAGrup()
hihagiCanvis := CERT
mentre hiHagiCanvis fer
    calcularDistancies()
    hiHagiCanvis := reassignarElementsAGrup()
    recalcularCentroides()
fimentre

```

Algorisme 1. Pseudocodi corresponent a l'algorisme K-Means

Cal fer notar que normalment els centroides inicials s'escullen de manera aleatòria, provocant que aquest algorisme pugui donar resultats diferents en diverses execucions, ja que la configuració cau en un mínim local dins de l'espai de cerca. Existeixen diverses variants d'aquest algorisme que també s'han provat basades en el *k-medoids algorithm*, com és el PAM (Partition Around Medoids) (Kaufman & Rousseeuw, 1987), amb l'objectiu de reduir la sensibilitat que demostra l'algorisme estàndard de *K-Means* quan es detecten valors molts grans de certs atributs que poden distorsionar el resultat, tot i la seva normalització.

Una variació del *k-medoids algorithm* és l'algorisme CLARA (Clustering LARge Applications) (Kaufman & Rousseeuw, 1990), el qual estén les capacitats de l'algorisme anterior, millorant el resultat quan el conjunt de casos és molt gran.

6.5.3.2 Algorisme X-Means

Una variant de l'algorisme anterior és l'algorisme *X-Means* proposat per Pelleg (Pelleg & Moore, 2000). Aquest algorisme intenta evitar alguns dels problemes que presenta el *K-Means*. Concretament, el més important és que aquest algorisme fa una estimació eficient del nombre d'agrupacions necessàries per al conjunt de dades existent. A part, l'algorisme millora el temps computacional respecte a fer diverses repeticions del *K-Means* i evita en bona mesura els mínims locals amb què es troba l'algorisme anterior.

L'algorisme principal consisteix en executar un *K-Means* tradicional i després preguntar-se si cadascun dels centroides es poden dividir en dos. Es torna a executar l'algorisme anterior per aquesta nova configuració i després es filtren els centroides resultants segons certs criteris de qualitat. El fet que es repeteixi l'algorisme *K-Means* per separat per a cada centroide obtingut inicialment, permet que es pugui paral·lelitzar l'algorisme i baixar el temps total de computació.

6.5.3.3 Algorisme WITT

S'ha vist anteriorment que un dels problemes principals que aporta l'algorisme *K-Means* i les seves variants és que està pensat per treballar únicament amb atributs continus. Si això ho portem a un entorn docent on no totes les dades que es tenen d'un alumne són dades numèriques, els algorismes anteriors resulten insuficients en certes ocasions.

Per evitar aquest problema s'ha cregut convenient introduir una altra possibilitat d'agrupar els conjunt de casos de treball quan aquests contenen atributs discrets: l'algorisme WITT.

L'algorisme WITT (Hanson & Bauer, 1989) (Hanson, 1990) és un algorisme inductiu (el coneixement s'adquireix després de realitzar una inferència deductiva sobre la informació subministrada) no supervisat i incremental, tot i que en la seva versió més robusta no és incremental. La principal aportació de WITT respecte els altres algorismes de la seva classe és el fet d'evitar la gran restricció en la que les agrupacions estan formades únicament per

elements que tenen unes condicions especials, necessàries i suficients com per crear una agrupació a part.

Els quatre trets diferencials de l'algorisme són:

- Les agrupacions tendeixen a tenir instàncies les característiques de les quals no són ni necessàries ni suficients.
- Les agrupacions tenen una distribució entre els seus membres, és a dir, alguns seran més representatius i altres menys.
- Les agrupacions poden ser representades mitjançant les correlacions i les relacions entre els atributs que les descriuen.
- Les agrupacions apareixen per contrast amb la resta.

L'algorisme es basa en el concepte de taula de contingència. Per a tots els parells possibles d'atributs existents, es crea una taula de contingència consistent en una matriu de co-ocurrències del parell d'atributs dins de cada agrupació.

Un exemple concret podria ser el següent: disposem de dues característiques d'un alumne com són la classe de teoria (suposant només dues classes possibles) i l'indicador de repetidor que anteriorment havíem avaluat. Una taula de contingència com la de la Taula 12 ens indicaria el resultat d'analitzar la situació d'una agrupació feta amb 4 alumnes:

Taula 12. Exemple de taula de contingència per a un *cluster* de 4 alumnes.

	Repetidor	No repetidor
Classe A	0	1
Classe B	0	3

De la mateixa manera es calculen doncs les taules de contingències de totes les combinacions de parells d'atributs existents.

Seguidament, l'algorisme WITT calcula la cohesió de cada agrupació, tot avaluant com són de similars els casos que es troben dins d'un *cluster*. La cohesió es defineix:

$$C_c = \frac{W_c}{O_c} \quad (6)$$

on W_c és la cohesió interna del grup i O_c representa la cohesió mitjana entre aquest grup i la resta. Com es pot veure, l'algorisme tracta de maximitzar aquest valor, fent que la cohesió interna del grup sigui elevada i la cohesió entre grups sigui baixa.

Per calcular la cohesió interna d'un grup es fa servir l'expressió (7) següent:

$$W_c = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N D_{ij}}{\frac{N(N-1)}{2}} \quad (7)$$

on N és el nombre d'atributs i D_{ij} és la distribució associada a la taula de contingència entre els atributs i i j que es calcula mitjançant l'expressió (8).

$$D_{ij} = \frac{\sum_{m=1}^{card(i)} \sum_{n=1}^{card(j)} f_{mn} \log(f_{mn})}{\left(\sum_{m=1}^{card(i)} \sum_{n=1}^{card(j)} f_{mn}\right) \left(\log\left(\sum_{m=1}^{card(i)} \sum_{n=1}^{card(j)} f_{mn}\right)\right)} \quad (8)$$

on f_{mn} és la freqüència en la co-ocorren el valor m de l'atribut i i el valor n de l'atribut j . L'expressió $card(i)$ indica el nombre d'atributs diferents que pot tenir l'atribut i .

Com es pot observar, D_{ij} pot prendre valors entre 0, quan no hi ha cap característica comuna, o 1, quan tots els atributs són iguals. En l'exemple de la taula anterior, el valor de D_{ij} és de 0,59, atès que:

$$D = \frac{(0 \log(0)) + (1 \log(1)) + (0 \log(0)) + (3 \log(3))}{(4 \log(4))} = 0,59$$

Finalment, cal calcular quina és la cohesió entre els diferents grups, és a dir, el valor de O_c . Aquest valor es basa en la mètrica B_{ck} :

$$O_c = \frac{\sum_{k=1}^L (k \neq c) B_{ck}}{L-1} \quad (9)$$

on en aquest cas, L és el nombre total de clusters i B_{ck} és funció del valor W_c indicat anteriorment que calculava la cohesió interna d'un grup:

$$B_{ck} = \frac{1}{W_c + W_k - 2W_{c \cup k}} \quad (10)$$

De manera gràfica, a la Figura 44 es pot entendre fàcilment què s'entén per cohesió interna i cohesió entre grups (representades pels valors W_c i O_c respectivament).

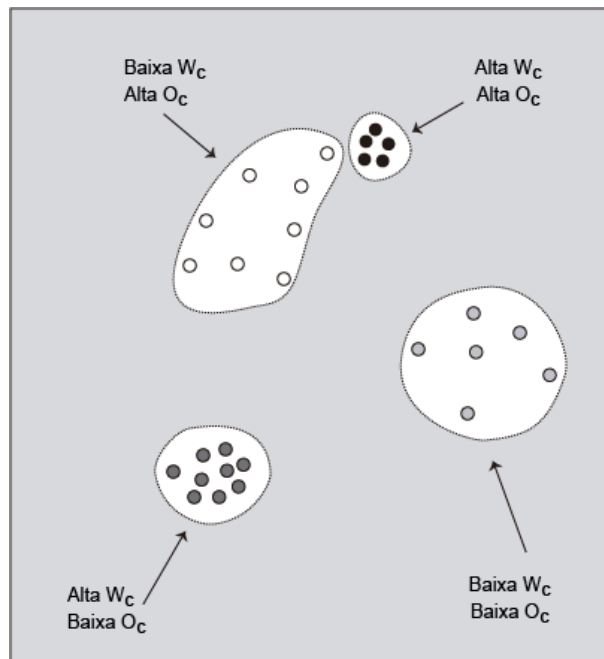


Figura 44. Exemple gràfic dels valors de cohesió interna i externa entre els diferents *clusters*.

L'algorisme es divideix en dues etapes diferenciades. En una primera fase es creen les categories i agrupacions inicials, mitjançant un algorisme de *preclustering*. Aquesta etapa de *preclustering* és senzillament un algorisme *greedy* on la funció d'avaluació és la distància entre els objectes. La finalització d'aquest algorisme depèn d'un paràmetre (F), que cal definir de manera manual. A continuació una segona part de refinament entra en joc, la qual incorpora tres moments importants: afegir una instància a una agrupació, crear noves agrupacions i fusionar dues agrupacions.

En l'algorisme de refinament els paràmetres $T1$, $T2$ i $T3$ permeten intercedir en l'algorisme per definir com d'iguals o diferents són dos atributs o els valors que prenen. Per exemple, en la implementació realitzada s'ha valorat la diferència conceptual que hi ha entre que un alumne vagi a la classe A o B, que clarament no és la mateixa al fet que sigui repetidor o no d'una assignatura. L'ajustament d'aquests paràmetres permet definir la rellevància dels atributs en aquest sentit i influir en la manera de decidir la cohesió dels elements.

Tot i que el paràmetre $T1$ només ens indica quan cal reagrupar *clusters* entre si, és especialment rellevant la relació entre els paràmetres $T2$ i $T3$. Quant més proper sigui $T2$ respecte $T3$, més *especials* han de ser els objectes (més condicions necessàries i suficients han de complir) per formar una agrupació independent. Després de diverses proves amb les dades existents, s'han definit els valors indicats a la Taula 13 per aquests paràmetres de configuració.

Taula 13. Valors definits per als paràmetres de configuració de l'algorisme WITT

Paràmetre	Valor
F	0,5
T1	1,35
T2	0,17
T3	0,55

L'Algorisme 2 descriu la versió final adaptada de l'algorisme de WITT que s'ha utilitzat en la clusterització del conjunt de casos de treball.

```

preClusteringInicial()
finalitzar:=FALS
repeteix
    cercarCasMesSimilarACluster()
    si cohesio > T1 llavors
        afegeixCasACluster()
    sinó
        preClustering()
        si existeixClusterSignificatiu (T2) llavors
            afegeixCluster()
        sinó
            prefusionaClusters()
            si cohesio > T2 llavors
                fusionaClusters()
            sinó
                finalitzar:=CERT
        fisi
    fisi
fisi
fins que finalitzar=CERT

```

Algorisme 2. Adaptació de l'algorisme WITT utilitzada per integrar atributs discrets als algorismes de *clustering*

L'aplicació d'aquest algorisme s'ha combinat amb la dels algorismes de clustering per atributs exclusivament numèrics amb la finalitat de millorar l'efectivitat del sistema.

6.5.3.4 Algorisme NormWITT

Un dels problemes que es van detectar i que es veuran posteriorment en els resultats obtinguts, és que l'algorisme WITT presenta certes dificultats a l'hora de tractar els valors resultants de cohesió quan aquests són molt baixos. En efecte, si comparem classes o agrupacions on la cohesió d'aquestes és molt petita, els valors de W_c i W_k aplicats a l'equació 10 provoquen una dispersió molt gran del resultat de B_{ck} , provocant una caiguda notable en el rendiment del sistema.

Per arreglar aquest problema s'han estudiat diferents alternatives que permetessin mantenir l'objectiu de l'algorisme però que a la vegada normalitzessin els resultats obtinguts. La solució final proposada i provada ha consistit en canviar dos aspectes clau de l'algorisme. En primer lloc es redefineix el càlcul de B_{ck} de la següent manera:

$$B_{ck} = W_{c \cup k} \quad (11)$$

Amb aquest canvi s'aconsegueix que la cohesió interna tingui més rellevància i per tant la cohesió mitjana entre els diferents clusters (O_c) també tingui un ventall de valors més gran. Per ajustar aquest augment de O_c s'ha modificat també el càlcul de l'equació 6, per tal que el valor final de C_c no quedés en un rang de valors massa petit i poc diferenciat entre clusters:

$$C_c = W_c - O_c \quad (12)$$

Com es veurà posteriorment a l'apartat 6.5.6, les variacions efectuades a l'algorisme original milloren significativament els resultats obtinguts, doncs eviten que, en cas que es trobin agrupacions amb cohesions molt baixes, es creïn nous *clusters* irrellevants per als valors d'entrada.

6.5.3.5 Aplicació dels algorismes d'agrupament per crear grups

Els algorismes d'agrupament que s'han establert s'apliquen a dos nivells dins del conjunt de casos existent. En primer lloc, es fa una primera agrupació dels alumnes segons la seva situació predita a partir dels mètodes de predicció comentats anteriorment, per després passar a un segon nivell que permeti crear els grups de classe ideals.

Durant la realització del primer nivell de *clustering*, en cas de dificultat en la classificació dels alumnes, s'han dissenyat dues polítiques (Vernet & Golobardes, 2003) (Vernet & Golobardes, An Unsupervised Learning Approach for Case-Based Classifier Systems, 2002) per establir si és necessari un nou refinament de cadascuna de les classes obtingudes:

- *Mean Clustering*. Es pren el centroide com a mesura central del *cluster* i s'analitza la proximitat dels elements continguts respecte aquest. Quan els elements continguts en el *cluster* estan prou cohesionats (a partir d'un llindar calculat a partir de la desviació estàndard), es dona el resultat per vàlid.
- *MKM Clustering*. En cas que en els elements analitzats del *cluster* es detectin possibles noves agrupacions entre ells, s'aplica recursivament un nou nivell de clustering que hem anomenat *Mean K-Means (MKM) Clustering*.

A la Figura 45 es pot observar gràficament la resolució d'aquest primer nivell de *clustering* del conjunt d'alumnes que formen l'aula virtual:

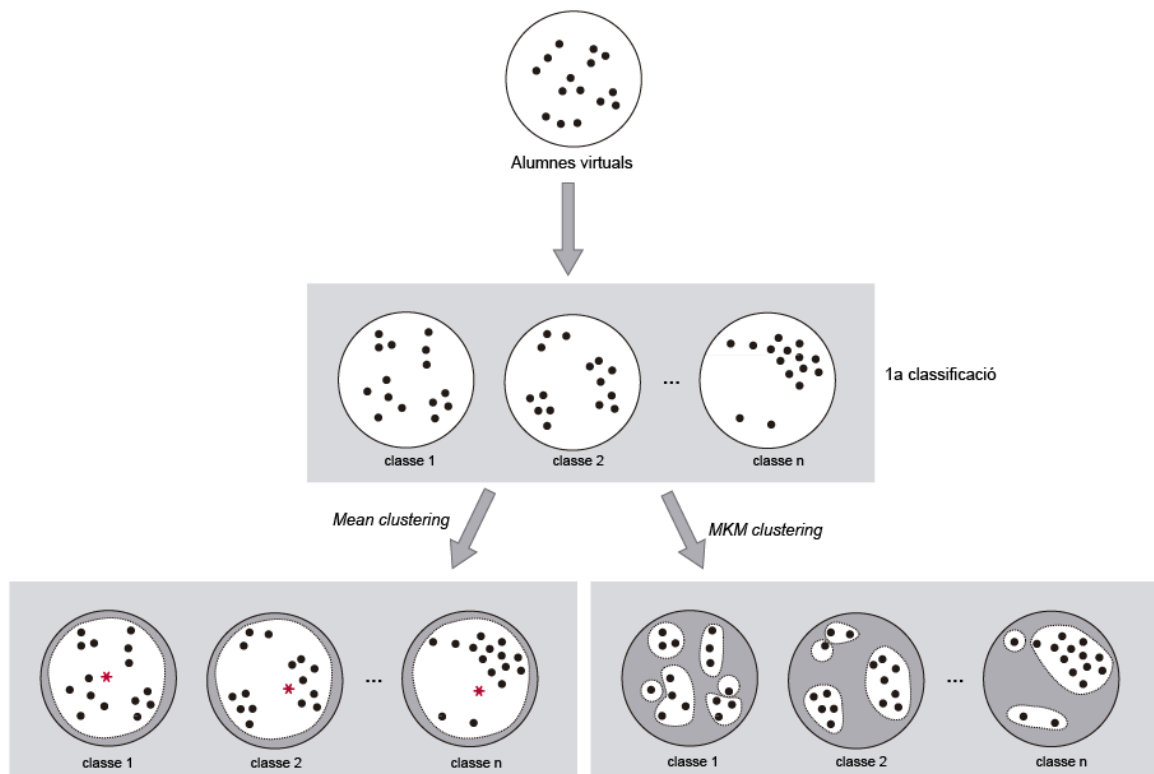


Figura 45. Esquema de com es realitza el primer nivell de *clustering* i el seu refinament posterior en cas de detectar subagrupacions.

Un cop detectats els diferents perfils dels alumnes, s'aplica el nou nivell de *clustering* per constituir els grups de treball inicials. Es pot configurar el sistema per tal de crear grups homogenis o heterogenis, així com el número d'alumnes per grup. En cas que siguin homogenis, es prenen els alumnes de la mateixa categoria (per exemple, tots els alumnes excel·lents), mentre que si són heterogenis, s'escull un alumne de cada tipus. Per a cada grup s'emmagatzema informació variada, tot fent un paral·lelisme al concepte d'*esfera* presentat a

(Golobardes, 1998) (Vernet & Golobardes, 2003). A l'hora de treballar amb la informació del grup, es pot treballar amb:

- Els valors concrets de les qualificacions o dades de cada membre del grup
- Els valors màxim i mínim de les qualificacions o dades del grup (*MaxMin Group*)
- El valor mitjà de les qualificacions o dades del grup (*Mean Group*)

A la Figura 46 es pot observar una representació dels elements existents en el grup. Com es pot veure, per cada alumne del grup es té informació de les diferents qualificacions de cada membre del grup.

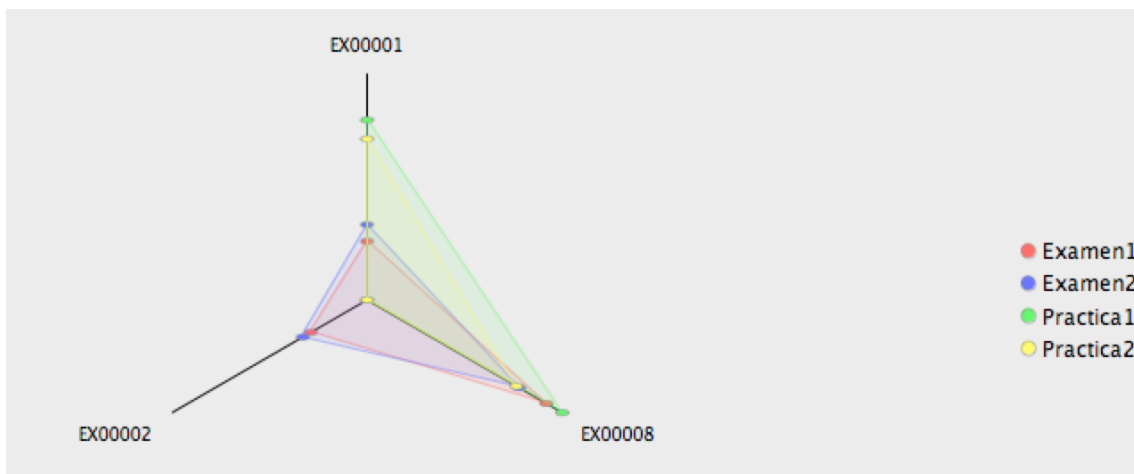


Figura 46. Representació gràfica del coneixement que posseïxen els membres d'un grup creat de manera automàtica.

Finalment, en la Figura 47 es veu l'esquema general de tot el sistema d'agrupament, des de la classificació inicial dels alumnes segons el perfil que presenten, fins a la confecció final dels grups.

Com es pot veure, el primer nivell de *clustering*, separa els alumnes en les diferents categories segons el seu nivell d'evolució, mentre que el segon nivell s'encarrega de formar els grups homogenis o heterogenis.

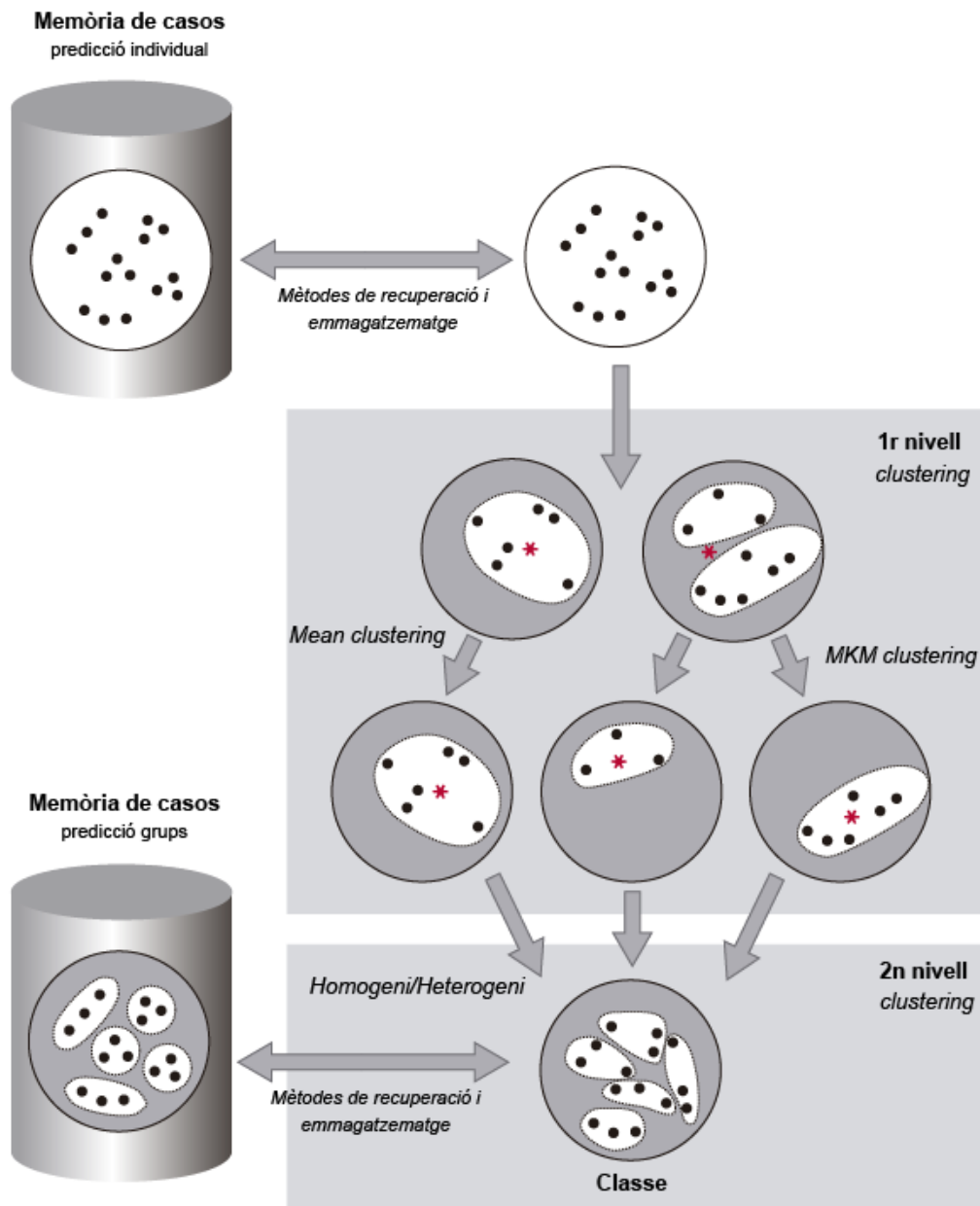


Figura 47. Esquema general del sistema de *clustering* a dos nivells.

6.5.4 Nous mètodes de recuperació

Un cop definits els mètodes utilitzats per a la nova estructuració dels casos, a continuació es veuran les tècniques que s'han dissenyat per ser aplicades a la Fase de Recuperació del cicle de CBR tenint en compte que la Memòria de Casos consta d'instàncies que representen a grups d'alumnes.

Tal com s'ha comentat en l'apartat anterior, en el sistema global cal distingir els dos moments de recuperació diferents: la recuperació quan s'està fent la predicció individual de l'alumne i la recuperació quan s'està fent la predicció grupal. A l'apartat 6.5.2 s'han analitzat les tècniques utilitzades per fer la recuperació en el primer cas i ara es veuran les introduïdes en la segona situació.

Cal recordar que un cop aplicada la *clusterització* a l'aula el que queda és un conjunt de M grups o *clusters* on cadascun d'ells està format N individus, corresponents als membres que formen el grup en qüestió. De la mateixa manera, a la Memòria de Casos grupal també hi haurà una situació idèntica. Així, suposant per exemple que els grups establerts estan formats per 3 alumnes, cada cas estaria format per 3 vectors de dades diferents com es representa a la Figura 48:

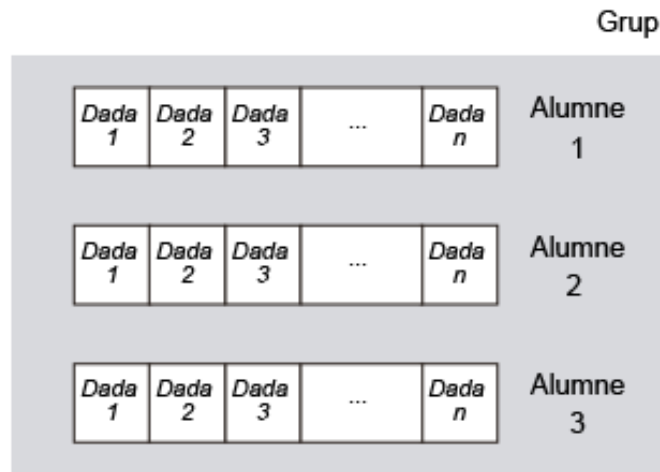


Figura 48. Exemple de la informació existent en un cas format per l'agrupació de 3 alumnes.

Les dades existents en un cas es poden tractar com a vectors independents, però si tenim en compte que els membres del grup haurien d'influir entre ells en la seva evolució, sembla més lògic pensar que els casos s'haurien de tractar com a sistemes matricials, on els components de cada individu influeixen en la globalitat del cas.

A continuació es veuran les diferents polítiques que s'han introduït per recuperar el cas o casos més propers del conjunt de casos de la Memòria de Casos respecte a cadascun dels grups existents a l'aula virtual.

6.5.4.1 Recuperació en una memòria clusteritzada

S'ha vist doncs que és interessant aprofitar la informació implícita que porta la confecció d'un grup. És a dir, val la pena no observar les dades recollides de cada alumne de manera

individualitzada, sinó tenir en compte amb quins companys ha estat agrupat per tal de poder recuperar quin és el cas més similar de la Memòria de Casos.

Per tal de tenir en compte tota la informació procedent del grup s'ha tractat la informació del grup en format de matriu $M \times N$, on M correspon al nombre d'estudiants del grup i N és el nombre de característiques disponibles de cada alumne. Aquestes característiques s'han unificat pels diferents jocs de dades en l'entorn docent en les característiques definides a la Taula 8.

Trobar la similitud entre dues matrius X i Y no és una tasca senzilla. La política usada per comparar dues matrius dependrà molt del que s'estigui perseguint en cada situació. En el cas que ens ocupa, s'han descartat algunes opcions a causa que aquestes no usaven les dades de la matriu com a vectors de dades i s'han introduït i provat les següents funcions de similitud:

- *Distàncies de Hamming i Euclidiana.* Aquestes distàncies es calculen a partir de les equacions 13 i 14 respectivament.

$$d(X, Y) = \sum_{i=1}^m \sum_{j=1}^n |x_{ij} - y_{ij}| \quad (13)$$

$$d(X, Y) = \sqrt{\sum_{i=1}^m \sum_{j=1}^n (x_{ij} - y_{ij})^2} \quad (14)$$

- *Diferència màxima entre els elements de les dues matrius (equació 15).* Aquesta distància només és vàlida quan totes les qualificacions recollides de l'alumne estan valorades sobre la mateixa nota màxima, o bé quan les dades han estat normalitzades prèviament.

$$d(X, Y) = \max_{1 \leq i \leq m} \max_{1 \leq j \leq n} |x_{ij} - y_{ij}| \quad (15)$$

- *Recompte de les posicions en la matriu on la seva diferència amb el valor corresponent de l'altra matriu supera un cert llindar.* Cal que es compleixi la mateixa condició que abans i cal que el llindar estigui fixat en funció del rang de valors possibles per cada nota.
- *Variante de la norma de Frobenius.* a partir de la norma de Frobenius (equacions 16 i 17, equivalents), s'ha definit la distància de Frobenius com el valor absolut de la diferència entre els dos valors calculats (equació 18). El valor $Tr(X)$ correspon a la *Traça* de la matriu X , mentre que X^H correspon a la seva matriu transposada conjugada.

$$F(X) = \|X\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |x_{ij}|^2} \quad (16)$$

$$F(X) = \|X\|_F = \sqrt{\text{Tr}(XX^H)} \quad (17)$$

$$d(X, Y) = |F(X) - F(Y)| \quad (18)$$

Existeixen altres possibilitats per al càlcul de distàncies entre matrius, però com s'ha comentat anteriorment s'han descartat per no ser convenients amb les dades que s'han utilitzat. Com es veurà posteriorment a l'apartat de resultats, la distància euclidiana i la de Hamming són les funcions de similitud que han donat més bon resultat en el nostre domini d'actuació. El recompte de posicions també és una bona estratègia quan el nombre d'alumnes per grup i el nombre d'atributs és elevat.

6.5.4.2 Ponderació d'atributs

La correcta ponderació dels atributs és un factor clau a l'hora de recuperar correctament el cas o casos més similars. Com ja s'ha vist en la predicció individual de l'alumne, s'han usat diverses polítiques de ponderació d'atributs dependent del moment. S'ha vist anteriorment que les ponderacions es duen a terme de tres maneres diferents:

- Sense ponderació
- Ponderació a partir de la informació subministrada pel professor
- Ponderació a partir del resultat obtinguts pels Algorismes Genètics

Per realitzar la recuperació en la predicció grupal s'han utilitzat les mateixes polítiques definides anteriorment, tot modificant les funcions de similitud sobre matrius per tal de tenir en compte els pesos de cada atribut quan s'estaven comparant. Així, s'ha utilitzat un vector de pesos que multiplica directament un a un qualsevol consulta dels elements de la matriu de qualificacions.

$$W = \{w_1, w_2, \dots, w_n\} \quad (19)$$

Cada atribut ha estat considerat amb la mateixa ponderació per a tots els alumnes que formen l'agrupació.

6.5.4.3 Algorismes genètics

Els Algorismes Genètics són una tècnica d'Intel·ligència Artificial basada en la computació evolutiva. Es considera una tècnica de cerca pel tipus de problemes que resol, sobretot d'optimització i cerca de valors desconeguts.

La seva implementació es basa en fer *evolucionar* una població d'individus que representen possibles solucions al problema que s'està resolent. Aquests individus se solen codificar de manera binària i la seva adequació a la solució final es calcula mitjançant una funció de bonesa explícita. Els individus es fan evolucionar cíclicament (Figura 49) tot aplicant-los operadors genètics de recombinació i mutació amb la finalitat de millorar les aptituds de la població existent.

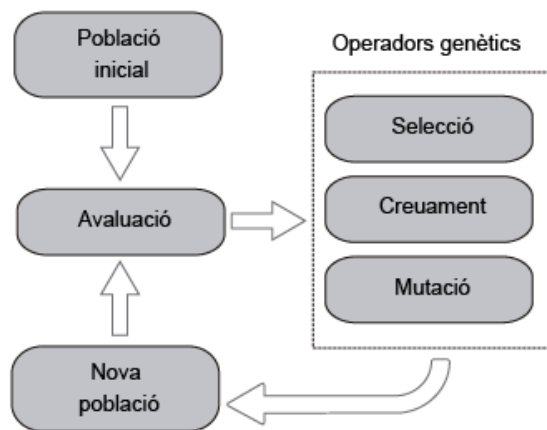


Figura 49. Cicle de vida d'un Algorisme Genètic.

Tot i que l'autor d'aquesta tesi ha aplicat els Algorismes Genètics en altres situacions, com ara en problemes pertanyents a entorns telemàtics (Zaballos, Vernet, & Selga, 2013) (Vallejo, Zaballos, Vernet, Orriols, & Dalmau, 2009) (Vallejo, Zaballos, Vernet, Cutiller, & Dalmau, 2008), els Algorismes Genètics aplicats a problemes educatius s'han limitat al descobriment de la millor ponderació dels atributs. Així, de la mateixa manera que en la predicció individual de l'alumne, els algorismes genètics s'han aplicat en la predicció grupal per descobrir el vector de pesos associat a les activitats de les que es té informació de cada alumne.

Els Algorismes Genètics poden classificar un problema tot treballant amb un conjunt de regles codificades de manera binària. Depenent de si cada individu de la població és per ell mateix un conjunt de regles, o pel contrari, cada individu és una regla i el conjunt d'ells forma el classificador, les aproximacions d'aquesta algorismes, tal com s'ha comentat anteriorment, s'anomenen aproximacions de Pittsburgh o de Michigan respectivament.

Per a la ponderació realitzada s'ha fet evolucionar un classificador de Pittsburgh amb les següents funcions de bonesa:

$$f_1 = \left(\frac{C}{T}\right)^2 \quad (20)$$

$$f_2 = \left(\frac{C - \beta N}{T}\right)^2 \quad (21)$$

On C és el nombre de casos classificats correctament, N és el nombre de casos no classificats, T és el nombre de casos totals i β és un factor que controla el fet que el conjunt de regles final de casos pugui acceptar més o menys casos no classificats.

Els resultats dels valors de ponderació s'obtenen a partir de la proporcionalitat de l'aparició dels atributs a ponderar en les regles resultants del classificador (Golobardes, Llorà, Garrell, Vernet, & Bacardit, 2000).

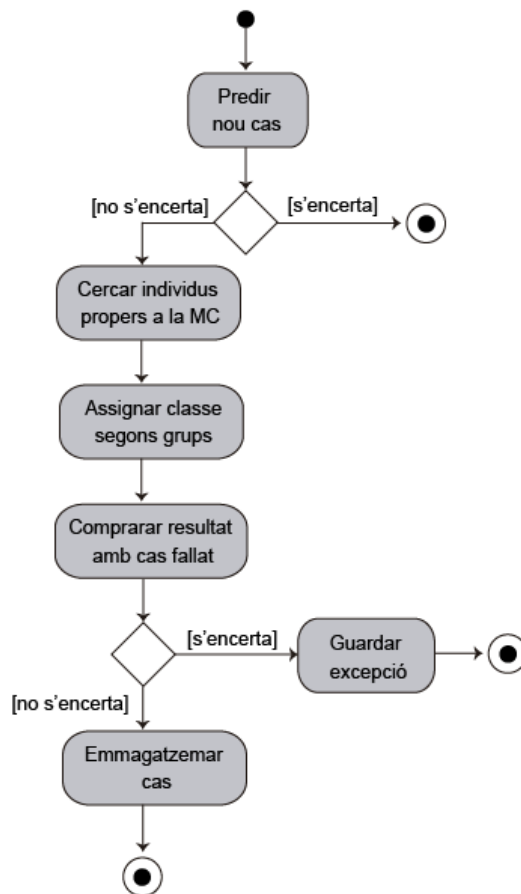
6.5.5 Aportacions a la fase d'emmagatzematge

La fase d'emmagatzematge és l'etapa on rau l'aprenentatge del sistema. Certament, l'efectivitat del model proposat dependrà en gran mesura que la informació emmagatzemada a la Memòria de Casos sigui prou bona per poder fer prediccions futures de nous casos.

Com s'ha comentat en apartats anteriors, un cop s'ha classificat un cas i en funció del seu resultat obtingut cal decidir si la informació resultant s'ha d'emmagatzemar en el sistema. En aquest sentit, s'han utilitzat tres polítiques principals (Vernet, et al., 2008) a l'hora d'executar aquesta decisió.

1. *Política Ultra Conservadora.* Els casos o situacions que són diferents als recollits a la Memòria de Casos s'emmagatzemen per un futur, hagin estat encertats o no. Aquesta política, inspirada en la política *DifSim* (Golobardes, 1998) presenta la problemàtica que la majoria dels casos s'acaben introduint a la Memòria de Casos. Això provoca que el sistema degeneri i tendeixi a una recuperació de casos posterior cada vegada més lenta.
2. *Política Conservadora.* Només s'emmagatzemen a la Memòria de Casos aquells casos que s'han fallat evitant així una sobrecàrrega de casos inútils. Això hauria de garantir que casos iguals o similars en el futur puguin ser predits arran d'aquesta nova informació introduïda al sistema.
3. *Política Flexible.* La política anterior garanteix que casos similars al cas emmagatzemat es puguin encertar amb la nova informació. Quan es falla un cas, però, caldria preguntar-se perquè ha passat tal error. Sovint és a causa que la informació recuperada

no és prou similar al nou cas i per tant, a vegades, la predicció falla. Però, la informació que ens portaria a la resposta correcta existeix a la Memòria de Casos? En una memòria *clusteritzada*, on cada cas està format per un conjunt d'individus, es podria donar el cas que el grup més similar a l'actual doni com a resultat una classificació incorrecta. Però, analitzant amb més detall, potser es pot arribar a confeccionar la nova informació amb la ja existent al sistema. Si se cerquen individualment els membres del grup dins de la Memòria de Casos, es podria descobrir si la combinació d'alguns d'ells respondria correctament a la classificació. Si és el cas, no val la pena emmagatzemar el nou cas i el sistema s'hauria de limitar a apuntar-se aquesta situació, reorganitzant la Memòria de Casos perquè això sigui possible. L'Algorisme 3 esquematitza aquest raonament.



Algorisme 3. Diagrama d'activitats representant la Política Flexible a la Fase d'Emmagatzematge del Raonament Basat en Casos.

Les proves realitzades amb les dades provinents de l'entorn educatiu han demostrat que l'ús de la Política Flexible en la Fase d'Emmagatzematge comporta una disminució del 23% de la Memòria de Casos respecte a l'ús de la Política Conservadora, amb la consegüent millora en els temps d'accés a la informació. En el proper apartat es mostraran els resultats amb més detall de

cadascuna de les polítiques aplicades al cicle de CBR quan la Memòria de Casos està formada per grups.

6.5.6 Resultats obtinguts del sistema complet

Tot i que en l'apartat 6.5.2.2 ja s'han avançat els resultats obtinguts en la predicció individual de l'alumne, a continuació s'exposaran els resultats més rellevants obtinguts amb el sistema complet, que permetin corroborar el correcte funcionament del mòdul complet.

Com es veurà a continuació, per tal de testejar algunes de les tècniques introduïdes s'han utilitzat no només jocs de dades provinents d'un entorn educatiu, sinó que s'han aprofitat altres dominis i entorns amb característiques similars en els que els Grups de Recerca hi han treballat en general.

Els resultats mostraran primer el funcionament de les diferents polítiques d'agrupament, per passar posteriorment a l'anàlisi de les funcions de similitud proposades. Finalment s'analitzaran les polítiques d'emmagatzematge dissenyades. Els resultats obtinguts que es mostren a continuació s'han realitzat seguint la metodologia de proves comentada a l'apartat 6.3, essent la *10-fold stratified cross validation* la tècnica de validació creuada més utilitzada i de la que es mostren els resultats en aquest apartat.

Malgrat els tres algorismes d'agrupació proposats permeten classificar els conjunts de dades amb un nombre de classes configurable, el nombre de classes que el sistema utilitza per defecte en l'entorn educatiu és 5, corresponent als diferents nivells de coneixement assolits per un alumne en un moment determinat.

Taula 14. Resultats obtinguts en l'aplicació de les polítiques de primer nivell de *clustering* a diversos jocs de proves. Per cada joc de proves es mostra el percentatge d'efectivitat, el temps de CPU en milisegons i la desviació estàndard, en cada política analitzada. Es marquen amb un ✓ els resultats significativament millors.

Ref.	CBR			Mean Clustering			MKM Clustering		
	Ef(%)	CPUt	std	Ef(%)	CPUt	std	Ef(%)	CPUt	std
BC	96.14	56.34	1.45	96.28 ✓	0.13 ✓	1.87	96.71 ✓	31.96 ✓	1.53
BI	83.15	199.63	3.55	79.07	0.27 ✓	4.66	81.40	1.52 ✓	3.76
GL	69.16	33.60	7.32	53.74	0.51 ✓	7.01	70.79 ✓	30.52 ✓	8.70
IO	90.03	99.61	4.28	81.77	0.45 ✓	5.07	90.31 ✓	1.60 ✓	5.38
IR	95.33	6.00	3.06	92.67	0.20 ✓	2.00	97.33 ✓	1.50 ✓	3.27
MA	62.50	90.41	13.73	64.81 ✓	0.23 ✓	9.42	63.89 ✓	65.69 ✓	9.86
SO	82.21	198.71	6.99	70.67	1.83 ✓	7.19	82.93 ✓	89.20 ✓	7.73
VE	66.90	125.06	4.33	84.04 ✓	0.54 ✓	4.39	65.60	2.00 ✓	3.75

Analitzant les dues polítiques aplicades en el primer nivell de *clustering* respecte al CBR tradicional, a la Taula 14 es pot veure com en tots els casos la introducció dels algorismes d'agrupament permet una recuperació molt més ràpida de la informació, tot mantenint satisfactòriament en general l'efectivitat del sistema. Tot i que hi ha alguns jocs de dades que presenten resultats significativament pitjors en la tècnica de *Mean Clustering* (GL, IO i SO) això no passa quan s'aplica la política *MKM Clustering*. Aquestes proves s'han realitzat, no només en l'entorn educatiu, sinó que s'han volgut provar en diferents jocs de dades provinents de l'*UCI repository* i de dades mèdiques (vegeu apartat 6.3). Per comprovar l'efectivitat de les agrupacions s'han avaluat els casos integrats en cada *cluster*, tot verificant la seva pertinença o no a la categoria assignada.

A la Taula 14 es mostren els resultats corresponents a l'efectivitat del sistema i al temps d'execució en la recuperació, tant per al sistema CBR tradicional, com per a les polítiques aplicades al primer nivell de *clustering*. Es pot observar que les polítiques aplicades milloren el temps d'execució del CBR tradicional i mantenen els resultats per a la majoria de jocs de dades. Cal tenir en compte que alguns d'aquests jocs de dades, com s'ha comentat a l'apartat 6.3, tenen dades que s'han de classificar en més de 5 classes. Això provoca que els valors d'efectivitat estiguin d'acord a aquesta circumstància.

Si ens centrem en els resultats obtinguts en l'entorn educatiu els resultats són els que es mostren a la Taula 15. En aquest cas s'aprofundeix en l'anàlisi mostrant cadascun dels resultats, tenint en compte el tipus d'algorisme emprat per a la realització del *clustering* i la política de refinament aplicada al primer nivell d'agrupació.

Taula 15. Resultats obtinguts per comprovar la qualitat del *clustering* realitzat en el primer nivell. Per cada combinació d'algorismes es mostra el percentatge d'eficàcia, el temps de resolució en milisegons i la desviació estàndard.

	Mean Clustering			MKM Clustering		
	Ef (%)	CPUt	std	Ef(%)	CPUt	Std
K/X-Means	54.83	0.23	6.95	70.75	0.41	5.21
K/X + WITT	52.27	0.99	4.05	67.92	1.20	4.48
K/X + NormWITT	56.25	0.98	4.34	73.02	1.15	4.54

Com es pot observar a la Taula 15 els algorismes utilitzats són l'algorisme *K-Means* (*X-Means* fixat a un límit de 5 *clusters*) i la combinació d'aquests algorismes amb els algorismes WITT i *NormWITT* pel tractament dels atributs discrets. Els algorismes WITT i *NormWITT* incrementen considerablement el temps de CPU atès el seu funcionament complex. Es pot veure com la variació dissenyada NormWITT aconsegueix millorar els resultats en evitar cohesions molt petites. Si ho comparem amb els resultats de la Taula 9 es pot veure que els resultats obtinguts

per la política *Mean Clustering* són significativament pitjors que usant una política sense algorismes d'agrupament. Tot i així, l'aportació que ens dona aquesta política és la velocitat d'execució, atès que la classificació es basa amb la comparació de només els centroides.

