Concept Discovery and Argument Bundles in the Web of Experiences

by

XAVIER FERRER ARAN

A dissertation presented in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy in Computer Science

Advisor:
Prof. Enric Plaza Cervera

Tutor:
Dr. Ricardo Toledo Morales

April 4, 2017

Universitat Autònoma de Barcelona
Departament de Ciències de la Computació

Abstract

Millions of people interact and share interesting information every day in the Social Web. From daily conversations to comments about products in e-commerce sites, the content generated by people in these sites is huge and diverse. Among the wide diversity of user-contributed content on the web, there is a particular kind that has the potential of being put to good use by intelligent systems: human experiences. People very often use other people's experiences before making decisions, and when these kind of human experiences are expressed and recorded on the web, they can be shared with by large number of people.

Nevertheless sometimes this content is not easily accessible, so a person trying to book a hotel may read a few reviews over a few hotels –but cannot possibly read them all. There is a clear need for an in-depth analysis of this kind of information, based on textual expressions of human particular experiences

Our approach, in the framework of the *Web of Experiences*, aims at acquiring practical knowledge from individual experiences with entities in the real world expressed in textual form. Moreover, this knowledge has to be represented in a way that facilitates the reuse of the experiential knowledge by other individuals with different preferences. Our approach has three stages: First, we extract the most salient set of aspects used by the individuals to describe their experiences with the entities in a domain. Second, using the set of extracted aspects, we group them in concepts to create a concept vocabulary that models the set of issues addressed in the reviews. Third, using the vocabulary of concepts, we create a bundle of arguments for each entity. An argument bundle characterizes the pros and cons of an entity, aggregating practical knowledge from judgments written by individuals with different biases and preferences. Moreover, we show how argument bundles allow us to define the notions of user query and the satisfaction degree of a bundle by a user query, proving that argument bundles are not only capable of representing practical knowledge but they are also useful to perform inference given a set of user preferences specified in a query.

We evaluate the argument bundles of our approach with the Amazon score ratings and the camera characterizations of Dpreview. We show that pro and con arguments are very close to those listed in Dpreview. Evaluating entity rankings, we show that Dpreview and our approach give congruent rankings, while Amazon's is not congruent neither with Dpreview's or ours.

Acknowledgements

This wonderful trip is dragging to an end. After an intensive period of writing, today is the day: I am finally finishing my thesis.

Those past 4 years had been a period of intense scientific learning and personal cultivation. I have learned a lot about science, meet a lot of interesting people, and traveled around quite a lot. I have visited many countries and conferences, and presented my research in some of them. And although I have been through some difficult times (long nights without having much sleep, periods of stress before submissions and presentations...), I would go through all of it over again. Well, maybe not *all* of it, but you get the point.

Honestly, I think going for a PhD is one of the most important decisions that I've taken in my life, and I am not disappointed. I'm happy with the results of my research, and I'm happy about the endless hours I've spent in this work.

First of all I want to thank my advisor Enric Plaza, because this work would not have been possible without his support and advice. Thank you for guiding me trough the process of becoming a scientist. Thank you for all those discussions and ideas, for your patience and help with The Writing (I've started to like it!), and for always having time for me. You have taught me to enjoy doing research, and that there is no better reward than a work well done.

Another cornerstone of this work are the visits to other research groups. I would like to thank Nirmalie Wiratunga for hosting me during three months in the Robert Gordon University of Aberdeen (Scotland). I spent a wonderful and fruitful time working under your supervision. I would also like to thank Mirjam Minor for hosting me in the Goethe University of Frankfurt am Main (Germany). Although my German did not improve much, I had a great time in Frankfurt. The short research periods with you and your groups were important steps for the maturity of this thesis. Thank you for helping me when I was in need, and for pointing me to the correct questions. I would also like to thank the researchers and friends I met in both universities; thank you Lei, Liang, Sofie, Felix, and specially Elaine, for being an awesome coauthor and a better person.