La combinació dels dos algorismes, *K/X Means* i *NormWITT* millora els resultats globals. D'altra banda, també es confirma que el refinament del primer nivell de *clustering* utilitzant la política *MKM Clustering* fixada a 5 agrupacions millora en gran mesura els resultats obtinguts a la política basada en el *Mean Clustering*.

Analitzant com de prop o de lluny han estat els algorismes de classificar correctament els alumnes en cada agrupació, la Taula 16 analitza el percentatge d'encerts i errors en funció de la classe assignada. Així, la matriu expressa quin percentatge de casos s'ha classificat en cada classe definida i a quina classe pertanyien realment aquests casos. La matriu s'ha construït a partir dels resultats obtinguts per a la millor combinació d'algorismes.

Taula 16. Matriu d'encerts entre la classificació proposada per l'algorisme de primer nivell i la classe real a la que pertanyia l'alumne.

Real/Predit	NP	[0-4)	[4-7)	[7-9)	[9-10]
NP	88.95	7.23	3.72	0.10	0.00
[0-4)	25.75	63.28	7.85	3.03	0.09
[4-7)	13.43	22.45	50.45	12.46	1.21
[7-9)	0,51	2,76	22.20	66.25	8,28
[9-10]	0,00	0,00	0,00	53,63	46.37

Si observem la Taula 16 es pot comprovar com l'algorisme de classificació de primer nivell encerta generalment amb la classificació esperada. A més, quan falla no es queda molt lluny de la classe real. Es veu també clarament com l'algorisme mai classifica un alumne excel·lent com a dolent i viceversa. Les categories que més dubtes generen són les que diferencien entre el No Presentat i el Suspens, o entre el Suspens i l'Aprovat just.

S'ha provat també de donar llibertat a la política *MKM Clustering* per veure quants *clusters* proposava per a la classificació dels casos sense cap límit d'agrupacions fixat. A la Taula 17 s'ha analitzat en primer lloc amb el joc de dades variat comentat anteriorment. Per cada joc de dades s'expressa el nombre de classes reals i el nombre de classes que proposa una política *MKM Clustering* sense cap limitació. El valor n_i indica que el nombre d'agrupacions coincideix amb el nombre de casos contemplats per aquella partició. Dels resultats obtinguts en aquesta taula es pot concloure que la política *MKM Clustering* tendeix a trobar el màxim d'agrupacions diferents possibles encara que tinguin cohesió entre elles, cosa que recomana fixar sempre el nombre de clusters màxim a priori.

Taula 17. Nombre d'agrupacions proposat per cada classe obtinguda per la política MKM Clustering sense cap restricció de màxims als diferents jocs de dades exposats a l'apartat 6.3.

Joc de dades	Classes	Nombre de <i>clusters</i>
BI	2	28-16
BC	2	$27-n_i$
GL	7	$20-n_i-10-n_i-20-n_i-10$
IO	2	24-6
IR	3	20-4-10
MA	2	n_i-40
SO	2	$25-n_i$
VE	4	25-20-35-35

Finalment, si analitzem el percentatge correcte de classificació pel conjunt de dades en un entorn educatiu segons la política *MKM Clustering*, el resultat és el que es mostra a la Figura 50. Com es pot confirmar, a mesura que es força la classificació amb un nombre majors d'agrupacions, el percentatge d'encerts disminueix.

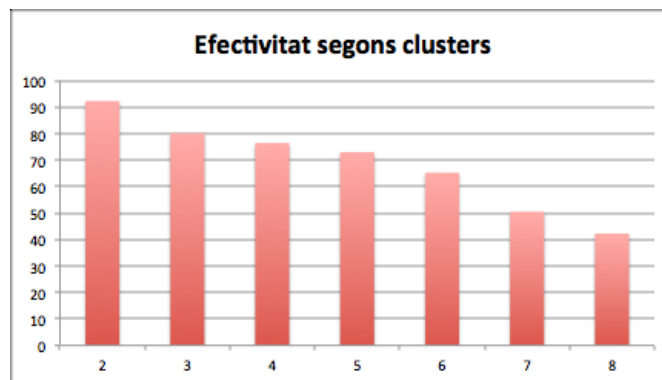


Figura 50. Disminució del percentatge d'efectivitat a causa de l'increment d'agrupacions realitzat sobre les dades de l'entorn educatiu.

Un cop s'ha realitzat el segon nivell de *clustering* corresponent a la formació dels grups de treball, el sistema entra en la fase de la predicció grupal pròpiament dita. A l'apartat 6.5.4 s'han proposat diverses polítiques de recuperació adaptades a aquesta nova organització de la Memòria de Casos. La Taula 18 mostra els resultats obtinguts tot aplicant les polítiques de recuperació grupal de Hamming (equació 13), Euclidiana (equació 14), Màxima (equació 15), Recompte de valors superant l'indiar i la distància calculada a partir de la norma de Frobenius (equació 18). Per dur terme aquest experiment s'han agrupat els alumnes de 3 en 3 i s'ha provat a partir d'agrupacions amb alumnes homogenis i heterogenis. Per a la funció de similitud Recompte de Valors s'ha fixat un llindar de 0,4 punts.

Taula 18. Resultats obtinguts en les diferents funcions de similitud grupals en funció de si els grups establerts són de caire homogeni o heterogeni. Els resultats estan expressats en percentatge d'encerts sobre la predicció a 2 classes.

	Hamming	Euclidiana	Màxima	Recompte	Frobenius
Grups homogenis	87.75	92.25	75.24	82.79	70.35
Grups heterogenis	81.38	87.12	66.72	70.93	62.25

Per tal d'analitzar la importància del llindar en la funció de similitud de Recomppte de Valors, s'ha testejat la seva rellevància a partir de diverses proves, els resultats de les quals es representen a la Figura 51. Com es pot observar, el millor comportament es detecta quan el llindar pren un valor entorn de 0,4. Aquest és el valor que s'ha fixat per al sistema.

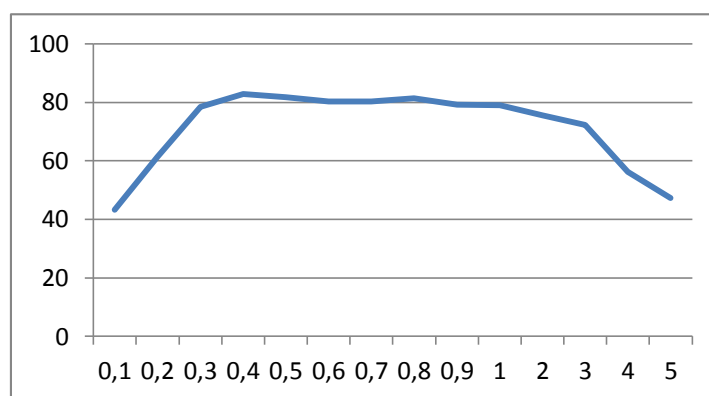


Figura 51. Efectivitat de la funció de similitud de Recomppte de Valors en funció del llindar discriminant utilitzat per aquest càlcul.

Observant els valors de la Taula 18 es pot veure que les millors funcions de similitud són les funcions Euclidiana i de Hamming. Quan els grups que es formen són homogenis els resultats obtinguts milloren significativament. Això és a causa que la tendència del grup és molt més marcada i el sistema identifica aquesta circumstància. D'altra banda també es va detectar que si el nombre d'alumnes o el nombre d'atributs creix, la funció de similitud de Recomppte de Valors es comporta millor atès que això permet una discriminació millor entre les matrius.

Les polítiques aplicades a la Fase d'Emmagatzematge permeten no només mantenir el percentatge d'encerts, sinó que a la vegada aconseguir un bon estalvi de la mida de la Memòria de Casos. S'observa que en un entorn com l'educatiu la Política Ultra Conservadora (PUC) és inviable, atès que quasi bé mai es trobaran dues situacions idèntiques i, per tant, el sistema hauria d'emmagatzemar la pràcticament totalitat dels nous casos que li arribin. Per contra, la Política Conservadora (PC), que només emmagatzema els casos fallats permet un estalvi considerable sense empitjorar en general els resultats del sistema.

La Política Flexible (PF), tot i introduir una càrrega extra al rendiment del sistema, permet conservar la fiabilitat de la PC aconseguint un estalvi mig del 23% en la mida de la Memòria de Casos. La Taula 19 mostra els resultats obtinguts aplicant una distància Euclidiana en la recuperació dels grups més similars, en funció de cadascuna de les polítiques d'emmagatzematge que s'han comentat. Per cada política aplicada s'indica l'efectivitat obtinguda en la classificació dels grups i el percentatge d'increment de la Memòria de Casos respecte la memòria original, tenint en compte que un 90% de la Memòria de Casos s'ha utilitzat per test (essent per tant un 10% l'increment màxim possible de la MC).

Taula 19. Percentatges d'efectivitat (predicció a 2 classes) i d'increment de la Memòria de Casos en funció de la Política d'Emmagatzematge aplicada: Política Ultra Conservadora (PUC), Política Conservadora (PC), Política Flexible (PF)

	PUC	PC	PF
Efectivitat	93,40	92,94	92,75
Increment de la MC	9,80	2,82	2,17

Com a resum podríem concloure que l'ús de la combinació dels algorismes K/X-Means més l'algorisme *NormWITT*, tot recuperant la distància grupal a partir de la funció matricial euclidiana, és la millor combinació per fer la predicció global. La ponderació dels atributs és un factor cabdal per a la bona efectivitat del sistema i la introducció dels algorismes genètics o d'un expert humà és indispensable per ajustar el sistema. D'altra banda, afegint-hi una Política Flexible en l'emmagatzematge es permet un increment progressiu de l'efectivitat del sistema i un millor rendiment del sistema, quant a volum d'informació gestionada.

L'ús d'aquesta mena d'algorismes ha denotat també un augment de velocitat en la recuperació de la informació fent bastant plausible el seu ús en temps real, en l'arquitectura proposada en el capítol anterior.

6.6 Resum i conclusions

En aquest capítol s'ha pogut comprovar que el Raonament Basat en Casos és una tècnica vàlida per realitzar les prediccions necessàries en el Mòdul pedagògic del sistema TICVA, tant quan s'està predint l'evolució d'un alumne de manera individualitzada, com quan aquest forma part d'un grup més ampli.

D'altra banda, les tècniques dissenyades i introduïdes a cadascuna de les fases del cicle de CBR permeten millorar els resultats tradicionals d'un sistema CBR tradicional.

S'ha verificat que la ponderació d'atributs és un aspecte clau en totes les funcions de similitud de la fase de recuperació i es fa indispensable quan els atributs que formen un cas són activitats acadèmiques, que, sovint, tenen dificultat i rellevàncies diferents. El mètode de ponderació

introduït amb els Algorismes Genètics ha demostrat ser un mètode robust i fiable i els resultats confirmen la seva efectivitat. D'altra banda, la ponderació efectuada en un entorn matricial també resulta correcta i classifica en general correctament els nous casos.

Així com les funcions de similitud aplicades a una Memòria de Casos tradicional no són gaire diferents de les utilitzades per la majoria d'investigadors, sí que ha calgut dissenyar i adaptar noves funcions de similitud que serveixin per comparar casos més complexos, els quals estan definits a partir de l'agrupació de diferents individus. En efecte, la introducció de cinc funcions de similitud vàlides per aquest entorn ha permès assolir amb èxit la recuperació dels grups d'estudiants més similars al grup tractat.

D'altra banda, l'organització doble de la Memòria de Casos, tenint la possibilitat d'utilitzar-la a nivell individual o grupal, és una interessant aportació que obre força vies d'investigació. Pot ser rellevant estudiar quina hauria de ser la millor manera de fer conviure les dues Memòries per tal d'optimitzar al màxim el procés complet del sistema. Actualment, el rendiment del sistema és més que acceptable, atesos els temps indicats en l'apartat de resultats.

Finalment, la definició de polítiques d'emmagatzematge que permetin optimitzar la informació inclosa a la Memòria de Casos aporta de nou un millor rendiment del sistema. En particular, la definició de la Política Flexible evita una sobrecàrrega destacable que altres polítiques més conservadores introdueixen.

S'ha observat el correcte funcionament del sistema de predicció global que permetrà al sistema TICVA tenir la informació necessària per fer la tutorització de la classe a partir de les dades establertes. Així, el sistema podrà en temps real, analitzar nous resultats que vinguin dels grups d'alumnes, executar novament el sistema global de predicció i proposar canvis en els grups que informaria al professor de l'assignatura. Aquest mòdul final es comentarà amb més profunditat a les línies futures de treball d'aquesta tesi.

Capítol 7

Conclusions i línies de futur

S'ha vist com al llarg dels diferents capítols d'aquesta tesi s'han anat resolent els objectius que s'havien proposat a la introducció. Com ja s'apuntava quan es descrivia la Metodologia d'Investigació a l'inici d'aquesta tesi, cadascun dels objectius ha tingut un cicle de recerca força independent que ha permès que es poguessin validar cadascun d'ells sense tenir cap dependència directa de la resta. Això no és res nou, doncs passa en molts altres àmbits d'investigació. No obstant això, la visió completa de la tesi i l'encaix indispensable que han de tenir els diferents capítols han estat treballats de tres maneres diferents, que de ben segur el lector ja haurà intuït:

1. En primer lloc, essent coherent amb el fil argumental principal de la tesi, presentat ja a la introducció i que dona el lligam de fons a tot el treball realitzat.
2. En segon lloc, havent-hi un creuament de les publicacions obtingudes en els diferents objectius, que permet deixar palès que el treball entre capítols està relacionat.
3. I, per últim, redactant les conclusions definides per cadascun dels objectius de tal manera que servissin sempre com a introducció per al següent capítol i, per tant, per al següent objectiu.

És per aquest motiu d'independència en la recerca de cada objectiu, que l'autor creu que és millor presentar les conclusions d'aquesta tesi separades per cadascun dels objectius inicialment definits. D'altra banda, cadascuna de les conclusions que s'enumeren a continuació desperta segur una visió futura i una manera natural de continuar el treball en cadascuna de les línies de recerca. Per aquest motiu també, les línies de futur d'aquesta tesi es relacionaran directament amb les conclusions que s'han anat extraient de cada objectiu, permetent així al lector veure quina seria la intenció de l'autor en cas de tenir continuïtat.

7.1 El naixement de l'*e-learning*

Com el lector recordarà, el primer objectiu de la tesi perseguia descobrir com la tecnologia ha marcat la història de l'*e-learning* amb la introducció de les TIC i el profund canvi obligat en el model pedagògic. A més, es volia analitzar les tendències de futur d'aquest sistema d'aprenentatge.

Al llarg de la consecució d'aquest objectiu, s'ha vist com els models pedagògics han anat variant al llarg dels temps. La teoria *instructivista*, on el professor era el centre d'atenció de l'ensenyament ha donat pas a un model on el centre és l'alumne: el *constructivisme*. Es constata que les particularitats de cada individu són respectades quan s'elaboren els recursos formatius i el propi alumne n'és del tot partícip.

Tot i així, es pot concloure, que no és fins ben entrada l'era digital que s'ha reformulat el concepte d'aprenentatge. La incursió de les tecnologies web al dia a dia de les persones, ha produït un moment de canvi social que també ha afectat l'educació. La Web 2.0 crea el concepte de comunitat virtual, on l'educació és també una experiència social. És l'anomenat *connectivisme*.

I és justament el desplegament d'aquesta teoria connectivista el que possibilita, recolzat per noves iniciatives d'*e-learning* el *boom* dels Entorns Virtuals d'Aprenentatge (EVA). Tot i així, també es pot observar que moltes institucions aposten per un model mixt, el *blended learning*, que combina el mode presencial i el virtual en proporcions raonables.

En qualsevol cas, la Web segueix evolucionant i ja es parla de les generacions 3.0 i 4.0. Generacions on els agents intel·ligents formaran part de la vida educativa. És per tant la Intel·ligència Artificial una eina indispensable en l'educació i en tots els àmbits de la vida. I és que mentre l'autor està escrivint aquestes línies, s'anuncia que Google publica parcialment els seus sistemes de raonament intel·ligent, mitjançant la pàgina *TensorFlow*, per tal que cadascú pugui definir on vol aplicar la IA en la seva vida.

Aquest objectiu ha sigut indispensable per fomentar la base teòrica necessària per arribar a la Hipòtesi 1 d'aquesta tesi.

7.2 De l'Analítica de l'Aprenentatge a la *Big Learning Data*

Si recordem el segon objectiu, aquest es dividia en dues parts diferenciades: d'una banda entendre l'aparició del concepte de la *Big Learning Data* i, de l'altra, introduir el model de referència de l'Analítica de l'Aprenentatge (LA).

Tal i com s'ha anat veient en el capítol 3, l'interès per disposar de les dades que es generen durant el procés d'aprenentatge ha anat en augment i són molts els investigadors que les volen

analitzar. És el naixement de la *Big Learning Data*. S'ha conclòs que si es realitza aquesta anàlisi de manera correcta es pot extreure un coneixement molt valuós en els sistemes d'aprenentatge en línia.

S'ha vist que la LA implementa un metodologia per aconseguir aquest objectiu, basada en tres passos concrets: preprocessament, acció sobre les dades i postprocessament. A mesura que les dades van passant per aquest procés van convertint-se en coneixement cada vegada més útil per als sistemes d'aprenentatge. Tanmateix, la tecnologia aplicada en aquest procés també ha anat variant i ara per ara destaquen algunes com la mineria de dades, l'anàlisi estadística, la visualització de la informació i l'anàlisi social de l'aprenentatge.

Certament, veient que cada vegada l'ésser humà és més digitalment social, sembla que aquest coneixement a tractar anirà creixement en els propers anys i caldrà experts especialitzats per cada àmbit d'aplicació.

També, relacionat amb aquest objectiu, s'ha introduït el model de referència de la LA, que mostra les quatre dimensions crítiques que presenta l'anàlisi de l'aprenentatge.

Les següents publicacions obtingudes són les més destacades com a resultat de la consecució d'aquest objectiu:

- Vernet, D., Nicolas, R., Golobardes, E., Fornells, A., & Garcia-Piquer, A. (2010). Work in Progress - Intelligent Tutoring System Framework for the Acquisition of Knowledge and Competences. *39th ASEE/IEEE Frontiers in Education Conference*. Washington.
- Canaleta, X., Vernet, D., Vicent, L., & Montero, J. (2014). Master in Teacher Training; a Real Implementation of Active Learning. *Computers in Human Behaviour*(31), 651-658.

7.3 Els Sistemes Tutors Intel·ligents

El tercer objectiu principal d'aquesta tesi pretenia estudiar a fons els Sistemes Tutors Intel·ligents (STI): quins han estat els seus antecedents, com han evolucionat aquests fins arribar als propis STI i quina n'és l'estructura estàndard.

En el quart capítol d'aquesta tesi s'ha presentat l'evolució dels sistemes d'aprenentatge assistits per ordinador. S'ha estudiat la seva evolució des dels CAI (Instrucció Assistida per Ordinador) fins a la seva evolució posterior en els ICAI (Instrucció Intel·ligent Assistida per Ordinador), primer punt d'encontre entre l'alumne i la Intel·ligència Artificial.

Així, amb l'aplicació de la IA, els SEI (Sistemes Educatius Intel·ligents) han evolucionat cap als STI. El naixement del Sistemes Tutors Intel·ligents va ser l'exponent màxim de la introducció de la IA en l'educació, tot i que s'ha pogut comprovar que estan pensats només en el guiatge individual de l'alumne. Amb l'aparició de la tecnologia web aquests sistemes han anat migrant cap als AIWBES (Sistemes Educatius Adaptatius i Intel·ligents Basats en Web). Aquests entorns faciliten d'alguna manera l'aprenentatge col·laboratiu tot i que la tutoria col·lectiva o grupal segueix essent inexistent a l'actualitat.

S'ha estudiat a fons l'estructura interna d'un STI, analitzant cadascun dels mòduls que el formen, per tal de veure-hi punts de millora per poder proposar posteriorment una arquitectura que els solventi.

Per contra, no s'ha entrat a fons en veure quina estructura pedagògica segueixen els STI. Hi ha força estudis al respecte (Nkambou, Frasson, & Gauthier, 2003) (Nwana, 1990) i una bona línia de futur en aquest sentit consistiria en estudiar quina relació directa hi ha entre el tipus de tutorització que ofereix el sistema (a què o qui es fa més incís, si és més important el contingut o el professor, ...) i la seva arquitectura intrínseca.

Com a resultat d'aquest objectiu es destaquen les següents publicacions:

- Vernet, D; Canaleta, X. (2004). Sistema Tutor Avanzado. *Actas de las X Jornadas de Enseñanza Universitaria de la Informática (JENUI'2004)*, (pp. 211-215). Alicante.
- Vernet, D. (2004). *Diploma d'Estudis Avançats*. Barcelona.
- Vernet, D; Golobardes, E. (2003). An Unsupervised Learning Approach for Case-Based Classifier Systems. *Expert Update, The Specialist Group on Artificial Intelligence*, 6(2), 37-42.

7.4 El sistema TICVA

El penúltim objectiu d'aquesta tesi consisteix en fer una proposta d'arquitectura d'un sistema d'aprenentatge en línia intel·ligent que permeti suplir les limitacions que tenen els AIWBES en general en el tractament de comunitats virtuals de manera no individualitzada.

Així, en el capítol 5 s'ha proposat l'arquitectura TICVA (Tutorització Intel·ligent de Comunitats Virtuals d'Aprenentatge) que busca introduir els sistemes tutors intel·ligents cap a la següent generació web, la 4.0, on els agents intel·ligents han de tenir paper rellevant i convertir la web en un entorn simbiòtic.

En primer lloc, s'ha constatat la validesa de l'arquitectura contrastant la seva conceptualització amb el model de referència de l'anàlisi de l'aprenentatge. En aquest sentit, s'han respost les

preguntes del model de Chatti que permeten la validació del sistema. S'ha conclòs per tant que és possible definir un model que abasti les necessitats d'un STI i que inclogui els problemes observats en aquest.

S'han desenvolupat cadascun dels mòduls seguint coherentment les respostes donades anteriorment, i s'ha presentat un model d'arquitectura detallat estructurat per mòduls per una més fàcil implementació. Al final del capítol es conclou que les tecnologies web necessàries per implementar-lo depenen del moment propi d'implementació, mentre que la implementació de la part intel·ligent ja es pot dur a terme.

En el Mòdul del Coneixement de Domini és important analitzar com emmagatzemar el material docent. Com s'ha comentat, el model competencial introduït pel procés de Bolonya requereix que les activitats estiguin dirigides a un resultat d'aprenentatge. És per això que s'ha distribuït el coneixement de tal manera que es tingui en compte aquest factor. Un immersió més a fons en aquest sentit, seria una bona línia d'investigació com ja s'ha fet per exemple a (Badaracco, 2013).

El Mòdul Expert s'encarrega de la correcció dels exercicis dels alumnes i de donar suport al professor en la revisió del material lliurat per l'alumne. En aquest mòdul s'obren grans possibilitats de recerca, atenent la diversitat d'enfocaments que es poden proposar. Així, el mòdul pot consistir senzillament en una automatització del procés de correcció (per exemple, per qüestionaris, exàmens tipus test, etc.) però també es pot anar més enllà i veure'l com un possible agent intel·ligent amb la capacitat de *raonar* com de correcte és un lliurament.

La informació de l'alumne o del grup, tant de lliuraments com de resultats i retroaccions, queda emmagatzemada en el Mòdul de l'Estudiant o Grup. A diferència dels Sistemes Tutors Intel·ligents clàssics, l'arquitectura proposada amplia la potència d'aquest mòdul i incorpora particularitats especials del concepte *grup*. Això implica una separació clara de les dues zones que caldrà tenir en compte en el moment de la seva implementació.

S'ha vist com el Mòdul Pedagògic és el més important de tot Sistema Tutor Intel·ligent. Incorpora l'habilitat de poder conèixer en quin estat es troba l'alumne o un conjunt d'alumnes. Al capítol 6 s'ha presentat una proposta d'implementació de part d'aquest mòdul, enfocada en aquella encarregada de fer una predicció de l'evolució de l'alumne en el seu procés d'aprenentatge. L'aproximació que s'ha fet és una de les possibles però en altres estudis també s'ha pogut veure que l'evolució de l'alumne es pot entendre com un conjunt d'estats dins d'un camí d'aprenentatge o *learning path*, com per exemple el treball presentat per Falmagne (Falmagne, 1993). Tal com proposa aquest estudi podria ser interessant veure el procés d'aprenentatge com un procés estocàstic de Markov, amb un conjunt d'estats definits, i poder-se fer preguntes tals com: Quins punts del procés d'aprenentatge són els més fàcils per l'alumne? Quin és el camí d'aprenentatge més habitual d'un alumne? O, quan triga l'alumne normalment en passar d'un estat a un altre? Poder respondre a aquests tipus de preguntes seria molt important i permetria millorar, entre altres coses, el disseny de les activitats i la seva presentació de cara a l'alumne.

El coneixement potencial generat per la *Big Learning Data* és també un altre camí de recerca. La necessitat d'un mòdul com el *Filtered Reporting*, amb la capacitat de recollir la informació i generar informes i estadístiques pels usuaris, ha de ser combinat amb una capacitat d'anàlisi i de descoberta de nou coneixement que pugui ser usat pel Mòdul Pedagògic del sistema. Així, s'apunta com a treball futur la investigació relacionada amb els processos de mineria de dades i anàlisi de variables. Certament, l'anàlisi de les interaccions d'un alumne amb la resta de companys de la seva aula virtual pot arribar a conèixer el nivell d'*aprenentatge social* que ha assolit.

La consecució d'aquest objectiu ens ha portat a comprovar la Hipòtesi 1 de la tesi, la qual ha produït les següents publicacions destacades:

- Vernet, D; Zaballos, A; Martin de Pozuelo & Caballero. (2015). High Performance Web-of-Things Architecture for the Smart Grid Domain. *International Journal of Distributed Sensor Networks*, In press.
- Vernet, D., Canaleta, X., Pallàs, G., Navarro, J., & Zaballos, A. (2013). Setting up and Tutoring the Working Groups of a Virtual Learning Community. *Journal of Research and Practice in Information Technology*, 45(3), 91-107.
- Vernet., Canaleta, X., & Pallàs, G. (2012). Tutorización Inteligente de Comunidades Virtuales de Aprendizaje. *Actas del XIV Simposio Internacional de Informática Educativa*. Andorra La Vella.

7.5 La predicció i la tutorització

Finalment, el darrer objectiu consistia en l'anàlisi, el disseny i la implementació del submòdul de predicció del Mòdul Pedagògic i la implicació que aquest tenia amb la tutorització posterior de l'alumne.

S'ha vist com la tècnica de Raonament Basat en Casos (CBR), per la seva particularitat, s'ha manifestat com una aproximació vàlida per desenvolupar la part intel·ligent del sistema de predicció. No obstant això, ha calgut fer intervenir altres tècniques en la representació interna de la Memòria de Casos i en diferents fases del seu cicle de vida.

Les polítiques d'agrupament introduïdes pels algorismes de *clustering* proposats han permès reorganitzar la Memòria de Casos per tal d'admetre la seva configuració grupal. La combinació del *K-Means Clustering*, *X-Means Clustering* i l'algorisme WITT han permès tractar individus amb diferents atributs, numèrics o discrets. Les millores aplicades a l'algorisme WITT amb la introducció de la variant *NormWITT* per evitar l'efecte de cohesió mínima, han contribuït a donar estabilitat en el seu funcionament.

La fase de recuperació del cicle de CBR ha estat treballada des de dos vessants diferents: des del punt de vista de cas individual i des del punt de vista grupal. A nivell individual s'ha incidit en la importància en la ponderació dels atributs en les funcions de similitud per aconseguir una millor exactitud en la predicció de l'evolució de l'alumne. Així, la introducció dels Algorismes Genètics en la ponderació d'aquests atributs ha resultat una aproximació vàlida, que ha refermat i millorat la ponderació que podria donar el professor de manera empírica.

D'altra banda, s'han proposat una sèrie de funcions de similitud vàlides per poder treballar amb informació grupal i fer comparacions coherents. Així, s'han introduït distàncies més innovadores com el Recompte de Valors o la distància de Frobenius, que tot i no millorar l'efectivitat de les distàncies més típiques entre conjunts matricials, han resultat ser un bon punt d'inici prometedor per altres camins d'investigació futurs. En aquest sentit, saber recuperar casos complexos en situacions com les que s'han vist, obre un ventall de possibilitats continuïstes de recerca en aquesta línia.

Una altra línia d'investigació oberta és la que es refereix a la convivència de les dues maneres d'organitzar la Memòria de Casos. En aquesta tesi s'ha dissenyat una estructura que permet recuperar la informació indistintament des de cadascuna d'elles, per separat, però es podria treballar en la possibilitat de crear un convivència *fusionada* entre els dos plantejaments: fer que la Memòria de Casos sigui única, i comparteixi tant casos complexos com senzills.

Pel que a la fase d'emmagatzematge s'han dissenyat tres polítiques diferents per realitzar l'aprenentatge del sistema, resultant la Política Flexible (PF) la més rellevant i aconseguint un estalvi de memòria considerable. Es pot concloure que aquest factor és clau per al rendiment del sistema, ja que afecta a la velocitat i a la seva execució en temps real, i que, tenint en compte que la Memòria de Casos sempre es va alimentant de tota l'experiència viscuda, aquesta hauria de ser una línia de treball clara com a continuació d'aquesta tesi.

D'altra banda, les dades que s'obtenen del submòdul de predicció són clau per a la tutorització de l'alumne. En aquest sentit, actualment només es poden utilitzar per a la creació d'alertes i recomanacions concretes, però el camp d'investigació que s'obre per tal d'aprofitar tot el valor que poden proporcionar en la tutoria de l'alumne, és clarament vast.

La implementació de la resta de mòduls proposats al capítol 5, també formen part del treball futur per aconseguir desenvolupar un sistema complet: un sistema que permeti integrar la màgia de la futura Web 4.0 i l'educació a distància. Sens dubte, un repte deliciós.

Els resultats obtinguts durant la realització d'aquest objectiu han permès verificar la Hipòtesi 2 de la tesi, tot destacant les següents publicacions:

- Vernet, D; Golobardes, E. (2003). An Unsupervised Learning Approach for Case-Based Classifier Systems. *Expert Update, The Specialist Group on Artificial Intelligence*, 6(2), 37-42.

- Golobardes, E; Vernet, D; Salamó, M. (2000). Prediction in an Educational Environment Using Case-Based Reasoning. *Proceedings of the Learning'00*. Madrid.
- Vernet, D & Golobardes. (2003). Adding Unsupervised Learning in the Case-base Organisation. *Actes del 6è Congrés Català d'Intel·ligència Artificial, CCIA 2003.*, 100, p. 307-315. Palma de Mallorca.
- Vallejo, A., Zaballos, A., Vernet, D., Orriols, A., & Dalmau, J. (2009). A Traffic Engineering Proposal for ITU-T NGNs Using Hybrid Genetic Algorithms. *The International Journal on Advances in Internet Technology*, 2(1), 162-172.
- Zaballos, A., Vernet, D., & Selga, J. (2013). A Genetic QoS-Aware Routing Protocol for the Smart Electricity Networks. *International Journal of Distributed Sensor Networks*, 2013(135056), 1-12.
- Vernet, D., Nicolas, R., Golobardes, E., Fornells, A., Garriga, C., Puig, S., & Malvehy, J. (2008). Pattern Discovery in Melanoma Domain Using Partitional Clustering. *Congrés Català d'Intel·ligència Artificial, CCIA. Frontiers in Artificial Intelligence and Applications.*, (pp. 323-330). Sant Martí d'Empúries.
- Vernet, & Golobardes, E. (2002). An Unsupervised Learning Approach for Case-Based Classifier Systems. In U. o. B. Lees (Ed.), (pp. 39-46). Cambridge
- Vernet, D., Salamó, M., Vallespi, C., Camps, J., Golobardes, E., & Bacardit, J. (2001). ¿Cómo predecir la evolución del alumno? *Actas de las VII Jornadas de la Enseñanza Universitaria de la Informática, JENUI 2001.* 1, pp. 335-340. Palma de Mallorca: Universitat de les Illes Balears. Servei de Publicacions i Intercanvi Científic.

7.6 Objectius de publicacions i resum d'aportacions

Dels objectius marcats inicialment quant a publicacions, es pot veure que s'han aconseguit les publicacions a les revistes indexades que s'havien proposat: JRPIT, CHB i IJDSN (2), juntament amb d'altres no indexades. Malauradament, des que es va enviar la publicació a la revista JRPIT fins que aquest va ser publicat, aquesta revista va deixar de indexada ISI, a partir de l'any 2012.

En total doncs, s'han aconseguit 3 publicacions en revistes indexades, 6 en revistes no indexades i 37 en altres publicacions (majoritàriament, publicacions en congressos). A la introducció d'aquesta tesi es pot observar la relació directa entre les publicacions obtingudes i els objectius plantejats.

Les aportacions principals realitzades per aquesta tesi se citen a continuació:

- a) Recull d'estats de l'art de diferents àmbits. S'ha repassat i agrupat la bibliografia existent en cadascuna de les temàtiques següents:
1. Modalitats d'e-learning i entorns virtuals d'aprenentatge
 2. Evolució de l'e-learning i la seva tendència futura
 3. L'Analítica d'Aprenentatge i la Big Data a l'Educació
 4. Evolució dels Sistemes Tutors Intel·ligents
 5. Tècniques d'Intel·ligència Artificial aplicades als Sistemes Tutors Intel·ligents i al món educatiu en general
- b) Definició d'un nou model conceptual de Tutorització Intel·ligent a partir del model de Chatti:
- A partir de les mancances que s'han detectat en els sistemes tutors intel·ligents actuals basats en web, s'han respost les preguntes del model proposat per Chatti, que corresponen a les quatre dimensions crítiques que cal abordar a l'hora de dissenyar models que analitzin el procés d'aprenentatge, i s'han enllaçat directament amb la implicació que tenen en els diferents mòduls i actors que formen part del sistema. Partint de la resposta a aquestes preguntes s'ha definit un nou model conceptual de Tutorització Intel·ligent.
- c) Definició de l'estructura i funcionament d'un sistema que implementa aquest model conceptual (TICVA):
- Seguint el model conceptual definit per l'aportació anterior, s'ha dissenyat una proposta d'arquitectura que pretén incorporar tota la potència de coneixement que ofereix la *Big Learning Data* i la possibilitat de gestionar-la mitjançant la definició i disseny dels mòduls corresponents.
- d) Aportacions destacables referents a l'aplicació d'Intel·ligència Artificial en el Mòdul Pedagògic d'un Sistema Tutor Intel·ligent. A continuació se citen les aportacions realitzades més importants que l'autor vol destacar:
1. La introducció d'Algorismes Genètics en la ponderació dels atributs d'un Sistema de Raonament Basat en Casos per tal de detectar automàticament la rellevància de cadascun d'ells.
 2. La creació de noves funcions similitud en la fase de recuperació d'un sistema CBR que permeten millorar l'efectivitat del sistema.
 3. La definició de noves polítiques d'emmagatzematge en el cicle de CBR que permeten millorar el rendiment del sistema.
 4. La definició d'una nova organització de la Memòria de Casos, podent-se utilitzar de manera tradicional o en format grupal.
 5. La definició d'una nova manera d'utilitzar la Memòria de Casos en un entorn grupal aplicant algorismes d'agrupació a 1 i 2 nivells de profunditat i posant-la en pràctica en la classificació a nivell grupal.

6. L'optimització de la velocitat en la recuperació dels casos en un Sistema de Raonament Basat en Casos sense perdre efectivitat en el sistema i reduint el temps total de classificació.
7. La modificació de l'algorisme WITT d'agrupament per a atributs discrets per tal d'estabilitzar-lo per cohesions de *clusters* molt baixes.
8. La integració de funcions de similitud matricials dins de la fase de recuperació amb l'objectiu de comparar casos complexos.

Referències

- Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning: Foundation Issues, Methodological Variations, and System Approaches. *AI Communciations*, 7, 39-59.
- Aghaei, S., Ali, M., & Khosravi, H. (2012). Evolution of the World Wide Web: From Web 1.0 to Web 4.0. *International Journal of Web & Semantic Technology (IJWesT)*, 3(1).
- Agrawal, R., & Ramakrishnan, S. (1994). Fast Algorithms for Mining Association Rules in Large Databases. *Proc. 20th int. conf. very large data bases, VDLB., 1215*, 487-499.
- Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. *ACM SIGMOD international conference on Management of data.*, 94-105.
- Aha, D. (1998). *Feature weighting for lazy learning algorithms*. Washington: Naval Research, Laboratory, Navy Center for Applied Research in Artificial Intelligence.
- Albó, J., Vernet, D., Canaleta, X., & Vilasís-Cardona, X. (2013). LSMaker: A Robotic Platform for Engineering Education. *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS 2013)*. Beijing.
- Alsina, R., Canaleta, X., Montero, J., & Vernet, D. (2004). La Interdisciplinarietat com a Recurs de Millora de l'Aprenentatge dels Alumnes de Primer d'Enginyeria. *Llibre d'actes del 3r Congrés Internacional 'Docència Universitària i Innovació' (CIDUI'2004)*. Girona.
- Amela, V. (2010). *Sistema tutor intel·ligent adaptatiu para laboratorios virtuales y remotos*. Valencia: Universidad Politécnica de Valencia.
- Anderson, J., Corbert, A., Koedinger, K., & Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. (I. Lawrence Erlbaum Associates, Ed.) *The Journal of the Learning Sciences*, 4(2), 167-207.
- Anderson, T., & Dron, J. (2011). *Three generations of Distance Education Pedagogy*. Obtenido de <http://www.irrodl.org/index.php/irrodl/article/view/890/1826>
- Ankerst, M., Breunig, M., Kriegel, H., & Sander, J. (1999). OPTICS: ordering points to identify the clustering structure. *Int. Conf. on Management of Data*, (págs. 49-60).
- Aslan, B., Galip, O., Ozlem, I., & Mustafa, M. (2014). Effect of Bayesian Student Modeling on Academic Achievement in Foreign Language Teaching. *Kuram Ve Uygulamada Egitim Bilimleri*, 14(3), 1160-1168.

- Badaracco, M. (2013). Sistema Tutor Inteligente basado en Competencias (STI-C). Propuesta de Arquitectura y Diagnóstico. *Tesi doctoral*.
- Baker, R., Pardos, Z., & Gowda, S. (2011). Ensembling Predictions of Student Knowledge within Intelligent Tutoring Systems. *User Modeling, Adaptation, and Personalization*, 6787, 13-24.
- Beck, J., Stern, M., & Haugsjaa, E. (1996). *Applications of AI in Education*. ACM Crossroads.
- Belghith, K., Nkambou, R., & Kabanza, F. (2012). An Intelligent Simulator for Telerobotics Training. *IEEE Transactions on Learning Technologies*, 5(1), 11-19.
- Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*. US Department of Education.
- Blaxter, L. (2001). *How to Research*. Buckingham: Open University Press.
- Bologna Declaration. (1999). The European Higher Education Area. *Joint declaration of the European Ministers of Education*.
- Bourdeau, J., & Grandbastien, M. (2010). Modeling Tutoring Knowledge. *Advances in Intelligent Tutoring Systems*, 123-143.
- Briggs, S. (2014). Big Data in education: Big potential or big mistake? *InformedED. Open Colleges*.
- Brown & Burton. (1982). Pedagogical, natural language and knowledge engineering techniques in Sophie. *ITS*, 227-282.
- Brown, J., & Burton, R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 155-191.
- Brusilovsky. (1999). Adaptive and Intelligent Technologies for Web-based Education. *Künstliche Intelligenz*, 4, 19-25.
- Brusilovsky, P., & Peylo, C. (2003). Adaptive and Intelligent Web-based Educational Systems. *International Journal of Artificial Intelligence in Education*, 156-169.
- Burns, H., & Capps, C. (1988). *Foundations of intelligent tutoring systems: An introduction*. Lawrence Erlbaim Associates.
- Cabero, J. (2006). Bases pedagógicas del e-learning. *Revista de Universidad y Sociedad del Conocimiento (RUSC)*, 3(1).

- Camarena, S. (2015). *Entrevista a Carlos Slim*. Obtenido de Diario el País: http://internacional.elpais.com/internacional/2012/10/12/actualidad/1350043165_165522.html
- Campbell, N. (1980). Robust procedures in multivariate analysis. En *Robust covariance estimation*.
- Camps, J., Garrell, J., Golobardes, E., & Vernet, D. (2003). Diseño de Funciones de Similitud para el Razonamiento Basado en Casos usando Programación Genética: estudio con problemas sintéticos. *Proceedings of MAEB 2003*, (pp. 409-416). Gijón.
- Canaleta & Vernet. (2003). Gestión Académica y Protección de Datos. *Proceedings of JENUI'2003, IX Jornadas de Enseñanza Universitaria de la Informática.*, (pp. 151-158). Madrid.
- Canaleta & Vernet. (2004). Propuesta de Créditos ECTS para la Asignatura de Programación de Primer Curso de Ingeniería. *Proceedings of JENUI'2004, X Jornadas de Enseñanza Universitaria de la Informática*. Madrid.
- Canaleta & Vernet. (2011). Introducción de las Tecnologías de la Información y la Comunicación en los Programas Universitarios para Mayores. *Actas del IV Congreso Iberoamericano de Universidades para Mayores (CIUUMM 2011)*. Alicante.
- Canaleta, & Vernet, D. (2004). TIC versus mTIC. *Actas de las V Jornadas de Informática y Sociedad*, (pp. 61-67). Bilbao.
- Canaleta, Vernet, & Navarro. (2013). Metodología on Demand para el Desarrollo de la Asignatura de Sistemas Operativos Avanzados. *Actas de las XIX Jornadas de la Enseñanza Universitaria de la Informática (JENUI 2013)*. Castelló de la Plana.
- Canaleta, Vernet, & Zaballos. (2011). Propuesta de Evaluación Pura por Competencias par Estudios de Ingeniería Informática. *Actas de las XVII Jornadas de Enseñanza Universitaria de Informática, Jenui 2011*. Sevilla.
- Canaleta, Vernet, D., Alsina, R., & Montero, J. (2004). La Interdisciplinariedad como Recurso Aplicable en la Mejora de la Docencia en Primer Curso de Ingeniería. *Proceedings del XIX Symposium Nacional de la Unión Científica Internacional de Radio (URSI2004)*.
- Canaleta, Vernet, Vicent, & Montero. (2012). La Mejora Continua del Profesorado en el Sistema Educativo Actual. *Actas del XIV Simposio Internacional de Informática Educativa*. Andorra La Vella.
- Canaleta, Vernet, Vicent, & Montero. (2014). Master in Teacher Training; a Real Implementation of Active Learning. *Computers in HUMAN Behaviour*(31), 651-658.

- Canaleta, X., Vernet, Vicent, & Montero. (2014). Master in Teacher Training; a Real Implementation of Active Learning. *Computers in Human Behaviour*(31), 651-658.
- Canaleta, X., Navarro, J., Solé, X., & López, P. (2014). Método no Formal para la Evaluación de la Docencia Aplicada al Grado de Ingeniería Informática. *Actas de las XX Jornadas de la Enseñanza Universitaria de la Informática (JENUI 2014)*. Oviedo.
- Canaleta, X., Ros, P., Vallejo, A., Vernet, D., & Zaballos, A. (2008). A System to Extract Social Networks Based on the Processing of Informaton Obtained from Internet. *Proceedings of the 11th International Conference of the Catalan Association for Artificial Intelligence(CCIA 2008)* (págs. 283-289). IOS.
- Canaleta, X, Zaballos, & Vernet. (2012). Evaluación del Aprendizaje, Aprendizaje de la Evaluación. *VII Congreso Iberoamericano de Docencia Universitaria, Ensino Superior: Inovação e Qualidade na Docência. VII Congresso Iberoamericano de Docência Universitária: Livro de Atas, Porto, Junio 2012*. Porto.
- Carbonell, J. (1970). AI in CAI: An artificial intelligence approach to computer assisted instruction. *IEEE Transaction on Man-Machine Systems*.
- Carr, B., & Goldstein, I. (2012). *Overlays : a theory of modelling for computer aided instruction*. AI Memo 406.
- Cataldi, Z., & Lage, F. (2009). Sistemas tutores inteligentes orientados a la enseñanza para la comprensión. *EDUTECH, Revista Electrónica de Tecnología Educativa*(28).
- Chatti, Dyckhoff, Schroeder, & Thüs. (2012). *A Reference Model for Learning Analytics*.
- Cheeseman, P., & Stutz, J. (1996). Bayesian classification (autoclass): Theory and results. *Advances in Knowledge Discovery and Data Mining*, 153-180.
- Clancey, W. (1991). Intelligent tutoring systems: A tutorial survey. *Applied Artificial Intelligence: A Sourcebook*. McGraw-Hill.
- Costa, G., Alves, N., & Rodrigues, T. (2013). *Moral reasoning in e-learning generations: from 1.0 to 4.0*. Obtenido de http://www.academia.edu/10848301/Moral_reasoning_in_e-learning_generations_From_1.0_to_4.0
- Cukier, R. (2013). *Big data: La revolución de los datos masivos*. Turner Publicaciones.
- Dawson, S. (2010). ‘Seeing’ the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology*, 41(5).

- De Jong, K. (1975). *An analysis of the behaviour of a class of genetic adaptive systems*. Michigan: Tesi Doctoral.
- Devedzic, V. (2006). *Semantic web and education*. Springer.
- Díaz, V., & Brown, M. (2012). *Learning Analytics .A Report on the ELI Focus Session*. Educause Learning Initiative.
- Dolenc, K., & Abersek, B. (2015). Tech8 Intelligent and adaptive e-learning system: Integration into Technology and Science classrooms in lower secondary schools. *Computer & Education*, 82, 354-365.
- Ester, M., Kriegel, H., & Xu, X. (1995). A database interface for clustering in large spatial databases. *Knowledge Discovery and Data Mining*, 94-99.
- Falmagne, J. (1993). Stochastic Learning Paths in Knowledge Structures. *Journal of Mathematical Psychology*, 37(4), 489-512.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., & Utuhursamy, R. (1996). *Advanced in Knowledge Discovery and Data Mining*. AAAI Press.
- Fernández, E. (2009). *U-Learning: El futuro está aquí*. Madrid: RA-MA Editorial.
- Ferrance, E. (2000). *Action Research*. Providence, RI: Brown University.
- Fisher, D. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, 139-172.
- Fornells, A., Golobardes, E., Vernet, D., & Corral, G. (2006). Unsupervised Case Memory Organization: Analysing Computational Time and Soft Computing Capabilities. *Lecture Notes in Computer Science Advances in Case-Based Reasoning, ECCBR 2006, LNCS/LNAI(4106)*, 241-255.
- Garrell, J., Golobardes, E., Bernadó, E., Llorà, X., Salamó, M., Camps, J., . . . Rios, J. (2002). Aprendizaje Artificial en Sistemas de Clasificación, Predicción y Diagnóstico Automático. *Proceedings del Workshop de Minería de Datos y Aprendizaje, VIII Iberoamerican Conference on Artificial Intelligence IBERAMIA '2002.*, (pp. 127-136).
- Gartner. (2014). *Gartner's 2014 Hype Cycle for Emerging Technologies Maps the Journey to Digital Business*. Obtenido de <http://www.gartner.com/newsroom/id/2819918>
- Gasevic, D. (2012). *Learning Analytics - A new discipline and linked data*. Obtenido de <http://es.slideshare.net/dgasevic/learning-analytics-a-new-discipline-and-linked-data>

- Gertner, A., & Vanlehn, K. (2000). Andes: A Coached Problem Solving Environment for Physics. *Lecture Notes In Computer Science. Proceedings of the 5th International Conference on ITS.*, 133-142.
- Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley.
- Golobardes, E. (1998). *Aportacions al raonament basat en casos per resoldre problemes de classificació*. Tesis Doctoral. Universitat Ramon Llull.
- Golobardes, E, Llorà, X., Garrell, J., Vernet, D., & Bacardit, J. (2000). Genetic Classifier System as a heuristic weighting method for a Case-Based Classifier System. *Proceedings del 3r Congrés Català d'Intel·ligència Artificial*, (págs. 132-141).
- Golobardes, E, Llorà, X., Garrell, J., Vernet, D., & Bacardit, J. (2000). Genetic Classifier System as a Heuristic Weighting Method for a Case-Based Classifier System. *Butlletí de l'Associació Catalana d'Intel·ligència Artificial*(22), 132-141.
- Golobardes, E, Nieto, M., Salamó, M., Camps, J., Martí, J., & Vernet, D. (2001). Generació de funcions de similitud mitjançant la Programació Genètica pel Raonament Basat en Casos. *Proceedings del 4t Congrés Català d'Intel·ligència Artificial*, (pp. 100-107).
- Golobardes, E, Nieto, M., Salamó, M., Camps, J., Martí, J., & Vernet, D. (2001). Generació de Funcions de Similitud mitjançant la Programació Genètica pel Raonament Basat en Casos. *Butlletí de l'Associació Catalana d'Intel·ligència Artificial*(25), 100-107.
- Golobardes, E, Vernet, & Salamó. (2000). Prediction in an Educational Environment Using Case-Based Reasoning. *Proceedings of the Learning'00*. Madrid.
- Gómez, D. (2015). Análítica visual en el e-learning. *Revista internacional de Informació y Comunicación*, 23(3).
- González, C. (2004). *Intelligent Systems in Education: A review of current research lines*. Obtenido de http://www.uv.es/RELIEVE/v10n1/RELIEVEv10n1_1.htm
- Greller, W., & Drachsler, H. (2012). *Translating Learning into Numbers: A Generic Framework for Learning Analytics*. Obtenido de Open University of the Netherlands: http://ifets.info/journals/15_3/4.pdf
- Gros, B. (2011). *Evolució i reptes de l'educació virtual. Construint l'e-Learning del segle XXI*. Barcelona: Editorial UOC.
- Han, J., & Kamber, M. (2000). *Data mining: Concepts and techniques*.

- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Francisco: Elsevier.
- Hanson, S (1990). Conceptual clustering and categorization: Bridging the gap between induction and causal models. *Machine Learning: An Artificial Intelligence Approach*, 235-268.
- Hanson, S., & Bauer, M. (1989). Conceptual clustering, categorization, and polymorphy. *Machine Learning*, 343-372.
- Hartigan. (1975). *Clustering Algorithms*. New York: John Wiley & Sons.
- Hartigan, J., & Wong, M. (1978). Algorithm AS139: A k-means clustering algorithm., 28, p. 100-108.
- Hays, C. (2004). *What Wal-Mart Knows About Customers' Habits*. Obtenido de The New York Times: <http://www.nytimes.com/2004/11/14/business/yourmoney/what-walmart-knows-about-customers-habits.html>
- Hinneburg, A., & Keim, D. (1998). An efficient approach to clustering in large multimedia databases with noise. *Knowledge Discovery and Data Mining*, 58-65.
- Holland, J. (1975). *Adaption in Natural and Artificial Systems*. Michigan: Ann Arbor.
- Jaques, N., Conati, C., & Harley, J. (2014). Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System. *Intelligent Tutoring Systems, ITS, 8474*, 29-38.
- Jarusek, P., & Pelanek, R. (2012). Modeling and Predicting Student Problem Solving Times. *SOFSEM 2012: Theory and Practice of Computer Science*, 637-648.
- Karrer, T. (2007). *Understanding E.Learning 2.0*. Obtenido de <https://www.td.org/Publications/Newsletters/Learning-Circuits/Learning-Circuits-Archives/2007/07/Understanding-E-Learning-20>
- Kaufman, J., & Rousseeuw, P. (1990). *Finding groups in data: An introduction to cluster analysis*.
- Kaufman, L., & Rousseeuw, P. (1987). Clustering by means of medoids. *Statistical Data Analysis Based on the L1-Norm and Related Methods*, 405-416.
- Koedinger, K., D'Mello, S., & MCLAughlin, E. (2015). Data mining and education. *Wiley Interdisciplinary Reviews-Cognitive Science*, 333-353.

- Kurzweil, R. (2009). *The intelligent universe*. Obtenido de Soft Technics: <http://softtechnic.blogspot.com/2009/09/intelligent-universe.html>
- Lathan, A., Crockett, K., & Mclean, D. (2013). Profiling Student Learning Styles with Multilayer Perceptron Neural Networks. *IEEE International Conference on Systems, Man, and Cybernetics*, 2510-2515.
- Lichman, M. (2013). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. *University of California, School of Information and Computer Science*. Irvine, CA.
- Long, P., & Siemens, G. (2011). *Penetrating the Fog: Analytics in Learning and Education*. Educause. Obtenido de Educause.
- Markoff, J. (2006). *Entrepreneurs see a web guided by common sense*. Obtenido de The New York Times: http://www.nytimes.com/2006/11/12/business/12web.html?ei=&_r=0
- Marr, B. (2015). *Big Data: 20 Mind-Boggling Facts Everyone Must Read*. Recollit de Forbes: <http://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/>
- Martí, J., Español, J., Golobardes, E., Freixenet, J., García, R., & Salamó, M. (2000). Classification of microcalcifications in digital mammograms using case-based reasoning. *International Workshop on digital Mammography*.
- Martin, J., & VanLehn, K. (1995). Student assessment using Bayesian nets. *International Journal of Human-Computer Studies*(42).
- Masie, E. (2014). *Big Learning Data*. ASTD.
- Mayer, R., & Moreno, R. (2000). A Learner-Centered Approach to Multimedia Explanations: Deriving Instructional Design Principles from Cognitive Theory. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*.
- Merz, C. J., & Murphy, P. M. (1997). *UCI Repository for Machine Learning Databases*. (D. o. University of California, Productor) Obtenido de <http://www.ics.uci.edu/>
- Ministerio de Educación Cultura y Deporte. (2004). *El Sistema Educativo Español*. Madrid.
- Montes, L. (2015). *Más big data en la receta: El médico diagnostica con el 10% de la información*. Obtenido de <http://www.elmundo.es/economia/2015/10/30/56325a6e46163f1f5e8b457b.html>

- Moravec, J. (2009). *Toward Society 3.0: A new paradigm for 21st century education*. Obtenido de <http://www.slideshare.net/moravec/toward-society-30-a-new-paradigm-for-21st-century-education-presentation>
- Myneni, L., Narayanan, N., & Rebello, S. (2013). An Interactive and Intelligent Learning System for Physics Education. *IEEE Transactions on Learning Technologies*, 6(3), 228-239.
- Navarro, J., Canaleta, X., Vernet, D., Solé, X., Jimenez-Ruano, V., & Costa, N. (2014). Motivación, Desmotivación y Daños Colaterales. *Actas de las XX Jornadas de la Enseñanza Universitaria de la Informática (JENUI 2014)*. Oviedo.
- Navarro, J., Sancho-Asensio, A., Zaballo, A., Jiménez-Ruano, V., Vernet, D., & Armendáriz-Iñigo, E. (2014). The Management System of INTEGRIS: Extending the Smart Grid to the Web of Energy. *The 4th International Conference on Cloud Computing and Services Science (CLOSER 2014)*.
- Nesbit, J., Adesope, O., & Liu, Q. (2014). How Effective are Intelligent Tutoring Systems in Computer Science Education? *IEEE International Conference on Advanced Learning Technologies*, 99-103.
- Nicolas, R., Vernet, D., Golobardes, E., Fornells, A., De la Torre, F., & Puig, S. (2009). Applying Distance Metric Learning in a Collaborative Melanoma Diagnosis System with Case-Based Reasoning. *Proceedings of the 14th UK Workshop on Case-Based Reasoning at The 29th SGA International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Cambridge.
- Nicolas, R., Vernet, D., Golobardes, E., Fornells, A., Puig, S., & Malvey, J. (2009). Improving the Combination of CBR Systems with Preprocessing Rules in Melanoma Domain. *Workshop Proceedings of the 8th International Conference on Case-Based Reasoning*. Seattle.
- Niemeijer, K. (2014). *The ABC's of Big Data*. Obtenido de Foresightinvestor: <http://foresightinvestor.com/articles/411823-the-abc-s-of-big-data>
- Nkambou, R., Frasson, C., & Gauthier, G. (2003). CREAM-Tools: An Authoring Environment for Knowledge Engineering in Intelligent Tutoring Systems, in *Authoring Tools for Advanced Technology Learning Environments: Toward cost-effective adaptive, interactive, and intelligent educational software*. Kluwer Publishers, 93-138.
- Nwana, H. (1990). Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, 4(4), 251-277.

- Oloruntoba, S. (2015). *S.O.L.I.D: The First 5 Principles of Object Oriented Design*. Obtenido de Scotch.io: <https://scotch.io/bar-talk/s-o-l-i-d-the-first-five-principles-of-object-oriented-design>
- Olsen, J., Alevan, V., & Rummel, N. (2015). Toward Combining Individual and Collaborative Learning Within and Intelligent Tutoring System. *Artificial Intelligence in Education, AIED, 9112*, 848-851.
- Pelleg, D., & Moore, A. (2000). X-Means: Extending K-Means with Efficient Estimation of The Number of Clusters. *Proceedings of the 17th International Conf. on Machine Learning*, 727-734.
- Quinlan, J. (1986). Induction on decision trees. *Machine Learning*, 81-106.
- Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.
- Ramírez, M. (2009). ¿Qué es realmente un profesor? *Cuadernos de Educación y Desarrollo*, 1(2).
- Reddy, P., & Sasikumar, M. (2014). Student Model for an Intelligent Language Tutoring System. *IEEE International Conference on Advanced Learning Technologies*, 441-443.
- Riesbeck, C. K., & Schank, R. C. (1989). *Inside Case-Based Reasoning*. Hillsdale: Lawrence Erlbaum Associates.
- Romero, C., & Ventura, S. (2010). *Handbook of Educational Data Mining*. CRC Press.
- Salamó, M., Golobardes, E., Vernet, D., & Nieto, M. (2000). Weighting Methods for a Case-Based Classifier System. *Proceedings of the Learning'00*. Madrid.
- Salamó, M., Rodríguez, I., López-Sánchez, M., Puig, A., Balocco, S., Taulé, M., & Rodríguez, J. (2015). GradeForeseer: Recurso docente para la predicción de notas del alumnado de informática. *Actas de las XXI Jornadas de la Enseñanza Universitaria de la Informática*, 89-96.
- Salinas, J. (2005). La gestión de los entornos virtuales de formación. *Seminario Internacional: La Calidad de la Formación en Red en el Espacio Europeo de Educación Superior*. Universitat de les Illes Balears.
- Salomé, I., & Suñé, F. (2011). *La Escuela 2.0 en tus manos. Panorama, instrumentos y propuestas*. Madrid: Ediciones Anaya Multimedia.
- Sánchez, E. (2015). *Big Data al servicio de la educación*. Obtenido de Learning World. Euronews.: <http://es.euronews.com/2015/05/22/big-data-al-servicio-de-la-educacion/>

- Sancho-Asensio, A., Solé, X., Montero, J., Navarro, J., Canaleta, D., & Vernet, D. (2014). Herramienta de Soporte para la Formación de Grupos de Trabajo en Entornos de Aprendizaje Colaborativo. *9a Conferencia Ibérica de Sistemas y Tecnologías de Información*.
- Sangrà, A., & Stephenson, J. (2012). *Modelos pedagógicos y el e-learning. Fundamentos del diseño técnico-pedagógico en e-learning*. Barcelona: Editorial UOC.
- Santamaría, F. (2012). *Learning Analytics – Análisis del aprendizaje*. Obtenido de <http://fernandosantamaria.com/blog/2012/09/learning-analytics-analisis-del-aprendizaje-2/>
- Schank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Self, J. (1990). Theoretical Foundations for Intelligent Tutoring Systems. *Journal of Artificial Intelligence in Education*, 1(4), 3-14.
- Seoane, A., & García, F. (2010). *Introducción al eLearning*. Obtenido de <http://repositorio.grial.eu/bitstream/123456789/27/6/introelearning.pdf>
- Sevindik, T. (2011). Prediction of student academic performance by using an adaptive neuro-fuzzy inference system. *Energy education science ant technology part b-social and educational studies*, 3(4), 635-646.
- Shah, F. (1997). *Recognizing and Responding to Student Plans in an Intelligent Tutoring System: Cirsim-Tutor*. Phd Thesis. Illinois Institute of Technology.
- Sheikholeslami, G., Chatterjee, S., & Zhang, A. (1998). WaveCluster: A multi-resolution clustering approach for very large spatial databases. *Proc. 24th Int. Conf. Very Large Data Bases, VLDB*, 428-439.
- Siemens & Downs. (2004). *Connectivism: A Learning Theory for the Digital Age*. Obtenido de Elearnspace: <http://www.elearnspace.org/Articles/connectivism.htm>
- Siemens, G. (2010). *What Are Learning Analytics?* Obtenido de <http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/>
- Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Buckingham, S., Ferguson, R., . . . Baker, R. (2001). *Open Learning Analytics: an integrated & modularized platform*. SOLAR: Society for Learning Analytics Research.
- Silva, J. (2011). *Diseño y moderación de entornos virtuales de aprendizaje*. Barcelona: Editorial UOC.

- Simon, J. (2007). *Campusvirtual UB: un nou entorn d'ensenyament-aprenentatge*. Obtenido de Quaderns de Docència Universitària, n.º 9: http://diposit.ub.edu/dspace/bitstream/2445/1681/1/QDU_9.pdf
- Smith, A., Min, W., & Mott, B. (2015). Diagrammatic Student Models: Modeling Student Drawing Performance with Deep Learning. *User Modeling, Adaptation and Personalization, 9146*, 216-227.
- Smola, A., & Sch, B. (1998). A tutorial on support vector regression. *NeuroCOLT2 Technical Report NC2-TR-1998-030*.
- Solé, X., Navarro, J., Sancho-Asensio, A., Zaballos, A., Jimenez-Ruano, V., Canaleta, X., & Vernet, D. (2014). Saggiarius: Una Herramienta para Potenciar el Trabajo Colaborativo en Entornos Virtuales de Aprendizaje. *9a Conferencia Ibérica de Sistemas y Tecnologías de Información*.
- Stevens, A., & Collins, A. (1977). The goal structure of a Socratic tutor. *Proceedings of a National ACM Conference*.
- Teten, D. (2007). Web 3.0: Where are we headed? *The Web 2.0 New York Conference*. New York.
- Tiffin, J., & Rajasingham, . L. (1997). *En busca de la clase virtual. La educación en la Sociedad de la Información*. Barcelona: Ediciones Paidós Ibérica.
- Tow, D. (2010). *Future of the web*. Obtenido de Future web: <http://futurewebblog.blogspot.com.es/2010/06/future-of-web.html>
- UNESCO Institute for Information Technologies. (2012). *LEARNING ANALYTICS. IITE Policy Brief*. Unesco.
- Urretavizcaya, M. (2001). Sistemas Inteligentes en el ámbito de la Educación. *Dept. Lenguajes y Sistemas Informáticos. Facultad de Informática*.
- Vallejo, A, Zaballos, A., Canaleta, X., & Dalmau, J. (2008). End-to-end QoS management proposal for the ITU-T IMS/NGN architecture. *Proceedings of SoftCOM 2008*, (págs. 147-151).
- Vallejo, A, Zaballos, A., Vernet, D., Cutiller, D., & Dalmau, J. (2008). Implementation of Traffic Engineering in NGNs using Hybrid Genetic Algorithms. *Proceedings of The Third International Conference on Systems and Networks Communications* (p. 262-267). Sliema: IEEE Xplore.

- Vallejo, A, Zaballos, Vernet, Orriols, & Dalmau. (2009). A Traffic Engineering Proposal for ITU-T NGNs Using Hybrid Genetic Algorithms. *The International Journal on Advances in Internet Technology*, 2(1), 162-172.
- Vanlehn, K. (1988). *Student Modelling. Foundations of Intelligent Tutoring systems*. New Jersey: Lawrence Erlbaum Associates.
- Vernet, D & Canaleta, X. (2004). Sistema Tutor Avanzado. *Actas de las X Jornadas de Enseñanza Universitaria de la Informática (JENUI'2004)*, (pp. 211-215). Alicante.
- Vernet, D & Canaleta, X. (2011). Los Clickers como Soporte en las Aulas para Mayores. *Actas del IV Congreso Iberoamericano de Universidades para Mayores (CIUUMM 2011)*. Alicante.
- Vernet, D & Canaleta, X. (2011). LSMaker: un Proyecto Interdisciplinar. *Actas de las XVII Jornadas de Enseñanza Universitaria de Informática, Jenui 2011*. Sevilla.
- Vernet, D & Golobardes, E. (2003). Adding Unsupervised Learning in the Case-base Organisation. *Actes del 6è Congrés Català d'Intel·ligència Artificial, CCIA 2003.*, 100, p. 307-315. Palma de Mallorca.
- Vernet & Golobardes, E. (2003). An Unsupervised Learning Approach for Case-Based Classifier Systems. *Expert Update, The Specialist Group on Artificial Intelligence*, 6(2), 37-42.
- Vernet, D. (2004). *Diploma d'Estudis Avançats*. Barcelona.
- Vernet, D, & Golobardes, E. (2002). An Unsupervised Learning Approach for Case-Based Classifier Systems. In U. o. B.Lees (Ed.), (pp. 39-46). Cambridge.
- Vernet, D, Canaleta & Zaballos. (2012). Plataforma interdisciplinar de soporte para prácticas. *Actas de las XVIII Jornadas de la Enseñanza Universitaria de la Informática*. Ciudad Real.
- Vernet, D, Canaleta, X, & Pallàs, G. (2012). Tutorización Inteligente de Comunidades Virtuales de Aprendizaje. *Actas del XIV Simposio Internacional de Informática Educativa*. Andorra La Vella.
- Vernet, D, Canaleta, Pallàs, Navarro, & Zaballos. (2013). Setting up and Tutoring the Working Groups of a Virtual Learning Community. *Journal of Research and Practice in Information Technology*, 45(3), 91-107.

- Vernet, D., & Canaleta, X. (2004). La Biometría y su Legalidad. *Actas de las V Jornadas de Informática y Sociedad (JIS'2004)*, (pp. 475-480). Bilbao.
- Vernet, D., Salamó, M., Vallespi, C., Camps, J., Golobardes, E., & Bacardit, J. (2001). ¿Cómo predecir la evolución del alumno? *Actas de las VII Jornadas de la Enseñanza Universitaria de la Informática, JENUI 2001. 1*, pp. 335-340. Palma de Mallorca: Universitat de les Illes Balears. Servei de Publicacions i Intercanvi Científic.
- Vernet, D, Nicolas, Golobardes, Fornells, & Garcia-Piquer. (2010). Work in Progress - Intelligent Tutoring System Framework for the Acquisition of Knowledge and Competences. *39th ASEE/IEEE Frontiers in Education Conference*. Washington.
- Vernet, D, Nicolas, R., Golobardes, E., Fornells, A., Garriga, C., Puig, S., & Malvehy, J. (2008). Pattern Discovery in Melanoma Domain Using Partitional Clustering. *Congrés Català d'Intel·ligència Artificial, CCIA. Frontiers in Artificial Intelligence and Applications.*, (pp. 323-330). Sant Martí d'Empúries.
- Vernet, D, Zaballos, Martin de Pozuelo, & Caballero. (2015). High Performance Web-of-Things Architecture for the Smart Grid Domain. *International Journal of Distributed Sensor Networks*, In press.
- Villagrà-Arnedo, C., Gallego-Durán, J., Molina-Carmona, R., & Llorens-Largo, F. (2015). Sistema de predicción para la Asistencia en el seguimiento del aprendizaje. *Actas de las XXI Jornadas de la Enseñanza Universitaria de la Informática*, 74-80.
- Wang, W., Yang, J., & Muntz, R. (1997). STING: A statistical information grid approach to spatial data mining. *VLDB Journal*, 186-195.
- Witten, I., & Frank, E. (2000). DataMining: practical machine learning tools and techniques with Java implementations. *Morgan Kaufmann Publishers*.
- Xiong, R., & Donath, J. (1999). *PeopleGraden: Creating data portraits for users*. MIT Media Laboratory.
- Zaballos, A., Vernet, D., & Selga, J. (2013). A Genetic QoS-Aware Rountig Protocol for the Smart Electricity Networks. *International Journal of Distributed Sensor Networks*, 2013(135056), 1-12.

Apèndix A: Índex de figures

Figura 1. Mòduls que formen un Sistema Tutor Intel·ligent clàssic.	5
Figura 2. Cicle de vida del Raonament Basat en Casos.	7
Figura 3. Cicle de metodologia de recerca proposat per Ferrance.	12
Figura 4. Tipologia de les tecnologies en la docència.	21
Figura 5. Tecnologia transmissiva. El model pedagògic està centrat en el professor: és aquest qui manipula la tecnologia educativa i transmet el coneixement a l'alumne com si fos un instructor.	21
Figura 6. Metodologia interactiva. L'activitat està centrada en l'alumne: Mitjançant la interacció amb la tecnologia educativa i amb l'orientació del docent, l'alumne crea el seu propi coneixement.	22
Figura 7. Tecnologia col·laborativa. L'activitat d'aprenentatge se centra en el grup d'alumnes o comunitat. Aquests, de manera col·laborativa interaccionen amb la tecnologia educativa i construeixen un coneixement grupal. El docent actua com a orientador o facilitador.	22
Figura 8. Representació de l'aparició dels diferents llenguatges i plataformes de programació, noves tecnologies i el moment associat d'evolució de la Web.	28
Figura 9. Cicle per a les tecnologies emergents (Gartner, 2014)	38
Figura 10. Dimensions de la <i>Big Learning Data</i> .	39
Figura 11. Cadena de valor de la <i>Big Learning Data</i> .	40
Figura 12. Procés de l'analítica de l'aprenentatge.	43
Figura 13. Model de referència de l'analítica de l'aprenentatge. Adaptat de (Chatti i altres, 2012).	44
Figura 14. Mètodes aplicats a la LA en funció del nombre de publicacions científiques consultades.	48
Figura 15. Representació visual de la participació en un fòrum (Xiong & Donath, 1999). A l'esquerra s'observa un grup amb una veu dominant, amb un gran nombre de respostes en blau. A la dreta un grup més democràtic, amb un gran nombre de posts inicials en magenta. L'alçada de les flors denota quant de temps ha estat l'usuari al tauler.	49
Figura 16. Diagrama SNA que mostra les connexions entre alumnes dins d'un entorn virtual (Dawson, 2010).	49
Figura 17. Mòduls que formen un Sistema Tutor Intel·ligent i les interaccions que s'hi donen.	55

Figura 18. Superposicions del model d'estudiant	56
Figura 19. Model bàsic de xarxa bayesiana que permet fer un seguiment de l'adquisició de coneixement.	57
Figura 20. Tècniques d'adaptació emprades en els AIWBES procedents de tecnologies clàssiques (STI i HA). Diagrama realitzat en base a (Brusilovsky & Peylo, 2003) i (González, 2004)	61
Figura 21. Els cinc grups de tecnologies AIWBES modernes i les seves tècniques principals. Diagrama elaborat a partir de (Brusilovsky & Peylo, 2003).	63
Figura 22. Conceptualització de l'arquitectura TICVA.	67
Figura 23. Model conceptual del TICVA que respon a les quatre dimensions crítiques plantejades per Chatti.	69
Figura 24. Arquitectura bàsica del sistema TICVA dissenyat.	73
Figura 25. Arquitectura completa del sistema TICVA dissenyat.	74
Figura 26. Arquitectura detallada del Mòdul de Comunicació dissenyat.	75
Figura 27. Arbre de coneixement clàssic, pre-Bolonya (A = Assignatura, U=Unitat o Tema, A=Activitat, RA = Resultat d'Aprenentatge).	76
Figura 28. Nova representació del coneixement després dels canvis introduïts pel model de Bolonya.	76
Figura 29. Arquitectura detallada del Mòdul del Coneixement del Domini dissenyat.	77
Figura 30. Arquitectura detallada del Mòdul Expert dissenyat.	78
Figura 31. Arquitectura detallada del Mòdul <i>Filtered Reporting</i> dissenyat.	79
Figura 32. Arquitectura detallada del Mòdul de l'Estudiant/Grup.	81
Figura 33. Arquitectura detallada del Mòdul Pedagògic.	83
Figura 34. Exemple del resultat d'una classificació utilitzant l'algorisme ID3	90
Figura 35. Representació gràfica del conjunt de dades provinents d'un entorn educatiu.	92
Figura 36. Imatge digital d'una mamografia i la seva segmentació posterior.	92
Figura 37. Tipologia de xarxa exemple utilitzada per optimitzar l'enrutament de paquets en el problema relacionat amb l'entorn telemàtic.	93
Figura 38. Cicle bàsic del Raonament Basat en Casos proposat per Aamodt i Plaza	95

Figura 39. Tendències detectades en les dades d'ensenyament eliminant completament aquelles dades que no depenen de l'esforç de l'alumne. Per cada regla obtinguda s'indica el seu suport i la seva confiança.	106
Figura 40. Correlacions existents entre les dades corresponents a l'examen 1 i a l'examen 2 del conjunt de dades d'ensenyament.	107
Figura 41. Retall del diari sud-americà Clarín on es fa ressò dels resultats obtinguts pel sistema de predicció implementat.	108
Figura 42. Exemple de diagrama de Voronoi	110
Figura 43. Exemple de funcionament de l'algorisme <i>K-Means</i> . Els punts marcats amb una X corresponen als centroides de cadascun dels <i>clusters</i> . Imatge en línia: Learn by Marketing.	111
Figura 44. Exemplificació gràfica dels valors de cohesió interna i externa entre els diferents <i>clusters</i> .	115
Figura 45. Esquema de com es realitza el primer nivell de <i>clustering</i> i el seu refinament posterior en cas de detectar subagrupacions.	118
Figura 46. Representació gràfica del coneixement que posseeixen els membres d'un grup creat de manera automàtica.	119
Figura 47. Esquema general del sistema de <i>clustering</i> a dos nivells.	120
Figura 48. Exemple de la informació existent en un cas format per l'agrupació de 3 alumnes.	121
Figura 49. Cicle de vida d'un Algorisme Genètic.	124
Figura 50. Disminució del percentatge d'efectivitat a causa de l'increment d'agrupacions realitzat sobre les dades de l'entorn educatiu.	130
Figura 51. Efectivitat de la funció de similitud de Recompte de Valors en funció del llindar discriminant utilitzat per aquest càlcul.	131

Apèndix B: Índex de taules

Taula 1. Publicacions relacionades amb la tesi classificades segons el tipus de publicació i l'objectiu associat	13
Taula 2. Evolució de l' <i>e-learning</i> des dels seus orígens fins a l'actualitat. Taula desenvolupada a partir de les aportacions de (Gros, 2011) i (Seoane & García, 2010).	27
Taula 3. Evolució de l' <i>e-learning</i> . Taula construïda a partir de (Gros, 2011) (Seoane & García, 2010) (Moravec, 2009) i (Teten, 2007).	30
Taula 4. Generacions d' <i>e-learning</i> vs exemples tecnològics (Costa, Alves, & Rodrigues, 2013)	32
Taula 5. Analítica acadèmica i analítica de l'aprenentatge. Adaptat de (Siemens, et al., 2001).	41
Taula 6. Paradigmes principals d'Intel·ligència Artificial aplicada a l'educació (Brusilovsky & Peylo, 2003).	53
Taula 7. Conjunts de dades de l'UCI repositori més utilitzats per a la validació dels algorismes. Per cadascun d'ells, s'indica la seva etiqueta abreujada, el nombre d'instàncies, el nombre d'atributs per instància, quants atributs són nominals i quants numèrics i el nombre de classes de classificació.	93
Taula 8. Conjunt de característiques mínimes escollides per tal de realitzar la predicció de l'evolució d'un alumne.	103
Taula 9. Eficàcia mitjana, màxima i mínima assolida amb les dades disponibles, en la predicció a 5 classes, tenint en compte les diferents ponderacions aplicades als atributs.	105
Taula 10. Ponderacions resultants obtingudes usant Algorismes Genètics i informació d'un professor expert.	105
Taula 11. Eficàcia mitjana, màxima i mínima assolida amb les dades disponibles, tenint en compte la ponderació proposada pels Algorismes Genètics i classificant a l'alumne en només dues classes (aprovat o suspès), emprant <i>10-fold crossvalidation</i> .	107
Taula 12. Exemple de taula de contingència per a un <i>cluster</i> de 4 alumnes.	113
Taula 13. Valors definits per als paràmetres de configuració de l'algorisme WITT	116
Taula 14. Resultats obtinguts en l'aplicació de les polítiques de primer nivell de <i>clustering</i> a diversos jocs de proves. Per cada joc de proves es mostra el percentatge d'efectivitat, el temps de CPU en milisegons i la desviació estàndard, en cada política analitzada. Es marquen amb un ✓ els resultats significativament millors.	127

Taula 15. Resultats obtinguts per comprovar la qualitat del *clustering* realitzat en el primer nivell. Per cada combinació d'algorismes es mostra el percentatge d'eficàcia, el temps de resolució en milisegons i la desviació estàndard. 128

Taula 16. Matriu d'encerts entre la classificació proposada per l'algorisme de primer nivell i la classe real a la que pertanyia l'alumne. 129

Taula 17. Nombre d'agrupacions proposat per cada classe obtinguda per la política MKM Clustering sense cap restricció de màxims als diferents jocs de dades exposats a l'apartat 6.3. 130

Taula 18. Resultats obtinguts en les diferents funcions de similitud grupals en funció de si els grups establerts són de caire homogeni o heterogeni. Els resultats estan expressats en percentatge d'encerts sobre la predicció a 2 classes. 131

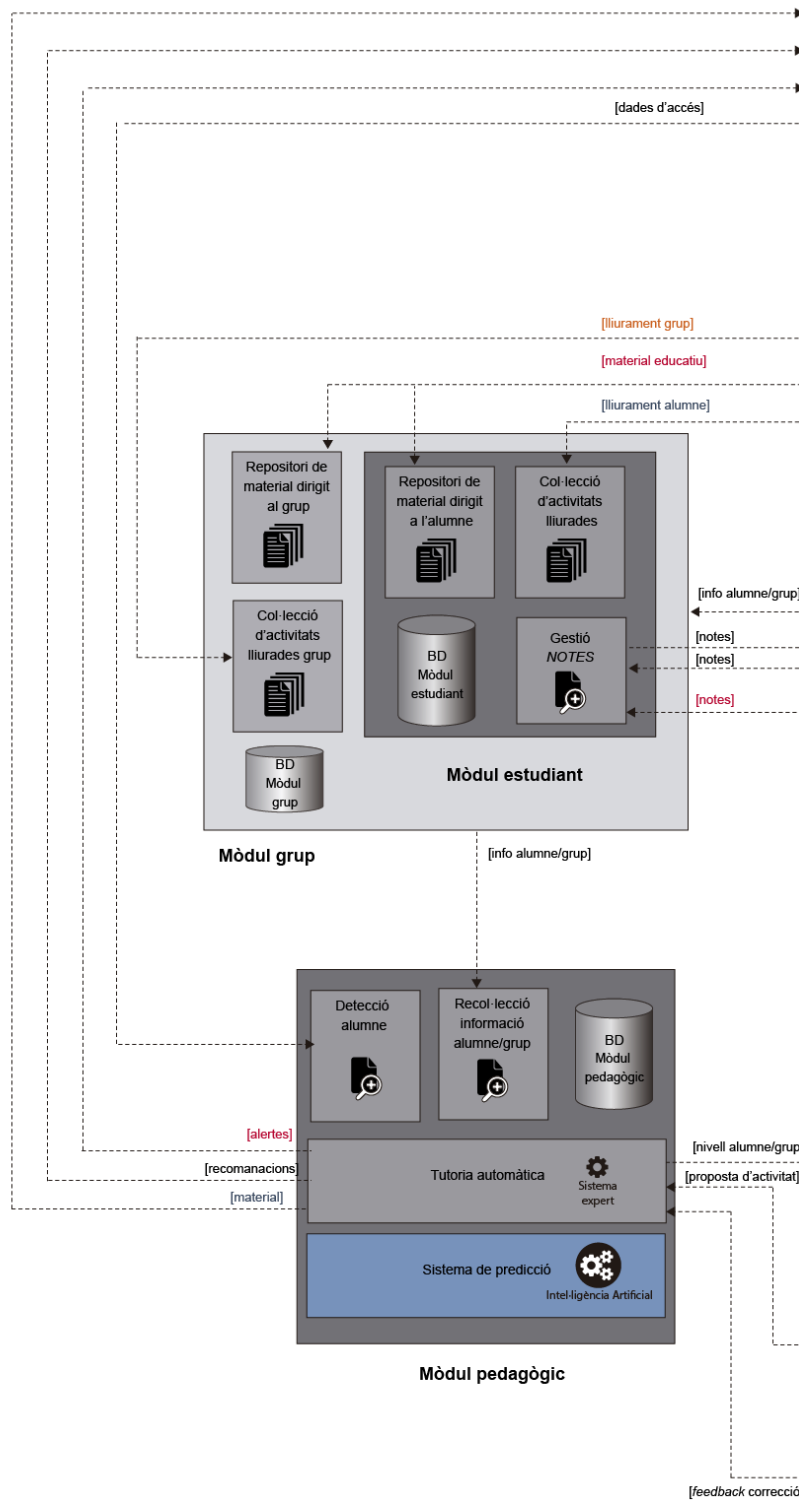
Taula 19. Percentatges d'efectivitat (predicció a 2 classes) i d'increment de la Memòria de Casos en funció de la Política d'Emmagatzematge aplicada: Política Ultra Conservadora (PUC), Política Conservadora (PC), Política Flexible (PF) 132

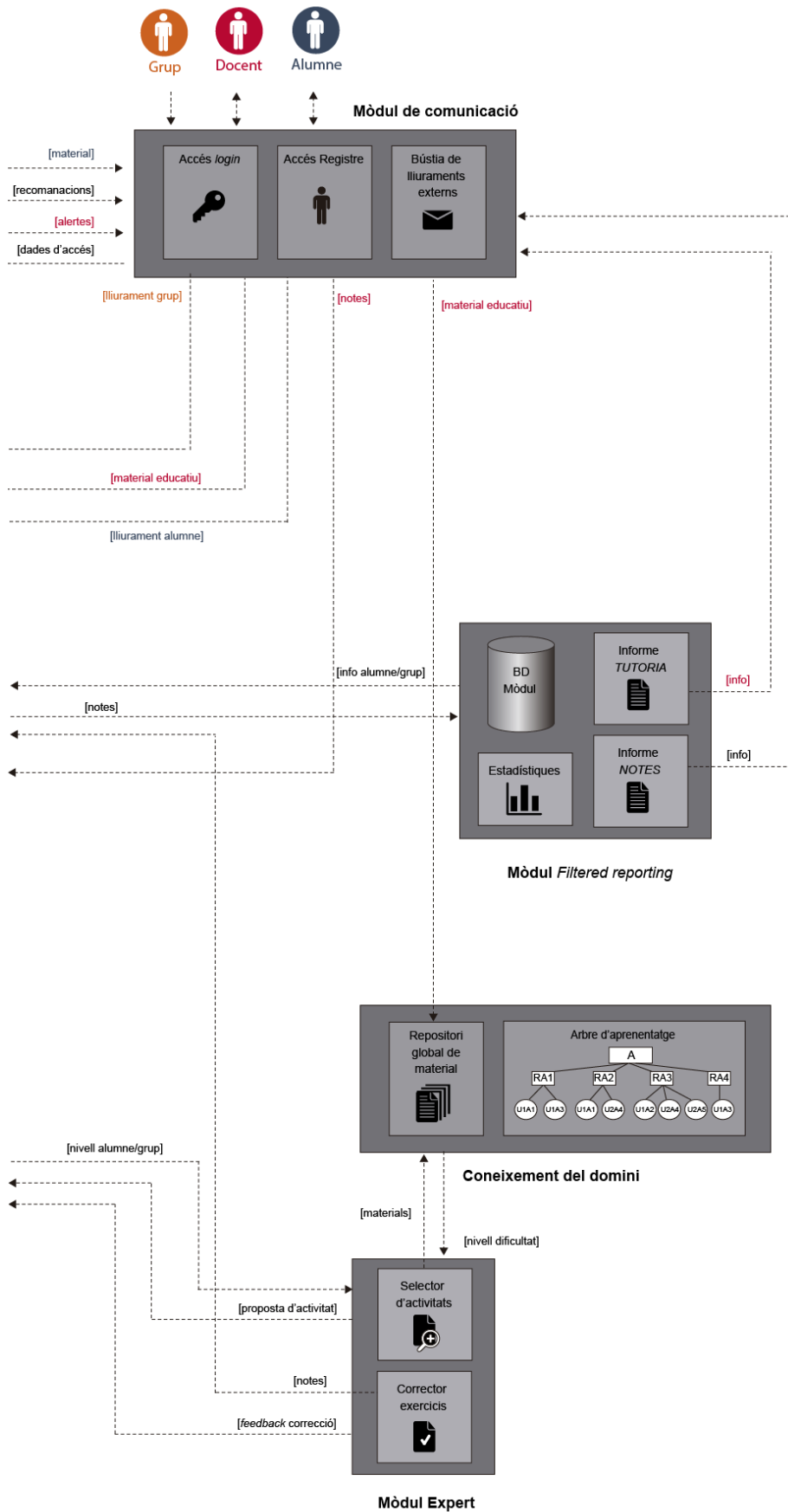
Apèndix C: Acrònims utilitzats

AA	Analítica Acadèmica
AHS/HA	Sistema Adaptatiu Hipermedia
AIWBES	Sistema Educatiu Intel·ligent i Adaptatiu Basat en Web
ARFF	Format de Fitxer Atribut-Relació
CAI	Instrucció Assistida Per Ordinador
CBR	Raonament Basat en Casos
CEO	Chief Executive Officer
CMS	Sistema de Gestió de Continguts
CSCL	Aprentatge Col·laboratiu Suportat per Ordinador
DM	Mineria de Dades
EAO	Ensenyament Assistit per Ordinador
EVA	Entorn Virtual d'Aprentatge
GA	Algorismes Genètics
GRETEL	Grup de Recerca en Technology Enhanced Learning
GRITS	Grup de Recerca en Internet Technologies and Storage
GRSI	Grup de Recerca en Sistemes Intel·ligents
IA	Intel·ligència Artificial
ICAI	Instrucció Intel·ligent Assistida per Ordinador
LA	Analítica de l'Aprentatge
LCMS	Sistema de Gestió de Continguts d'Aprentatge
LCS	Sistema de Co-Aprentatge

LMS	Sistema de Gestió de l'Aprenentatge
LOMCE	Llei Orgànica per a la Millora de la Qualitat Educativa
MKM	Mean K-Means
MOP	Paquet d'Organització de Memòria
NN	Veí Més Proper
PC	Política Conservadora
PF	Política Flexible
PLE	Entorns d'Aprenentatge Personal
PUC	Política Ultra Conservadora
SEI	Sistema Educatiu Intel·ligent
SNA	Anàlisi Social de l'Aprenentatge
STI/ITS	Sistema Tutor Intel·ligent
TIC/ICT	Tecnologies de la Informació i Comunicació
TICVA	Tutorització Intel·ligent de Comunitats Virtuals d'Aprenentatge
WEKA	Entorn Waikato d'Anàlisi de Coneixement

Apèndix D: Detall de l'arquitectura TICVA





Annex: Publicacions més rellevants

A continuació s'adjunten les publicacions més rellevants que l'autor vol destacar, corresponents a les següents referències:

Vernet, D., Canaleta, X., Pallàs, G., Navarro, J., & Zaballos, A. (2013). Setting up and Tutoring the Working Groups of a Virtual Learning Community. *Journal of Research and Practice in Information Technology*, 45(3), 91-107.

Vernet, D., Zaballos, Martin de Pozuelo, & Caballero. (2015). High Performance Web-of-Things Architecture for the Smart Grid Domain. *International Journal of Distributed Sensor Networks*, In press.