A great part of this monograph has been developed in the Artificial Intelligence Research Institute (IIIA-CSIC) situated in the campus of the Universitat Autònoma de Barcelona. I really loved this place, and since the first day I felt at home. Here I met wonderful scientists and friends, we went for beers and discussed about really important stuff, such as *link prediction in evolutionary* *graphs*. While writing this, I realized that I had crossed paths with a lot people in the IIIA-CSIC institute. Some were already ending their PhD's when I arrived, and helped me a lot with their experience, specially during the first year. Some others were here only for a short period of time, but became an important part of my day-to-day life in the institute. I would like to say thanks to all of them. To Jesus, Marc, Pere, Mari Carmen, Toni, Pablo, Arturo, Paula, Ewa, Kemo, Dani and Sergi, and the rest of thesis-mates and young scientists Axel, Sergi, Albert and Toni. I would also like to thank Tito, Lluís, Pere, Jesus, Jar, Roberto, Julian, Felix, Andrew, Dagmar, Carles, Ramon, Ana, and the other staff members, who were always there when I needed a hand.

I want to thank as well all my friends from Terrassa, the city I lived in during my PhD. To all my *caramerdes* and *packs*, wherever you are, thank you for the good times. And thanks to all my *mancors* university friends, Sandro, Polete, Samu, Boji, Edgard, Iban, Pajaro, Anna, and Eslí, because those years where some of the best years of my life.

Finally, these last lines are for my family. To my mother Carme, my father Jordi, and my brother, Albert. Thank you for always being here for me, and for your unconditional love and support. And to Laia, my soulmate, thank you for taking care of me during this journey and for coping with me during those long, sleepless nights.

– Xavier Ferrer Aran

This work has been partially supported by AGAUR Scholarship (2013FI-B 00034), Project Cognitio TIN2012-38450-C03-03 and NASAID (CSIC Intramural 201550E022).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The social web, with forums, review sites, networking sites and social media, offers a platform where millions of people interact, share interesting information, and socialize every day. The amount of social content and information generated by people in these sites is huge and diverse: from daily conversations to comments about products in e-commerce sites. In this work, we are interested in user-generated content expressing people's experiences while acting in the real word, i.e. expressing their experiences while traveling, hiking, watching movies, taking photos or meeting with friends.

People very often use other people's experiences before making decisions: it is a common behavior to ask a friend for some hotel or restaurant, or what to visit in a city that a friend has already visited. When deciding to buy a new digital camera, we may ask friends which camera are they using and what is their experience: is the price fair? Do they enjoy shooting with that camera more than with the old one? Is there anything they particularly dislike about it?

We constantly reuse other people's experiences in our daily activities for our own personal purposes. Nowadays, we have textual records of experiences by other people —granted, people we do not know and whose trust may be a concern. However, strength in numbers may compensate for debatable trust, since taking into account hundreds or even thousands of recorded experiences about entities in a domain (e.g. hotels or cameras) is likely to help us discover well supported pros and cons for each one of the entities.

Nevertheless, the web is now captured by the "search and browse" paradigm, so a person trying to book a hotel may read a few reviews over a few hotels — but cannot possibly read them all! That is why in practice we, individually, can only use 5-star scoring or similar rankings that summarize this plethora of information. The particular experiences of the people writing reviews are largely ignored and substituted by asking them to evaluate a few fixed topics with a numeric score. There is a clear need for an in-depth analysis of this kind of information, based on textual expressions of human particular experiences, that goes beyond the 5-star scoring system. However, this is not an easy task since it would mean a global agreement about how to represent all different kinds of user-generated content we might find in the web.

A new approach, called "The Web of Experiences" [Plaza, 2008] was proposed in the context of the International Conference on Case-based Reasoning (ICCBR). This approach proposed to enlarge the paradigm of Case-based Reasoning (CBR), based on solving new problems by learning from past experiences, and include all forms of experiences about the real world expressed in the web as user-contributed content [Plaza, 2009]. As other researchers were interested in exploring these ideas, three workshops were organized in collaboration with ICCBR in 2009 and 2010 to start up this topic, the *WebCBR-09: Reasoning from Experiences on the Web* and *WebCBR-10: Reasoning from Experiences on the Web*, and a related workshop in collaboration with ICCBR 2012, the *TRUE Workshop* focused in Traces for Reusing Users' Experiences - Cases, Episodes, and Stories.

1.1 The Web of Experiences

The core idea proposed by the Web of Experiences approach is that among the wide diversity of user-contributed content on the web there is a particular kind of content that has the potential of being put to good use by intelligent systems: human experiences. This user-generated content does not merely reflect an opinion or a belief of an individual about a certain fact, but describes (in some format) practical experiences in the real world.

For instance, when a person has bought a camera p or has experienced a stand in a hotel h, and writes a review on a web site, the comments of that person concerning both experiences are not merely expressing general opinions or beliefs, but facts about his actual dealings with that camera or that hotel. If the camera has a low quality built-in flash or the hotel's staff did not attend the person's requests, the comments are statements that certain facts occurred. Furthermore, comments such as "p's built-in flash is poor" and "h's staff is quite unfriendly" are also specific facts that may indicate, if similar comments are repeatedly expressed by a number of people, a recurring pattern in the experiences of people when dealing with p and h.

When these kind of human experiences are *expressed and recorded* on the web, they can be shared with by large number of people. These recorded experiences share *practical knowledge* concerning a wide variety of real world objects and situations. This practical knowledge is different from theoretical knowledge, such as that which can be provided by the Semantic Web. For instance, the Semantic Web approach can offer theoretical knowledge about hotels as in the statement "Hotel h is three stars, according to European Standards", which means that some authority has classified hotel h this way because h satisfies certain properties adjudicated to that category. Although this knowledge also provides evidence on the quality and features of hotel h, it is in fact knowledge about the *three star hotel category* rather than specifically about the hotel h.

Theoretical knowledge is, by definition, about a category or a class of objects

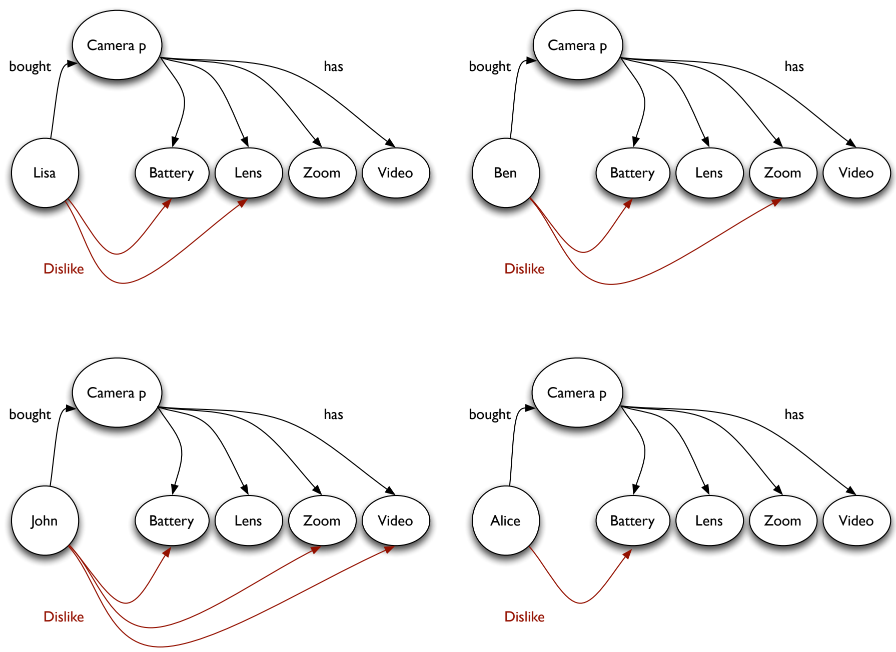


Figure 1.1: Cluster of relationships among various experiences with camera p.

or situations — while practical knowledge is mostly about specific instances (of objects or situations). What is there about instances (that is not in a class)? One answer is provided by the Case-based Reasoning, that works with a *case base* (a memory containing a collection of instances). CBR retrieves, analyzes and reuses cases from the case base in order to decide the solution or outcome of a new problem. The question "What is there about instances", may now be reformulated as this: *What is represented in a case?* Basically a case can be seen as a concrete *cluster of relations* among a collection of instances (that includes how people *use* that object or instance). Most of these relations are outside the purview of a theoretical/semantic definition of a category. Experiences, on the other hand, being concrete, are precisely those *clusters of relations among instances.*

Human experiences, when expressed, essentially provide a description of how people have used an object — and therefore a description of relevant relations of that object with its (physical and conceptual) environment. If we can find patterns on these clusters of relations among instances originating from a number of different people, we can gain new, practical knowledge from their experiences in that domain.