Canaleta, X., **Vernet, D.**, Vicent, L., & Montero, J. (2014). Master in Teacher Training; a Real Implementation of Active Learning. *Computers in Human Behaviour*(31), 651-658.

Zaballos, A., **Vernet, D.**, & Selga, J. (2013). A Genetic QoS-Aware Routing Protocol for the Smart Electricity Networks. *International Journal of Distributed Sensor Networks*, 2013(135056), 1-12.

Vernet, D., Nicolas, R., Golobardes, E., Fornells, A., & Garcia-Piquer, A. (2010). Work in Progress - Intelligent Tutoring System Framework for the Acquisition of Knowledge and Competences. *39th ASEE/IEEE Frontiers in Education Conference*. Washington.

Vernet, D.; Golobardes, E. (2003). An Unsupervised Learning Approach for Case-Based Classifier Systems. *Expert Update, The Specialist Group on Artificial Intelligence*, 6(2), 37-42.

Setting up and Tutoring the Working Groups of a Virtual Learning Community

David Vernet, Xavi Canaleta, Joan Navarro and Agustin Zaballos

La Salle, Universitat Ramon Llull – c/ Quatre Camins 30, Barcelona, Spain

Emails: {dave, xavic, jnavarro, zaballos}@salle.url.edu

Gemma Pallàs

Universitat Oberta de Catalunya, Barcelona, Spain

Email: gemmaps@uoc.edu

Collaborative work has emerged as a hot research topic in Virtual Learning Communities since it may considerably improve the knowledge quality and experience of students. However, this novel approach makes the assessment process challenging (i.e., interactions between virtual students, their achievements, and their profiles have to be properly addressed). The purpose of this paper is to propose a comprehensive Intelligent Tutoring System for Virtual Learning Communities that relies on artificial intelligence techniques which are able to manage the specificities of the collaborative working groups that arise in this domain. These specificities can be summarized in the following four goals: 1) conduct an individualized tracking of every student upon the collected data from his/her profile and daily work, 2) configure the classroom to maximize the performance of all its members, 3) automatically obtain the teacher's feedback about the class operation and possible anomalies, and 4) monitor the working groups behaviour and achievements automatically to redirect their operation when necessary. The framework of the proposed system is described, a proof of concept is presented, and a new virtual student profile, named as bystander, is identified in preliminary experimentations.

Keywords: Intelligent tutoring systems; collaborative learning; virtual learning communities; Web 2.0; constructivism; artificial intelligence; bystanders

ACM Classifications: K.3.1 – Computer Uses in Education; K.3.2 – Computer and Information Science Education; I.2.6 – Learning

1. Introduction

Over recent years, learning and knowledge acquisition techniques have experienced considerable progress and development. Such advances have been mainly fostered by the constant incorporation of several theories which have emerged in the domain of Psychology and which tend to be targeted at the portrayal and analysis of the mental phenomena involved in the learning process. Therefore, current learning theories in Pedagogy are based on conceptual systems that describe how individual brains acquire knowledge.

As far as the learning theory is concerned, two major thinking trends can be distinguished in the Pedagogy field: instructivism and constructivism. Although both paradigms refer to the degree of involvement of the formative influence agent during the educational process and assess up to

Copyright© 2013, Australian Computer Society Inc. General permission to republish, but not for profit, all or part of this material is granted, provided that the JRPIT copyright notice is given and that reference is made to the publication, to its date of issue, and to the fact that reprinting privileges were granted by permission of the Australian Computer Society Inc.

Manuscript received: 19 January 2013

Communicating Editor: Francisco José García Peñalvo

what extent students have the flexibility to exploit their own resources in order to achieve specific learning aims, there are important differences between them that should be noted.

Instructivism leaves the learning objectives and definition of strategies exclusively to the teacher (also referred to as educative entity). Therefore, instructivist theories foster instructor agents to predefine the student's learning goals, methodologies, and pace. Hence, this approach, which entails a unique teaching pattern for all students, does not consider their individual specificities. This situation might create a significant issue in the development of the learning process that should be addressed properly.

From the instructivist point of view, knowledge exists independently from the student. According to this approach, the main goal of students is to passively accept the information transmitted by the teacher, which is known as the Teacher-centred model. Contrary to this perception of education, there is another trend which confirms that there are students who are capable of building their own knowledge on their own. Actually, this hypothesis is the backbone of what is known as the constructivist learning approach, and leads to what has been coined as the Student-centred model.

Indeed, constructivist theories focus on the individual context of every student and establish a conceptual framework of the individual as the starting point in the learning process. Thus, according to this idea, educative agents advise students during their learning process although they do not exclusively and rigidly define their specific progression. This advisory task aims to assist students by emphasizing personal experiences, managing the capabilities developed up to date, and providing them with the necessary tools to achieve the proposed learning outcomes. Thereby, with the utilization of this novel paradigm – aimed at fostering the student's personal abilities – the educator becomes a facilitator (Stephenson and Sangrà, 2012) rather than an instructor.

Hence, the constructivist natural learning process consists of inviting individuals to analyze a new environment or landscape, and later requesting them to compare their new findings with previously gained knowledge and experiences. In this way, with the proper guidance provided by the teacher, the students are able to develop a unique and personalized knowledge model, potentially different from the one built by their peers, and moulded according to their individual competencies.

The purpose of this paper is to make headway on constructivist techniques applied in the educational field. More specifically, it presents a novel technological framework that integrates an intelligent tutoring system with a collaborative learning environment. Preliminary obtained results with this approach have permitted the early identification of a new conflictive student profile. The consequent rapid correction of these situations has encouraged practitioners to work in this direction.

The remainder of this paper is organized as follows. Section 2 stresses the advantages of applying technology to the educational domain and it also highlights the major lessons extracted from existing collaborative learning environments. Section 3 reviews the key concepts of tutoring systems, and Section 4 presents the proposed system. Then, Section 5 details some experiences collected on its implementation on a pilot course. Finally, Section 6 ends the work with some conclusions.

2. ICT and Collaborative Learning Environments

In recent years, society has incorporated a considerable number of new technologies that have provided new forms of interactions and development. This has become particularly relevant in the educational field, where the introduction of Information and Communication Technologies (ICTs)

has leveraged a constructivist view of education and fostered the idea of collaborative learning. This section reviews the influence of ICTs in education and how their application has led to virtual learning communities. In addition, the main characteristics that these virtual learning environments must have in order to successfully carry out collaborative work are detailed.

2.1 ICT in Education

In the last decades of the twentieth century, Information and Communication Technologies (ICT) were included in every area of society. Along with the phenomena of globalization, the incorporation of ICTs has led to a global reshape of the educational system and, thus, to the concept of education.

While in the past, education was conceived as an act of instruction, aimed at homogeneously training individuals according to a specific set of cultural values, nowadays this idea has evolved considerably. Certainly, the continuous introduction of ICTs in this field has aligned the educational paradigm with the society in which it is integrated. Hence, these technological tools have assisted, and actually facilitated, the introduction of useful concepts, such as experimentation or socialization, in classrooms.

However, despite the several relevant advances that have emerged so far, such developments have been mainly focused on improving the teaching materials rather than enhancing the teaching model itself. In fact, a wide variety of ICT-based tools have been proposed in the literature, which have been integrated in instructivist educational systems. However, most of them fail to fully exploit their potential since they find themselves strictly constrained by the educational system model. Recent contributions on this area (Vernet *et al*, 2012), envisage that the main reason for such issue relies on the fact that they are considered as mere training supplements instead of useful paths to acquire new knowledge.

In addition, from the educational model point of view, it is worth mentioning the rising interest in the application of the constructivist theories in educative domains. Nevertheless, there is still a latent difficulty in adapting both methodologies and trainers to these novel ideas, which makes their implementation difficult.

Certainly, the instructivist approach has been broadly deployed in the educational field so far. Actually, there are very few organizations that still support constructivist theories nowadays. Accordingly, most of the tools provided by Knowledge and Information Technologies applied to the educational field are targeted at emulating the instructivist learning approach by modeling the traditional in-person teaching model. Therefore, most of the cutting-edge technological tools developed in recent years, such as tutoring systems, literally translate the traditional instructivist educative model and ignore the appealing features that these new technologies may bring to the constructivist field which would enable it to take advantage of such advances.

Certainly, the integration of ICT in the educational domain forces practitioners to reformulate the concept of learning and its execution. Indeed, these new technological tools based on ICTs that support and enrich the learning process have opened a new way to conceive the learning process referred to as e-learning, which is different from the traditional learning approach. As García Peñalvo (2008) defined, e-learning is a process aimed to ensure that a set of competences are obtained by students using today's technologies, specially the web-based ones. In addition to technology, other facets have to be properly addressed in this new educational model in order to guarantee its success. These include structured contents, an assessment method able to cope with

the specificities of this new learning paradigm and an effective collaborative environment to offer appropriate services.

In fact, e-learning is also in line with constructivist theories that suggest the creation of learning groups in order to achieve a comprehensive learning experience. Thanks to the ICT-based tools encompassed by e-learning, traditional group learning strategies have found fundamental support for their implementation and development. These have been coined as Collaborative Learning Environments. Considering their importance in the current context, their main characteristics are detailed in what follows.

2.2 Collaborative Learning Environments

From a sociocultural perspective, the emergence of ICT in education has led to a reformulation of the learning concept. In fact, today's society is based on the interaction between individuals, which results in an interleaved set of knowledge networks (Martín-Moreno, 2004). In the same way, educational communities have integrated the concept of social networks in their methodologies, which has led to the concept of collaborative learning.

In recent years, internet has become a platform for mediation (Sotomayor, 2010) where geographically distanced people share and exchange knowledge. Similarly, existing learning environments have integrated some tools commonly found on the Web 2.0, such as social networks or negotiation and debate instruments into their frameworks. These tools facilitate the conception of a new form of social network, coined as alumni network, as a collaborative learning environment.

Virtual learning communities base their operations on the principle of collaboration – which should be distinguished from cooperation. Although the terms collaboration and cooperation are often randomly exchanged to designate the type of interaction, there are several authors (Dillenbourg, 1999; Panitz, 2012) that clearly differentiate these two concepts in the sense that they differ in regards to how the task is being carried out. In fact, although both approaches agree with two or more individuals solving an activity together, cooperation involves a division, a distribution, and a subsequent assembly of content portions, while collaboration entails a joint construction of the final content, which forces members to trade information and thus requires specific platforms for discussion.

Generally speaking, a collaborative learning environment can best be seen as a virtual space where a group of individuals interact and develop a joint activity. Typically, this interaction is conducted through two modules – considering the nature of the support (i.e., internet) more modules might be developed. These modules are devoted to performing specific functions that provide the collaborative learning environment with enhanced tools to achieve its goal. The two main modules are detailed below and depicted in Figure 1.

2.2.1 Working Platform

This is a mandatory module that must exist in any collaborative environment. This platform enables students to interact with each other and build their own knowledge upon what they are currently learning. There are several virtual working platforms in use, such as the so-called Learning Management Systems (also referred to as LMS).

On the contrary, other working platforms are used as data repositories, which enable students to have a place to post their progress and share it with the rest of the group. These tools range from Portfolio systems (e.g., Mahara) to Version Control Systems (also referred to as VCS) which are used when a group is involved in a progressive project.

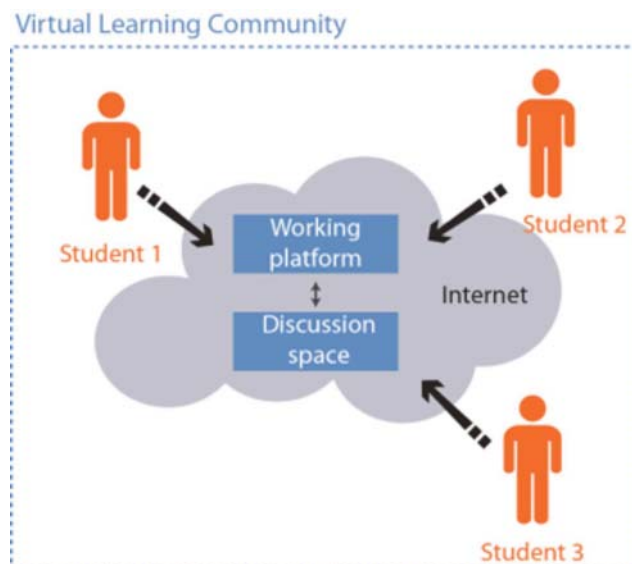


Figure 1: Group work through a collaborative system representation

Note that in order to take full advantage of the versatility featured by these working platforms, an advanced proficiency and a deep knowledge of them is required. In general terms, this issue does not entail an additional difficulty in the learning process since (1) most of the students already have them or (2) can rapidly catch up with them thanks to their predisposition in the digital environment.

2.2.2 Discussion Space

The existence of a discussion space in a collaborative learning environment is essential: this module enables students belonging to the same group to exchange their views and ideas. Fortunately, the aforesaid Web 2.0 provides several solutions to achieve such commitment: social networks (e.g., Facebook, Twitter, Google+), electronic mail, or even some mechanisms which have already been implemented in the working platform.

It has been previously shown that the usage of collaborative learning communities entails a significant advance towards the constructivist paradigm, since the student gains a better control of the learning process.

3. Tutoring Systems

Recent progress on the integration of ICT tools in the educative field has eased the development of what have been coined as tutoring systems. As a part of an ever changing and evolving technology, this kind of student-aid instrument has benefited from the latest improvements in this area, such as the application of artificial intelligence (Pao-Hua *et al*, 2009). As detailed below, the provision of certain capabilities to tutoring systems has led to qualitative improvements. In addition, it can be noted that such enhancements have led to an opening of the traditional educative method, thus enabling a teaching model conception which is more consistent with the current social moment.

This section reviews such advances by (1) describing the concept of traditional tutoring systems and (2) analyzing the main features of intelligent tutoring systems.

3.1 Traditional Tutoring Systems

Traditional tutoring systems are aimed at helping students during the learning process. This is achieved by the educative agent giving them a structured and predefined guidance on the knowledge acquisition phase. In fact, it is commonly considered that ICT-based tools applied to the educational area are merely complementary instruments of the master class, which is generally carried out on a face-to-face basis. Therefore, these tools are generally considered as a small portion of this technology.

In fact, migrating instructivist procedures into the technology-based educational tool spectrum implies a limitation of the capabilities of the resulting system: the potential provided by these tools remains unexplored because of the excessive rigidity of these traditional education methods. Hence, common teaching practices, such as sequentially delivering rigid contents following a predefined order – determined by either the designer or the lack of personalized interaction – drive most of the ICT tools applied to the educational area to behave as secondary teaching complements that poorly help the learning process (Pérez *et al*, 2001).

3.2 Intelligent Tutoring Systems

In the specific field of tutoring systems, the conversion of the instructivist paradigm to that of constructivism has been achieved mostly thanks to the application of artificial intelligence techniques (Hung *et al*, 2003). This has led to the emergence of a novel approach in education that has been recently coined as intelligent tutoring systems. These ICT-based systems are targeted to guide every student throughout their learning process, as a physical tutor would do (Brusilovsky and Peylo, 2003), but accurately adapting themselves to the specific characteristics of every individual. Indeed, these techniques have provided a significant degree of personalization to the tutoring systems, permitting a better adaptability in terms of behaviour and supplying them with enough capabilities to assess the specific characteristics of the student they are interacting with.

Typically, these systems are composed of five modules (Self, 1990), whose interactions are depicted in Figure 2. Although the most interesting modules are the first two, every module is described in what follows:

1. Student module. It is devoted to collecting all the historic data concerning the student's grades and characteristics. In this module all the information related to the person, referring to either academic outcomes or aptitude/competence assessment, is stored. The more data the system has on the student the better it may perform.
2. Pedagogical module. It is able to forecast the student's evolution and infer a possible set of alternatives to compensate for the deficiencies observed. Actually, significant improvements have been made towards this goal in previous works (Vernet *et al*, 2001; Vernet *et al*, 2010; Golobardes *et al*, 2000) using artificial intelligence.
3. Communication module. It is aimed at connecting the student, the teacher, and the system through a user interface.
4. Domain knowledge module. It is targeted at collecting all the information and available resources concerning the subject or course to be taught (e.g., syllabus, assignments, exercises, schemas, diagrams, exams). The storage technique used to efficiently manage all this information has been discussed already (Vernet *et al*, 2010).

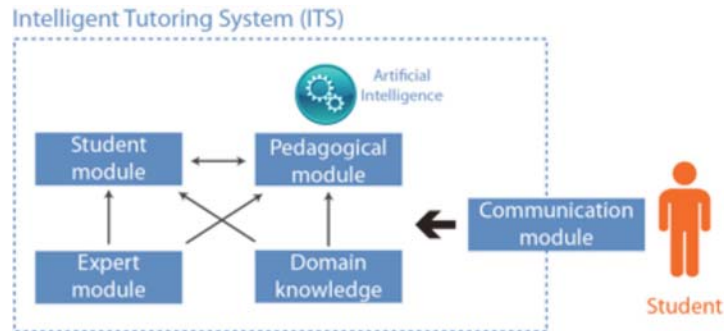


Figure 2: Traditional modules of a tutoring system

5. Expert module. It is able to process the answers that every student has delivered to the system and verify whether they are correct or wrong.

Certainly, intelligent tutoring systems are amazingly effective tools to perform personalized student education since they are based on the continuous analysis of student evolution and thus can select the learning methodology that best fits the learner. Although the aforementioned adaptability of these systems plays a key role from the constructivism point of view, they fail to obtain the maximum performance of this paradigm. Specifically, according to the current social moment, they are unable to equally integrate the student's abilities and the learning outcomes that he or she could reach if interaction with other individuals was considered (Johnson and Johnson, 1986), which is actually one of the goals of the constructivist paradigm (Wilson, 1998). Hence, if intelligent tutoring systems symbolize highly personalized technological teaching tools – and thus individualized – then, collaborative learning environments promote a new learning model based on a system of sharing and exchanging.

4. An ITVLC System

It is quite clear that the interaction with other like-minded individuals envisages a great chance to build specific knowledge networks. However, some challenging questions arise from this statement: How are the networks organized so that the knowledge gained is optimal? How does an individual select a given set of partners from the virtual community in order to efficiently improve their knowledge and/or skills? Can artificial intelligence somehow help to determine this choice?

These questions have been behind the need to converge the two most relevant educational technologies proposed so far: intelligent tutoring and collaborative learning environments. This union might provide virtual communities with specific tools to improve their performance and orientation. Therefore, the aforesaid intelligent tutoring systems are no longer confined to overseeing and guiding the education of a single student, but extend their domain to all the individuals of a community and the relationships established between them.

The purpose of this section is to present a novel intelligent tutoring system of virtual learning communities to exploit this idea. First, a system description is provided and later, its technical features are analyzed.

4.1 Challenges

The intrinsic gap between teachers and students found in virtual learning communities opens new

challenges in terms of learning monitoring. Despite the latest efforts in this field to include new technologies to reduce this gap, they have not generally led to an automated management of virtual classrooms.

Hence, the success of this type of virtual learning relies not only on obtaining the maximum amount of information about the classroom, but on effectively processing this vast amount of data as well. This information becomes especially relevant when 1) configuring the working groups and 2) conducting corrective measures when faults are detected.

The system proposed herein is designed to cope with the following needs which have been broadly detected in the e-learning domain:

1. Students' intelligent monitoring.
2. Classroom's intelligent tracking.
3. Teacher's feedback inclusion.
4. Working groups' optimal configuration to make their control and supervision possible.

In what follows, each one of these challenges is elaborated and different techniques are proposed to face them, which may result in future lines of work.

4.2 System Description

As already mentioned, the proposed tutoring system covers not only the individual's personalized learning, but aims to go further and give individual attention to students in relation to the other members of the virtual community. That is, the system is aimed at tutoring the student considering that the collaboration with other members of the classroom may help him or her to reach the proposed learning objectives efficiently.

The proposed system conceives virtual communities as classrooms where different working groups of several students may cohabitate, which allows for the performance of intelligent tutoring in three key elements: the class, the group, and the student. That is, in the same way that the community can support the development of individual training, the student may have a positive influence on the acquisition of skills and development process of other partners, and by extension, the entire class.

Moreover, thanks to the intelligent tutoring, the system is able to monitor and assess both the performance and the operation of the class, so that the teaching staff can track the progress of their students at anytime. Indeed, this represents a significant difference with existing approaches because educational technology no longer operates as a mere complement to traditional education. In this system, collected data are used to dynamically reformulate the contents of the lectures at anytime, which permits their adaption to the specific needs of the class and its members.

Figure 3 depicts a general layout of the proposed system. It can be seen that this approach integrates concepts from both intelligent tutoring systems and collaborative learning. This integration permits the usage of Web 2.0 techniques to effectively collect and analyze data. In this way, this data will be used to extract valid conclusions and ease the teaching entity to apply changes in different facets of the class, either individually or within a community scope.

In a virtual learning environment, guiding alternatives are dramatically increased: teachers no longer have to closely monitor each member of the class – which is time consuming and thus unfeasible given the overcrowding experienced nowadays – since software agents are responsible for monitoring both student progression and the role it plays in a group context. Obviously, early

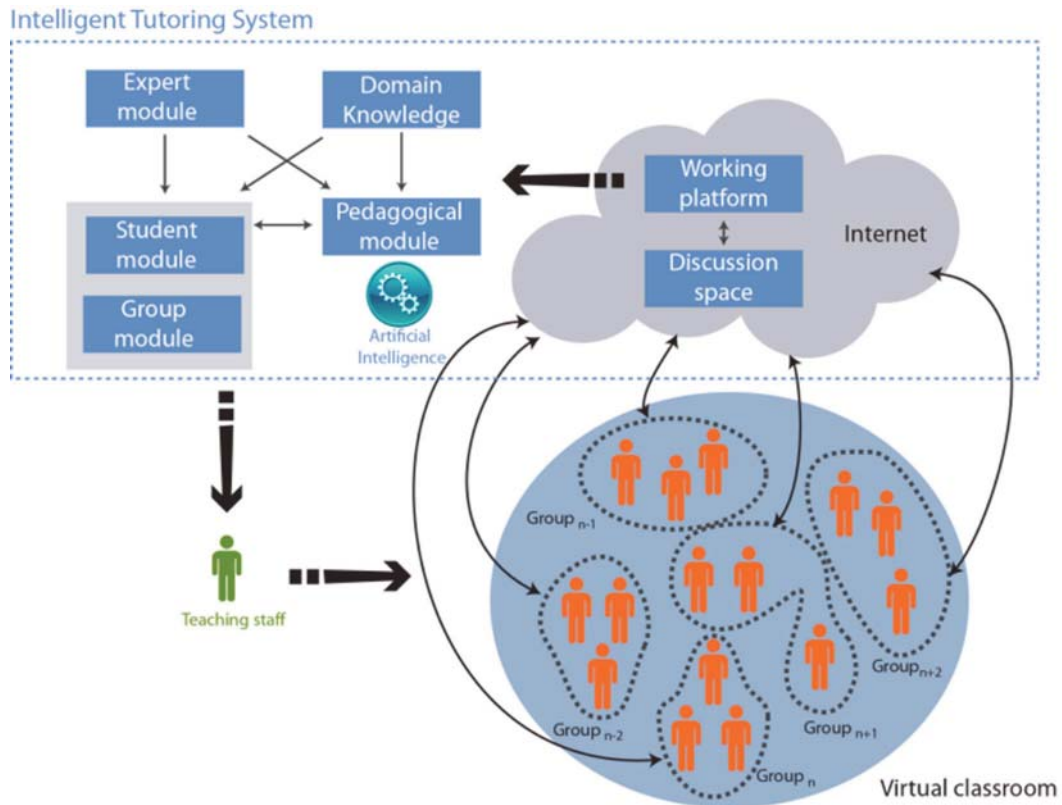


Figure 3: General layout of an intelligent tutoring system of a virtual learning community.

detection is essential to correct any deviations or problems arisen. However, is it possible to detect these problems before they occur?

From the perspective of the human brain, it is extremely difficult to detect whether and when a student's path may be truncated if we rely solely on the knowledge that teachers have concerning the student's academic history, personality, or academic results pattern. Hence, a higher degree of analysis and prediction can be reached if all this information is effectively processed through a computer with artificial intelligence support.

From a pragmatic point of view, early detection of possible patterns indicating a decrease in the academic performance may be of great interest to correct the behaviour of a student in time. Furthermore, the following question arises: can it be useful as a guide for the overall trend of class performance as well? Intuitively, the answer seems to be positive. Just as an intelligent tutor can detect incorrect behavioural patterns in individuals or groups, it could also analyze the personal profiles of students belonging to a community and propose associations which, based on the knowledge gained from past experience with successful outcomes, could be more productive and beneficial to the class as a whole.

The following sections describe the basic characteristics of the proposed system in detail.

4.2.1 Intelligent Monitoring of the Student

The fundamentals of the aforementioned individualized tutoring system are not lost when it is applied on a global scale. Although the primary target is focused on improving the global performance of the group, individualized tutoring is still important and it is feasible to reach such commitment. Therefore, the proposed system permits both individual student tutoring and an accurate tracking of the degree of skills of the student.

The student module housed in our system collects the information regarding every student. This information is constantly processed by the teaching module, which will be responsible – through the usage of artificial intelligence techniques – to infer the most effective and relevant actions (e.g., provide specific exercises to reinforce the knowledge concerning some topics, reallocate students in a group whose members can better complement their shortcomings, publish the assessment results or forecast them to the teaching staff to help with personal tutoring).

4.2.2 Intelligent Monitoring of the Class

Symmetrically to the student element, the tutoring system internally monitors the student groups operation as well. Hence, it continuously tracks and analyzes the interaction between individuals and evaluates the progression of their collective work. According to such analysis, it proposes specific strategies to correct a malfunctioning group.

These strategies range from assigning reinforcement collective tasks to breaking up a given group, including the reallocation of a student to another group if it is found that this change may improve the work performance of the whole class.

4.2.3 Teacher's Feedback

The proposed tutoring system is able to periodically provide the teaching entity with comprehensive data that reflects the operation of key elements in this learning environment: the class as a whole, the groups, and the individuals. Such statistics permit validating the correct execution of collaborative learning, locating those elements with a divergent behaviour, and the detection of possible conceptual shortcomings in a general perspective. In this way, the teaching entity can reformulate the strategies to be used in their lectures.

4.2.4 Creating Working Groups

One of the most appealing novel features of this system is the fact that, by taking the individual profiles of each student, the system is able to autonomously propose the student groups or associations that would maximize the chance of overall success. The creation and design of these groups is not only intended to optimize their individual performance, but also to ensure that the other groups operate properly, which increases the performance of the class as a whole.

4.3 Technical Features

After overviewing the general features of the proposed system, a technical description of the system deployment is provided. Taking into account the nature of the context in which the interaction takes place, the tutoring system has been implemented in a web environment, thus taking benefit from the in-built tools of such framework.

The main module that manages the whole system has been named the teaching module. It contains the tracking system and detects all the possible situations that may happen in the classroom. Note

that it is strongly related to the student and group modules, where the knowledge acquired concerning these elements is stored.

This module uses artificial intelligence techniques to cover both student and group entities in order to provide the aforementioned facilities. More specifically, the system applies these techniques in the following cases: student type elements, group type elements, and interactions between these modules and the teaching staff. Although presenting the teaching module is not the main objective of this article, it forms a fundamental part of the system. Its main features are detailed in the next section.

4.3.1 Relationship to the Student Model

In this first case, the student module collects all available individual information from every student and generates a data set to compare each profile.

The teaching module manages this profile using an artificial intelligence technique called Case-Based Reasoning (also referred to as CBR) (Riesbeck and Shanck, 1989). This technique is aimed at solving problems using the analogy concept, i.e., comparing the new problem with other ones previously solved in order to apply a solution that already worked in the past. Recall that these previously solved problems are stored in a case base at an earlier stage (also referred to as CB or case memory), which hosts the system experience.

The Case-Based Reasoning follows the lifecycle (Aamodt and Plaza, 1994) shown in Figure 4. The major goals of each step of CBR are briefly detailed as follows:

- a) Retrieval phase. Given a new case, all cases stored in the case base are reviewed; the ones sharing the most similarities with the new case are retrieved.
- b) Reuse phase. It is aimed at adapting the set of cases, which the retrieval step has selected, to the new scope and demands. Recall that with traditional pure classification problems, this step is

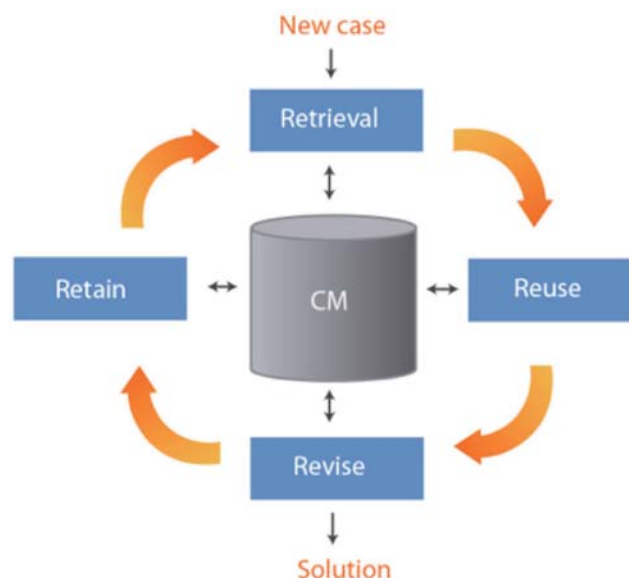


Figure 4: Case Based Reasoning traditional lifecycle

not necessary since the previous phase has provided a set of possible final solutions. However, in this case this step cannot be ignored due to the fact that the retrieved cases must be adapted to the new situation challenged by the new case.

- c) Review phase. It is devoted to assessing whether the proposed solution is valid or not for the new case. Usually, such achieved decision is taken with the interaction of a human expert. However, if the expert is not available, this step can be carried in an automatic fashion. In the proposed system, the human expert is expected to be the teacher whose expertise is used to review the proposals suggested by the system.
- d) Retain phase. This phase implements the system's learning process. That is, it decides whether the proposed solution has to be stored in the case base for future cases or not. That is, every time the system discovers a new student with novel features, it stores their associated information to the case memory in order to boost the performance of the system in further situations.

4.3.2 Relationship to the Group Module

In this second case, the module collects group information concerning the class. From a technical perspective, this data is processed with clustering techniques. So far, practitioners have developed an overwhelming amount of clustering techniques which are suitable for each situation according to the data nature they are classifying (Han and Kamber, 2006).

The proposed Intelligent Tutoring for Virtual Learning Communities (ITVLC) system uses a modified version of the MKM technique, already presented in Vernet and Golobardes (2003), in order to run the clustering process. The key to success in the use of this strategy is to come up with an efficient way to compute the similarity of two of the input elements. Although several similarity functions have been broadly explored in the literature for general data, very few of them fit with the concrete specificities of the educational context. In fact, this distance function has to consider all these concrete attributes that are related to the teaching domain and with the particular features of students.

More specifically, from the educational point of view, it can be concluded that there are three key metrics that enable the reliable computation of the distance (i.e., similarity) between two students: the social profile, the academic profile, and the learning profile. Each one of these parameters is elaborated in the next section.

Social distance takes into account the proximity of two individuals given their personal interests, knowledge, and profile within society. Hence, this metric is aimed at promoting the cohesion and collaboration between the members of a group. Note that although this distance parameter can drive a group composition that promotes the interaction between group members, it is still unable to ensure an overall performance since it ignores some of the relevant knowledge acquisition parameters of every member.

Therefore, the second distance metric is introduced: the academic profile, which is devoted to covering the aforementioned issue. This parameter embraces all grades, competencies, and objective academic data of each student collected during previous periods. Hence, from the distance point of view, this metric emphasizes those students with similar academic profiles. Taking such information into account, the system ensures that the preliminary knowledge contribution of every group member when facing a given assignment will be pretty similar, thus minimizing the risk of finding important performance differences among them. However, this might derive on an unfavourable outcome: since all similar academic profiles are grouped together, heterogeneous

clusters are inexistent and thus, groups with lower academic profiles increase their chances of failure substantially.

In order to correct this negative effect, it is still mandatory to consider all the specific strategies used by each student to acquire knowledge, which is known as the learning profile (also referred to as learning style). In fact, several learning profiles can be identified through the Felder-Silverman Learning Style Model (FSLSM) (Felder and Silverman, 1988). Therefore, after discovering the profile corresponding to every student through the Soloman questionnaire (Soloman and Felder, 2005) the Group Module of the presented ITVLC system suggests several student combinations according to these new findings (i.e., student profiles). Note that this strategy has already been shown to be successful in the past (Peña *et al*, 2002; Hong and Kinshuk, 2004).

To summarize, the profiles provided by this model are elaborated in what follows:

- a) Reflective student vs active student: It distinguishes those students that are more involved in the class from the ones that tend to have reflective and introspective behaviours.
- b) Inductive student vs deductive student: It has to do with the information acquisition process and classifies students according to the type of reasoning used to reach a particular goal: either induction or deduction.
- c) Visual student vs verbal student: It differentiates students according to their sensitivity to the different modes of presenting information.
- d) Sensorial student vs intuitive student: It identifies students according to how they perceive information: either through their senses or reasoning themselves.
- e) Sequential student vs global student: It classifies students according to their knowledge acquisition process: through a sequential procedure that requires a logical progression of small incremental steps or through a comprehensive understanding that requires a holistic view.

The combination of different distance metrics enables the fine creation of different group types with different mixes of student profiles. In this way, a greater heterogeneity on the learning styles is obtained, which ensures a better coexistence of different strategies in order to reach a common goal.

4.3.3 Relationship with the Teaching Staff

As far as the relationship with the teaching staff is concerned, the proposed ITVLC system provides two types of information: relative to the initial group configuration and relative to the academic paths, either from individuals or groups. Each information type is elaborated in what follows.

On the one hand, the Pedagogical Module provides the best initial group configurations which maximize the performance of the whole class based on the groups proposed by the ITVLC system and using the aforementioned clustering techniques. Note that one of the most appealing features of the MKM technique is that it allows the configuration of the default number of groups in the class and thus, the amount of students belonging to each group. In addition, while new data is added to the system, it is continuously verifying that the current group configuration still matches the selected optimal criteria, which enables the teaching staff to keep track of the groups' evolution and reorganize them if necessary.

On the other hand, the teaching staff are periodically informed by the system of some actions concerning the groups' arrangement in order to correct the poor record of a given group, or maximize once again the overall performance of the class.

Although this continuous assessment and revision of groups' configuration provides a considerable degree of robustness in the learning achievement, some challenging questions arise concerning the groups loyalty: how often should the recommendations provided by the system concerning groups' reconfiguration be considered? What should be the response time for a group to correct deficiencies in their working strategy? And last but not least, from an individual perspective, is it positive for a student to deal with different partners throughout the learning process? Indeed, answers to these questions are still open research lines nowadays. After conducting some previous experimentation using the proposed ITVLC system, we have seen that this tool successfully contributes to the improvement of group performance and effectiveness and thus, may help with the resolution of these questions. The following section describes our new findings in these preliminary experiences and elaborates some near-future directions based on this approach.

5. On the Application of the ITVLC System in Groups' Management

In fact, the proposed ITVLC system unveils a broad range of hypotheses concerning the tutoring of virtual learning communities that may lead to years of complex experimentation in order to obtain relevant results. Nevertheless, some initial trials have already been carried out concerning the group management in virtual learning communities.

As stated by Yang and Zhang (2010), the effectiveness of working groups is directly based on the outcome of the following strategies:

- a) Setting up a study group properly.
- b) Establishing clear cooperative learning goals and responsibilities.
- c) Developing a sense of community and cooperation skills among group members.
- d) Choosing the right learning content for the team.
- e) Establishing a rational cooperative learning evaluation mechanism.

According to these indications, the purpose of this section is to detail up to what extent previous experiences using the herein presented ITVLC system – aimed at improving the global performance of the class – have met these guidelines.

As far as the global performance of the class is concerned, it is worth mentioning that critical problems might be encountered when the tasks are not uniformly distributed among the group members. After applying our ITVLC system in some virtual communities, we have detected a new odd category of student: behind an apparently shy appearance, these students remain unnoticed and do not actively participate in the collective task resolution, shifting their burden to other members. Fortunately, the psychology domain has already detected this profile, and labeled such students as bystanders. However, this student category is very difficult to identify in Virtual Communities since its detection demands a deep and accurate group activity tracking analysis.

However, the bystander effect has been successfully detected by the proposed ITVLC system as mentioned above. More specifically, we have run it over the last years on the subject "Web Projects" of the Computer Engineering Degree in which students are assigned a task to be solved in groups of three using a framework with version control as Working Platform. Our prototype version of the ITVLC system has found that, on average, 33% of groups have at least one bystander.

Indeed, the proposed ITVLC system successfully detects such challenging profiles from the academic records and learning profiles. We have seen that the above information delivered to the teaching staff, especially that which refers to the reallocation of a student to another group based

on an accurate selection process, is surprisingly valuable (i.e., it includes enough objective arguments) when considering whether to accept the suggestions provided by the system.

Although these experiments are aimed at assessing the management of working groups, we have already seen that the usage of the proposed ITVLC system successfully meets strategies a) and c) proposed by Yang and Zhang (2010).

Note that, after this initial success, a further experimentation might be required in order to explore the ability of this ITVLC system to meet the remaining strategies. However, the positive experiences described herein envisage the goodness of our approach and encourage practitioners to work towards this direction.

6. Conclusions

This paper introduces the concept of intelligent tutoring of virtual learning communities. The proposed approach combines the power of intelligent tutoring systems with the educational learning process by using new technologies and Web 2.0 features.

The presented system enables the teaching staff not only to monitor each student individually, but also to boost group performance. In fact, most existing intelligent tutoring systems that consider group interactions are focused on the individual behaviour of each member, rather than addressing the effect of individual student's profiles on other classmates as our system does.

We have found that it is very important to consider the different student profiles in order to properly build the working groups. Therefore, it is mandatory to take into account not only previous academic records, but also student learning and social profiles.

Overall, this paper has addressed four key aspects to manage a virtual learning community: 1) a smart student monitoring, 2) an intelligent classroom tracking, 3) an automated feedback system for the teaching staff, and 4) an exhaustive control of the group performance in order to adjust groups and detect possible faults.

Additionally, the proposed system poses new research challenges: firstly to improve the overall classroom performance as well as that of each individual student and, secondly, to achieve the total integration of all the students working in a group and to avoid the "bystander" scenario.

Nowadays, beyond the proof of concept explained in this paper, a physical system is being developed based on the modules herein described in pursuit of this novel concept.

References

- AAMODT, A. and PLAZA, E. (1994): Case-Based Reasoning: Foundations Issues, Methodological Variations, and System Approaches. *AI Communications*, 7: 39–59.
- BRUSILOVSKY, P. and PEYLO, C. (2003): Adaptive and Intelligent Web-Based Educational Systems. *Int'l Journal Artificial Intelligence in Education* 13: 156–169.
- DILLENBOURG, P. (1999): Collaborative Learning. Cognitive and Computational Approach. New York: Pergamon Earli, 1999.
- FELDER, R.M and SILVERMAN, L.K. (1988): Learning and teaching styles in engineering education. *Journal of Engineering Education* 78(7): 674–681.
- GARCÍA PEÑALVO, F.J. (Ed) (2008): Advances in E-learning: Experiences and Methodologies. Information Science Reference (formerly Idea Group Reference), ISBN 978-1-59904-756-0, Hershey, PA, USA.
- GOLOBARDES, E., VERNET, D. and SALAMO, M. (2000): Prediction in an Educational Environment using Case-Based Reasoning. *Proceedings of Learning'00 (IEEE)*.
- HAN, J. and KAMBER, M. (2006): Data mining: Concepts and techniques. *The Morgan Kaufmann Series in Data Management Systems*. Jim Gray, Series Editor Morgan Kaufmann Publishers.

- HONG, H. and KINSHUK (2004): Adaptation to student learning styles in web based educational systems. In CANTONI, L. & McLOUGHLIN C. (Eds.) *Proceedings of EDMEDIA 2004 – World Conference on Educational Multimedia, Hypermedia & Telecommunications, AACE*, 491–496.
- HUNG, D., CHENG, D. and TAN S. (2003): A Social-Constructivist Adaptation of Case-Based Reasoning: Integrating Goal-Based Scenarios with Computer-Supported Collaborative Learning. *Educational Technology*, 30–35.
- JOHNSON, R.T. and JOHNSON, D.W. (1986): Action research: Cooperative learning in the science classroom. *Science and Children*, 24: 31–32.
- MARTIN-MORENO, Q. (2004): Aprendizaje colaborativo y redes de conocimiento. *Actas IX Jornadas Andaluzas de Organización y Dirección de Instituciones Educativas*. Granada, 15–17: 55–70.
- PANITZ, T. (2012): Collaborative versus cooperative learning: a comparison of the two concepts which will help us understand the underlying nature of interactive learning. <http://home.capecod.net/~tpanitz/tedsarticles/coopdefinition.htm>. Accessed: 14-June-2012.
- PAO-HUA, C., MENQ-JIUN, W., PI-HSIANG, L. and KUANG-KU, C. (2009): Accessing e-Learners' Knowledge for Personalization in e-Learning Environment. *Journal of Research and Practice in Information Technology*. 41(4): 295–318.
- PEÑA, C.I., MARZO, J.L., DE LA ROSA, J.LI. and FABREGAT, R. (2002): Un sistema de tutoría inteligente adaptativo considerando estilos de aprendizaje. *IV congreso iberoamericano de informática educativa*. IE2002, Vigo (España).
- PÉREZ, T.A., GUTIERREZ, J., LOPEZ, R., GONZALEZ, A. and VADILLO, J.A. (2001): Hipermedia, adaptación, constructivismo e instructivismo. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, 12: 29–38.
- RIESBECK, K. and SCHANK, R.C. (1989): *Inside Case-Based Reasoning*. Lawrence Erlbaum Associates, Hillsdale, NJ, US.
- SELF J. (1990): Theoretical Foundations for Intelligent Tutoring Systems. *Journal of Artificial Intelligence in Education*, 1(4): 3–14.
- SOLOMAN, B.A. and FELDER, R.M. (2005): Index of Learning Styles Questionnaire. <http://www.engr.ncsu.edu/learningstyles/ilsweb.html>. Accessed 14-Jan-2013.
- SOTOMAYOR, G. (2010): Las redes sociales como entornos de aprendizaje colaborativo mediado para segundas lenguas. In *EDUTEC, Revista Electrónica de Tecnología Educativa* 34.
- STEPHENSON, J. and SANGRÀ, A. (2012): Modelos pedagógicos y e-learning. <http://es.scribd.com/doc/85600394/Modelos-Pedagogicos-learning>. Accessed 14-June-2012.
- VERNET, D., SALAMÓ, M., VALLESPI, C., CAMPS, J., GOLOBARDES, E. and BACARDIT, J. (2001): ¿Cómo predecir la Evolución del Alumno? *Actas de las VII Jornadas de la Enseñanza Universitaria de la Informática (JENUI'2001)* 335–340.
- VERNET, D. and GOLOBARDES, E. (2003): An Unsupervised Learning Approach for Case-Based Classifier Systems. *Expert Update. The Specialist Group on Artificial Intelligence*. 6(2): 37–42.

Biographical Notes

David Vernet is a researcher and he is developing his work at La Salle in Barcelona, Spain (Ramon Llull University). He is currently the Program Coordinator of the computer science degree and he teaches programming methodology and web programming in this faculty. His main research interests are artificial intelligence, intelligent tutoring systems and learning communities, and he has numerous important publications that support his work in this field. He holds a BSc and an MSc degree in computer science from Ramon Llull University and he is currently a PhD candidate in the same university.



David Vernet

Xavi Canaleta has participated as a researcher in two research groups inside La Salle: The Intelligent System Group and the Group of Distributed Systems in which several publications in national and international congresses are carried out. Since 2009, he assumed the position of Coordinator of Secondary School Teacher Education MA Programme with the specialty of technology and helped with the design of its



Xavi Canaleta

syllabus. As a teacher of this master, he continues studying and contributing with the competence evaluation and ICT application in the classrooms. He also organized several practical workshops about the ICT introduction as useful tools in the learning process.