Figure 1.1 shows four clusters of relations describing the experiences of four persons (Lisa, Ben, John, and Alice) regarding a specific camera p. So, Lisa particularly disliked the battery and lens of that camera; Ben disliked the battery as well as the zoom of camera p; John disliked the battery, the zoom and the lens; and finally, Alice did not like the battery either. Analyzing these four experiences about camera p, we may observe a common pattern: they did not like the battery of camera p. This common pattern is relevant practical knowledge that will, most likely, be repeated in experiences of other people with the same camera p.

The web usually organizes people interested in the same domain into websites. In our approach, those contributing users are considered members of a community of practice[1]. The practical knowledge present in the written experiences created by a community of practice may be useful for other people interested in the same domain. Before the Internet, one option was to ask friends and family on their experiences concerning the camera they own. Now we have a new option, that is to say reading user-contributed reviews about cameras in one or several websites on digital photography.

However, since experiential content is treated as documents in the web, users typically first need to use a search engine (such as Google or Yahoo), or an internal search engine (e.g. search inside a forum for the posts that may talk about the topic of interest), to find a *relevant content* on that domain. Then, users need to browse a large collection of *found items*, perform a cursory read of them to filter out those (seemingly) irrelevant, and carefully read just a few of them (seemingly more relevant) to finally take a decision. Moreover, the users have to reuse this relevant content, that might be dispersed in dozens of pages, without any support that facilitates their task — the users may need to copy and paste the information pieces found relevant in those web pages, or print all those pages and annotate the important issues on them, and finally integrate that information and take a decision.

This process is inefficient if it is to be manually processed by each individual user considering the vast number of information available on the Web. Conscious about the problem of finding relevant content, some websites focus on improving information in the web by developing better search and retrieve techniques [Spink et al., 2001]. However, this solution only partially solves the user's tasks on identifying relevant content, but it does not solve the problem of how to reuse this content. Furthermore, this reuse implies that the content the user is interested in finding directly depends on the purpose of that user.

This retrieval-based approach is called the "Search and Browse" paradigm in [Plaza, 2009], and the Web of Experiences is intended as a new paradigm that focuses on the user task:

> " *[The Web of Experiences] is not about finding something (in the Web), it's about doing something (in the world), taking an action in the world, and the Web is merely used to take a more informed decision or action. For this reason my emphasis is on reusing the experiential knowledge provided by others for the actual purposes of a final user.* "

[1] By community of practice we mean a group of people who share a concern or a passion for something they do, and learn how to do it better as they interact regularly. The three key elements of a community of practice are: the domain, the community, and the practice.

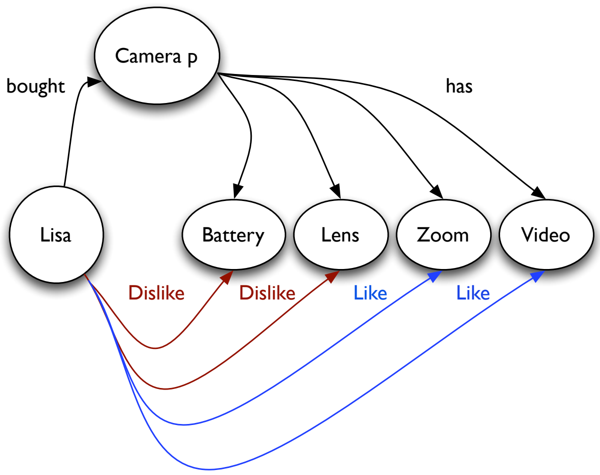


Figure 1.2: Lisa's positive and negative experiences with respect to camera p's aspects.

How can we help people making a more informed decision based on the experiences of other people? As we will show, A.I. based support can improve both the user workload and the outcome quality.

1.2 Framing the Problem

As we have seen, the practical knowledge contained in user experiences on a given domain may be useful for other people interested in taking more informed decisions in that domain. Buying a new camera, finding a good cycling route trough the mountains, staying in a nice hotel, or learning how to cook a healthy meal, are tasks that can benefit from the practical knowledge of other people's experiences.