Joan Navarro received the MSc graduate degree in telecommunications engineering in 2008 and the BSc in telematics engineering in 2006 from the Universitat Ramon Llull, Barcelona, Spain (URL). He is currently an Associate Lecturer and PhD student and member of the “Grup de Recerca de Sistemes Distribuïts i Telemàtica” at the same university. His main research interests are focused on the area of distributed systems and data management specifically in the context of Big Data and cloud computing. Throughout these years, he has participated in many public and private funded R&D projects and contributed to several conferences.



Joan Navarro

Agustín Zaballos received the PhD degree in telecommunications engineering in 2012 and the MSc degree in electronics engineering in 2000 from the Universitat Ramon Llull, Barcelona, Spain (URL). He has been an Assistant Professor in the computer science department in URL since 1999 and Project Manager of the R&D Networking Area since 2002. His research is focused on real-time communication protocols, ICTs for smart cities and education for the smart cities.



Agustín Zaballos

Gemma Pallàs holds a computer science degree from the Universitat Oberta de Catalunya, Barcelona, Spain. She is currently completing a multimedia degree in the same university and her areas of interest are: interfaces designing, web interaction and digital photography. She has developed several projects related to human-computer interface and web interaction. She is currently involved in different web projects and she has been the project manager in the definition of various designs of mobile interfaces.



Gemma Pallàs

High Performance Web-of-Things Architecture for the Smart Grid Domain

David Vernet, Agustín Zaballos, Ramon Martin de Pozuelo, Víctor Caballero
Engineering Department
La Salle – Universitat Ramon Llull (URL)
Barcelona
{dave,zaballos,ramonmdp,vcaballero}@salleurl.edu

Abstract — The increasing complexity in the management of Smart Grids is an essential factor in the creation of new technological infrastructures capable of managing the different devices involved in the network. This network has been converted into an example of the Internet of Things. In this regard, the Web of Things enables an improvement in the processing of these data. Besides, the large amount of data in the Smart Grid domain means that a high performance architectural design able to manage concurrently the entire information processing ability. This paper presents an initial approach for a new architecture and the first results after the system implementation.

Index Terms—Internet of Things, Web of Things, Sensing, High Performance Web

I. INTRODUCTION

The recent growth of the Internet has fostered the interaction of many heterogeneous technologies under a common environment (i.e., the Internet of Things - IoT). Smart Grids entail a sound example of this situation where several devices from different vendors, running different protocols and policies, are integrated in order to reach a common goal: bring together energy delivery and smart services. This proposal deploys an IoT-based infrastructure that enables machine-to-machine interactions between small and resource-constrained devices on the Smart Grid domain based on HTTP protocol. It extends the IoT concept by providing a bidirectional human-to-machine interface, inspired by the Web of Things (WoT), which results in a ubiquitous energy control and management system (i.e., uniform access to all devices of the Smart Grid) coined as Web of Energy (WoE) [1].

The recent advances on this domain have led to effective architectures that support this idea from a technical perspective, but fail to provide powerful tools to assist this new environment. Hence, the purpose of our technological proposal is to research a novel unified and ubiquitous sensor management interface that uses the advantages featured by the Web of Things to manage the Smart Grid. Therefore, this work opens a new path from the Internet of Things to the WoT and results in a new concept coined as the WoE [1]. Overall, the Web of Energy links all domains and permits a bidirectional communication between electricity domain and the application domain. More concretely, the main objective is to carry out a proof-of-concept of an open web-based interface that isolates the electricity grid domain from its utility functions, and relies on (1) a distributed storage layer to support the massive amount of data generated by the grid. The database proposed is an open-source document database that

provides high performance, high availability and automatic scaling [2], and (2) a Southbound REST-full API that permits an easy and seamless management of the distributed storage layer and the smart objects connected to the system.

II. THE SMART GRID AS AN IOT

Recent advances on Smart Grids have explored the feasibility of considering the power electrical distribution network as a particular case of the IoT [3][5]. Certainly, this specific domain poses appealing challenges in terms of integration, since several distinct smart devices (also referred to as Intelligent Electronic Devices or IEDs) from different vendors — often using proprietary protocols and running at different layers — must interact to effectively deliver energy and provide a set of enhanced services and features (also referred to as smart functions) to both consumers and producers (prosumers) such as network self-healing, real-time consumption monitoring and asset management [7]. Although the latest developments on the IoT field have definitely contributed to the physical connection of such an overwhelming amount of smart devices [4], several issues have arisen when attempting to provide a common management and monitoring interface for the whole Smart Grid [7][8].

Indeed, integrating the heterogeneous data generated by every device on the Smart Grid (e.g., wired and wireless sensors, smart meters, distributed generators, dispersed loads, synchrophasors, wind turbines, solar panels and communication network devices) into a single interface has emerged as a hot research topic [5][6]. So far, some experimental proposals [3] have been presented to face this issue by using the Web of Things (WoT) concept to access a mashup of smart devices and directly retrieve their information using reasonably thin protocols (e.g., HTTP, SOAP) [9]. However, the specific application of these approaches into real-world environments is fairly dubious

because (1) they may open new security breaches [10][12] (i.e., end-users could gain access to critical equipments), (2) there are no mature electric devices implementing WoT-compliant standards available in the market [7] and (3) industry is averse to include foreign modules (i.e., web servers) on their historically tested and established — but poorly evolved — proprietary systems [13].

Therefore, the authors of this paper explore a new way to overcome these issues through the European project INTElligent Electrical GRId Sensor Communications (INTEGRIS) [7] and Future INtErnet Smart Utility ServiCEs project (FINESCE) [11]. This work provides a management interface for the Smart Grid inspired by the WoT. Continuing the work done in these projects, the aim is to implement an ICT infrastructure—based on the IoT paradigm—to handle the Smart Grid storage and communications requirements [14] to manage the whole Smart Grid and link it with end-users using a WoT-based approach, which results in a new bridge between the IoT and WoT. This proposal, which takes the pioneering new form of the WoT, is targeted at providing a context-aware and uniform web-based novel environment to effectively manage, monitor, and configure the whole Smart Grid. Moreover, conducted developments prove the feasibility and reliability of our approach and encourage practitioners to further research in this direction and to envisage new business models [15][16].

The open IoT-based infrastructure presented in our Web of Energy proposal will provide new tools to manage energy

infrastructures at different levels, from IoT-based infrastructure enabled machine-to-machine interactions between small and resource-constrained devices on the Smart Grid domain. Thus, we have extended the IoT concept by providing a bidirectional human-to-machine interface—inspired by the WoT—that results in a ubiquitous energy control and management system coined as Web of Energy [1]. This proposal will combine the web-based visualization and tracking tools with the Internet protocols, which enables a uniform access to all devices of the Smart Grid. In order to provide such an effective and reliable management interface to address the heterogeneous nature of devices residing on the grid, we will continue the deployment of an intelligent subsystem devoted to (1) learn from the real-world events, (2) predict future situations, and (3) assist in the decision making process.

The tools developed can be provided in an open format, available to anyone in the research community, and able to contribute and enhance the platform building new modules for managing other resources. As a first step, the platform will be demonstrated through the management of Smart Metering resources, based on the formats of DLMS COSEM and IEC 61850, but the whole system will be designed for a much wider scope where every utility, enterprise, public administration, or any organization or single prosumer can build their own sensor application on top of it (Figure 1).

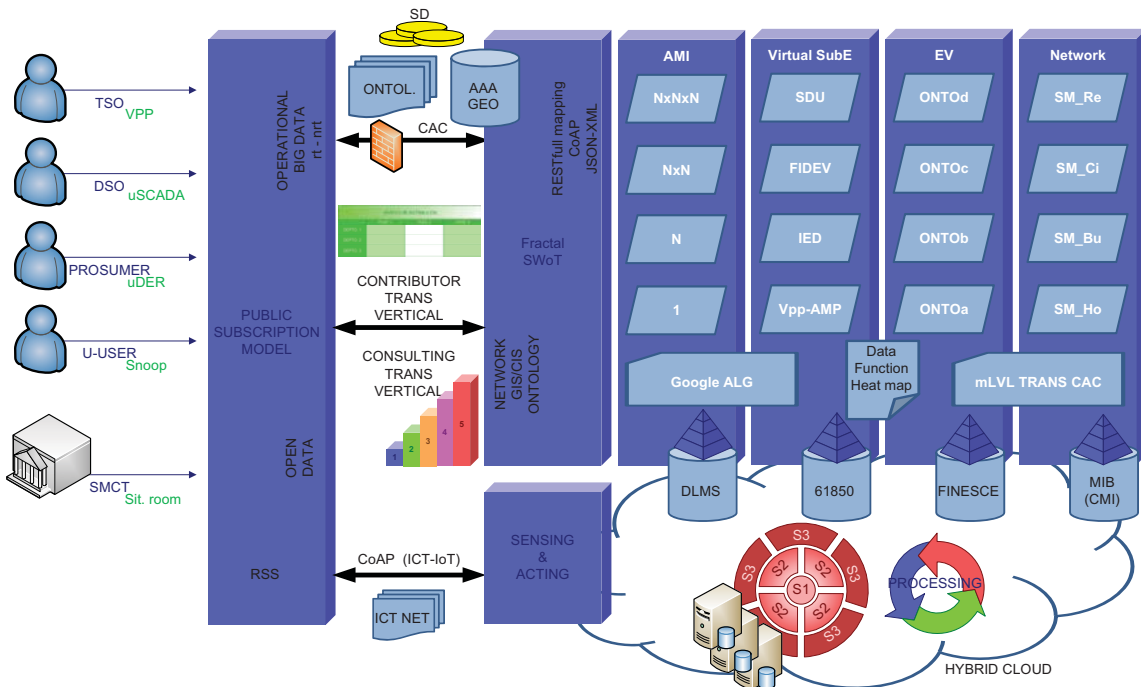


Fig. 1. WoE interactions

A. WoE presentation and specified functions

The concept of Smart Grids has been considered in recent years as an appropriate answer to address new challenges in the energy domain. However, additional resources are still needed in challenges such as the proactive operation of the grid, efficient integration of demand into grid operation, integration of renewable generation or maximum network reliability to obtain applicable solutions. The WoE will combine energy domains and ICT technologies with the objective of building an interoperable platform for the coordinated planning, operation and settlement of future distribution and access to networks by integrating demonstrable solutions in real user environments based on web of things technologies.

The current control of electricity distribution networks, including the Advanced Distribution Automation (ADA) and Demand Side Management (DSM), is (1) centralized, (2) silo oriented (fragmented by applications), (3) using specific communication technologies for each purpose but there is a lack of integration among them and (4) DSO owned ICT systems tend to be used [17]. So, in practice the situation is still centralized and fragmented but with existing solutions and trials allowing the integrated management and the distribution of selected granular functions. This contrasts with the distributed and fractal nature of the future Smart Grids, which can only be based on standardization, flexibility, distributed systems and communication among all the actors [13][11].

Furthermore, the increased use of renewable and distributed generation means the operation and management of the electric power system must change radically. Increased levels of automation, distributed intelligence and on-line data mining and management are required to deliver the network control functions, reducing reconfiguration and the restoration times. Reconfiguration of smart grids addresses new challenges during normal operation and also for restoration and management of crisis situations [37][10]. The connection of end-users (prosumers) to the energy market will facilitate the installation and connection of devices that offer grid services that will help to mitigate capital and operational costs of the grid modernization required for energy transition, and to minimize environmental impact, thus ensuring lower electricity prices for everything involved. New benefits will be generated and shared in a fair way between all actors, from aggregators to industrial end-users and citizens.

A coordinated vision of the grid will provide mechanisms to tackle the challenges mentioned above showcasing them through real and modern applications. WoE will provide synergies at different levels: Smart Grids and other smart networks, individuals and communities as a first approach to a Software Defining Utility (SDU). The potential benefits of the WoE are framed by: a) innovative communications, acquisition and processing platform based on the extensive use of the "Real Time Services" concept providing open and interoperable access; b) metering integration platform based

on multiplatform web technologies; c) Advanced Medium/Low voltage control centre integrating real-time grid information coming from devices to provide a clear view of the current and the near-future status of the grid thanks to a high performance environment; d) an energy services market platform. The performance quantification of that WoE concept will be the key to accelerate the implementation of new policies, market rules, and emerging smart grid programs [15][16].

WoE outlines some of the challenges in improving the resiliency of the electrical grid and proposes an approach that makes use of advanced sensor technology (advanced sensors are needed to improve the knowledge of state), analytics, and agile control in a Smart Grid [38][39]. Furthermore, WoE proposes a smart grid supervision infrastructure, which can deliver real-time and high-performance notifications on a global scale for transferring measurements from different distributed sensors and take actions over the grid via different communication protocols, informing the different stakeholders (e.g. producers, consumers, aggregators, system operators, etc.) within the adequate time frame [17][6]. We propose to use a distributed infrastructure based on web of things and implement a novel service platform for facilitating distributed control, auto healing and power grid control.

In addition, WoE technology will tackle the implementation of smart real time distributed monitoring platforms enabling the data fusion and knowledge extraction for the different faults detection and prevention schemes. Intelligent HTTP based sensors will provide a new source of relevant distribution status information, including loadings, voltage profiles, harmonics and outage conditions which, combined with equipment condition data, such power frequency interference signatures, will provide predictive perspectives of potential equipment failures. This platform requires large storage technologies that can hold the massive amounts of information that will be generated by the millions of sensors implemented in the Smart Grids and that will make the information available in negligible times to the system or systems demanding it (Figure 2).

Smart grids need monitoring strategies based on decentralized and uncoupled architectures (service oriented, multi-agent), supported by real time middlewares (DDS) capable of dealing with huge amounts of information at different time scales, process events (complex event processing) and discover sequences of them (sequence event discovery) working together with OLAP and Data Mining solutions. A multi-agent conception of the grid is necessary to deal with coordination and optimization requirements for an efficient and safe network. Pervasive web monitoring devices with communication capabilities increase security of the network as a whole but imply new cyber-security issues and make optimal restoration and reconfiguration more complicated [18]. Currently, heuristics, meta-heuristics and learning methods are mature and available. Restoration and reconfiguration techniques including auto healing will benefit from these reasoning engines.

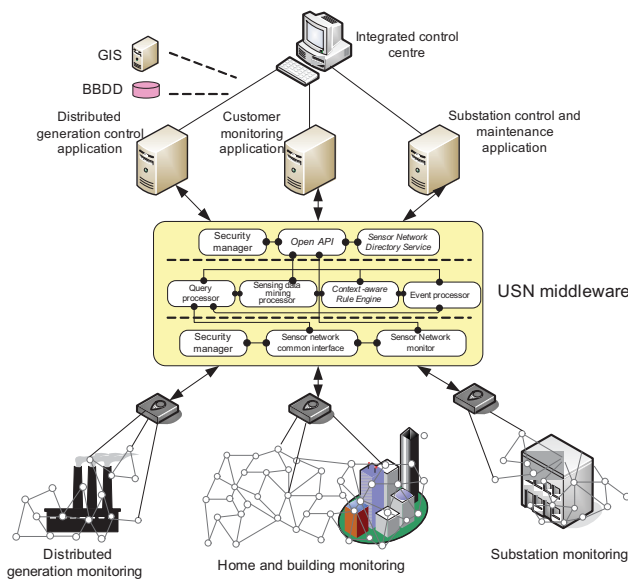


Figure 2. Open API architecture

III. THE WEB OF THINGS

Day by day, the number of connected Things is growing exponentially. The latest data shared by Gartner finds that 4.9 Billion of Things will be accessible during 2015 and this figure will increase to 25 Billion in 2020 [19]. The way to access these devices from a single platform is undoubtedly one of the biggest headaches for researchers. In this regard, standardized solutions provided by the rapid evolution of the Internet have laid the foundation of what we call the Web of Things. Therefore, WoT architectures aim to integrate everyday objects with web technologies. Those devices should be able to communicate with each other using existing web standards. Prerequisites for those Things are minimal processing and communication capabilities. WoT researchers try to define and delimit concepts (e.g. what is a Thing? [20]) implied on those envisioned architectures and solve some problems that arise when every Thing may sense or actuate on every Thing.

The architecture presented by Dominique Guinard in his thesis [21] proposes a good basis to start other WoT designs. Guinard defines WoT general prerequisites, he also defines what a Thing is and a Virtual Object is in a WoT architecture and finds solutions to problems like how to discover and find Things and how those Things can connect and push information to a server. In [21] a layered architecture (see Figure 3) that consists of four layers that address four main problems is proposed:

- Device Accessibility Layer: how to enable consistent access to all kinds of connected objects?
- Findability Layer: how do we find their services to

- integrate them into composite applications?
- Sharing Layer: how we preserve privacy?
- Composition Layer: how do we get closer to end-users?

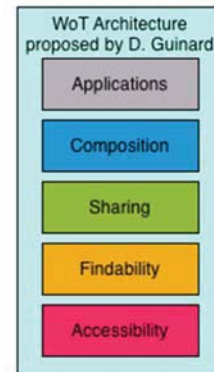


Figure 3: Layered architecture [21]

One of the main problems of the WoT architecture is to standardize communication protocols between different Things. Indeed, there is still no clear standard defined for this purpose and there are different options available. In the next section, we present the best placed and supported protocols by researchers and which protocols are also valid/eligible alternatives.

IV. WOT PROTOCOLS

It would be easy for WoT to use only one protocol, but the heterogeneity of devices composing the WoT makes it unfeasible. Thus, different communication protocols should be considered. The use of each one depends on the final proposed solution.

A. Preferred candidates by the community

1) MQTT

As stated in its official webpage [29], MQTT stands for MQ Telemetry Transport. It is a publish/subscribe (pub/sub), extremely simple and lightweight messaging protocol, designed for constrained devices and low-bandwidth, high-latency or unreliable networks. The design principles are to minimize network bandwidth and device resource requirements whilst attempting to ensure reliability and some degree of assurance of delivery. These principles also contribute in the protocol ideal of the emerging “machine-to-machine” (M2M) or “Internet of Things” world of connected devices, and for mobile applications where bandwidth and battery power are at a premium. MQTT was introduced by IBM and Eurotech companies.

MQTT is a protocol that uses a pub/sub model, connecting publishers and subscribers via a broker (server). Its headers are small and therefore their overhead is minimum. MQTT can also work over SSL for security reasons, but SSL adds an

extra overhead to the communication. As publishers and subscribers connect via a broker, the use of a centralized server leads to a SPF (Single Point of Failure).

2) *CoAP*

CoAP [30] was specified and standardized by the CoRE (Constrained RESTful Environments) group in IETF, the Constrained Application Protocol (CoAP) is a specialized web transfer protocol for use with constrained nodes and constrained networks, such as those that will form the Web of Things. This protocol shares several similarities with HTTP like its REST architectural style but instead of using TCP it uses UDP to achieve its goals.

As it is a request/response protocol like HTTP, both WoT servers and constrained devices or gateways should act as servers and clients at the same time to ensure bidirectional communication at any time. For example, a constrained device using this protocol may fire an event to the WoT server and the WoT server may request something to the constrained device. Proxies between HTTP and CoAP will achieve interoperability between HTTP and CoAP clients. Translation between CoAP and HTTP protocols is easy and straightforward as equivalences of response codes, options and methods are present in both protocols. Security is achievable using DTLS and a variety of key management methods.

B. Other Candidates

1) *DDS*

From [31], “DDS (Data Distributed Service) is an API specification and an interoperability wire-protocol that defines a data-centric publish-subscribe architecture for connecting anonymous information providers with information consumers”. DDS follows a decentralized pub/sub model. It differs from MQTT model in the following two key points:

- DDS protocol starts to operate on top of the link level layer of the OSI model creating a Common Data Bus where every device can connect in a decentralized manner. This protocol also defines several QoS options.
- As a decentralized protocol, it does not have a SPF like the broker in MQTT.

DDS has only implementations for C, C++ and Java and has a higher learning curve compared to MQTT. On the other hand, MQTT clients are implemented for several languages.

2) *XMPP*

XMPP [40] (originally named Jabber) is a protocol for person-to-person communication based on XML. Its main use is for chat communication but since the growth of the IoT concept, the XMPP Standards Foundation is working on defining extensions (XEP) for use in the IoT [41]. These

extensions aim to specify standards for a wide variety of communication types between IoT devices such as Control, Discovery, Multicast, Pub/Sub message types, or many others. They use EXI [42] (compressed XML) to reduce the size of messages, as XML is known to produce larger file/message sizes than other text based formats. Even though XEPs for the IoT are not as much as popular as MQTT or CoAP, it is worth to keep track of them as they are growing fast and may be a basis of the format WoT messages are defined.

3) *AMQP*

AMQP [A1] is a message-centric binary wire protocol that uses a centralized broker. AMQP is built on top of the TCP layer (at least, it is assumed to work on top of TCP). Authentication and encryption is made available through SASL and TLS respectively. As AMQP was created by businesses-to-businesses, it provides transactional modes of operation that allow it to take part in a multi-phase commit sequence. The key feature of AMQP is that it was designed for interoperability between vendors. It mandates the behaviour of the messaging provider and client to the extent that implementations from different vendors are interoperable.

Third party implementations of AMQP clients exist for several languages. Although AMQP is a great opponent for MQTT, the latter seems more suitable to build a proof of concept for the WoT architecture we have been talking about in this section.

V. SOME WoT IMPLEMENTATIONS (RELATED WORK)

This section is going to review some WoT implementations of this architecture. Guinard joined the EVERYTHING platform and engine [22]. This engine allows Things (they call them “things”), to be connected to this platform through a RESTful API. They describe two types of things:

- Unconnected/Tagged: They are encoded in a 1D/2D bar code or a NFC/RFID tag and users can interact with them by scanning the tag.
- Connected: Those things can interact with the RESTful API of the EVERYTHING engine, they can be sensed and/or actuated.

This engine offers the creation of applications that represent remote client applications (like those used in social networks like twitter or facebook). Thanks to his THNG-Push technology (currently in beta), the engine provides a publish/subscribe MQTT M2M broker where WebSockets wrapping MQTT are used to allow communication with browsers. They are also working on adding CoAP support to this technology. Node.js and javascript libraries are available to facilitate the use of their API.

In [23], the authors propose and implement a holistic web architecture for the Internet of Things. They point out key features and capabilities of a holistic architecture and they use a layered model with an abstraction layer for communication among devices.

Node-RED [24] is a visual tool for wiring hardware devices, APIs and services. From the Node-RED front page: “Node-RED provides a browser-based flow editor that makes it easy to wire together flows using the wide range nodes in the palette. Flows can be then deployed to the runtime in a single-click.” Also in beta, this tool aims to provide users with a visual manner connecting things. Node-RED is built on Node.js and customized functions between nodes can be created within the editor using Javascript. There is an EVERYTHING Node-RED integration library to add some functionality of the EVERYTHING platform to this tool.

Octoblu [25] is an open-source cloud platform (public, private or hybrid) built to connect people, devices and systems through a great variety of protocols like MQTT, CoAP, HTTP(S) and WebSockets using a RESTful API. It also offers a very powerful visual tool for connecting things (nodes). This tool also allows developers to program their own nodes as Javascript functions. Node.js and Javascript libraries are also available to facilitate the use of their API.

Neura [26] and TempoIQ [27] are platforms for collecting sensor data through a RESTful API. While Neura is more person-oriented, TempoIQ is a general-purpose data collector. TempoIQ (former TempoDB) also offers tools for monitoring and analyze this sensor data.

Finally, in [28] the concept of storage registration is introduced for the WoT. In this storage approach a web client announces its interest of storing some sensor data to the server. The server will store the data until the web client requests removal or an expiration time is reached. This prevents the server from reaching its storage limit.

VI. PROPOSED ARCHITECTURE

The purpose of this section is to present the architecture and announce the key parameters used to link the layers. From our real-world experiences collected during the INTEGRIS and FINESCE [11] projects, we have found that dividing the Smart Grid into these logical layers poses some critical difficulties arisen from the fact that typically, IEDs are closed devices that do not allow implementing custom developments (e.g., security or information-exchange protocols)—as novel experimental devices do. Therefore, we proposed a new device coined as I-Dev [17][14][18] that behaves as a frontier between these two layers and implements (1) a communications subsystem that allows heterogeneous network coexistence, (2) a security subsystem that provides a reliable and secure low layer communications infrastructure, (3) a distributed storage subsystem that smartly stores all data generated by IEDs, and (4) a cognitive subsystem that is aware of all events arisen from any subsystem of the network.

In Figure 4 an architecture is proposed in order to allow Things to communicate with each other. The proposal explanation will be divided into the following sections: specifications, used protocols and software design.

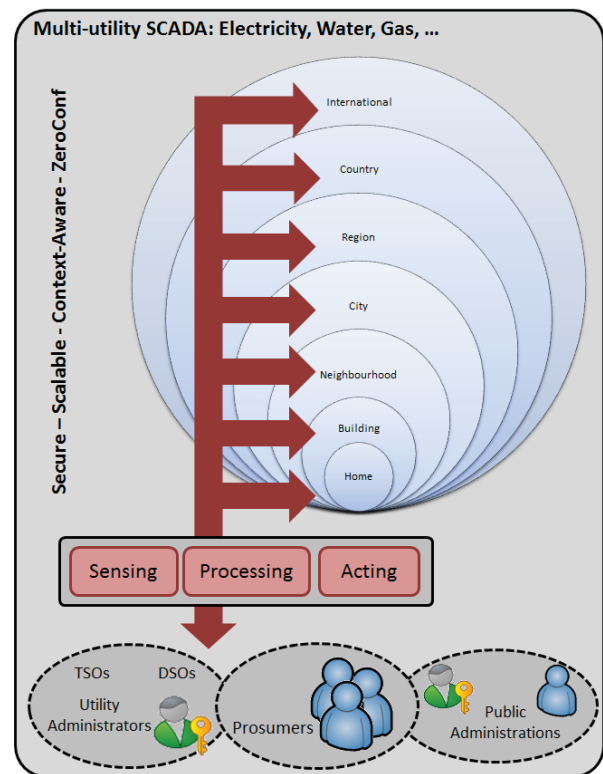


Figure 4: Functional architecture

A. Specifications

In this proposal, a Thing can be understood as every device that has minimum communication, processing and storage capabilities so it is able to be sensed or actuated and can communicate with another device to send or receive data. The proposed architecture has those goals to satisfy:

- As standard Web protocols such as HTTP or WebSockets use resources that may not be available on some constrained Things, this architecture will connect them using a variety of protocols (but ideally one) like CoAP or MQTT that are more suitable for constrained devices.
- Every Thing must be using an open standard protocol and understandable by the architecture to connect to the WoT. Gateways serving as a proxy for those Things that use proprietary or non-understandable protocols should be used.
- Things that cannot connect to the WoT infrastructure directly should connect to a gateway instead and let that gateway do the connection for them.
- This architecture will allow developers to interact with Things without knowing the protocol they are really using to communicate; hence there is a need to provide developers with a communication abstraction layer.
- It must allow on-demand deployment due to the fact that the number of Things will increase and therefore more resources will be needed to connect them.

- This architecture should be able to balance the load on its servers. Load balancers will prevent or minimize bottlenecks.
- Things will become virtual objects for the architecture and those virtual objects could be aggregated and linked to form other virtual objects.
- Eventually, every virtual object will be accessible by a RESTful URI.
- The architecture must have authorization of Things to interact with them to preserve owner’s privacy.
- Third-party applications must have the architecture authorization to interact with Things.
- Every Thing must have an owner.
- Things can be queried and discovered.

In Figure 5 we can see how Things are connected to servers directly or via a gateway/aggregator being part of the WoT.

B. Protocols

While the use of one protocol like HTTP for all the Web of Things would be desirable, it is obvious that it is not achievable due to the heterogeneity of devices. So, a WoT architecture would need interoperability with a variety of protocols to interconnect devices.

1) HTTP and WS

As the Web uses HTTP to communicate, is mandatory for the architecture to support this protocol. It will also provide a RESTful communication using URIs to identify resources and methods to actuate or sense on them.

With IPv4 still in use and incapable of addressing all internet connected devices, internet connection to personal computers, for example, is available through port translation mechanisms. This translation makes them incapable of receiving data asynchronously from a server, hence the use of some sort of mechanism to allow them to receive data in an asynchronous manner is required. In addition, as proposed in [2], WebSockets [13] protocol can overcome this constraint by providing a mechanism for browser-based applications that need two-way communication with servers.

2) MQTT and CoAP

As described in the previous section, MQTT and CoAP seem to be very suitable protocols for the WoT world. We propose to use these protocols but the architecture should be capable of understanding more protocols (e.g. DDS).

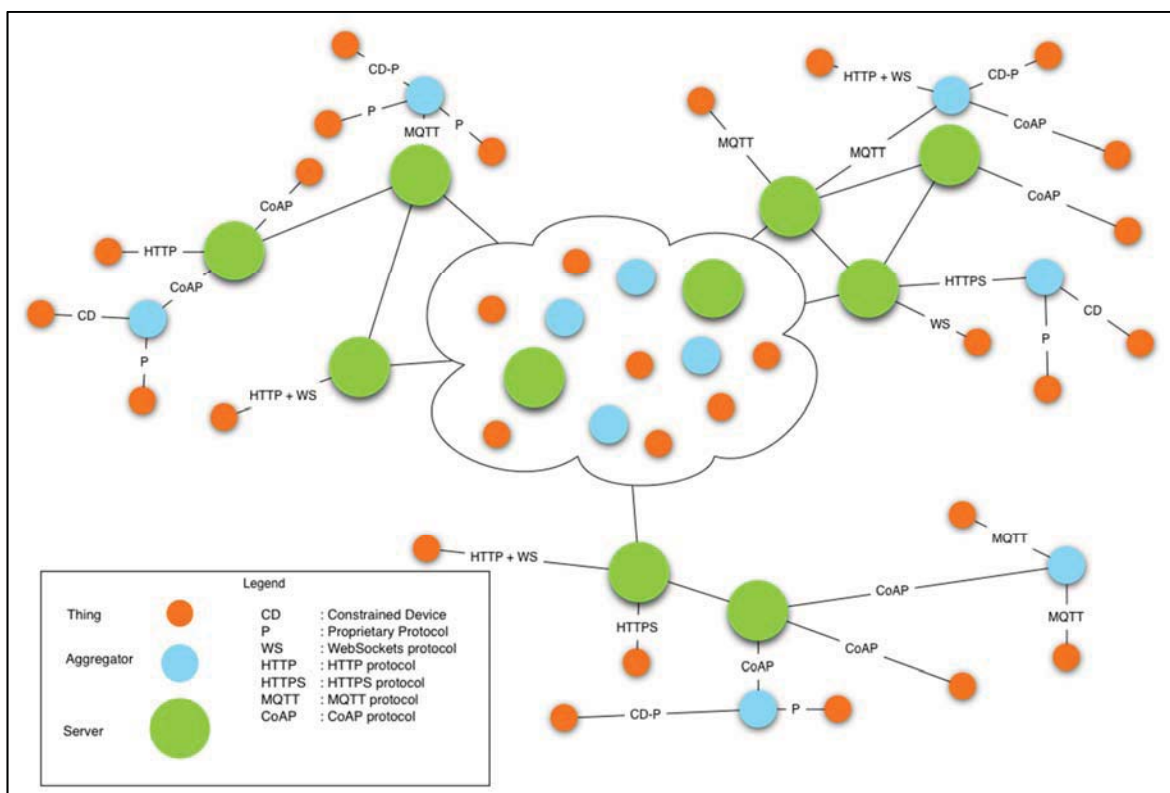


Figure 5: The proposed WoT architecture

C. Layered and modular design

To be able to seamlessly intercommunicate these protocols and help developers implement and provide different implementations of functional units, (e.g., how a protocol is handled, adding a new protocol to the framework, how a VO is accessed, etc.) the architecture relies on a layered and modular design. It is composed of five main layers (Fig. 6), they are:

- **WoT Protocol Abstraction Layer:** The purpose of this layer is to provide an abstraction mechanism for developers to interact with Things.
- **REST to VO/Query:** The goal of this layer is to map dynamically generated URIs to virtual objects (VO). It is an interface to sense or actuate on VOs, that is, Things. It is also an endpoint for querying Things like the temperature on a specific location.
- **Auth:** This layer is responsible for requesting and granting access between Things. As mentioned in section IV, every Thing must have an owner, therefore, WoT servers must request access to Things and VOs must request access to other VOs.
- **Virtual Object Space:** This is the space for virtual objects. Things will become virtual objects in the virtual world. Those VOs could be then sensed, actuated and aggregated.
- **Proxy Layer:** This abstraction layer will serve as a proxy for servers to communicate. A protocol for this communication must be defined for the architecture.

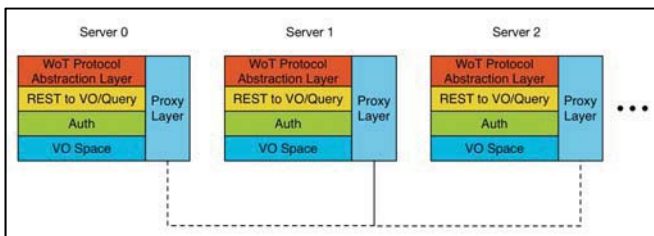


Fig. 6 – Layered and modular design

Those layers will be backed up by the database and the reasoner [33] modules. As the WoT architecture will generate URIs to address Things, a distributed database will be suitable for routing request based on their URI. Given a query, servers can first search in this database if this resource has been discovered beforehand. If this were the case, no discovery protocols would be needed; otherwise, a discovery protocol would start a search to find the resource and once discovered, the resource URI would be stored in the distributed database. Servers may also have local storage to implement the storage registration mechanism.

The WoT architecture presented in this paper is based on the architecture proposed by D. Guinard. This section is focused

on the architecture but we will highlight main similarities between two architectures.

As Guinard stated in his thesis, the upper layers presented in Fig. 7 do not hide lower layers and instead, they are development layers where users with different technical knowledge can develop applications on top of them.

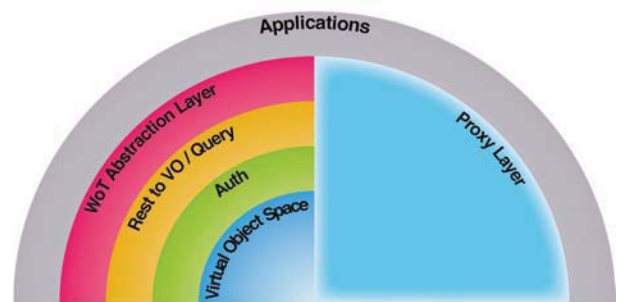


Figure 7: Applications can be developed over the upper layers

On the other hand, the definition of each of the layers of the proposed architecture and its correspondence with the Guinard’s proposal would be:

a) *Virtual Object Space*

The layer that shares more similarities in Guinard’s architecture is the Composition Layer, but in his approach this layer is a Physical Mashup where web services and enabled smart devices can create composite applications. In the proposed system and as in [20], this layer serves the purpose of digitally enabling smart devices for their use by the architecture and by WoT users. Then, applications built on the upper layer of the architecture (e.g. Physical Mashups), could use these digitally enabled smart devices.

b) *Auth*

This layer shares the same purpose as the Sharing layer exposed in Guinard’s thesis. The goal is to preserve the privacy of each Thing, allowing Things to have owners that share the capabilities of their Things. A great approach to this authentication and authorization layer would be the same as the one proposed in [21].

c) *REST to VO / Query*

The goal of this layer is to dynamically generate meaningful, RESTful URIs for VOs as they are queried for the first time. Once the device has a meaningful base URI, its capabilities

can be exposed through expanding its URI and the common HTTP verbs. Querying Things will involve complex semantic processing and a process such as the described in [22].

d) *WoT Protocol Abstraction Layer and Proxy Layer*

As several protocols can join the WoT architecture, there is a need to unify those protocols at the entrance of the proposed architecture (WoT Protocol Abstraction Layer), translating those protocols to an internal language. This will allow the development of the inner layers of the architecture regardless of the protocol Things are using to connect to the architecture. This layer would accomplish the functions of the Accessibility Layer proposed by Guinard. In figure 8, the correspondence between the layers defined by Guinard and the layers of the proposal is presented.

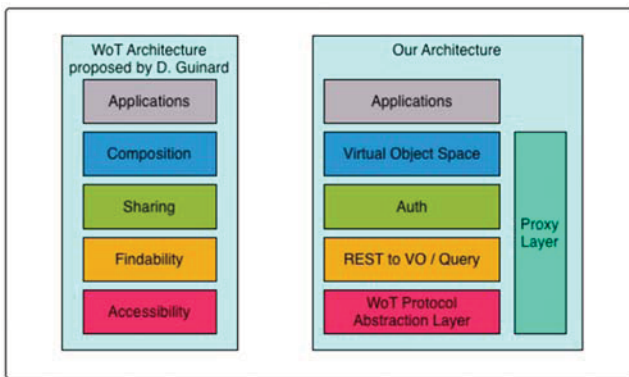


Figure 8: Comparison between the proposed architecture and Guinard’s one

As the inner architecture should be agnostic from the outside protocols (those used by Things and the architecture to exchange information), there is a need to develop an internal format and mechanism (Proxy Layer) to pass messages between different nodes of the architecture. The next section shows a proof of concept of this idea. Firstly, the used dataset for testing is described and, secondly, the discovery algorithm is depicted. Finally, two implementation approaches are presented and the results obtained are discussed.

VII. PROOF OF CONCEPT

A. Dataset

The data used for performing the tests of this proof of concept come from public data from the National Statistics Institute of Spain (<http://www.ine.es>) concerning cities/towns and inhabitants of Spain’s municipality. Specifically, all the inhabitants of the Catalonia region have been accounted (about 7.5 millions), taking into account the following two assumptions:

- In each house live an average of three people.
- Only 1% of households have some functional and Internet-controllable device.

Each house has been assigned to a city, and then a device is randomly assigned to each house. Different types of sensors are chosen, including temperature sensors, electricity meters, etc.

For each city it has been possible to obtain geolocation data (longitude and latitude) and it has been geographical represented on the map. In Figure 9 a snapshot (from the created application, called Semantic Web of Energy) can be observed. The control of these devices is displayed, indicating in color the answer to a request for consultation (Discovery of Thing).

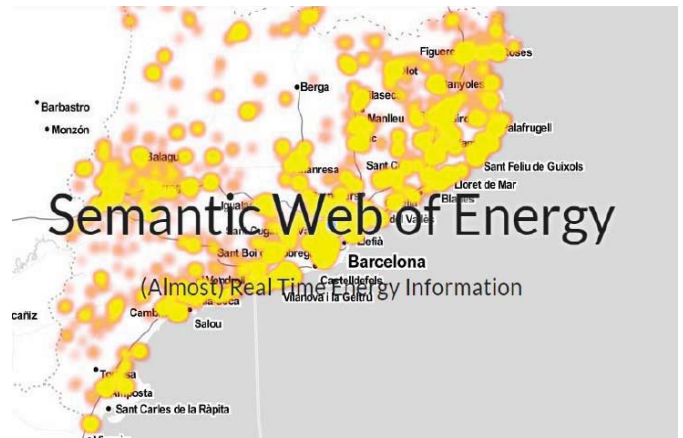


Figure 9: Snapshot of the representation of the devices response.

B. Algorithm

Based on the assumptions outlined above a first version of the architecture presented in this paper is implemented. In this sense we have implemented the first “Discovery of Things” functionality. It is simulated by the example of how to access a temperature sensor using the steps described in the following algorithm:

1. User A sends a request R1 to its local dispatcher node D1 (located in Madrid).
`P1 = {"action": "GET", "what": "temperature", "loc": "Carrer de Sants"}`
2. D1 processes de address (“Carrer de Sants”), searches for the Barcelona IP address and send a response to user A.
`R1 = {"request": {"what": "ws-conn" "to": "@IP-BarcelonaNode"}}`
3. User A sends a request P2 to the node with the IP address @IP-Barcelona.
`P2 = {"request": {"what": "ws-conn"}}`
4. Barcelona’s D2 dispatcher receives P2 and sends a response R2.
`R2 = {"uri": {"method": "GET", "uri": "/ws/id/1"}}`

5. User A connects to @IP-Barcelona/ws/id/1 via WebSockets (WS1 connection) and resends P1 through this connection. This connection will be managed by N1 node (server).
6. N1 processes the requests and searches for an address to forward the request and receive the temperature of Carrer de Sants. N1 has the private address of an aggregator/gateway (A1) that can provide the result of the query. N1 sends a CoAP request P2 to A1.

```
P2 = { "action": "GET", "what": "temperature" }
```
7. A1 processes the request and sends a response R3 to N1. R3 = { "uri": { "method": "GET", "uri": "/temperature" }, "data": { "value": 15, "unit": "celsius" } } .
A1 also saves an URI map to its local database: saveUriMap: "(GET)/temperature" -> (function to get data)
8. N1 receives R3 response and saves to its local database: saveUriMap: "/carrer-de-sants/" -> (GET)@IP-A1/temperature.
N1 also saves a query map to the distributed database: saveQueryMap: "Carrer de Sants, temperature" -> "(GET)@IP-N1/carrer-de-sants/temperature".
N1 sends a R4 response to user A.
R3 = { "uri": { "method": "GET", "uri": "/carrer-de-sants/temperature" }, "data": { "value": 15, "unit": "celsius" } } .
Moreover, if N1 has users that need the same data, sends R4 to these users too.
9. User A receives the requested data and a URI to identify the resource and made future requests to it.

If the same user A wants to request the same data again, he will send an URI request ((GET)@IP-N1/carrer-de-sants/temperature) directly to N1 and as query and uri maps are now present, the response will return faster as less processing time will be needed.

If user B sends the same request P1 to D1, D1 will respond saying that user B must request a WebSocket connection to N1 and send a request to this URI "(GET)@IP-N1/carrer-de-sants/temperature".

As shown, the framework is capable of handling the discovery of the new object and further processing and storage in the system for future reference.

C. Implementation

in this section two implementation approaches are compared:

1) First approach

As a lot of services found in the Web 2.0 are implemented using the PHP language¹ and in order to ease Web of Things applications at every layer for experienced PHP developers, we have developed a prototype implementing the algorithm presented in this section using the PHP language. Although we have succeeded in developing basic functionalities using PHP, several problems have arisen during the implementation.

PHP was born to serve CGIs, e.g. serve HTTP requests with dynamic content. The main workflow where PHP has been used consists in three steps: load, execute and die. For this reason, few efforts have been made to solve problems such as memory leaks or the fact that executing one PHP statement requires more low-level instructions than the actually needed ones.

In recent years, PHP has undergone some changes in its usages and performance:

- With the arrival of Node.js [43], some PHP developers started to implement libraries with the objective of allowing PHP users to create servers in this language as in React PHP [44], instead of relying on web servers like Apache HTTP.
- Compiled PHP frameworks such as the Phalcon Framework [45] with a high-performance boost on execution time have motivated the release of PHP 7, solving memory leaks and decreasing the number of low-level instructions needed to execute a PHP statement.

In order to build the prototype, we have used the Phalcon Framework and React PHP to boost performance. The Phalcon Framework presents high speed in performing operations and React PHP presents a novel manner for building PHP applications using the reactor pattern and asynchronous programming. However, the immaturity of React PHP and the lack of asynchronous libraries in PHP present an obstacle for developers to use even basic technologies such as MongoDB.

In this way, PHP has succeeded in building and fast-prototyping Web applications thanks to its low learning curve and its dynamic typing. However, its use at low-level/core layers in a WoT/IoT architecture where program correctness is crucial, facilitates the appearance of execution time errors.

2) Second approach

After evaluating other solutions in the market, the Scala [46] language has been selected for a re-implementation of the prototype. Although this language usage is not wider than PHP, Scala is experiencing an adoption growth².

Moreover, this language presents key characteristics that make it suitable for building a future WoT prototype architecture. Scala has also been chosen due to its research community. In

¹ <http://trends.builtwith.com/framework>

² <http://www.indeed.com/jobtrends?q=scala&l=>

fact, Scala was born at the EPFL (École Polytechnique Fédérale de Lausanne) thanks to Martin Odersky [46]. Therefore, continuous investigation is being made in order to optimize speed and provide better API's.