However, an important issue about the Web of Experiences is that experiences are not cases, as in classical CBR, where a case has usually the form *(situation, outcome)*. In the Web of Experiences, such records of experiences are a collection of situations without a specifically designated element that is "the solution" or "the outcome". That is to say, we have records of individual experiences that describe certain facts in relation with entities, facts that can be positive and negative from the individual point of view, but that do not explicitly express the individual goals, or preferences, or the outcome.

For instance, in Figure 1.2 we show an individual experience description with camera p. That experience description contains certain aspects of camera p judged positive or negative, but do not explicitly express the individual goals or preferences of this individual or how they differ from those of Lisa, Ben, John or Alice in Figure 1.1. What we have is the result of the individual goals or preferences in judging the interaction of Lisa, Ben, John and Alice with specific aspects or properties of the entity.

Since records of experiences are not simply of the form *(situation, outcome)*, a valid question is which *forms of experience* may be found [Plaza, 2008]:

> “ *How many different* forms of experience *are there? [...] are there a small collections of forms of experience that could be characterized and reused? Which are they and how to find them? [...] It is essentially an empirical matter to be settled only after trying to develop systems that reuse experiential knowledge on the web.* ”

We identify four different forms of experiences, one for each class of task commonly known in Artificial Intelligence: classification, regression, planning and configuration. These types of tasks are distinguished and characterized by having solutions of different forms. Classification tasks assume there is an enumerated collection of known solutions, and where the goal is, given a problem description as input, selecting one of the solutions (or equivalently, an ordering over the collection of solutions). Thus, selecting a camera among a collection of cameras in a shop, or selecting a hotel in a city, are examples of classification tasks. Regression tasks are characterized by a numerical attribute understood as the solution attribute (e.g. price), and where the solution is to predict the numeric value of the solution attribute given a problem description. Planning and configuration are synthetic tasks, i.e. the complete set of possible solutions are not known beforehand, since their number is huge; synthetic tasks, however, do have a enumerated collection of known *solution elements*, from which solutions can be constructed. Planning tasks assume there is a set of known solution elements called actions, where the goal is to create a solution composed by a sequence of actions (or a partially ordered collection of actions) given a problem description. Configuration (or design) tasks assume there is a set of known solution elements that can be composed in a complex structure, and where the solution is to build a concrete network of interconnected solution elements such that the requirements of a problem description are satisfied. Case-based configuration and design systems have developed techniques for this kind of task.

From Artificial Intelligence we know that methods for each of this kind of tasks are very different. There are methods adequate for planning and methods for classification, but there are no general methods for all types of tasks. Nevertheless, Artificial Intelligence methods for classification are expected to work in different domains that require classification. We assume that each type of task would require different methods to be developed inside the framework of the Web of Experiences. Therefore, in this monograph we have to choose one task in the framework of the Web of Experiences, and we choose the task of classification.

Assumption 1. *A goal of this research is to develop methods for the Web of Experiences amenable for the task of classification, i.e., we assume the domain has an enumerated collection of known solutions and, given a problem descrip-*

tion, the method has the goal of selecting one of the solutions (or equivalently an ordering over the collection of solutions).

It is clear that the focus on user-contributed experiential knowledge also poses a practical constraint: the form in which experiences are expressed has to feel easy and natural to the people integrating a community of practice; otherwise, far less content will be contributed by them. There are types of experiences easier to express in textual format, such as product reviews, and other experiences easier to express in other formats such as video, for instance a game replay or a cooking recipe. Clearly, experiences that are expressed as textual content are the most widespread mode of sharing experiences on the web nowadays. Therefore, in addition to focus only on classification tasks, a further assumption is that we will work with experiences expressed as textual user-contributed content.

Assumption 2. *A goal of this research is to develop methods for the Web of Experiences where individual experiences are expressed as user-contributed content in one or several web resources that can be considered a community of practice, i.e. where domain, community and practice can each be clearly discerned.*

To perform experiments and evaluate our approach we need to choose a domain with a large quantity of textual user-contributed content in which authors express their practical experiences about real-world entities. Furthermore, user-contributed content needs to be easily accessible. Therefore, we experiment with the domain of digital photography, using the practical experiences of users contributed to Amazon.com in the form of product reviews.