3) Comparison table

A comparison table between PHP and Scala is presented below (Table I). Data presented in this comparison table have been extracted from the authors' experimentation. Speed has been extracted from benchmarks done in [50].

Table I. Benchmarking PHP vs Scala languages

Feature	PHP	Scala
Natural workflow	Load, execute, die	Always running on a JVM
Speed	1	28 times faster
Backed by research	No	Yes
Libraries	PHP, C extensions	JVM compliant libraries
Distributed libraries	As C extensions	Akka and other JVM libraries
Async libraries/constructs	ReactPHP	Akka and Futures
Typing	Dynamic	Static

VIII. CONCLUSIONS AND FURTHER LINES

A new framework approach and a proof of concept have been presented in this paper. It has been shown that the proposed architecture is feasible and that the implementation of successive parts can be made using this design.

Although the results exposed are promising, we have realized that PHP lacks libraries and implementation for the most relevant WoT protocols. If there is a valid implementation, it only covers basic features of the protocol.

Moreover, the reactor approach was performed to prototype the architecture using ReactPHP and it was found that there are not libraries to connect the most commonly used databases in distributed systems like Redis or MongoDB, or the WoT protocols.

For the reasons exposed, it seems that PHP is not mature enough for the purpose, that is, to develop a holistic architecture for the Web of Things. A reimplementing of the architecture using the Scala language has been made, speeding up its performance and opening up the possibility to take advantage of robust libraries and frameworks built on top of JVM compliant languages.

Heterogeneity, parallelization and distribution as explained in [47] are also a key characteristic of a WoT architecture. More work has to be done to fully achieve these characteristics. The Actor Model [48] seems well suited to build an architecture

with such characteristics as asynchronous messaging, location transparency, distribution and concurrency as its core principles.

ACKNOWLEDGEMENTS

This work received funding from the European Union's 7th Framework Program under the FLICT-2011 Grant number 604677 - FINESCE (Future INternet Smart Utility ServiCEs).

REFERENCES

- [1] Navarro, J., Sancho, A., Zaballos, A., Jiménez, V., Vernet, D., Armendáriz-Iñigo E. "The Management System of INTEGRIS - Extending the Smart Grid to the Web of Energy", Proceedings of the 4th International Conference on Cloud Computing and Services Science pp: 329-336 (2014).
- [2] MongoDB database [available in Internet]: <https://www.mongodb.org/>
- [3] D. Guinard, V. Trifa, F. Mattern, and E. Wilde, "From the Internet of Things to the Web of Things: Resource-oriented architecture and best practices," in *Architecting the Internet of Things*, 2011, pp. 97–129.
- [4] Cheng Bo, Cheng Xin, Zhai Zhongyi, Zhang Chengwen and Chen Junliang "Web of Things-Based Remote Monitoring System for Coal Mine Safety Using Wireless Sensor Network". *International Journal of Distributed Sensor Networks*, Volume 2014 (2014), pp:1-14.
- [5] A. Zaballos, A. Vallejo, and J. M. Selga, "Heterogeneous communication architecture for the Smart Grid," *IEEE Network*, vol. 25, no. 5, pp. 30–37, 2011.
- [6] Agustín Zaballos, David Vernet, and Josep M. Selga, "A Genetic QoS-Aware Routing Protocol for the Smart Electricity Networks". *International Journal of Distributed Sensor Networks*, Volume 2013 (2013), pp:1-12.
- [7] INTEGRIS. (2011) INTEGRIS FP7 Project INTElligent Electrical Grid Sensor communications. ICT-Energy-2009 call (number 247938). [Online]. Available: <http://fp7integriss.eu>.
- [8] S. Aman, Y. Simmhan, and V. K. Prasanna, "Energy management systems: State of the art and emerging trends," *IEEE Communications Magazine*, vol. 51, no. 1, pp. 114–119, 2013.
- [9] D. Guinard, V. Trifa, S. Kamoukos, P. Spiess, and D. Savio, "Interacting with the SOA-Based Internet of Things: Discovery, query, selection, and on-demand provisioning of Web Services," *Services Computing, IEEE Transactions on*, vol. 3, no. 3, pp. 223–235, 2010.
- [10] D. Zeng, S. Guo, and Z. Cheng, "The Web of Things: A survey (invited paper)," *JCM*, vol. 6, no. 6, pp. 424–438, 2011.
- [11] FINESCE. (2015), European Union's 7th Framework Program under the FLICT-2011 Grant number 604677. Available: <http://www.finesce.eu/>.
- [12] E. Bou-Harb, C. Fachkha, M. Pourzandi, M. Debbabi, and C. Assi, "Communication security for Smart Grid distribution networks," *IEEE Communications Magazine*, vol. 51, no. 1, pp. 42–49, 2013.
- [13] V. C. Gungor, D. Sahin, T. Kocak, S. Ergüt, C. Buccella, C. Cecati, and G. P. Hancke, "A survey on smart grid potential applications and communication requirements," *IEEE Trans. Industrial Informatics*, vol. 9, no. 1, pp. 28–42, 2013.
- [14] J. Navarro, A. Zaballos, A. Sancho-Asensio, G. Ravera, and J. E. Armendáriz-Iñigo, "The information system of INTEGRIS: INTElligent Electrical GRId sensor communications," *Industrial Informatics, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2012.
- [15] Ataul Bari, Jin Jiang, Walid Saad and Arunita Jaekel "Challenges in the Smart Grid Applications: An Overview". *International Journal of Distributed Sensor Networks*, Volume 2014 (2014), pp:1-11.
- [16] Jesús Rodríguez-Molina, Margarita Martínez-Núñez, José-Fernán Martínez and Waldo Pérez-Aguilar "Business Models in the Smart Grid: Challenges, Opportunities and Proposals for Prosumer Profitability", *Energies* 2014, pp:1-30.
- [17] Josep M. Selga, Guiomar Corral, Agustín Zaballos, Ramon Martín de Pozuelo "Smart Grid ICT Research Lines out of the European Project INTEGRIS". *Network Protocols and Algorithms*, Vol 6, No 2 (2014).

High Performance Web-of-Things Architecture for the Smart Grid Domain

- [18] Andreu Sancho-Asensio, Joan Navarro, Itziar Arrieta-Salinasc, José Enrique Armendáriz-Íñigo, Virginia Jiménez-Ruano, Agustín Zaballos, Elisabet Golobardes, "Improving Data Partition Schemes in Smart Grids Via Clustering Data Streams" Expert Systems with Applications, Volume 41, Issue 13, 1 October 2014, pp:5832–5842.
- [19] "Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020" STAMFORD, Conn., December (2013). Available: <http://www.gartner.com/newsroom/id/2636073>
- [20] Oriwoh, E., & Conrad, M. (2015). 'Things' in the Internet of Things: Towards a Definition. *International Journal of Internet of Things*, 4(1), 1-5.
- [21] Guinard, D. (2011). *A Web of things application architecture* (Doctoral dissertation, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 19891, 2011).
- [22] EVERYTHING - Every Thing Connected. (n.d.). Retrieved April 13, 2015, from <https://evrything.com>
- [23] Tracey, D., & Sreenan, C. (2013, May). A holistic architecture for the internet of things, sensing services and big data. In *Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium on* (pp. 546-553). IEEE.
- [24] Node-RED. Retrieved April 13, 2015, from <http://nodered.org>
- [25] Octoblu | We make all APIs, Platforms and Devices talk to each other. Easily. Retrieved April 14, 2015, from <http://octoblu.com>
- [26] Neura. Retrieved April 14, 2015, from <http://www.theneura.com>
- [27] TempoIQ. (n.d.). Retrieved April 14, 2015, from <https://www.tempoi.com>
- [28] Bovet, G., & Hennebert, J. (2013, June). A web-of-things gateway for knx networks. In *Smart Objects, Systems and Technologies (SmartSysTech), Proceedings of 2013 European Conference on* (pp. 1-8). VDE.
- [29] MQTT. Retrieved March 27, 2015, from <http://mqtt.org/>
- [30] CoAP, A. P. (2012). Coap: An application protocol for billions of tiny internet nodes.
- [31] Data Distribution Service Portal. Retrieved April 6, 2015, from <http://portals.omg.org/dds/content/data-distribution-intro>
- [32] RFC 6455 - The WebSocket Protocol. Retrieved March 27, 2015, from <https://tools.ietf.org/html/rfc6455>
- [33] Christophe, B., Verdot, V., & Toubiana, V. (2011, September). Searching the 'Web of Things'. In *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on* (pp. 308-315). IEEE.
- [34] Kegel, D. The C10K problem. Retrieved March 31, 2015, from <http://www.kegel.com/c10k.html> - nb.epoll
- [35] Web Server Performance Comparison. Retrieved March 31, 2015, from http://wiki.dreamhost.com/Web_Server_Performance_Comparison
- [36] Schmidt, D. C. (1995). Reactor: An object behavioral pattern for concurrent event demultiplexing and dispatching.
- [37] Y. Oualmakran, J. Meléndez, S. Herraiz, M. López-Perea, E. González, "Survey on knowledge based methods to assist fault restoration in power distribution networks", ICREPQ, 2011. Energy
- [38] Craig Miller. "The Fractal Grid: Achieving Grid Security, Reliability, and Resiliency Through Advanced Analytics and Control". National Rural Electric Cooperative Association. December 2013.
- [39] Craig Miller, Maurice Martin, David Pinney, and George Walker. "Achieving a Resilient and Agile Grid". National Rural Electric Cooperative Association. April 2014.
- [40] The XMPP Standards Foundation. (n.d.). Retrieved June 22, 2015, from <http://xmpp.org>
- [41] Tech pages/IoT XepsExplained. (n.d.). Retrieved June 22, 2015, from http://wiki.xmpp.org/web/Tech_pages/IoT_XepsExplained
- [42] Efficient XML Interchange Working Group. (n.d.). Retrieved June 22, 2015, from <http://www.w3.org/XML/EXI/>
- [43] "Node.js." Node.js. Web. 25 June 2015. <<https://nodejs.org>>.
- [44] "ReactPHP". Web. 25 June 2015. <<http://reactphp.org/>>.
- [45] "A Full-stack PHP Framework Delivered as a C-extension." Phalcon PHP. Web. 25 June 2015. <<https://phalconphp.com>>.
- [46] "The Scala Programming Language." The Scala Programming Language. Web. 25 June 2015. <<http://www.scala-lang.org>>.
- [47] Odersky, Martin, Philippe Altherr, Vincent Cremet, Burak Emir, Sebastian Maneth, Stéphane Micheloud, Nikolay Mihaylov, Michel Schinz, Erik Stenman, and Matthias Zenger. *An overview of the Scala programming language*. No. LAMP-REPORT-2004-006 (2004).
- [48] Miorandi, Daniele, Sabrina Sicari, Francesco De Pellegrini, and Imrich Chlamtac. "Internet of things: Vision, applications and research challenges." *Ad Hoc Networks* 10, no. 7, pp: 1497-1516 (2012).
- [49] Hewitt, Carl, Peter Bishop, and Steiger Richard, 'A Universal Modular Actor Formalism for Artificial Intelligence', in Proceedings of the 3rd International Joint Conference on Artificial Intelligence (Morgan Kaufmann Publishers Inc., 1973).
- [50] The Computer Language Benchmarks Game [Available online] <http://benchmarksgame.alioth.debian.org> (2015).



Master in Teacher Training: A real implementation of Active Learning



Xavi Canaleta^a, David Vernet^a, Lluís Vicent^{b,*}, José Antonio Montero^a

^a Engineering Department, La Salle, Ramon Llull University, Barcelona, Spain

^b La Salle Open University, La Massana, Andorra

ARTICLE INFO

Article history:

Available online 24 October 2013

Keywords:

Active Learning
Project-based learning
Competence evaluation method
Technical impact

ABSTRACT

Teaching methods based on lectures often result in student passivity rather than pro-activity with the only goal of the student being to pass the final exam. Consequently, content retention is temporary and true learning is not achieved. Lack of student motivation can be solved by using Active Learning methodologies: serious games, Project-Based Learning (PBL), blended learning, etc. In addition, these methodologies enhance the development of the competences of students and provide a better evaluation of outcomes, provided adequate tools are used. However, apart from this, a more profound use of Information and Communications Technology (ICT) is needed to improve the effectiveness of learning and to create a learning system adapted to our current society. Learning Management Systems (LMS), social networks and blogs are indispensable for the efficient application of innovative learning methodologies. The Master in Teacher Training (MTT) provides an ideal case study as the most relevant features of Active Learning are present. The main contributions to this learning environment have been the incorporation of Active Learning methodologies (using real scenarios and project-based collaborative learning), an innovative method for assessing the Master and the impact of technology in learning. All of these have resulted in very positive results in relative to academic marks, surveys and entrepreneurship.

© 2013 Published by Elsevier Ltd.

1. Introduction

Until recent times the traditional teaching model considered learning to be a process in which the student accumulates knowledge and skills; and the teacher is the person in charge of transferring this knowledge. Thus the traditional teaching model is essentially based on the actions of the teacher, with the student taking a passive role, mostly limited to listening and taking notes.

Present day cognitive science defends a model which is clearly very different to the traditional one, based on the knowledge that we have acquired on the way people learn. Authors such as [Vygotsky \(1978\)](#), [Piaget and Inhelder \(1973\)](#) or [Gardner \(1983, 1993\)](#) have laid the foundations for a new model associated with the way we learn. According to the new focus, all learning should be based on individual students actively constructing their own knowledge ([Biggs, 1996](#)). This idea promotes the application of different teaching methodologies in the classroom, in which the role of the teacher changes substantially with respect to that of a traditional lecture. When strategies such as cooperative learning, case studies, problem or project-based learning are applied in the classroom, the main task of the teacher is to guide students through this

learning process, helping them to discover for themselves the knowledge they must learn, and using the tools which best suit the character and personality of each individual student. See [Feden & Vogel \(2003\)](#) Feden & for more detail on these methodologies.

It could be said that in the teaching–learning binomial, the traditional teaching model is based on teaching, while the current model focuses its attention on learning. In a lecture the message that the teachers provide is the objective of the class, whereas when the more active strategies mentioned in the previous paragraph are employed the aim is for the student to experience a complete learning experience.

In an ideal situation the training of future teachers should include all of the previously mentioned techniques. Furthermore, in order to make the course more effective, some of these training techniques should be applied in the course itself so that future teachers not only acquire the knowledge of the strategies, but experience them for themselves.

In an effort to achieve an effective learning experience in the area of teacher training, different Active Learning strategies have been applied. Innovate teaching methodologies have been employed and a competence-based evaluation system has been created. This new evaluation system should be able to validate whether the learning objectives have been acquired or not. None of this would have been possible without the intervention of the ICTs in the process.

* Corresponding author.

E-mail addresses: xavic@salleurl.edu (X. Canaleta), dave@salleurl.edu (D. Vernet), vicent@uols.org (L. Vicent), montero@salleurl.edu (J.A. Montero).

2. Framework

The Master in Teacher Training (MTT) is a professional MA program for teachers who teach in Secondary Education (*Educación Secundaria Obligatoria* – ESO), Baccalaureate and Vocational Training (*Formación Profesional* – FP). It has been the program in Spain since 2009 and replaced the former teaching diploma (called CAP – *Certificado de Aptitud Pedagógica*), which ceased to be valid in October 2009.

The Master's Degree in Teacher Training in Secondary Education, Vocational Training and Language Training is a certified MA program whose main objective is to train participants to be competent professionals in the field of pedagogy and didactics so that they can competently perform their teaching duties in a variety of environments.

The course has 60 ECTS-credits¹ in accordance with Article 5 of the Order ECI/3858/2007². The general structure of Master should include at least the following modules:

- A basic Multidisciplinary Module of 12 ECTS-credits applicable to all specialties of the Master.
- A Specific Module of 24 ECTS-credits for each of the specialties.
- An internship of 16 ECTS-credits including a Specialty-Specific Master's Dissertation.

The Master in Teacher Training is essentially a powerful tool that should serve for the training of future teachers of Secondary Education, Baccalaureate or Vocational Training. But the effectiveness of this tool is directly related to the way these studies are designed, implemented and developed. If qualified teachers fall into the trap of past teaching habits and commit the same mistakes (a poor development of the contents or a short period of placement in a school, for example) the MTT may lose its effectiveness. Thus the success of the training depends on each university rather than the government regulations in force.

The first important contributions made in the field of MTT were the modifications to the overall structure of this Master in order for it to be validated by ANECA, *Agencia Nacional de Evaluación de la Calidad y Acreditación* (National Agency for Quality Evaluation and Accreditation) and thereafter, by AQU, *Agència per a la Qualitat del Sistema Universitari de Catalunya* (Agency of the Quality of the University System of Catalonia).

In line with the demands of the regulation ECI/3858/2007, the overall structure of the three modules proposed an adjustment in the division of the ECTS-credits which reduced the number of credits obtained in the specific module subjects in order to achieve a more homogeneous structure. The end result can be seen in Table 1. This adaptation permitted the standardization of the workload of the materials to be studied (both the generic and the specific module) to 5 ECTS-credits. Additionally, the expansion of the number of module credits of Practicum and Master's Thesis to 20 ECTS-credits led to improved educational work placements, considered essential for the training of future teachers.

The dual role of the students makes this Master unique. They are both students and teachers and in this active training MA program they will develop individual teaching strategies. This is where the learning goal can really be exploited and makes the

Table 1

MA program structure (Ramon Llull University).

Multidisciplinary module (15 ECTS)	
Personal Development, Education and Society	
	<i>Subject 1 (5 ECTS)</i>
	Learning and Personality Development
	<i>Subject 2 (5 ECTS)</i>
	Processes and Contexts
	<i>Subject 3 (5 ECTS)</i>
	Society, Family and Education
Specific Module (25 ECTS)	
Advanced Training for the Teaching of Technology	
	<i>Subject 4 (10 ECTS)</i>
	Course Training Contents: Technology
	4.1: Objectives and Location Technology
	(5 ECTS)
	4.2: Advanced Supplementary Curriculum Contents of Technology
	(5 ECTS)
	<i>Subject 5 (10 ECTS)</i>
	Learning and Teaching Technology
	5.1: Technology in the Social Context
	(5 ECTS)
	5.2: Learning Methodologies and Evaluation Methods
	(5 ECTS)
	<i>Subject 6 (5 ECTS)</i>
	Teaching Innovation and Introduction to Educational Research
Practicum and Master's Dissertation (20 ECTS)	
	<i>Subject 7 (14 ECTS)</i>
	Placement in a School
	<i>Subject 8 (6 ECTS)</i>
	Master's Dissertation

program the ideal scenario for applying Active Learning (Biggs, 1985), the awareness and the understanding of the phenomenon of learning itself, because the course objective for these students is the learning process itself.

3. Learning methodology

A notable aspect in the implementation of MTT is the learning methodology used in the development of the Master. If our aim is for the future teacher to apply Active Learning strategies in order to obtain an effective learning process, then he or she should be trained using the same techniques. The MTT had the great advantage of starting from scratch. It did not inherit previous curriculum. It was therefore the ideal environment to implement a major methodological change.

The level of satisfaction expressed by the students is correlated by the teacher's perceptions of the effectiveness of group work (Oakley, Hanna, Kuzmyn, & Felder, 2007). Therefore the training of teachers in these methodologies is essential.

The main features of the learning methodology applied on the MTT are the following:

- The Master in Teacher Training of Ramon Llull University develops its materials promoting cooperative learning.
- Collaborative Project-Based Learning Methodologies are used.
- Final exams have been replaced by class presentations of projects (and continuous evaluation).
- Projects are subject to review and evaluation by teachers and other students.

A great number of examples (Mitchell, Canavan, & Smith, 2010) show that the use of PBL guarantees both academic success and the satisfaction of the students.

However, special attention must be given to the learning strategy which was initially employed in the Master in the academic year 2009–2010 and which has been annually reviewed and improved upon until the current course 2012–2013. This

¹ ECTS: European Credit Transfer and Accumulation System. One academic year corresponds to 60 ECTS-credits that are equivalent to 1500–1800 h of study in all countries irrespective of standard or qualification type and is used to facilitate transfer and progression throughout the European Union.

² ECI/3858/2007 of 27 December, establishing the requirements to verify official university degrees which enable qualified graduates to practice in the professions of Teacher of Secondary School Education, Vocational Training and Language Teaching. BOE n. 312, 2007.

experience unites 3 subjects of the specific module of the Master: Technology in the social context, Learning methodologies and evaluation methods and Teaching innovation and introduction to educational research (Subjects 5.1, 5.2 and 6 of Table 1). Each of the subjects develops its own content and activities but all of them are focused on a common transversal project shared by the three subjects: the design and implementation of a didactic unit for the subjects of Secondary Education, Baccalaureate and Vocational Training.

In this way, each subject provides support to the part of the project which is related to the contents which it develops. Subject 6 (Teaching Innovation and Introduction to Educational Research) aims to promote/boost teaching innovation within the format of the unit. Subject 5.1 (Technology in the Social Context) is centred on the incorporation of ICT in the development of the teaching content. Finally, subject 5.2 (Learning Methodologies and Evaluation Methods) provides support to design a suitable evaluation methodology.

The academic marks of the first semester and the academic background of the MA students (architecture, engineering, communication, etc.) are the factors considered in the formation of the groups. Our aim is to make these groups as heterogeneous as possible in order to maximise collaborative learning. ECOA software (Canaleta & Solé, 2012) has been used in the group formation process.

These heterogeneous groups are the same for the three subjects (5.1, 5.2 and 6). Each subject has a specific study program in order to develop the common transversal project. Furthermore, a series of short support activities are incorporated into the program to enable students to acquire different techniques, tools and contents. Checkpoints are set throughout the semester to monitor the progress of the project.

Finally, the group has to write a report in which all aspects of their project are fully explained. As will be seen in later evaluation methods employed, each group will present their work in front of the other groups and the directors of the three subjects and they will also upload all the teaching material they have generated to the LMS so that it can be evaluated.

4. Evaluation

Evaluation, as a tool to measure to what degree the student has acquired the contents, is a fundamental part of the teaching–learning system.

This section describes the evaluation in three stages. First, the overall evaluation of the Master which is dictated by the Ministry of Education and where there is little room for manoeuvrability. Secondly, the evaluation of each module which uses a completely innovative method is described. Finally, a detailed description of the evaluation of each subject is provided and an explanation of the procedure for re-taking competences in the case of a student not passing.

4.1. Global evaluation of the MTT

As explained in the MTT study plan and according to current legislation, the MA program must be globally evaluated according to the 3 numerical grades awarded for each module. The final global mark is obtained by averaging these marks weighting them according to the number of ECTS-credits the subject has. The following formula is used to calculate the final mark, *MTT*, where *MM* is the mark for the Multidisciplinary Module, *SM* is the mark for the Specific Module and *PMD* is the mark for the Practicum and the Master's Dissertation.

$$MTT = 15/60 \cdot MM + 25/60 \cdot SM + 20/60 \cdot PMD$$

For this formula to be applied, students must have passed each of the modules separately. For example, if a student obtains a mark of 8.4 for the Multidisciplinary Module, 7.5 for the Specific Module and 9.0 for the Practicum and the Master's Dissertation, the final mark of the MTT will be 8.2. Marks are out of 10, following the Spanish grading system.

In the event of a student not passing any of the three modules, he or she re-takes the failed module according to the procedures which will be explained later. These mechanisms are not based on written exams, but on activities which enable the student to acquire the previously failed competences. It should be noted that the MA program runs from October to June and the re-take period is from July to September.

The evaluation of the MTT based on the independent marks of the 3 Modules could be considered a limitation. In fact, alternative methods were investigated with the aim of creating a more global evaluation of the whole MTT program, instead of a partial evaluation separated by Modules. However, this would affect the mechanisms of repeating and consequently be more of a hindrance than a help for students who do not pass one or more of the MTT modules.

4.2. Pure competence evaluation for each module of the MTT

The evaluation of each module is based on a system labelled pure competence evaluation (Canaleta, Vernet, & Zaballo, 2011). In the study program of the Master each module was assigned a set of competences to be developed, and in turn, evaluated (Table 2 shows the set of the 12 specific competences which the student has to acquire). The competences are incorporated into the different subjects which make up the module. Hence every subject must assess the competences assigned to it. The importance of each competence varies depending on the subject into which it has been incorporated. Matrices have been created which determine the weight of each competence within the subject (see Table 3). The final mark of a competence is based on the weighted average of all the marks of each subject. Finally, in order to achieve the final grade of a module, the arithmetical average of all the competences evaluated in that module is calculated. At this point, the competences are not weighted given that all are considered essential and that no single competence is more important than another.

In order to fully understand the evaluation method we have chosen the specific module subject of Technology. The five coordinators of each subject and the coordinator of the Specific Module reach a consensus of which subjects will be included in each module and how they will be weighted. Table 3 shows the distribution and weighting of the competences. In this example of the Specific Module described, the final mark of the module is calculated by averaging the 12 competences.

$$SM = \frac{\sum Ei}{12} \quad i \in \{18, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56\}$$

This process is also used for the Multidisciplinary Module and the Practicum and Master's Dissertation.

The subjects do not have their own individual mark. While subjects are not passed or failed, the competences assigned to them may be passed or failed. The subject is only the means to facilitate the acquisition of competences within the module. Therefore, there is no point in awarding the subject a mark.

4.3. Competence evaluation for each subject

Once the evaluation method has been defined for the module as a whole, the evaluation process and procedure for each competence associated to a specific subject must be refined. Changes and improvements have been made to this process over the past three years of the MTT and although each subject of the MA program is

Table 2
Description of the competences for specific module.

Competences required for completion of specific module
E18) Knowledge of the context and the situations where the different subject contents are applied and used
E46) Acquisition of knowledge of the current education system and how it works
E47) Location of technology within the state education system
E48) Affirmation and clarification of the differences and similarities between technology subjects and the rest of the subjects in the education system
E49) Knowledge of what science and technology is and an awareness of the ethical repercussions of their development and how they are related
E50) Recognition of the main difficulties in the learning process of technology subjects and the ability to design activities which enable students to learn and are of a varied nature
E51) Transformation of subject knowledge into technological teaching proposals to be applied in the teaching–learning process of technology
E52) Ability to arise and stimulate interest and detect vocations in the technological area
E53) User knowledge of the main functions of a computer for teaching and learning
E54) Ability to use the computer and the variety of resources available to teach science and technology
E55) The planning, development and evaluation of computer-based activities which facilitate the acquisition of competences from the area of technology
E56) Evaluation of the new technologies (ICTs) as useful and essential tools to find out information and to carry out the simulation of experimental situations

Table 3
Distribution of competences and weighting in Specific Module.

Competence	Subj 4.1 (%)	Subj 4.2	Subj 5.1	Subj 5.2	Subj 6
E18	30				70
E46	100				
E47	100				
E48	100				
E49		100			
E50	15	30		35	20
E51	10	20	20	25	25
E52	30	35			35
E53		40	60		
E54		30	40	15	15
E55		20		40	40
E56		40	60		

Table 4
Evaluation of competences in subject 5.1.

Competence E54	Weight (%)
Activity 3. Classroom presentation software	20
Activity 3. Online presentation software	20
Activity 5. Classroom Response Systems	10
Global Activity. Teaching Unit	50
Competence E56	Weight (%)
Activity 1. Questionnaire on motivation	10
Activity 4. Blog posts	35
Activity 5. Social Networks	15
Activity 13. Valuation survey	40

free to determine its own evaluation method for the assigned competences, a common evaluation procedure has been agreed upon.

Initially, in order to develop the competences within a subject, the teacher designed a set of activities in which one or more of the assigned competences could be undertaken, and consequently evaluated. For each activity, the teacher had to specify which competences were being evaluated and how the marking was weighted. Table 4 illustrates an example of this with the evaluation of two competences from subject 5.1, Technology in the social context.

For the evaluation of the collaborative projects an evaluation method based on what has come to be known as Peer to Peer (P2P) inter-group and Peer to Peer intra-group is used. A combination of these two techniques has enabled us to meet a double objective: an effective evaluation of the learning has been obtained, but furthermore, the student has acquired a learning of the evaluation method (Canaleta, Zaballo, & Vernet, 2012). This is important as it is one of the intrinsic objectives of the MTT.

The Peer to Peer inter-group is a technique whereby each project group grades the projects of the other groups in the MA program. To do this, once all the projects have been submitted and duly presented, each group must agree upon the marks of the projects of the other groups. The teacher will provide students with a guideline and some pointers (generally through rubrics) to guide them through the evaluation process. Once all the marks have been submitted, the arithmetical average of all the marks provides the evaluator the final grade. It should be pointed out that the teachers of the subject have also graded the projects and carried out the evaluation according to the competences of the project. The final mark of each activity or project for each competence is the weighted average of the marks of the students and the teachers.

However, in order to obtain an individual mark for each member of the group, a Peer to Peer intra-group is employed. Once the project or group activity has been completed, each member of the

group has to divide 100 points between the other members of the group. This is obviously carried out in strict confidence. The points accumulated (see P2P# column in Table 5) are a multiplying factor to weigh the final individual mark of each student. Table 5 shows a summary of the results obtained by the 5 members of a group from the academic course 2011–2012. These are actually the final marks of the group project previously mentioned and evaluated in subject 5.1. As you can see, the final group marks for the competences E51, E53 and E54 remain independent as the P2P intra-group factor (P2P# column) is used as a multiplying factor. Here we can see how the poor evaluation of Student 5 from fellow group members led to his mark dropping from 9.2 to a final mark of 5.1 for the E51 competence ($9.2 \times 56/100 = 5.1$). Student 5 obtained 15, 23, 8 and 10 points (56 in total) from the other members of the group.

Evaluation is conceived as a stimulus for the learning process. If we consider that the evaluation should value the process of learning more than the product itself (McDonald, Boud, Francis, & Goncz, 2000) then this should be continuous throughout the period of study. However, this poses a problem when it comes to re-sitting. If a student fails a module, can he or she re-sit it during the same academic course? Common sense would say that should not be possible. In fact, there are certain tasks which cannot be re-taken if failed. For example, if a student fails the work placement carried out in an educational centre, he or she cannot re-take this part of the MA program until the following academic course.

In the rest of the MTT cases, a re-sit mechanism has been created based on reverse engineering. If a student fails a module, the competences which have caused him or her to fail are analysed. At this point, we can see in which subjects these competences were developed and evaluated. The teachers assigned to the subject will ask the student to do more activities which will enable him or her to improve the aforementioned competences. Given that these tasks will be carried out individually in the re-take period between July and September, the marking criteria may change due to the fact that the student did not carry out the work in a collaborative environment.

Table 5
Competence evaluation applying P2P inter-group^a and P2P intra-group[#].

	Teacher 1			Teacher 2			P2P Groups ^a			Final Group			P2P [#]	Final Student		
Students	E51	E53	E54	E51	E53	E54	E51	E53	E54	E51	E53	E54	P2P	E51	E53	E54
Student 1	9.5	10	9.5	9.5	10	10	8.5	8.8	8.9	9.2	9.6	9.5	107	9.8	10.0	10.0
Student 2	9.5	10	9.5	9.5	10	10	8.5	8.8	8.9	9.2	9.6	9.5	94	8.6	9.0	8.9
Student 3	9.5	10	9.5	9.5	10	10	8.5	8.8	8.9	9.2	9.6	9.5	121	10.0	10.0	10.0
Student 4	9.5	10	9.5	9.5	10	10	8.5	8.8	8.9	9.2	9.6	9.5	122	10.0	10.0	10.0
Student 5	9.5	10	9.5	9.5	10	10	8.5	8.8	8.9	9.2	9.6	9.5	56	5.1	5.4	5.3

5. Impact of ICT in learning

The ICTs take on a principal role in the effective and efficient implementation of Active Learning within the MTT. Technology itself does not improve learning but, with an effective application, it can (Vicent, Ávila, Riera, Anguera, Badia, & Montero, 2006). Technology is essential in order to successfully implement the methodology within this MTT program and to carry out the later evaluation of the modules and competences acquired.

Furthermore, in this Master program, another factor must be considered. Today's Secondary Education students are digital natives. Technology is integrated into their social context. It may even be inseparable from them. So, their learning environment includes ICTs. If we do not want a digital divide to exist between student and teacher, then the latter, who is not digitally native, should be familiar with the student's environment in order to integrate the learning process into this new context. This item will be covered in the sections that follow.

The framework for teacher knowledge is based on TPACK methodology (Mishra & Koehler, 2006). When this methodology is properly implemented with suitable resources and correct training, an extremely effective integration of the ICTs is achieved and a solid technological, pedagogical and content knowledge is acquired. This type of knowledge (TPACK, see Fig. 1) is actually the only real knowledge that a teacher has to have in order to teach with ICTs. It combines content knowledge with the most suitable pedagogical knowledge to teach such contents, using the most appropriate technology (which is not always the latest ICT) in order to make the learning experience as successful as possible.

5.1. The technological contents of the MTT

The profile of the student when it comes to digital competences has progressed over the last three courses of the Master. While the Master students cannot be considered digital natives, their grasp of ICT skills has improved. Subject 5.1, Technology in the social context, aims to train the future teachers in digital skills, to equip them with the user skills required to manage ICTs and to explore how the ICTs can be integrated into the learning environment of students.

Students are able to use a number different applications (such as Word processors, excel sheets, software presentations, video-editing, and image-editing) to carry out activities and projects with relative ease and without excessive technical problems. Other ICT applications which the average student cannot use well (blogs, Classroom Response Systems, interactive whiteboards, etc.) are included as part of the contents of subject 5.1 and through learning by doing the student learns the skills necessary to use them effectively, thus obtaining good training in ICT at the same time.

5.2. Impact of ICT in learning methodologies

Learning Management Systems (LMS) are a basic support tool for both e-learning and classroom based teaching, especially if Ac-

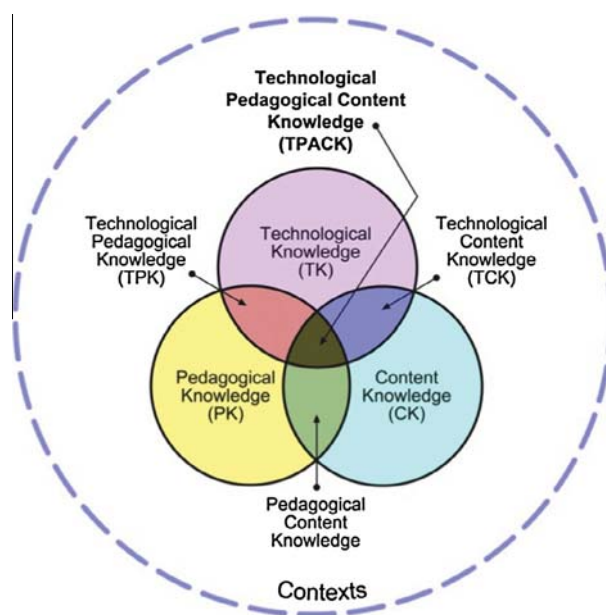


Fig. 1. TPACK model (reproduced by permission of the publisher, © 2012 by tpack.org).

tive Learning strategies, such as project-based learning in which interpersonal competences or skills are undertaken and evaluated, are to be developed (Vicent, Bou, Ávila, Riera, Montero, & Anguera, 2007).

We opted for the Moodle platform as the LMS to support our learning methodology. Throughout the development of the Master, the majority of Moodle applications were made available to the teacher and the student:

- Public forums to inform the students of the logistics and features of each subject.
- Private forums to facilitate communication and support in the project work of each group.
- Topics format to facilitate multimedia contents (documents, links, videos, etc.) and activities for the different sessions of each subject.
- Online repositories which enable students to deposit submissions and receive feedback on corrected work.
- Surveys to evaluate both the teachers and the subjects as structural features of the Master.

The ICTs also have a key role for the students in the learning process. The undertaking of tasks and the presentation of subject-related projects need the ICTs (Salomon, 2003). It is also true that technology has transformed the concept of the classroom (Pearlman, 2006) and has distanced it from the environment created for traditional lecturing, although possibly not quite enough yet.

However, it must be said that there is a growing tendency for students to use alternative ICT tools to those stipulated and provided by the academic institution to support the learning process. An obvious example would be the communication channels used between group members when carrying out group activities of projects. In theory, private forums were created by Moodle to monitor group progress while simultaneously provide members with a communication platform to carry out projects and associated tasks. However, over the course of the semester it became clear that students were using communication channels which were more natural to them, namely social networks (they preferred Twitter and Facebook), which they have on both their laptops and smartphones.

Another example of the change in ICT behavior of the students was the limited use of email. While the teaching staff tend to use email as the channel to inform students of any modifications in the planning or of any modifications to project work, it was observed that the students do not regularly read their emails. Research into this subject showed that email inboxes have become a repository for hundreds of trash mails of diverse origin and nature (spam, social network notifications, etc.) which the student is incapable of processing; therefore this communication channel is no longer valid. A Facebook group message is more effective than an email sent to a group of Master students.

Thus, it can be concluded that today's students use the same mechanisms for academic communication as they do for social communication. The students do not separate the social context from the learning or academic environment. This fact should make us question certain aspects of learning if we want to make it effective, given that learning no longer occupies a time or physical space, but is increasingly integrated into the social context.

5.3. Impact of ICT in evaluation

Moodle was used a support platform in the evaluation of the activities and projects. However, the Academic Information System of La Salle Engineering was not prepared to provide the pure evaluation of competences which was incorporated into the MTT program.

As a consequence, in addition to Moodle, an information system was designed to support this new evaluation system of the MTT. The System Manager MTT is software web which:

- Permits the definition of a Master, its modules, subjects and associated teachers.
- Permits competences to be assigned to each subject and calculates the percentages associated with each one.
- Facilitates each teacher with the ability to introduce qualifications for subjects and competence.
- Has different roles which provide different accesses to course data. There is a student profile which permits the student to consult both global marks and their breakdown from different views.

This system enables the student to see the exact breakdown of his or her marks and how the pure competence evaluation method works. The traditional academic management system, which was based on subjects, was prone to conceptual errors committed both by teachers when they uploaded marks and by students when they consulted and tried to interpret their results.

6. Results

The results obtained in the MTT indicate that the implementation of Active Learning is an effective learning model. The academic results of the three MTT courses 2009–2010, 2010–2011 and

2011–2012 are shown below. The results of the satisfaction surveys completed by the students on the quality of the subjects taught are also provided. Additionally, some cases of entrepreneurship as a result of the training received are described.

Table 6 shows the academic results of the students. The low drop-out rate in the three years which the MTT has been running is noteworthy. This could be a sign that the teaching methodology employed motivates the student and keeps him or her actively involved in the learning process. And, once the student becomes involved, the academic results are remarkable: zero failed students and high numerical qualifications (8.27 out of 10 on average in the last academic course for example). These academic results are the basis for the affirmation of the third point of McKinsey Report (Barber & Moursheed, 2007) which has always been a reference: "High performance requires every child to succeed". And, probably, could infer the second point in the report: "The only way to improve outcomes is to improve instruction".

We also carried out surveys to gauge the students' perception of the Specific Module over the past three academic years. The results show little variation in the results presented in Fig. 2 as a summary of the opinion of the students.

Fig. 2 shows how students rated the 5 subjects of the Specific Module of Technology. The results refer to the learning methodology and how up to date the contents are, as well as a global evaluation of the subject. Apart from the data observed, it should be noted that the three subjects included in the teaching strategy described in Section 3 received higher marks in teaching methodology than the other two (an average of 7.8 as opposed to 6.5).

Fig. 3 shows the perception of the time spent on each subject within the Specific Module. If we consider that the value of 5 would be the optimal workload with respect to the number of ECTS-credits, then we can conclude that the three subjects (5.1, 5.2 and 6) involved in this learning experience have a greater workload than the subjects (4.1 and 4.2) which were not considered (an average of 7.9 as opposed to 5.7). Even though these

Table 6
Academic marks of MTT program.

MTT course	Number Students	Drop-out rate	Failure rate (%)	Pass rate (%)	Mark average	Standard deviation
2009–2010	44	0.0% (0)	0.0	100	8.29	±0.55
2010–2011	26	4.0% (1)	0.0	100	7.94	±0.61
2011–2012	29	6.9% (2)	0.0	100	8.27	±0.66

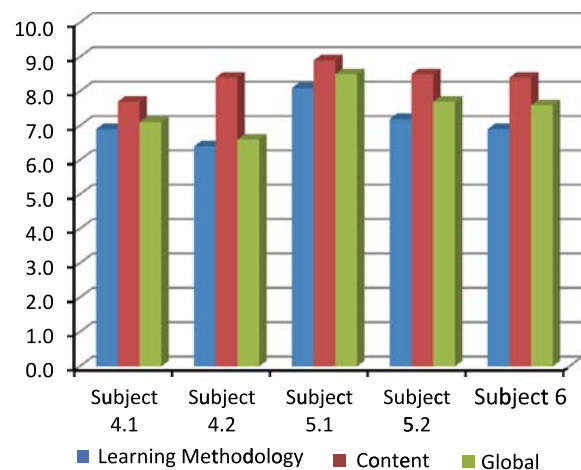


Fig. 2. Quality survey for the specific module in 2011–2012 course.

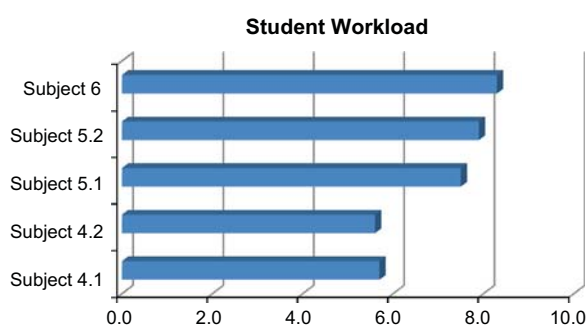


Fig. 3. Evaluation of student workload for the specific module.

subjects require more dedication, the students gave them a better global rating.

Finally, as well as training future teaching professionals, the MTT fosters innovation and entrepreneurship. There are two cases of entrepreneurship created by MTT students from the 2 previous academic courses which we consider worthy of mention:

- A student from the course 2010–2011 created Enhancing Ebooks SL³. Enhancing Ebooks is an editorial and digital producer of Apps for ipads and tablets in general, which develops and commercializes enhanced appbooks. As a producer of the apps, it aims to collaborate with authors in order to innovate and enrich the ways of communication and interacting with users, reader or consumers of the contents.
- Three students from the last course 2011–2012 have created Wårks⁴. Wårks Creative Contents describes itself as a team which creates innovative teaching contents in which one of the objectives is to maximise intrinsic motivation through elements of extrinsic motivation. They are currently working for the Universitat Oberta La Salle in the design of e-learning courses.

7. Conclusions

The Master in Teacher Training is an excellent opportunity to provide the future teaching staff for Secondary Education (ESO), Baccalaureate and Vocational Training with the ability to offer high quality teaching. The contents developed in the Master's program aim to ensure that the future professionals of the education sector are highly competent in areas such as new teaching methodologies, competence evaluation methods and teaching innovation.

There are currently a number of ICT tools which can be used to support collaborative project-based learning (Moursund, 2003). If PBL is to be implemented effectively then teaching staff must be properly trained in digital competences. As stated in the McKinsey Report (Barber & Mourshed, 2007) "The quality of an education system cannot exceed the quality of its teachers".

The MTT is not aimed at professionals currently teaching in educational centres. Future incorporations in the teaching staff of schools will have completed the Master, as part of their training. They should be considered as a drop of oil which can be spread over the education sector bestowing the new knowledge acquired wherever they go.

The new generations of students entering the education system will be even more integrated into the digital world. In this new reality the academic world is not divorced from the social context. There is an insertion of the learning environment within the social

context in which technology is a highly significant factor. This will make it necessary for all levels of the educational system to dominate the new social, educational and family models in order to tailor the academic processes to the learner and to make learning effective from infant school to university studies.

8. Future work

The huge quantity of academic data compiled from students in the MTT through the competence evaluations should permit us to obtain more knowledge about the students themselves. Research is being carried out in the area of educational data mining (Romero & Ventura, 2006) with the aim of processing the academic qualifications of past students in order to explore whether an analysis of the competences using Artificial Intelligence tools, as machine learning techniques, to classify students into different profiles or groups.

Another interesting idea to follow up on this work might be to study how effective the MA program has been for teachers who have recently completed the MTT and who are teaching at present. Data on their students' subsequent success could be used to evaluate the effectiveness of the MTT program.

The need to train teaching staff in the ICTs has opened another line of work: how to evaluate the degree of digital competence of the students before and after completion of the MTT. Different parameters are being defined to obtain a ratio which can quantify the ICT competences and allow us to see how they develop in the course of the learning process.

Finally, the efficiency of collaborative learning is based on the correct functioning of the work groups. The creation and monitoring of these groups is of vital importance and may be a key factor in the success of this type of learning. A current line of research (Vernet, Canaleta, & Pallàs, 2012) coincides with this opinion and aims to ensure the general effectiveness of the distribution of groups in the classroom.

Acknowledgments

The authors wish to thank the General Coordinator of the Master in Teacher Training, Dr. Pau López, and the Specialty Coordinator of Languages, Dr. Maria Gonzalez Davies, for their work, support and collaboration in the design and implementation of this Master and for their dedication over the past three academic years that have enabled the provision and implementation of these and other contributions in the field of educational innovation.

The authors would like to thank Lisa Kinnear (La Salle, Ramon Llull University) for the linguistic reviews of the paper.

References

- Barber, M., & Mourshed, M. (2007). *How the world's best-performing school systems come out on top*. McKinsey & Company.
- Biggs, J. B. (1985). The role of meta-learning in study process. *British Journal of Educational Psychology*, 55, 185–212.
- Biggs, J. B. (1996). *Enhancing teaching through constructive alignment*. Higher Education, vol. 32, issue 3. Kluwer Academic Publishers, pp. 347–364.
- Canaleta, X., & Solé, X. (2012). ECOA: Extracción de Conocimiento Original de los Alumnos, VII Congreso Iberoamericano de Docencia Universitaria, Ensino Superior: Inovação e Qualidade na Docência. Livro de Atas, Porto: VII Congresso Iberoamericano de Docência Universitária.
- Canaleta, X., Vernet, D., & Zaballos, A. (2011). *Propuesta de evaluación pura por competencias para estudios de ingeniería*, Proceedings of JENUI'2011, 3–10. Sevilla: XVII Jornadas de Enseñanza Universitaria de la Informática.
- Canaleta, X., Zaballos, A., & Vernet, D. (2012). *Evaluación del aprendizaje, aprendizaje de la evaluación*, VII Congreso Iberoamericano de Docencia Universitaria, Ensino Superior: Inovação e Qualidade na Docência. Livro de Atas, Porto: VII Congresso Iberoamericano de Docência Universitária.
- Feden, P. D., & Vogel, R. M. (2003). *Methods of teaching: Applying cognitive science to promote student learning*. Boston: McGraw-Hill.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.
- Gardner, H. (1993). *Multiple intelligences: The theory in practice*. Basic Books.

³ <http://enhancingebooks.com>.

⁴ <http://www.warks.es>.

- McDonald, R., Boud, D., Francis, J., & Gonczi, A. (2000). Nuevas perspectivas sobre la evaluación. *Boletín cinterfor*, no.149, pp. 41–72.
- Mishra, P., & Koehler, M. (2006). *Technological pedagogical content knowledge: A framework for teacher knowledge*. Michigan State University.
- Mitchell, J.E., Canavan, B., & Smith, J. (2010). Problem-based learning in communication systems: Student perceptions and achievement. *IEEE Transactions on education*, vol. 53, n. 4, pp. 587–594.
- Moursund, D. (2003). *Project-Based Learning Using Information Technology*, Eugene, Or.: International Society for Technology in Education (2nd ed.).
- Oakley, B. A., Hanna, D. M., Kuzmyn, Z., & Felder, R. M. (2007). Best practices involving teamwork in the classroom: Results from a survey of 6435 engineering student respondents. *IEEE Transactions on Education*, 50(3), 266–272.
- Pearlman, B. (2006). *New skills for a new century: Project-based learning teaches kids the collaborative and critical thinking abilities they'll need to compete*. Edutopia magazine.
- Piaget, J., & Inhelder, B. (1973). *Memory and intelligence*. London: Routledge and Kegan Paul.
- Romero, C., & Ventura, S. (2006). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*. s.l.: Elsevier, 2006, vol. Expert Systems with Applications.
- Salomon, G. (2003). Project-based learning: A primer. *Technology & Learning*, 23(6), 20–27.
- Vernet, D., Canaleta, X., Pallàs, G. (2012). Intelligent Tutoring on Virtual Learning Communities. International Symposium on Computers in Education (SIIE), Andorra.
- Vicent, Ll., Àvila, X., Riera, J., Anguera, J., Badia, D., & Montero, J.A. (2006). Do Multimedia Contents increase the effectiveness of learning? IEEE Frontiers in Education Conference. FIE 2006, San Diego, USA.
- Vicent, Ll., Bou, G., Àvila, X., Riera, J., Montero, J.A. & Anguera, J. (2007). Which are the best e-learning tools for an Engineering Degree in the European Higher Education Area? In: The 7th IEEE International Conference on Advanced Learning Technologies (pp. 882–886). IEEE Computer Society Press.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Research Article

A Genetic QoS-Aware Routing Protocol for the Smart Electricity Networks

Agustin Zaballos, David Vernet, and Josep M. Selga

Computer Engineering Department, La Salle Engineering, University Ramon Llull (URL), 08022 Barcelona, Spain

Correspondence should be addressed to Agustin Zaballos; zaballos@salle.url.edu

Received 22 April 2013; Revised 16 July 2013; Accepted 1 August 2013

Academic Editor: Ataul Bari

Copyright © 2013 Agustin Zaballos et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a QoS-aware routing protocol suitable for distribution of smart electricity grids based on heterogeneous machine to machine communications. The distribution Smart Grid needs high performance communication networks capable of handling QoS, an issue that is addressed by the present paper. The proposed algorithm is a merger between a genetic algorithm (GA) and Ticket-Based Routing (TBR), which is an on-demand routing protocol for ad hoc networks that provide quality of service. A suitable parameterization of the GA parameters is needed in order to use this protocol in the coming Smart Grid networks. The resulting routing protocol, named genetic algorithm with TBR algorithm for Smart Grids (GATAS), is an adapted intelligent evolution of the TBR. The performance of TBR has been improved by reducing the overhead of routing packets in the network and by minimizing the communication latency due to its on-demand behavior. Experimental evidence indicates that the likelihood of finding the optimum route using multiobjective dynamic metrics increases when the genetic algorithm is applied. In this paper, the main simulation results on the parameterization carried out are discussed, and the proposed attributes of the GA are described.

1. Introduction

The future of electrical utilities walks hand in hand with Smart Grids and their advantages. Smart Grids will save energy and will cope better with the unpredictable renewable energy supplies [1]. At present, utilities have to be prepared to face the increasing needs of their telecommunication infrastructures. In fact, one of the main challenges of the Smart Grid is to redesign the architecture of its communication network [2, 3]. The current utility grid scheme is relatively easy to operate, but the Smart Grid is much more complex. Its architecture is based on a decentralized scheme with elements logically identified but not geographically located. Future Smart Grids will manage great amounts of real-time information through a data network and will collect information from Intelligent Electronic Devices (IEDs) established for control purposes. This kind of data network is not exempt from the growing needs of quality of service (QoS) [4, 5]. Smart Grids are expected to face a drastic increase in information demand, communication, and various data such as voice, data, image, video, and multimedia communications, which will all have

to be accessed anywhere and at any time inside an M2M architecture [6, 7].

This work deals with the issues of utmost importance to achieve QoS-aware routing in wireless and wired sensor networks based on a genetic algorithm for the sensor networks of Smart Grids. A sensor network consists of distributed sensors that cooperatively monitor physical or environmental conditions. Those sensor nodes can be located anywhere in the network and form an ad hoc network, which does not require a communication infrastructure. In this environment, sensor networks must dynamically provide the necessary QoS depending on the type of information transmitted by sensor nodes in a multihop topology.

This paper presents a new algorithm that copes with these necessities. The genetic algorithm with TBR algorithm for Smart Grids (GATAS) evolves from an ad hoc QoS-aware routing protocol but uses a genetic algorithm (GA) [8, 9] to reduce the amount of routing traffic. The object of this paper is to propose a suitable parameterized GA integrated into a QoS-aware routing protocol for the Smart Grid ad hoc network. A QoS routing selects paths based on several QoS

metrics to satisfy specific requirements. This new routing protocol has been simulated using OPNET Modeler [10] in Smart Grid related scenarios [11] under the umbrella of the European project INTEGRIS [11]. The interdisciplinary project INTEGRIS addresses the development of an ICT infrastructure to handle the Smart Grid requirements. Thoroughly, INTEGRIS takes benefit from the profiles variety of its members and tackles the Smart Grid domain by proposing a global solution that considers (1) the QoS-aware communication network, (2) the ICT security issues, (3) the storage and distributed computation, and (4) a cognitive system as a self-containing block.

This paper is organized as follows. Section 2 briefly describes routing mechanisms for ad hoc networks and provides a general description of the network model used. Section 3 describes the fundamental topics involved in the work carried out and it also covers all the important design issues of our genetic QoS-aware ad hoc routing protocol for Smart Grid access networks. Section 4 introduces the characteristics of modeled routing nodes and simulation scenarios for the analysis and outlines the results obtained. Finally, Section 5 presents the conclusions of the paper.

2. Routing for Smart Grid's Data Network

Smart Grids will manage lots of real-time information through a data network, and they will collect information for control purposes from established IEDs. Smart Grid network control and monitoring are very important features in order to provide continuity [5, 12], QoS [4, 13], and security [14–16]. The future Smart Grid must be distinguished by self-healing and automation. Actually, international organizations, governments, utilities, and standardization organizations are becoming aware that the grid needs a modernization [3, 5].

Due to these circumstances, Smart Grid will be supported by highly heterogeneous data network with strict QoS constraints depending on the Smart Grid service to provide [3]. Therefore, one of the most important specifications required for Smart Grids is that regarding their communications. A framework for management of end-to-end QoS for all communications in the grid will be a must in the future [4, 7] and this specificity is something that is directly addressed by the proposal made in the paper. In fact, a suitable communication infrastructure increases the efficiency of the electric system to a much greater extent than automation without communication capacities could ever increase it.

There are several aspects that must be defined to obtain an algorithm that could be implemented in a real Smart Grid such as the detection of neighbors, the hierarchy of the network, the definition of which synchronization mechanism is used, the addressable elements in the network, or the address scheme used by the protocol to identify the nodes in the network [17, 18]. Furthermore, if the protocol is oriented to provide QoS, additional aspects have to be established, such as the QoS metric, the specification of the protocol to minimize the amount of bandwidth needed, and the load balancing scheme [4, 12].

Ad hoc network among objects is built and every sensor node may need to transmit information to other sensor nodes

and not only to the center node. If the network topology changes dynamically due to mobility and if the state information is inherently imprecise, the routing protocol must be optimized for ad hoc networking. Even if the network is wired or stationary, the network topology may change because of power network changes or degradation of channel characteristics, especially in the case of Power Line Communications (PLC) and also in the case of radio systems using common frequency bands. The main goals of a routing protocol for Smart Grids are simplicity, scalability, and energy efficiency. At present, topology changes due to node mobility are infrequent as sensor nodes are stationary in most applications [12, 17, 19].

2.1. Related Work. A routing protocol consists of two basic tasks: it has to collect the state information of the network and to keep it up to date. This paper is focused on the analysis of this first task inside the ad hoc network of Smart Grids and leaves the path repair functions for further study. Many alternative solutions have been proposed and analyzed to solve the need for a routing algorithm in ad hoc networks. The main features of well-known ad hoc protocols have been studied in depth [20].

In recent years, routing optimization in data networks has received considerable attention. There are several GAs in the literature that address different routing problems, such as multicasting routing problem [21], traffic engineering based on link weight optimization [22], or shortest path routing problem without providing QoS [23, 24]. Some of them are applied in a non-real-time background mode [23]. Far from the supposed full cooperation of the participating communication nodes, game theory has attracted the interest of researchers in the field of routing as well in order to monitor possible conflicting interests between communication domains [25]. Although the findings of these studies are relevant, those approaches are beyond the scope of this paper.

Recently, routing in Wireless Sensor Networks (WSNs) has been recognized as an important research area and much work has been carried out. As a result, a great number of studies have discussed the application issues of evolutionary computation techniques [26], clustering [27], and data mining [28] in this kind of networks. Some of them will be referenced along this paper. A survey of the main approaches to the application of evolutionary techniques in WSNs can be found in [29]. Related fields of knowledge worth to mention are those of Particle Swarm Optimization (PSO) and ant routing algorithms [30] that are similar to our proposal in that they use probes that explore the space but that differ in essential aspects such as the randomness of PSO versus the flooding-like nature of our proposal (GATAS).

To the best of our knowledge, our approach is the first real-time integration of a genetic algorithm with both routing parts: the routing algebra and the routing distribution mechanism for QoS-aware networks that focus on the Smart Grids necessities on QoS. An exhaustive study using multiple simulations to determine the routing multipath algorithm with the most adequate QoS behavior for High Voltage (HV) segments has been carried out in [18]. In this highly meshed network environment, where the communication devices are

very powerful, the main difficulty comes from improving the QoS behavior of the existing widely spread commercial routing protocols [13, 19]. However, the idea of designing a routing protocol appropriate for another segment of Smart Grids (medium and low voltage) is presented in this paper, in which routing protocols must operate under a set of constraints that traditional protocols do not typically consider. In this sense and given the similarities between Smart Grid networks and sensor networks, it is interesting to consider carefully the work done in the field of sensor networks.

Although there are many academic papers and well-known routing protocol implementations available based on ad hoc networks, studies and demonstrations carried out in [11, 31] formally discard all these protocols since they do not meet the minimum criteria needed for Low Power and Lossy Networks (LLNs) that are a class of networks in which both the routers and their interconnections are constrained [31]. If a protocol cannot meet these minimum criteria, then it cannot be used in several major Smart Grid application domains, and it is therefore unlikely to be a good candidate for use within a broader scope.

2.2. Network Model. In this section, the network model and the notation used for the routing algebra and policies are described. This notation is used to formally define the routing protocol behavior of GATAS, and it is based on Sobrinho's routing algebra [32]. An algebraic approach is very useful to both understand existing protocols and to explore the design space of future Internet routing protocols. The routing policy defines the elements used by the routing protocol to carry out the routing process (1). The routing policy (RP) is formed by

$$RP = \langle \Sigma, \oplus, L, \preceq \rangle. \quad (1)$$

Each element of this array (1) is defined in Table 1. Based on this representation, we propose the following model of a network, where vertex j is the destination and vertex i the origin of routing information (Table 2). The proposed notation is crucial for the protocol specification in order to define the information used and stored by the routing protocol. The objective is to avoid any confusion when different routing schemes and metrics are defined at a point in the future.

3. Description of the Proposal

3.1. Underlying QoS Routing Protocol Description. GATAS algorithm is based on a network layer on-demand routing algorithm known as Ticket-Based Routing (TBR) [33]. A ticket-based probing algorithm is an imprecise information model used to find a QoS-aware routing path in ad hoc networks. TBR is very interesting for our purposes as maintaining a consistent route table in Smart Grids has become increasingly challenging due to the number of nodes whose information has to be consistent and also because of the unpredictable changes in the actual topology mentioned in Section 2. It is often impossible to know a priori what kind of environment the protocol will find itself in. A QoS routing algorithm is, after all, a complex optimization problem. Therefore and in order to solve this complex problem, the

TABLE 1: Elements of the routing policy (RP).

Element	Description
Σ	It is the cost associated with a path, and it is known as the signature of the path.
\oplus	It defines the way to add a link cost to a path and to calculate the total cost. It is known as the metric operator.
L	It represents the cost associated with a link, and it is known as the label of the link.
\preceq	It is the precedence relationship, and it is used to decide which path is the best choice.

use of one of the best known techniques that has proved successful in these matters is proposed: a genetic algorithm. We evolve a QoS routing protocol using an artificial intelligent technique, and, for this purpose, reactive protocols are the most suitable kind of algorithms [20].

In the TBR routing protocol, the source node issues a certain number of tickets and sends these tickets in several probe packets to find a QoS feasible path. Each probe packet carries one or more tickets. This distributed QoS routing protocol probes multiple paths in parallel. The number of multiple paths searched is limited by the number of tickets issued by the source node in all the sent probe packets. State information maintained at intermediate nodes is used for more accurate route probing. If the available state information is not precise or if the QoS requirements are very stringent, more tickets are issued in order to improve the chances of finding a feasible path. In each probe, the probe state (signature's path and label's links) is recorded, including the path, the accumulated delay, and the accumulated cost of the path.

Figures 1 and 2 show a dialog example of the TBR mechanism as used by GATAS routing protocol. When source *Node 1* wants to find a path with some QoS requirements to the destination *Node 5*, it generates n tickets. Then, it has to distribute them among different probes delivered to every neighbor. In this example, *Node 1* only has one direct neighbor. The followed path by the probe $S_{n,p}$ is depicted in Figure 1 where the subindex n is the number of remaining probe's tickets and the sub-index p is the type of probe (probe request p or probe response r). When a probing message arrives at a neighbor, it may be split into multiple probes and forwarded again. The neighbor *Node 2* generates, in this example, two probes and distributes n tickets between its two neighbors (x tickets to *Node 4* and y tickets to *Node 3*). Each child probe will contain a subset of tickets from its parent. A probing message has to contain at least one ticket. In this way, when using a one-ticket probe, the node is not able to continue the splitting process any further, and the node can only forward it to one neighbor. When one probe arrives at the destination, the recorded path's signature is sent to the origin within a response probe (response probes $S_{y,r}$ and $S_{x,r}$ in Figure 2).

The study presented in this paper is focused on network level analysis (level 3). So, it is assumed that a link-level protocol assures that every node knows its neighbors, which

TABLE 2: Routing algebra notation.

Element	Description
i	It represents the origin node.
k	It represents a neighbor of node i , which has sent a routing advertisement to node i .
j	It represents the advertised destination of the routing information received.
λ_{ik}	It is the cost of the link from node i to node k .
σ_{kj}	It is the cost of the path from node k to node j advertised by node k .
$\tilde{\sigma}_{kj}^j$	It stands for the estimated cost of the path from node k to node j stored on the routing table of node i .
$\tilde{\sigma}_{ikj}^j$	It stands for the estimated cost of the path from node i to node j through the neighbor k stored in the routing table of node i .
$\tilde{\sigma}_{ij}^k$	It is the cost estimated of the path from node i to node j that node i guesses that is known by node k .
S_{ij}	It is a set of all the neighbor nodes of node i that are feasible successors to node j .
N_i	It is a set of all the neighbor nodes of node i .

Estimated values are the information received from the neighbors that can be potentially outdated due to network changes that have not yet been notified, as the routing protocol has not converged.

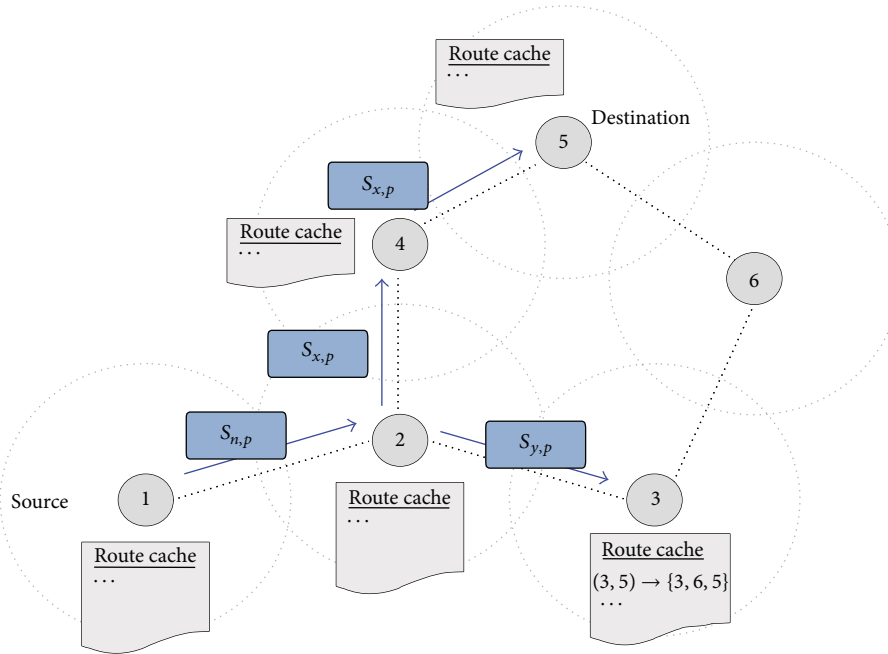


FIGURE 1: New search of a path (probes sending).

implies that a node detects within a finite time the existence of a new node or the loss of connectivity with a neighbor and all packets transmitted over an operational link are received correctly and in proper sequence within a finite time.

The study made in [31] discards all the analyzed ad hoc routing protocols, but it does not analyze TBR-based proposals, and, in fact, TBR and GATAS do not easily fit into any of the analyzed routing protocols. Baseline TBR table size is a function of the number of communicating pairs in the network, scaling with $O(\text{Destinations})$. As explained in [31], this is acceptable, and so TBR would pass the routing state criterion defined in [14]. As an on-demand protocol, TBR does not generate any traffic until data is sent; therefore, control and loss traffic is correlated with the data and so it receives a pass for the control traffic criterion. Furthermore,

TBR does not fail the link/node cost criterion because any QoS-aware metric can be used, and the router can indicate its willingness to route a packet to a destination.

Nevertheless, the criteria defined in [31] do not take into account the special behavior of the TBR algorithm. As the number of destination nodes and paths increases, the great number of probes needed to find different paths, especially if they require strict specifications of QoS, becomes a risk for the scalability of the protocol. This is because of the large number of routing packets that has to be transmitted in the whole network. Obviously, routing protocols must be able to send at least a very small amount of control traffic, in order to discover a topology. Nevertheless, this bootstrapping discovery traffic should be small, since most of the energy is consumed by both transmissions and receptions. This is why

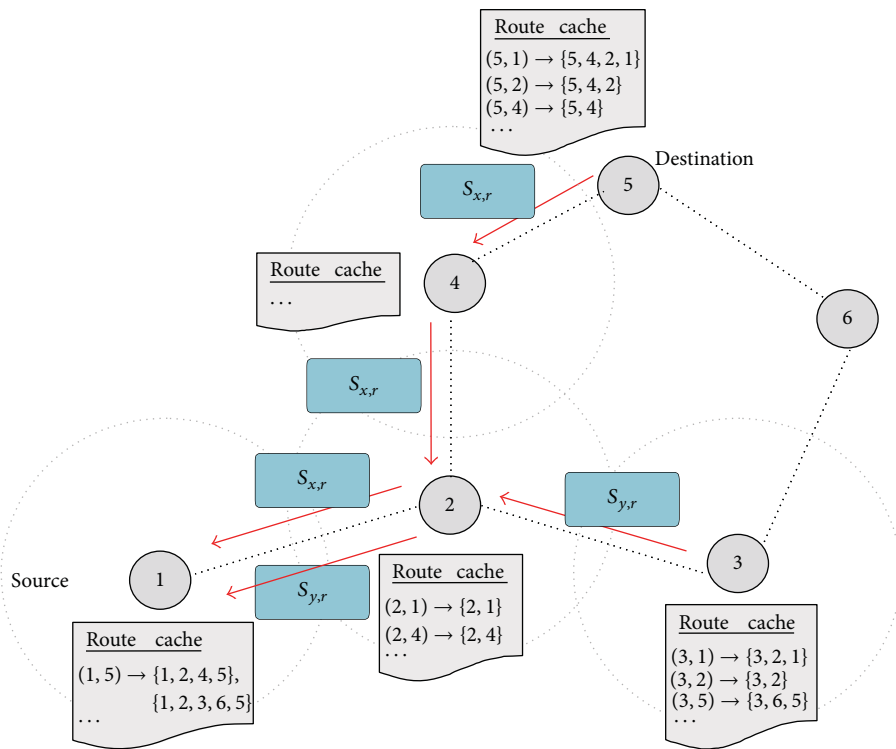


FIGURE 2: New search of a path (probes receiving).

evolutionary computation techniques are applied in GATAS to improve this TBR protocol aspect, although within the established limits in [31], it turns into the weakest point.

3.2. Underlying Genetic Algorithm Description. GATAS is a randomized forwarding TBR scheme that has been improved using a GA. The resulting QoS-aware routing protocol is decentralized and on-demand based. By using the GA, GATAS reduces latency and routing overhead due to the high number of tickets sent by baseline TBR for finding a valid route to the destination. It uses the parallelized multifocus population-based search provided by GA to search new solutions without any extra consumption of bandwidth. Each source node uses the GA, in an online manner, during the TBR's search phase for new routes, which meets the constraints of the communication to the destination. The main goal of this paper is that it bridges the gap between GA and QoS-aware routing protocol for a Smart Grid. The following section discusses the proposed solution.

3.2.1. Chromosome Codification. The first key element is to choose a proper codification of the chromosome, since it will determine the search space and the mobility through this search space. Several chromosome codification strategies have been successfully used in data network routing algorithms and topology control in ad hoc mesh networks. The network representation could be used in order to code the chromosome of the GA.

Two main strategies have been used to code a chromosome in ad hoc networks. For example, [26] proposed to

codify the complete tree of the network in the chromosomes because the sensor network is expressed by a tree network and the genes are expressed by the tree junctions. The main problem of this proposal is that a topology extraction mechanism is needed for that method, which minimizes the value of GA unless it is used in an offline manner inside large networks. In [34], a chromosome of the GA consists of sequences of positive integers that represent the IDs of nodes through which a route path passes. This second flavor has been chosen for on-demand GATAS routing protocol as this fits perfectly with the underlying QoS routing protocol. Internet protocol (IPv6) addresses could be used as the node ID.

Therefore, as shown in Figure 3, each GATAS chromosome is an existing path between the source node and the destination node. There are as many genes as intermediate nodes in the complete path. Thus, the size of the chromosome depends on the number of intermediate nodes. Chromosome genes are coded by the host-addressing part of the IPv6 address. A chromosome provides a possible routing solution, and the population is formed by individuals representing all the evaluated paths.

3.2.2. Fitness Formulation. GAs guide the search toward fitter solutions; therefore, the definition of a fitness function that identifies which are the goal solutions is the second key for success. The fitness function of GAs is generally the objective function that requires to be optimized. QoS-aware routing algorithms attempt to find an optimized path based on one or more QoS metrics. GATAS protocol can work whichever

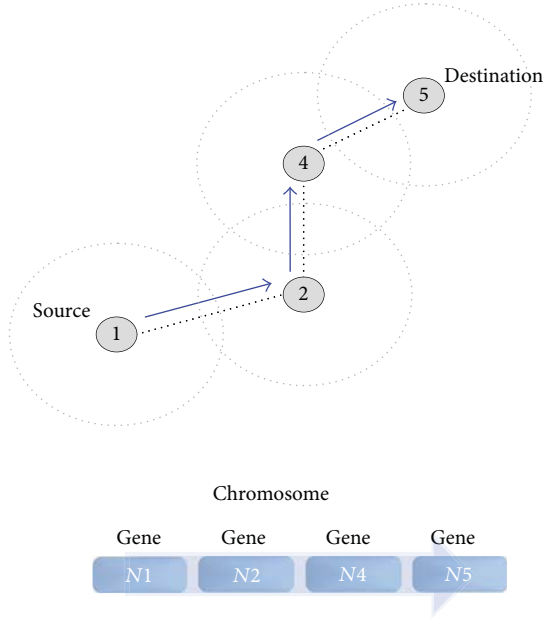


FIGURE 3: Chromosome coding.

metric strategy is used as GATAS copes with the drawback of multiobjective NP-complete problem [35]. The reason is that the cost of the path is computed in real time during the probe's trip from source node to destination due to its on-demand nature. GATAS routing capabilities are not limited by using a relax path criteria or by using concave metrics [22, 35]. While a feasible path can be selected using any shortest path algorithm [24], additional optimality constraint needs to be imposed to achieve a feasible QoS-aware path [12]. QoS metrics choice is a critical decision as the challenge for sensor ad hoc networks is to design a routing protocol that can adapt to the wide variety of conditions that may appear in Smart Grid networks over time. Metrics should be orthogonal to each other so that there is no redundant information among the metrics.

The results presented in this paper are based on a multipath dynamic delay metric (Table 3), the most important metric for several Smart Grid functions [5, 12, 13, 19], although any QoS metric strategy could be applied by using the proposed routing algebra (Table 4).

3.2.3. End Condition. The end condition is responsible for deciding when to stop the search for a better solution. Typically, in the real world, the application of the GA is run either for a fixed number of iterations or till the search has not been able to find better solutions for a number of iterations. In the parameterization of GATAS algorithm, several end conditions have been used. Actually, QoS-aware routing problem, given a source node s , a destination node d , a set of constraints C and an optimization goal, pursues finding the best feasible successor (2) or several conforming paths from s to d , which satisfies C (3), if multipath routing is desired. Since a QoS-aware routing protocol must search for routes with sufficient resources in order to satisfy the QoS

TABLE 3: Routing policy based on delay metric.

Element	Description
Σ	$\sigma \in R^+$ (real positive numbers)
\oplus	$\lambda_{ik} \oplus \sigma_{kj} = \lambda_{ik} + \sigma_{kj} \mid k \in N_i$
L	$\lambda \in R^+$, $\lambda = \text{delay}$
\leq	\leq

requirements of a flow, this is fairly a suitable end condition for real operation. This end condition must not increase unnecessarily the latency of the network. Although in order to carry out the parameterization of the GA attributes, other end conditions have been used such as certain number of GA iterations, however, they can hardly be used in a real scenario. In our simulation scenarios, the end condition could be configured during simulation time.

Consider the following:

$$S_{ij} = \{k \mid \tilde{\sigma}_{ikj}^i = \min \sigma_{ij}, \forall k \in N_i\}, \quad (2)$$

$$S_{ij} = \{k \mid \tilde{\sigma}_{ikj}^i \langle (\min \sigma_{ij} \cdot \gamma), \forall k \in N_i \rangle \mid \gamma \rangle 1. \quad (3)$$

3.2.4. Initial Population. The initial population should be supplied with sufficient variety of genetic material so that the genetic pressure could lead the population toward better individuals. Therefore, the existence of partial solutions is necessary for the success of the genetic search. The correctness of the path coded by the chromosome inside the actual network topology is critical, so it must be carefully verified by our methodology in order to avoid potential gibberish. However, if the initial population is randomly created as usually done in GA application, it is always necessary to check whether individuals are valid, and they can exist when a new generation is created or when a genetic operation is applied.

To avoid this problem, instead of a random generation, GATAS relies on the underlying QoS routing protocol to obtain suitable paths for the initial population and to avoid a mismatch between actual topology and new individuals. Thus, the existence of each individual of the initial population does not need to be verified by a topology extraction mechanism. The maximum number of initial individuals is limited by the number of available tickets issued by the origin of the path request process. In this way, the genetic algorithm cycle is activated by a source node, only when the search phase of the underlying QoS routing protocol is needed. It is important that GATAS uses the GA in an online manner like in real-time systems [12, 28]. The reason is that the path repair function could be activated by the neighbor discovery process and it is crucial that the repair function is not activated during the search phase as the routing protocol will converge slowly.

3.2.5. New Mutation Operator. The mutation operator is responsible for locally searching new solutions in the parent's neighborhood. In our case, mutation is applied gene by gene. Legacy mutation operator allows modifying any of the genes of a chromosome from the mutation probability. The mutation, as it has been defined in GATAS, requires the node

TABLE 4: Routing policy based on several additive metrics.

Element	Description
Σ	$\Sigma_{\text{delay}} \times \Sigma_{\text{hop}} \times \Sigma_{\text{bandwidthinversion}} : \langle \sigma_d, \sigma_h, \sigma_{bi} \rangle$
\oplus	$(\lambda_d, \lambda_{bi}) \oplus (\sigma_d, \sigma_h, \sigma_{bi}) =$ $\langle \lambda_d + \sigma_d, \sigma_h + 1, \lambda_{bi} + \sigma_{bi} \rangle$ $\lambda_d \in R^+, \lambda_d = \text{delay}$
L	$\lambda_{bi} \in R^+, \lambda_{bi} = \frac{1}{\text{bandwidth}}$
\preceq	$(\sigma_d, \sigma_h, \sigma_{bi}) \preceq (\sigma'_d, \sigma'_h, \sigma'_{bi})$ if and only if $(\sigma_d < \sigma'_d) \vee (\sigma_d = \sigma'_d \wedge \sigma_h < \sigma'_h) \vee$ $(\sigma_d = \sigma'_d \wedge \sigma_h = \sigma'_h \wedge \sigma_{bi} \leq \sigma'_{bi})$

to send a new one-ticket probe to obtain a mutated path. The generation of infeasible chromosomes, which violate the current network topology, is avoided by using the underlying QoS routing protocol.

In the process of Figure 4, the gene $N2$ of an individual, like path no. 10, has been selected for mutation. Therefore, the source node sends an m type one-ticket probe (mutation type probe $S_{1,m}$) to the mutation point, in this example the node $N2$. Then, node $N2$ uses the underlying QoS-aware routing protocol to search for another path by issuing a p type one-ticket probe (request type probe $S_{1,p}$) to the destination node $N5$. When the packet $S_{1,p}$ arrives at the required destination node, an r type one-ticket probe (response type probe $S_{1,r}$) is transmitted to the origin in order to advertise a new potential route to the destination $N5$. In that example, you can observe how a hypothetical individual such as $\{N1, N2, N4, N5\}$ mutates into the individual $\{N1, N2, N3, N6, N5\}$. GATAS generates one and only one new individual obtained from a successful mutation phase of each individual. This fact limits the number of extra routing packets overhead to one packet per chromosome mutation.

3.2.6. Crossover Operator. The crossover operator is responsible for the identification and reassembly of subsolutions with the aim of creating better solutions. In our specific problem, the crossover operator is used to expand the search space by finding paths unknown to origin during the search phase of the on-demand routing protocol. The flexibility provided by GATAS enables the application of different crossover schemes. In the experiments conducted herein, we applied 1-point crossover since we tried to avoid an overdisruptive approach in the reproduction phase. Nevertheless, we acknowledge that a more detailed study on which crossover operator would be the best is an interesting future line of research.

The crossover operator works as follows: to start with, the individuals for the crossover must be chosen, selected from the current population from the crossover probability. In addition to this, the crossover point between both individuals must also be chosen. The crossover process does not need any extra routing packet because it is locally executed in the source node during the search path phase of GATAS.

4. Genetic Algorithm Parameterization

In this section, we analyze in detail several results obtained throughout the OPNET simulations of GATAS in order to carry out the required parameterization of the underlying GA and the TBR for its use in the Smart Grid sensor networks.

4.1. Models for the Parameterization of the GA. In order to study the behavior of GATAS protocol in a Smart Grid scenario, we used the OPNET Modeler, which is a network simulation tool oriented to events. All the nodes of the network obey the state machine of Figure 5. The most important states of the finite state machine (FSM) are shown in Table 5.

The studies carried out in this paper are focused on the parameterization of the underlying QoS routing protocol and the underlying GA of GATAS. Furthermore, our study focuses on the analysis of the improvement using a GA compared with the underlying QoS routing protocol. The GA parameterization and the analysis carried out are specifically for ticketing issues, QoS metric, initial population, crossover probability, mutation probability, end condition, selection method, bandwidth requirements, routing overhead, and response time.

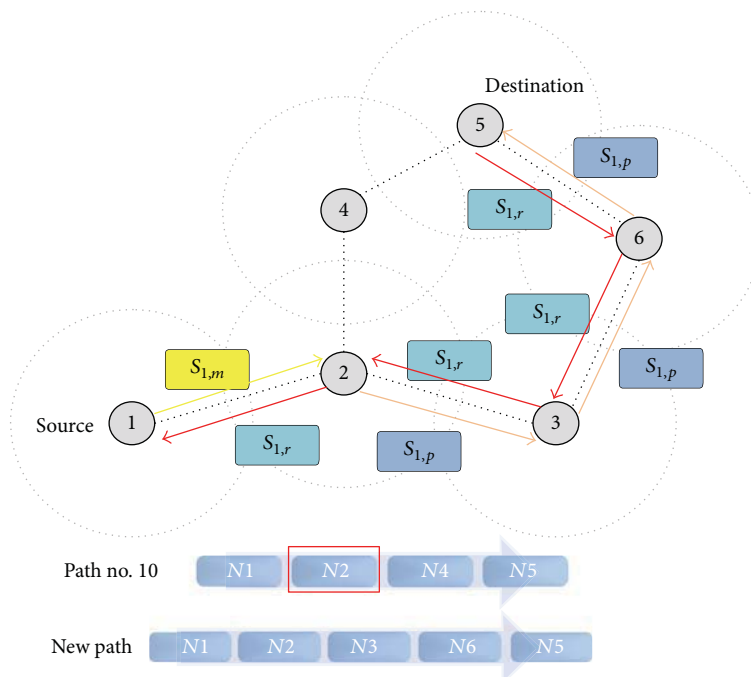
Simulated scenarios are based on networks that have more than 100 sensor nodes with reduced mobility. This behavior is likely to occur in the future Smart Grids based on heterogeneous Power Line Communications (PLC) plus wireless networks [7, 17]. For example, AMR systems can intelligently integrate the actions of all users connected to it in order to efficiently deliver sustainable electricity supplies using narrowband PLC and Zigbee communications [5, 12, 19]. The Smart Grid network topology may change due to channel characteristics. Disconnections and lowest bit rates may arise when there is significant interference from outside sources or other transmitting nodes. During these periods of time, network topology may change rapidly [17].

4.2. GATAS GA Parameterization. In this point, the decision of the most important parameters of the GA will be justified. Through Figure 6, the initial population of the GA and how many generations will be necessary for its correct operation (a possible end condition) could be determined. These simulations have been carried out by means of the Roulette Wheel Selection method with a crossover probability of 95% and a mutation probability of a 1%. These probabilities will be justified later and the evaluation of the selection method choice is out of the scope of this paper. Elitism is applied by copying the best individual to the next generation in every GA iteration.

In Figure 6, it is observed that up to an initial generation of 100 tickets, the system continues improving. From that value on, almost the same final result is obtained every time, and no additional improvement is obtained by issuing more initial tickets. The only exception that improves over 100 tickets is the TBR with 0 iterations because, since the number of tickets increases, there are more available routes. Thus, an initial population of 100 individuals could be a suitable one with acceptable bandwidth consumption. Figure 6 could also help us to choose the number of generations that our system

TABLE 5: Finite state machine of the GA.

State	Description
INIT	The FSM of a GATAS node begins in this state, where all the variables needed for the rest of the process are initialized. In this first state, the routing algorithm and the neighbor discovery algorithm are activated.
Wait	The process remains in the wait state waiting for any interruption to jump to one or another state depending on the interruption arrived at.
Hello	This state manages the neighbor discovery process. It operates in the link layer, and it is responsible for the discovery of other nodes directly connected to the origin, thereby determining the node address and its associated link metric (L). This process implements a basic keep-alive mechanism controlled by the routing protocol.
Routing	This is the state that manages the underlying QoS routing protocol as it has been succinctly described in Section 3.
GA, mutation, and solve	These states are all related with the underlying genetic algorithm as described in Section 3.

FIGURE 4: Mutation example (where $N2$ is the mutation point).

has to produce at most, and, therefore, it will set a candidate for an end condition based on the maximum number of iterations. Note that in most of the cases, it is nonsense to evolve the system more than 10 generations, because it does not practically improve the best obtained path. We must take into account that the optimal end condition for a routing algorithm is to obtain the desirable number of feasible paths that satisfy all the QoS constraints by allowing a premature convergence.

Figures 7 and 8 illustrate the crossover and mutation probabilities, and they clearly show that the best performance of the system is at a crossover probability of 95%. At a crossover probability of 10%, the system does not evolve as the final solution (the shortest path) is almost the same as the basic TBR gets. From the outset, there are three possible candidates as the best mutation probability: 1%, 10% and 20%. The mutation probability has a drastic effect on the usage of the network bandwidth. As stated before, a mutation requires

to send a new probe packet; this means that, if we increase the mutation probability by too much, the network could be collapsed by routing overhead. Note that with a probability of 20%, 20 new probes are sent for each GA iteration, for each destination by every origin node. This explains our decision to choose a mutation probability of a 1%.

The recapitulation of the achieved GA parameterization after the carried out simulations and experiments can be seen in Table 6.

4.3. Simulation Conclusions. Overall, the effectiveness of our GATAS routing algorithm scheme has been tested through a series of 50 simulation experiments. The bandwidth usage of our algorithm was determined by reckoning the number of packets that each node has generated in simulations. As expected, the network usage increases as the number of initial tickets is increased. The most outstanding peaks are in those nodes with many ad hoc interconnections, a fact that makes

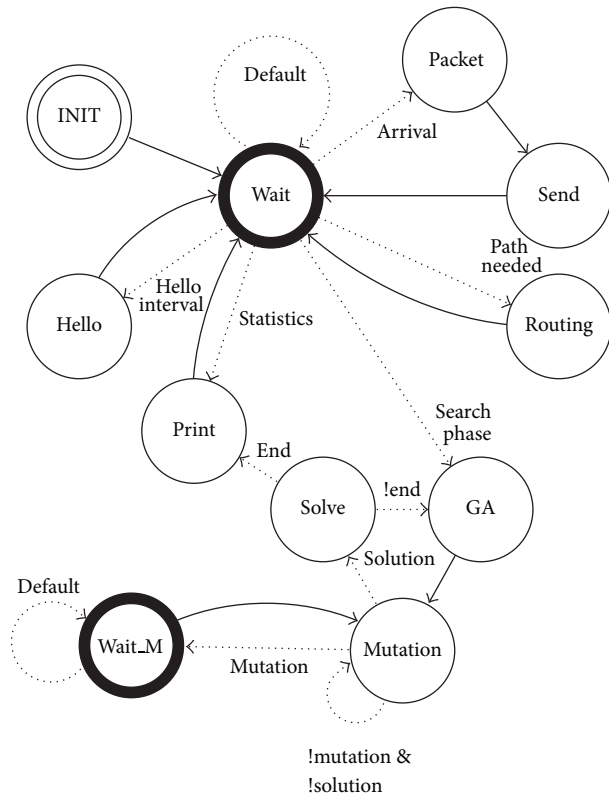


FIGURE 5: Finite state machine of the GATAS router OPNET processor.

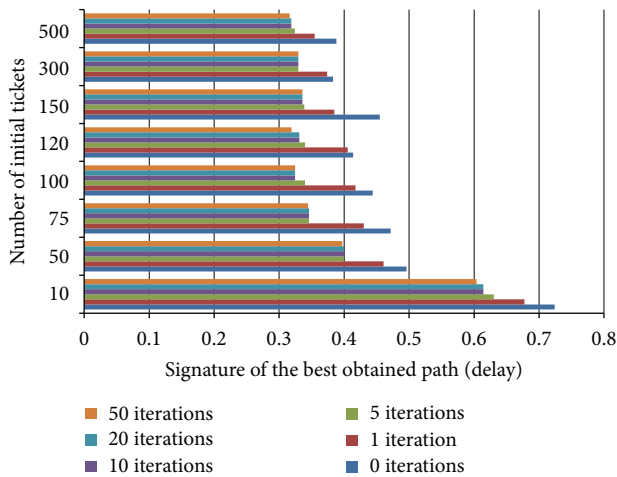


FIGURE 6: Genetic algorithm study for parameterization.

them have to deal with more tickets. At this point, three different case studies have been analyzed in order to assess the improvements introduced by the GA-based approach in the underlying QoS ad hoc routing protocol: TBR with 100 tickets, TBR with 1000 tickets, and GATAS algorithm with the final parameterization. The given convergence time results are standardized at 1 time unit.