Assumption 3. *A goal of this research is to experiment with the methods we develop for the Web of Experiences for an specific domain, for which we select digital photography, and a community of practice, for which we select the reviews of digital cameras available from Amazon.com.*

Digital photography is a domain with a large number of textual user-contributed content and big communities of practice (i.e. Amazon's), where people freely express their personal experiences with digital cameras. Differently from user reviews we can find in other domains, digital camera reviews tend to be detailed. Users express their experiences with respect to the numerous aspects of a camera; we find reviews that describe user personal experiences with issues like battery life, lens quality, image sharpness, or high definition video mode. Furthermore, this knowledge is interesting and reused on a daily basis by other people, for instance those interested in buying a new digital camera.

These assumptions are stated to clarify the limits of the problem we want to address inside the framework of the Web of Experiences. Given these assumptions, we turn now into the way we intend to tackle the problem of (1) analyzing and acquiring practical knowledge from user-contributed textual descriptions of individual experiences (in particular, reviews on digital cameras), and (2) structuring this practical knowledge in such a way that can be used in a classification task where, given a problem description from a concrete user, a solution to that problem is yielded (as a ranking of digital cameras adequate for the problem description).

1.3 Approach

We have seen that every textual experience is unique and provides a particular description of how someone has used an entity and what he has experienced with that entity. As such, different users may have different experiences with the same entity. In Figure 1.1, Lisa, Ben, John, and Alice, had different experiences with various aspects of a digital camera p; John and Ben have had a poor experience with the zoom of the camera p, while Lisa and Alice enjoyed it. In order to effectively reuse this experiential information, there is a need to find a way to acquire the practical knowledge implicitly present in the texts and based on people's interactions with an entity. For this purpose we introduce now the notion of *issue*.

Issues. *Issues* are the most salient properties of the entity as described in the textual reviews of people's experiences. In our approach issues are not inherent properties of the entity, they are rather conceived of as arising from the interaction of an individual person with an entity by which an experience of usage is incorporated in the mind of that person, and later it is expressed in textual form.[2] An issue can be expressed by a wide variety of textual descriptions, because different *lexical items*[3] can refer to the same issue (i.e. have the same semantic content or meaning).

A clear example of two words usually referring to the same issue are synonyms. For instance, 'picture' and 'pic' are two lexical items often used as synonyms to describe the same issue of the camera. Consequently, these two sentences "the pictures taken with this camera are great" and "the pics taken with this camera are great" have the same intended meaning. It would not make sense to consider both 'picture' and 'pic' as different issues, because they are, in fact, referring to "the image taken by a photographic camera".[4]

Now consider the words 'resolution' and 'megapixel' in the context of the next two sentences: "the resolution of the camera will meet your needs" and "with 12 megapixels, you will get detailed pictures". Although those words are not synonyms, both 'resolution' and 'megapixel' are used to describe the issue "the detail an image holds". Despite those two individuals used different words to describe their experiences, both words are, in fact, referring to the same camera issue as well.

---

[2] Our notion of issue is related to the notion of *affordance*, but is different in that affordance is applied to just those action possibilities that are readily perceivable on an entity by an actor. Affordances are more related to a good interface design that facilitates human-machine interaction. Issues may be affordances, but not necessarily: e.g. in the domain of digital cameras, if battery life is perceived as too short by an individual, this is an issue for us, but it is not an affordance in the traditional sense.

[3] A lexical item, or lexical unit, is a single word, a part of a word, or a chain of words that forms the basic elements of a lexicon or vocabulary.

[4] The process where lexically different concept representations are judged to have the same meaning receives the name of *semantic unification* in philosophy, linguistics, and computer science [Hagoort et al., 2009].

Therefore, to discover the issues described in the reviews of a product is equivalent to the task of identifying all semantically equivalent lexical items.