In the TBR scenario with 100 tickets, the shortest path that the TBR has achieved in the best simulated case study is

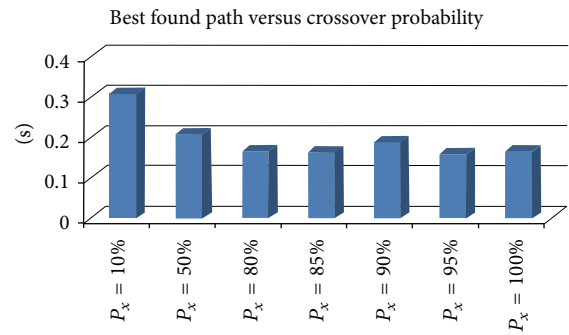


FIGURE 7: Crossover probability studies (average).

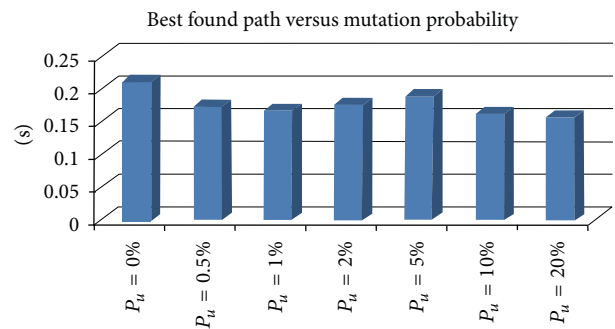


FIGURE 8: Mutation probability studies (average).

TABLE 6: GATAS genetic algorithm parameterization for smart grids.

Parameter	Value
TBR method	Random TBR
Initial population	100 individuals
Selection method	Roulette Wheel Selection
Elitism	Yes
End condition 1	Feasible QoS path found
End condition 2	10 GA iterations
Crossover probability	0.95
Mutation probability	0.01

obtained in a response time of 75 time units with a path cost of 0.248 seconds. On average, TBR with 100 tickets takes 94 time units to give us the best route it can get during the whole simulation time, and it has a path cost of 0.459 seconds. It is important to keep in mind that TBR with 100 tickets takes an average of 280 time units to send us all probes with valid paths with a maximum convergence time of 299 time units. Table 7 shows the results of the three case study scenarios. In it, the best path (metric delay in seconds) and the simulation real time needed to obtain the resulting feasible paths by using an orthogonal unit to the deployed network (OPNET simulation units [u]) can be noticed.

In the TBR scenario with 1000 tickets, we will find more and better routes than TBR with 100 tickets since there are more tickets. In contrast, the convergence time of the

TABLE 7: Results comparison.

	Best shortest path—simulation time	Average best feasible path—simulation time	Convergence time
TBR (100 tickets)	0.248 seconds—75 [u]	0.459 seconds—94 [u]	280 [u]
TBR (1000 tickets)	0.177 seconds—162 [u]	0.381 seconds—486 [u]	1762 [u]
GATAS	0.169 seconds—283 [u]	0.344 seconds—394 [u]	626 [u]

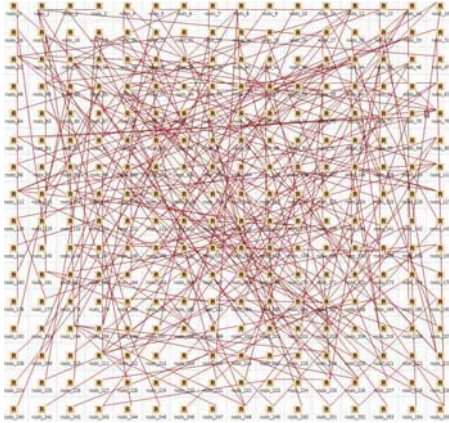


FIGURE 9: Example of an automatic generated scenario.

routing protocol is more than 6 times higher, and the routing overhead is more than 9 times higher.

GATAS response time is directly affected by the response time of TBR, since the GA is not activated until the TBR has completely generated the initial population of 100 individuals and these are either received or their timeout has expired. The mutation operation is the only one that adds delay to the convergence time since it is the only one that generates an extra probe routing packet which must be waited for in order to be processed. However, GATAS provides the best balanced ratio of best found feasible path to response time in the simulations carried out. Regardless of the convergence time of the TBR with 100 tickets, the GATAS results are the best by far in all the automatic generated scenarios by OPNET (e.g., Figure 9).

4.4. Future Comparisons. In future work, other ad hoc routing protocols must be compared with GATAS over a standard Medium Access Control (MAC) wireless level in order to evaluate its commercial utilization. To this respect, our main goal for further work is to develop the GATAS protocol in a general purpose development platform module for IEEE 802.15.4/6lowpan compatible Wireless Sensor Networks. The future objective is to assess the protocol designed to empirically monitor the consumption parameters and their convergence times. The comparison has to be made between the parameterized GATAS protocol and real mature protocol implementations which provide a feasible contrast in the created working environment.

The only algorithm that, so far, could be compared with the GATAS algorithm is the IPv6 Routing Protocol for Low Power and Lossy Networks (RPL) [36] as the IETF ROLL

Working Group focuses specifically on the IPv6 routing architectural, which has been recently specified by the IETF. For this reason, the comparison will be done when the first devices are launched in 2013-2014 and when we conclude our development. RPL intends to support a variety of low-cost network applications including industrial monitoring, building automation, connected homes, health care, environmental monitoring, urban sensor networks (e.g., Smart Grid), and asset tracking.

5. Conclusions and Further Work

GATAS routing protocol based on QoS-aware ad hoc routing has been presented as the evolution of its predecessor, the TBR algorithm. The study has concluded that this enhanced protocol based on evolutionary computation techniques improves many aspects in an M2M communication network such as optimal found path and convergence time. The main advantage is the increment of the network efficiency by minimizing routing overhead and by increasing the practicable bandwidth with the same resources.

Given that on-demand routing protocols for multihop ad hoc networks can result in increased packet latency, the paper has successfully applied GA to the existing TBR routing protocol to create the GATAS routing protocol that improves the mentioned latency aspect over that in TBR to better fit the Smart Grid requirements.

The design of an ad hoc QoS-aware routing protocol is more demanding than a shortest path routing protocol. The reason is the increase in the number and exigency of usable paths. On the other hand, the amount of routing information to transmit is greater. As our experiments demonstrate, GATAS is a better protocol to use for Smart Grids than TBR scheme. Overall, this paper presents the results of an incipient research work. In terms of CPU runtime and complexity, GATAS is comparable with TBR and other similar routing protocols with the difference that GATAS could be aware of multiple QoS metrics, the thing that is fit to the Smart Grid nature and requirements. In terms of energy saving, convergence time, overhead, and effectiveness, GATAS greatly improves its predecessor. Several trends were clearly visible in this study but the most important is that evolutionary techniques have been successfully applied to a QoS-aware ad hoc routing protocol for Smart Grids networks.

Conflict of Interests

The authors certify that there is no conflict of interests with any financial organizations regarding the material discussed in the paper.

Acknowledgments

The authors would like to thank “La Salle-URI” (University Ramon Llull) for their encouragement and assistance, especially L. Kinnear for the linguistic reviews of the paper. This work was partly supported by the EU’s seventh framework funding Program FP7 (INTEGRIS Project ICT-Energy-2009 under Grant 247938).

References

- [1] V. C. Gungor, D. Sahin, T. Kocak et al., “A survey on smart grid potential applications and communication requirements,” *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 28–42, 2013.
- [2] EPRI’s IntelliGrid initiative, Electric Power Research Institute, 2011, <http://intelligrid.epri.com/>.
- [3] U.S. Department of Energy, “Communication requirements of smart grid technologies,” 2010.
- [4] A. Vallejo, A. Zaballos, J. M. Selga, and J. Dalmau, “Next-generation QoS control architectures for distribution smart grid communication networks,” *IEEE Communications Magazine*, vol. 50, no. 5, pp. 128–134, 2012.
- [5] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, “A survey on smart grid communication infrastructures: motivations, requirements and challenges,” *IEEE Communications Surveys and Tutorials*, vol. 15, no. 1, pp. 5–20, 2012.
- [6] Q. Yang, J. A. Barria, and T. C. Green, “Communication infrastructures for distributed control of power distribution networks,” *IEEE Transactions on Industrial Informatics*, vol. 7, no. 2, pp. 316–327, 2011.
- [7] A. Zaballos, A. Vallejo, and J. M. Selga, “Heterogeneous communication architecture for the smart grid,” *IEEE Network*, vol. 25, no. 5, pp. 30–37, 2011.
- [8] D. E. Goldberg, *Genetic Algorithms in Search, Optimization & Machine Learning*, Addison-Wesley, Reading, Mass, USA, 1989.
- [9] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, MIT Press/Bradford Books edition, 1992.
- [10] OPNET University Program, 2011, <http://www.opnet.com/services/university/>.
- [11] INTEGRIS FP7 Project ‘INTElligent Electrical Grid Sensor communications’ ICT-Energy-2009 call (number 247938), <http://fp7integriss.eu/>.
- [12] J. Navarro, A. Zaballos, A. Sancho-Asensio, G. Ravera, and J. E. Armendariz-Iñigo, “The information system of INTEGRIS: INTElligent Electrical GRID sensor communications,” *IEEE Transactions on Industrial Informatics*, no. 99, p. 1, 2012.
- [13] H. Li and W. Zhang, “QoS routing in smart grid,” in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM ’10)*, pp. 1–6, December 2010.
- [14] Y.-J. Kim, M. Thottan, V. Kolesnikov, and W. Lee, “A secure decentralized data-centric information infrastructure for smart grid,” *IEEE Communications Magazine*, vol. 48, no. 11, pp. 58–65, 2010.
- [15] E. Bou-Harb, C. Fachkha, M. Pourzandi, M. Debbabi, and C. Assi, “Communication security for Smart Grid distribution networks,” *IEEE Communications Magazine*, vol. 51, no. 1, pp. 42–49, 2013.
- [16] P.-Y. Chen, S.-M. Cheng, and K.-C. Chen, “Smart attacks in smart grid communication networks,” *IEEE Communications Magazine*, vol. 50, no. 8, pp. 24–29, 2012.
- [17] H. Gharavi and B. Hu, “Multigate communication network for smart grid,” *Proceedings of the IEEE*, vol. 99, no. 6, pp. 1028–1045, 2011.
- [18] A. Zaballos, A. Vallejo, G. Ravera, and J. M. Selga, “Simulation and analysis of a QoS multipath routing protocol for smart electricity networks,” *International Journal on Advances in Networks and Services*, vol. 3, no. 3-4, pp. 417–429, 2011.
- [19] B. Sivaneasan, P. L. So, and E. Gunawan, “A new routing protocol for PLC-based AMR systems,” *IEEE Transactions on Power Delivery*, vol. 26, no. 4, pp. 2613–2620, 2011.
- [20] M. Ilyas, *The Handbook of Ad Hoc Wireless Networks*, CRC Press, Boca Raton, Fla, USA, 1st edition, 2003.
- [21] M. L. P. Bueno and G. M. B. Oliveira, “Multicast flow routing: evaluation of heuristics and multiobjective evolutionary algorithms,” in *IEEE Congress on Evolutionary Computation (CEC ’10)*, July 2010.
- [22] A. Riedl and D. A. Schupke, “Routing optimization in IP networks utilizing additive and concave link metrics,” *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 1136–1148, 2007.
- [23] A. Singh and S. Sundar, “An artificial bee colony algorithm for the minimum routing cost spanning tree problem,” *Soft Computing*, vol. 15, no. 12, pp. 2489–2499, 2011.
- [24] W. A. Chang and R. S. Ramakrishna, “A genetic algorithm for shortest path routing problem and the sizing of populations,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 6, pp. 566–579, 2002.
- [25] F.-N. Pavlidou and G. Koltsidas, “Game theory for routing modeling in communication networks—a survey,” *Journal of Communications and Networks*, vol. 10, no. 3, pp. 268–286, 2008.
- [26] L. Barolli, A. Koyama, and N. Shiratori, “A QoS routing method for ad-hoc networks based on genetic algorithm,” in *International Workshop on Database and Expert Systems Applications*, 2003.
- [27] D. Apiletti, E. Baralis, and T. Cerquitelli, “Energy-saving models for wireless sensor networks,” *Knowledge and Information Systems*, vol. 28, no. 3, pp. 615–644, 2011.
- [28] A. C. B. Garcia, C. Bentes, R. H. C. de Melo, B. Zadrozny, and T. J. P. Penna, “Sensor data analysis for equipment monitoring,” *Knowledge and Information Systems*, vol. 28, no. 2, pp. 333–364, 2011.
- [29] N. Guofang and L. Minqiang, “Evolutionary based approaches in wireless sensor networks: a survey,” in *Proceedings of the 4th International Conference on Natural Computation (ICNC ’08)*, pp. 217–222, October 2008.
- [30] V. Uchhula and B. Bhatt, “Comparison of different Ant Colony Based Routing Algorithms,” *International Journal of Computer Applications*, no. 2, pp. 97–101, 2010.
- [31] P. Levis, A. Tavakoli, and S. Dawson-Haggerty, “Overview of Existing Routing Protocols for Low Power and Lossy Networks,” IETF Draft-Letf-Roll-Protocols-Survey-07, pp. 1–27, 2009.
- [32] J. L. Sobrinho, “An algebraic theory of dynamic network routing,” *IEEE/ACM Transactions on Networking*, vol. 13, no. 5, pp. 1160–1173, 2005.
- [33] X. Li, W. Jun, and K. Nahrstedt, “The enhanced ticket-based routing algorithm,” in *Proceedings of the International Conference on Communications (ICC ’02)*, pp. 2222–2226, May 2002.
- [34] P. Sateesh and S. Ramachandram, “Scalability of network size on genetic zone routing protocol for MANETs,” in *Proceedings*

of the International Conference on Advanced Computer Theory and Engineering (ICACTE '08), pp. 763–767, December 2008.

- [35] Z. Wang and J. Crowcroft, “Quality-of-service routing for supporting multimedia applications,” *IEEE Journal on Selected Areas in Communications*, vol. 14, no. 7, pp. 1228–1234, 1996.
- [36] P. Thubert, T. Winter, A. Brandt et al., “RPL: IPv6 Routing Protocol for Low power and Lossy Networks,” *IETF Draft-Letf-Roll-Rpl-18*, pp. 1–160, 2011.

Work in Progress - Intelligent Tutoring System Framework for the Acquisition of Knowledge and Competences

David Vernet, Ruben Nicolas, Elisabet Golobardes, Albert Fornells and Alvaro Garcia-Piquer
 La Salle, Universitat Ramon Llull, dave@salleurl.edu, rnicolas@salleurl.edu, elisabet@salleurl.edu, afornell@salleurl.edu, alvarog@salleurl.edu

Abstract – Even though there are many tutoring systems to help students achieve their goals in terms of theoretical knowledge, as yet there is no system to foment the acquisition of the competences which form an integral part of university degree programs. This issue is crucial because the Higher Education System is changing in Europe. New educational models are being created to introduce competences which correspond specifically to degree programs. In this work-in-progress the general framework to develop an Intelligent Tutoring System (ITS) based on competences is presented. The system monitorizes the acquisition of theoretical knowledge and also the assessment of the competences related to the subject and it takes corrective actions when needed to fix a negative evolution of the student. The framework is divided into four main phases based on artificial intelligence techniques: construction, location, prediction and reinforcement. The main feature of the proposed framework is the fact that it promotes the academic development of the student in the future educational context by providing guidance and supervision to ensure the successful acquisition of both theoretical knowledge and the corresponding subject-related competences.

Index Terms – Assessment, Competences, Distance Learning, Intelligent Tutoring Systems.

INTRODUCTION

The new Higher Education Area has produced a number of changes in teaching and assessment methods. [1]-[2]. In the past, the weight of the learning process was borne by the teacher and the student was a participating user of the class. In the new teaching model, the student is at the centre of the learning process and the objective is not only based on the acquisition of knowledge, but also on competences (this word refers to a set of skills, knowledge and attitudes that an individual must possess in order to be capable of perform a specified job [1]).

This new model, however, presents us with a handicap which is that most distance-learning teaching tools have not been thought out for this new framework. Practically all the systems focus on the acquisition of knowledge and ignore the acquisition of competences. Some examples are [3]-[4].

In order to cover this shortfall we propose to create an intelligent tutoring system frame for the acquisition of Knowledge and Competences. The work is still in progress but we have started to define the most relevant features of this type of system.

The following sections define the structure of a system of this nature and its development framework. The framework is divided into four phases: construction, location, prediction and reinforcement. The following sections will describe each phase in greater detail. Firstly, however, we shall present the data which the system will work with. This is specific data which has to be collected and analyzed in order to carry out an evaluation of the competences.

DATA DEFINITION

One of the most important features of the new Higher Education Area is that each subject is assigned a series of competences which have to be acquired by the student and assessed by the teacher [5]. Taking things to extreme, we could define the final assessment of a subject as being the weighted assessment of these competences. In other words, the acquisition of a competence is computed in line with the knowledge areas which evaluate the competence. The final mark does not depend on one particular topic, but on a set of knowledge areas which evaluate the competence. The results obtained correlate with the acquisition of these competences which shape the final course mark.

Given this situation, the data used by the system is pre-determined. A set of competences is assigned to each knowledge area of the subject and a subsequent set of activities is defined for each competence. This enables us to assess the progress of the student who is acquiring this competence. Figure 1 shows the distribution in terms of areas of knowledge. The letters are coded as follows:

- S = Subject, the subject to teach.
- U_i = Unit of the subject ($1 \leq i \leq N$). In the traditional teaching a subject is divided in units.
- A_i = Activity associated to the unit in order to assess the learning of the student.
- C_i = Competence associated to activity. Each activity may be associated to more than one competence.

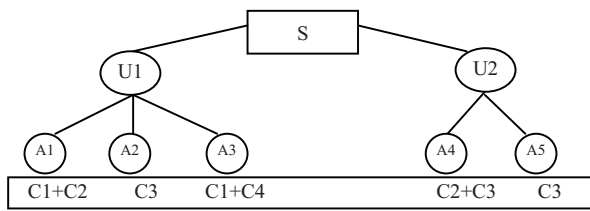


FIGURE 1
TRADITIONAL REPRESENTATION OF THE KNOWLEDGE TREE

As can be seen in Figure 1, the knowledge of the subject is represented according to the model of traditional teaching and way which the most important systems are currently using to represent it.

Our approach sets the competence model as the center of the knowledge to be taught. Figure 2 shows how the previously-mentioned knowledge tree would be transformed in the new teaching model.

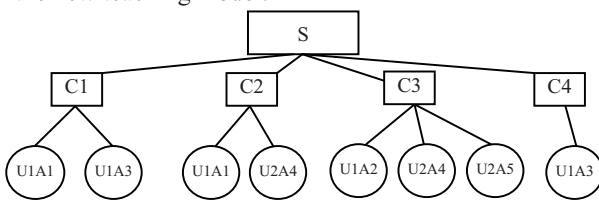


FIGURE 2
REPRESENTATION OF THE DATA IN OUR FRAMEWORK

The diagram shows how the tutoring system focuses on all of its operations on the competence and how the activities proposed throughout the learning process of the student are used to optimize the potential of each competence.

THE FRAMEWORK

The point of view used of relating the information is the operational base of our tutor system. The activities which are proposed to the students are closely linked to the competences they will obtain if they successfully complete the activity in question. The system automatically detects when a student does not acquire a specific competence and suggests new activities to enable the student to reach the required competence level. The most important data introduced into the system are the qualifications of the competences of each of the activities carried out.

The system can be divided into 4 clearly marked phases:

I. Construction

In the construction phase the knowledge tree of the subject to be taught is obtained. The nodes of this tree constitute units of the domain of knowledge which, at the same time, are divided into different activities. Unlike the *learning paths* proposed by Falmagne [6], the teachers of the subject design the activities to solve, associate competences to them and place them in the tree, depending on the competence in question. As we can see, in our system not only is theoretical knowledge placed on the tree but the competences achieved through the resolution of these activities are also added. If we observe the tree generated in

the traditional way, the competences are distributed in the tree to one or several scattered nodes.

II. Location

In the location phase the system assigns the student a state of knowledge and competences using an automatic algorithm extracted from different evaluations obtained through the activities proposed by the teachers. Thus, using the information gathered from the activities already carried out, the system then assesses the degree to which every competence is reached and generates the tuple (competence-achievement grade) for each of the competence designed for the course. This evaluation and location should be repeated more than once in accordance with the level of acquired competences and knowledge in the different fields. At this point the system can activate the predictive phase.

III. Prediction

The prediction phase is based on the performance of a student (exercise marks and acquired competences), and it predicts the evolution of the student at the end of course using Artificial Intelligence techniques.

IV. Reinforcement

In the reinforcement phase an intelligent selection of suitable activities enables the system to give the student activities which not only improve learning potential of a specific subject but also develop the tree of knowledge and competences.

CONCLUSIONS AND CURRENT WORK

A new framework approach for education has been presented. The new features of it are: 1) A new way to organize the knowledge of a subject and, 2) The design of an ITS framework using this new model. Currently, we are developing an ITS with this framework and integrating AI techniques.

ACKNOWLEDGEMENTS

We would like to thank the AGAUR for the support under grants 2010FI_B2 00084 and 2010FI_B 01084, and Lisa Kinnear for the English revision.

REFERENCES

- [1] Golobardes, E et al., "Guía para la evaluación de competencias en el área de Ing. y Arq.", *Guías de eval. de competencias, AQU Cat*, 2009.
- [2] Hager, P, Gonczi, A, "General issues about assessment of competence, Assessment and Evaluation in Higher Education", Vol 19 (1), University of Bath, Bath, 1994.
- [3] Hockemeyer, C., Held T., Albert, D., "A relational adaptive tutoring hypertext WWW-environment based on knowledge space theory", *Proceedings of CALISCE '98*, 1998, 417-423.
- [4] Wetzel, M., & Hanley, G. "Evaluation of MERLOT Tools, Processes, and Accomplishments". *Center for Usb. in Design and Assess.*, 2001
- [5] Bologna declaration, "The European Higher Education Area", *Joint declaration of the European Ministers of Education*, 1999.
- [6] Falmagne, J.-C. "A latent trait theory via stochastic learning theory for a knowledge space", *Psychometrika*, 53, 1989, 283-303.

An Unsupervised Learning Approach for Case-Based Classifier Systems

David Vernet and Elisabet Golobardes

Enginyeria i Arquitectura La Salle, Universitat Ramon Llull,
Pg. Bonanova 8, 08022 Barcelona, Spain
{dave,elisabet}@salleurl.edu

Abstract. Case-Based Classifier Systems obtain low accuracies on generalisation and higher waste on CPU time when the class distribution space is not well defined. This paper presents the Mean Sphere and the Mean K-Means approach based on unsupervised learning to improve the CPU time and to improve or maintain the accuracy. We use clustering in an unsupervised way to decide which is the representational space of each class. The concept of clustering is introduced in two levels. *First level* cluster the training data into *spheres*, obtaining one *sphere* for each class. *Second level* consists of clustering the spheres in order to detect the behaviour of the elements present in the sphere. In this level two policies are applied, Mean Sphere and Mean K-Means approaches. Experiments using different domains, most of them from the UCI repository, show that the CPU time is considerably decremented while maintaining, and sometimes improving, the accuracy of the system.

1 Introduction

Case-Based Reasoning systems are often faced with two main problems when a great number of features and cases exist in the case memory. The first problem is the reduction of the system performance because the system can not detect different behaviours on the data. The second problem is an increase in CPU time because the retrieval phase has to use all the information available.

This paper describes the Mean Sphere and the Mean K-Means (MKM) approach to organising the case memory. Our aim is twofold: (1) to reduce the CPU time and (2) to distinguish between different behaviours of the data, avoiding noisy instances. The organisation of the case memory proposed consists of applying two levels of clustering. Firstly, a construction of the *spheres* is done based on the class distribution of the cases present in the case memory. Later, a second level of clustering is applied using the results of the previous one. In the second level, each sphere contains a set of clusters obtained using the Mean Sphere approach or K-Means algorithm. Both approaches have been introduced into our ULIC (Unsupervised Learning In CBR) platform.

This paper is organized as follows: first, Section 2 presents the related work about clustering and unsupervised learning in general; the next section introduces the

unsupervised organisation approaches. Then, Section 4 explains the testbed and experiments used and the results obtained. Finally, Section 5 presents the conclusions and further work.

2 Related Work

This section summarises related work present in the literature for clustering methods and for different approaches used in CBR systems to organise the case memory.

First of all, most of the clustering methods are described in Hartigan's book [13].

There exist a large number of clustering algorithms. The choice of clustering algorithm depends on the type of data available and on the particular purpose and application [10].

In general, clustering methods can be classified in the following approaches:

The first approach is the *partitioning methods*. They consist of clustering training data into k clusters where $k < n$ and n is the number of objects in the data set. An example of this approach is *k-means algorithm* [12]. There are special variations to improve some aspects of the algorithm. The first variation is the *k-medoids algorithm* or PAM (Partition Around Medoids) [15]. In this algorithm, the objective is to reduce the sensitivity of the *k-means* algorithm when some extremely large values that distort the distribution of data are found. A variation of the *k-medoids* algorithm is the *CLARA algorithm* (Clustering LARge Applications) [16]. In this case, the algorithm extends the capabilities of the *k-medoids* algorithm so as to perform more efficiently when large data sets are explored.

The second approach is called *hierarchical methods*, which work by grouping data objects into a tree of clusters. The hierarchical decomposition can be formed as a bottom-up or top-down procedure.

Another approach considered are the *density-based methods*. The main objective of these methods is to discover clusters with an arbitrary shape. These typically regard clusters as dense regions of objects in the data space that are separated by regions of low density (representing noise). The most popular algorithms in this category are: DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [5], OPTICS (Ordering Points to Identify Clustering Structure) [2] and DENCLUE (DENsity-based CLUstEring) [14].

Grid-based methods use a multiresolution grid data structure that divides the space into a finite number of cells that form a grid structure on which all operations for clustering are performed. This method has the advantage of a constant processing time, independently of the number of data objects. We can identify in this group algorithms such as CLIQUE (Clustering High-Dimensional Space) [1], STING (STatistical INformation Grid) [23], and WaveCluster [22] (an algorithm that clusters using the wavelet transformation).

Finally, *model-based methods* use mathematical and probability models. These methods can be focused in two ways: firstly, as a statistical approach, and second as a neural network approach. Some examples of this approach are AUTOCLASS [3] and COBWEB [6].

One criticism directed at researchers that use conceptual clustering (similar to [20]) has been that the clustering of objects or events without a context, goal or some information concerning the function of the derived clusters is not likely to be useful for real-world problems [11]. Hanson proposes a different point of view and approach real-world problems with algorithms like WITT [11].

On the other hand, in the literature there exist different approaches in CBR to produce a new organisation of case memory. The most important approaches are the following. RISE [4] treats each instance as a rule that can be generalised. EACH [21] introduced the *Nested Generalized Exemplars* (NGE) theory, in which hyperrectangles are used to replace one or more instances, thus reducing the original training set. And finally, a method that avoids building sophisticated structures around a case memory or complex operations is presented by Yang and Wu [24]. Their method partitions cases into clusters where the cases in the same cluster are more similar than cases in other clusters. Clusters can be converted to new smaller case-bases. However, not all the approaches are focused on the organisation of the case memory in order to improve the case memory and, at the same time, the computational time.

3 Unsupervised Organisation Approaches

The spheres construction process performs the first level of the organisation of the case memory. The concept of *sphere* had been introduced in the CaB-CS [8] and exploited with success in preliminary work such as [9,17].

The success of this type of representation of the Case Memory is based on two aspects: first of all this representation greatly improves the speed of the CBR system, and secondly the spheres offer high reliability in the selection of the candidate cases.

Each case from the original case memory is distributed to one sphere depending on the class associated with the case (see Fig. 1.). One sphere contains a subset of cases from the original case memory (NC). All the cases that belong to the same sphere represent the same class. The union of all spheres is the whole set of cases in the original case memory. It follows that:

$$\sum_{i=1}^k n_i = NC \quad (1)$$

where n_i is the number of cases that belong to the sphere i , k is the number of different classes and NC is the number of cases in the training data.

3.1 Mean Sphere

The Mean Sphere uses the previously computed spheres to generate the centroid of each class (see algorithm in Fig. 2.).

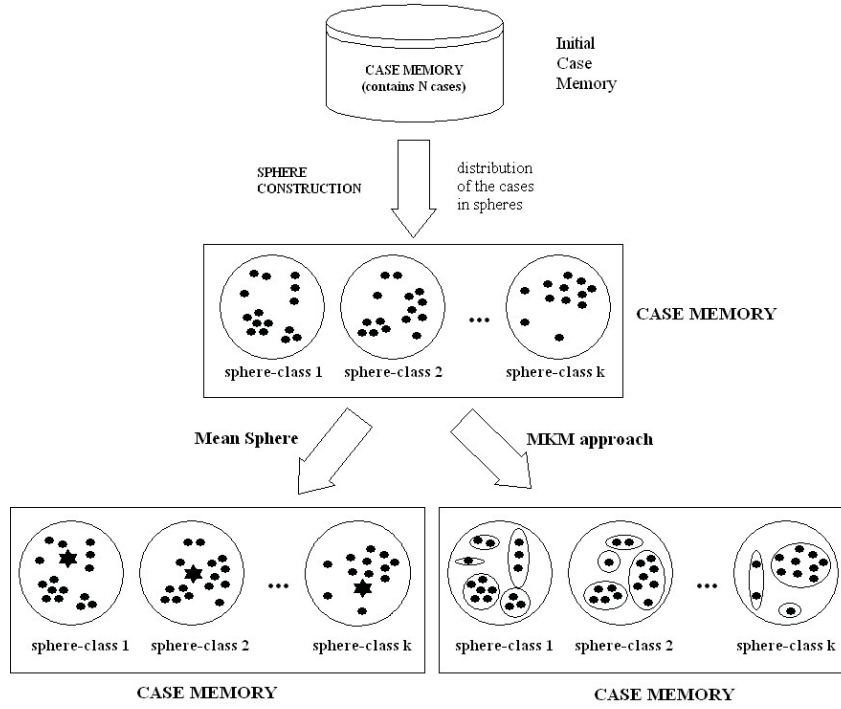


Fig. 1. Representation of the Case Memory for the two approaches. The symbol ● represents a case and the symbol ★ represents the centroid of the sphere.

Each Mean Sphere (of any class C) contains the mean value for each feature computed using the cases present in the sphere of class C .

```

1. for each sphere  $S_i$  in  $k$  different classes
2.   Let  $n_i$  be represented as  $N$  the number of cases of sphere  $i$ 
3.   for each attribute  $j$  in the set of all attributes
4.      $centroid\ S_{ij} = \sum_{c=1}^N \frac{a_{ci}}{N}$ 
5.   end for
6. end for
    
```

Fig. 2. Algorithm to compute the centroid ★ of each sphere. The variable a_{ci} represents the value of the attribute j in the c th case of the i th sphere.

Later, in the retrieval phase, we search for the sphere that represents the new case, we compute the *Euclidean Distance* between each centroid and the new case. We classify the new case depending on the class represented by the sphere selected.

3.2 Mean K-Means Approach

The *Mean K-Means (MKM)* approach takes the original spheres and obtains a new categorisation of the case memory by applying the *k-means clustering method* (Fig. 3.) internally in each sphere. This new categorisation also produces a smaller case memory. Thus, the CPU time is reduced.

1. Choose an initial partition of the cases into k clusters. This is random assignment to k clusters.
2. Compute the distance from every case to the mean of each cluster and assign the cases to their nearest clusters.
3. Recompute the cluster means following any change of cluster membership at step 2.
4. Repeat steps 2 and 3 until no further changes of cluster membership occur in a complete iteration. The procedure has now converged to a stable k-partition.

Fig. 3. K-Means algorithm modified to cluster each sphere.

For each sphere we obtain C_i clusters and we compute the centroid of each cluster in the sphere. Then, we select the nearest cluster to the new case in the same way as the Mean Sphere approach. The number of clusters for each sphere can be different.

The clustering method is configured independently of the class treated. Thus, we define the optimum number of clusters for each sphere.

Moreover, we can configure a class with n_i clusters (one cluster for each case). In this case we can combine the application of traditional Case-Based Reasoning in one class with clustering in the other classes. Therefore, a sphere with n_i clusters is equivalent to applying CBR in this class.

4 Experiments and Results

In this section we describe the data sets for testing the techniques proposed and the results obtained.

4.1 Testbed

In order to evaluate the performance rate, we use eight datasets. Datasets can be grouped in two ways: *public* and *private*. The datasets and their characteristics are listed in Table 1.

Table 1. Datasets and their characteristics used in the empirical study.

Dataset	Reference	Samples	Features	Classes	Inconsistent
1. <i>Biopsy</i>	BI	1027	24	2	Yes
2. <i>Breast-cancer (Wisconsin)</i>	BC	699	9	2	Yes
3. <i>Glass</i>	GL	214	9	6	No
4. <i>Ionosphere</i>	IO	351	34	2	No
5. <i>Iris</i>	IR	150	4	3	No
6. <i>Mammogram</i>	MA	216	23	2	Yes
7. <i>Sonar</i>	SO	208	60	2	No
8. <i>Vehicle</i>	VE	846	18	4	No

Public datasets are obtained from the UCI repository [19]. They are: *Breast Cancer Wisconsin*, *Glass*, *Ionosphere*, *Iris*, *Sonar* and *Vehicle*. Private datasets come from our own repository. They deal with *diagnosis* of breast cancer. Datasets are *Biopsy* and *Mammogram*. *Biopsy* [7] is the result of digitally processed biopsy images, whereas *Mammogram* consists in detecting breast cancer using the microcalcifications (μCa) present in a mammogram [18, 9]. In *mammogram* each example contains the description of several μCa present in the image; in other words, the input information used is a set of real valued matrices.

These datasets were chosen in order to provide a wide variety of application areas, sizes, combinations of feature types, and difficulty as measured by the accuracy achieved on them by current algorithms. The choice was also made with the goal of having enough data points to extract conclusions.

The configuration of the ULIC system for this paper is 1-Nearest Neighbour algorithm. Training cases are represented by *spheres*. We have not used weighting methods in order to test the reliability of our system. The retain phase is limited to the original training data. The learning process in the test is skipped in order to check the behaviour of the spheres.

4.2 Results

In this section we describe the results obtained by the ULIC system using both approaches.

In this paper we have only performed one distance function in order to test the new organisation of the case memory. Other similarity functions will be studied in further research.

The mean accuracy (mean percentage of correct classifications) is the result of 10 different executions of the stratified 10-fold cross-validation.

Firstly, in the Mean Sphere approach, we have a sphere for each class without internal clustering. We have computed the centroid values for each sphere. Therefore, we reduce all the cases that belong to a class to one case.

Secondly, in the MKM approach, we cluster each sphere in order to detect different behaviours of the data contained in each sphere.

In Table 2 we present the results obtained by the traditional 1-NN CBR, the Mean Sphere approach and the MKM approach. As we can observe, the results in general improve both the mean accuracy and the CPU time of resolution of one case.

Table 2. This table compares the mean percentage of correct classifications (%PA), standard deviation (std) and mean CPU time (CPUt) in milliseconds of the Retrieval phase using traditional CBR, a first level of clustering (Mean Sphere) and a second level of clustering (MKM approach). The results that improve the prediction accuracy or the CPU time of the traditional CBR are marked with a ✓.

Ref.	CBR			MeanSphere			MKM		
	%PA	CPUt	std	%PA	CPUt	std	%PA	CPUt	std
BC	96.14	56.34	1.45	96.28 ✓	0.13 ✓	1.87	96.71 ✓	31.96 ✓	1.53
BI	83.15	199.63	3.55	79.07	0.27 ✓	4.66	81.40	1.52 ✓	3.76
GL	69.16	33.60	7.32	53.74	0.51 ✓	7.01	70.79 ✓	30.52 ✓	8.70
IO	90.03	99.61	4.28	81.77	0.45 ✓	5.07	90.31 ✓	1.60 ✓	5.38
IR	95.33	6.00	3.06	92.67	0.20 ✓	2.00	97.33 ✓	1.50 ✓	3.27
MA	62.50	90.41	13.73	64.81 ✓	0.23 ✓	9.42	63.89 ✓	65.69 ✓	9.86
SO	82.21	198.71	6.99	70.67	1.83 ✓	7.19	82.93 ✓	89.20 ✓	7.73
VE	66.90	125.06	4.33	84.04 ✓	0.54 ✓	4.39	65.60	2.00 ✓	3.75

As we can observe, the MKM Sphere approach improves the prediction accuracy obtained by the Mean Sphere approach in some data sets. The reason for the greater efficiency of the Mean Sphere approach is that the size of the reduced case memory tends to be much smaller than in MKM. In the Mean Sphere approach the case memory is reduced to a unique case for each class. So, the CPU time is lower in the Mean Sphere approach for all data sets.

Table 3 shows the optimum number of clusters for each data set and the average accuracy, the standard deviation and the average CPU time of resolution for one case.

Table 3. Best configuration of the clusters. We indicate for each class the number of clusters generated. A number n_i indicates that there is a number of clusters equal to the number of cases in the sphere.

Dataset	Classes	Number of clusters
BI	2	28-16
BC	2	27- n_i
GL	7	20- n_i -10- n_i -20- n_i -10
IO	2	24-6
IR	3	20-4-10
MA	2	n_i -40
SO	2	25- n_i
VE	4	25-20-35-35

The number of clusters can be different for each sphere. Thus we can define different numbers of clusters in different classes. For each data set, we have determined with previous executions of the system which is the best combination of number of clusters in order to configure the system for further experiments.

5 Conclusions and Further Research

We have introduced two *unsupervised learning approaches* in a traditional Case-Based Reasoning System achieving our initial objectives.

It is important to keep in mind that the main goal was to maintain the prediction accuracy obtained by the traditional Case-based Reasoning improving the speed on CPU time of the Retrieval phase. As we have seen, we have improved the speed and we have also improved the results.

The results show that the clustering methods notably reduce the CPU time when resolving a new case. Moreover, MKM approach maintains or even improves the prediction accuracy obtained by the traditional Case-based Reasoning.

MKM approach obtains better prediction accuracies because this approach organises the memory case with a higher number of cases than the Mean Sphere approach. However, the CPU time is lower in the Mean Sphere approach.

On the other hand we can introduce in further work the idea of not predicting when the clustering method does not give a reliable result. In this way, we want to increase the reliability of the system, in particular when we are working in medical environments (like mammograms, biopsies and so on).

Another further objective is to add to the system other clustering algorithms that automatically detect the optimum number of clusters. In the current configuration we have tuned the best configuration for each problem.

Another aspect to take into consideration is the possibility of applying clustering methods with discrete data. Algorithms such as WITT and variations of k-means algorithms should be adapted in order to solve this question.

Finally, we want to improve the accuracy results by using weighting methods in the clustering algorithms.

Acknowledgements

This work is supported by the *Catalana Occidente and La Salle* award. We want to thank the *Ministerio de Sanidad y Consumo, Instituto de Salud Carlos III, Fondo de Investigación Sanitaria* for its support under grant number FIS 00/0033-02. The results of this work were obtained using the equipment co-funded by the *Direcció General de Recerca de la Generalitat de Catalunya (D.O.G.C 30/12/1997)*. Finally, we would like to thank Enginyeria i Arquitectura La Salle for their support of our AI Research Group.

References

1. R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan. Automatic subspace clustering of high dimensional data for data mining applications. In *Proceedings of ACM SIGMOD Conference on Management of Data*, pages 94-105, 1998.
2. M. Ankerst, M.M. Breunig, H. Kriegel, and J. Sander. OPTICS: ordering points to identify the clustering structure, pages 49-60, 1999.
3. P. Cheeseman and J. Stutz. Bayesian classification (autoclass): Theory and results. In *Advances in Knowledge Discovery and Data Mining*, pages 153-180, 1996.
4. P. Domingos. Context-sensitive feature selection for lazy learners. In *AI Review*, volume 11, pages 227-253, 1997.
5. M. Ester, H.P. Kriegel, and X. Xu. A database interface for clustering in large spatial databases. In *Knowledge Discovery and Data Mining*, pages 94-99, 1995.
6. D.H. Fisher. Knowledge acquisition via incremental conceptual clustering. In *Machine Learning*, pages 2:139-172, 1987.
7. J.M. Garrell, E. Golobardes, E. Bernadó, and X. Llorà. Automatic diagnosis with Genetic Algorithms and Case-Based Reasoning. *Elsevier Science Ltd. ISSN 0954-1810*, 13:367-362, 1999.
8. E. Golobardes. Aportacions al raonament basat en casos per resoldre problemes de classificació. *PhD thesis*, Enginyeria La Salle, Universitat Ramon Llull, june 1998
9. E. Golobardes, X. Llorà, M. Salamó, and J. Martí. Computer Aided Diagnosis with Case-Based Reasoning and Genetic Algorithms. *Journal of Knowledge-Based Systems* 15, pages 45-52, 2002.
10. J. Han and M. Kamber. *Data mining: Concepts and techniques*, 2000.
11. S.J. Hanson. Conceptual clustering and categorization: Bridging the gap between induction and causal models. In Y. Kodratoff and R.S. Michalski, editors, *Machine Learning: An Artificial Intelligence Approach (Volume III)*, pages 235-268, Kaufmann, San Mateo, CA, 1990.
12. J. Hartigan and M. Wong. A k-means clustering algorithm. In *Applied Statistics*, pages 28:100-108, 1979.
13. J.A. Hartigan. *Clustering Algorithms*. John Wiley and Sons, New York, 1975.
14. A. Hinneburg and D.A. Keim. An efficient approach to clustering in large multimedia databases with noise. In *Knowledge Discovery and Data Mining*, pages 58-65, 1998.
15. L. Kaufman and P.J. Rousseeuw. Clustering by means of medoids. In *Statistical Data Analysis Based on the L1-Norm and Related Methods*, pages 405-416, North-Holland, Y. Dodge.
16. L. Kaufman and P.J. Rousseeuw. *Finding groups in data: An introduction to cluster analysis*. John Wiley and Sons, 1990.
17. X. Llorà, E. Golobardes, M. Salamó, and J. Martí. Diagnosis of microcalcifications using Case-Based Reasoning and Genetic Algorithms. In *Proceedings of Engineering of Intelligent Systems (EIS2000)*, volume 1, pages 254-263, 2000.
18. J. Martí, J. Español, E. Golobardes, J. Freixenet, R. García, and M. Salamó. Classification of microcalcifications in digital mammograms using case-based reasoning. In *International Workshop on digital Mammography*, 2000.
19. C.J. Merz and P.M. Murphy. UCI Repository for Machine Learning Data-Bases [<http://www.ics.uci.edu/~mllearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science, 1998.
20. R.S. Michalski. Knowledge acquisition through conceptual clustering: A theoretical framework and an algorithm for partitioning data into conjunctive concepts. *Technical Report 1026*, Urbana, Illinois, 1980.
21. S. Salzberg. A nearest hyperrectangle learning method. *Machine Learning*, 6:277-309, 1991.

10 **David Vernet and Elisabet Golobardes**

22. G. Sheikholeslami, S. Chatterjee, and A. Zhang. WaveCluster: A multi-resolution clustering approach for very large spatial databases. *In Proc. 24th Int. Conf. Very Large Data Bases, VLDB*, pages 428-439, 24-27, 1998.
23. W. Wang, J. Yang and R.R. Muntz. STING: A statistical information grid approach to spatial data mining. *In The VLDB Journal*, pages 186-195, 1997.
24. Q. Yang and J. Wu. Keep it Simple: A Case-Base Maintenance Policy Based on Clustering and Information Theory. *In Proc. of the Canadian AI Conference*, pages 102-114, 2000.



Aquesta Tesi Doctoral ha estat defensada el dia ____ d_____ de 201__
al Centre_____

de la Universitat Ramon Llull, davant el Tribunal format pels Doctors i Doctores
sotasignants, havent obtingut la qualificació:

President/a

Vocal

Vocal *

Vocal *

Secretari/ària

Doctorand/a

(*): Només en el cas de tenir un tribunal de 5 membres