Judgments. Individuals usually describe their experiences concerning different issues of an entity by means of sentences with positive or negative polarity. For instance "the camera lens is outstanding", describes a positive experience with respect to the camera 'lens' issue. So, to deal with this crucial factor, we introduce now the notion of *judgments*. We call a sentence that expresses a positive or negative experience on a particular issue a *judgment* on that issue. We call those sentence judgments because they are considered determinations or sensible conclusions made by the users, based on their experiences, about the various issues of an entity.

Therefore, our hypothesis is that, for each textual description of an individual experience, we should be able to discover a set of positive and negative judgments, related to the different issues of the entity being discussed or evaluated. Thus, we assume that every user experience will have a set of positive and negative appreciations as experienced by interacting with that entity, and that the more salient appreciations will be reflected in the text written by each individual in a set of positive and negative judgments.

Arguments. We will now introduce the notion of *argument* applied to an entity. By analyzing the recurring patterns between the sets of user judgments about the various issues of an entity, we expect to discover those issues that play a positive and a negative role for that entity, as experienced by the users. Thus, our next hypothesis is that, for any entity, aggregating the judgments from the experience-based texts about that entity, we will be able to determine for which issues there are recurring patterns among the individual experiences. Whenever a recurring pattern about an *issue* of an entity is found, we consider that we may create an *argument* about the issue of that entity. Thus, an argument is a knowledge structure created by aggregating the knowledge found in the recurring patterns of user judgments concerning a specific issue. Moreover, if a recurring pattern is found over a set of positive judgments, we consider this to be a *pro* argument. Otherwise, if the recurring pattern is found over a set of negative user judgments, we consider this to be a *con* argument.

An argument about an entity can be seen as a reason in favor (pro argument), or against (con argument) preferring that entity over others on account of the issue of the argument. For instance, consider the experiences of Lisa, Ben, John, and Alice presented in Figure 1.1. They all disliked the battery of camera p. Suppose we analyzed the written judgments of those four users with respect to the issue 'battery' of camera p, together with other people's written judgments, and found this recurring pattern: 90% of them dislikes the battery of camera p. With this analysis, it would be reasonable to create a con argument on the 'battery' of camera p (that expresses the said recurring pattern).

Bundle of Arguments. Next, we will introduce the notion of *bundle of arguments* of an entity. A bundle of arguments of an entity characterizes that entity in terms of the recurring patterns of (positive and negative) judgments. Thus, the bundle of arguments of an entity will be build by grouping all arguments that can be created from recurring patterns of judgments found in the textual descriptions of user experience concerning that entity. Therefore, a bundle of arguments of an entity is the collection of arguments (positive and negative) for preferring that entity over others, embodying the practical knowledge of the textual expression of individuals experiences, to the degree that our methodology will be able to ascertain the salient issues, identify sentences that express judgments, estimate the degree of their positive and negative polarities, and accurately aggregate this knowledge in the form of argument bundles. Section 1.4 will explore these challenges in more detail.

1.4 Challenges

Related to the analysis and reuse of user experiences in textual form and its applications, there are many challenges that can be addressed. In this section, we introduce the specific challenges to be tackled in this thesis.

Challenge 1. *Identify the set of salient terms, in a given domain, used by people while writing about their experiences with a concrete entity of that domain.*

Specifically, since our experiments focus of the domain of digital photography, we are interested in identifying the most salient terms used by individuals when describing their experiences while using digital cameras. For instance, do people write a lot about the 'lens' of the camera? Or instead they focus in the 'image quality' of pictures? Do they often use 'pic' instead of 'picture'?

The set of salient terms (or lexical units) more often used by people while writing their experiences within a domain will constitute for us a vocabulary of that domain. Currently, literature on text analysis and mining calls these terms either 'aspects' or 'features', and more recently this task has been often been dubbed 'aspect extraction'. Therefore, from now one we will refer to these salient terms as aspects and our goal, given a domain with a corpus of texts, will be to create an aspect vocabulary (see Chapter 3).

The vocabulary used by people while describing their experiences does not need to be the same that can be found in the classical feature lists present in product catalog descriptions published by the companies selling the cameras. That is to say, camera reviews created by people may focus on a limited subset of the list official and technical features describing the camera by its manufacturer; moreover, camera reviews created by people may also focus on aspects not considered "features" in the classical sense.

For instance, in user reviews, people usually uses the word 'pic' to describe the photographic image obtained with a camera, while 'pic' is not usually found in the camera feature lists. Even the use of acronyms varies between professional and non-professional vocabularies. The word 'evf' (electronic viewfinder), for instance, is often found in user reviews while is not used to describe cameras in the common camera feature lists. Thus, we cannot simply use a pre-existing set of features (like camera features lists) to build an aspect vocabulary. Our goal in creating an aspect vocabulary is to *identify the aspects actually used* while describing the experiences of the users .

Challenge 2. *Identify user judgments and their polarity.*

As we have discussed in Section 1.3, users write judgments about the different issues of a product arising from their experiences while actually using a camera. When, for instance, a user writes a sentence saying that he or she likes the 'picture' taken by a camera, he is doing a positive judgment about the issue 'picture'.

Judgments are found within sentences that express a polarity, positive, or negative, with respect to an entity's issue. The given polarity, positive or negative, depends on the individual preferences, bias and expectations, and therefore a variety of responses concerning a specific issue on the same entity can be large. However, as we previously discussed, even if we do not know the individual variety of preferences, bias and expectations, we can analyze whether for particular issues of particular entities there are recurring patterns or not. Therefore, judgments are important constructs that express the experiential knowledge of the individuals in user reviews, and being able to identify and assess their polarity is an important challenge to be addressed.

However, we do not have explicitly defined the entity's issues in the text, we just have the aspects that are related to the entity. Therefore, our approach is that judgments can be identified by searching, in the user reviews, for sentences with a positive or negative polarity that refer to one or more aspects from the aspect vocabulary. Thus, the challenge is to adequately identify those sentences that refer to aspect vocabulary words by analyzing the semantic structure of a sentence (see Chapter 3), and then accurately assess its polarity value, by means of sentiment analysis techniques (see Chapter 4).

Challenge 3. *Discover the main issues addressed by people when writing their experiences and create a concept vocabulary.*

Issues are addressed in user reviews by means of judgments, but they can be expressed using a wide variety of words (aspects from the aspect vocabulary). Each aspect does not need to be an individual issue, but several aspects can refer to the same issue of a camera. For instance, consider the next sentences:

- "The electronic viewfinder of the Canon T3i is great."
- "Canon T3i features a new evf, improved from the previous Canon."

Both 'electronic viewfinder' and 'evf' are important aspects belonging to the digital camera domain. Aspect 'evf' is an acronym of 'electronic viewfinder', often used in camera reviews. Therefore, both sentences are expressing judgments with respect to the same issue of a camera: the viewfinder (that in this model happens to be electronic and not optical). It would not make sense to consider both words as different issues, because both words refer to the same concept “a camera viewfinder where the image captured by the lens is projected electronically onto a miniature display”.

Acronyms and dictionary synonyms are clear examples of how different aspects from the aspect vocabulary can refer to the same issue in the reviews. Identifying that ‘electronic viewfinder’ and ‘evf’ refer to the same issue does not suppose much of a problem, since we can use a dictionary or an acronym dictionary to learn that both words are related. However, since we are working with user experiences, people use their own vocabulary. And it is a challenge to identify that two or more *non-dictionary* related aspects refer to the same issue.

Moreover, there are optical viewfinders and electronic viewfinders: should they be considered as different or as a single issue? From a certain viewpoint, the issue is “the viewfinder”, since it is the same affordance of the entity with respect to the user. There is no predefined solution to this dichotomy, since some individuals may treat them as two distinct objects while other individuals may consider them two modalities of the same object. The challenge is thus deciding, beyond mere synonymy, which aspects are considered to be part of the same issue, and thus indistinguishable at our level of analysis of experiential knowledge.

Our approach is to create a vocabulary of concepts, where a concept models an issue addressed in the textual descriptions of experiences. Creating such vocabulary involves deciding, after analyzing the textual corpus, which aspects are treated as interchangeable or indistinguishable (at the level of user judgments). Those aspects that are indistinguishable, or very close, will be deemed to refer to the same issue (i.e. to have the same meaning, at least for our purposes in analyzing judgments). Consequently, those aspects will belong to the same concept in our model. Therefore, from a computational point of view, creating a vocabulary of concepts means determining a partition of the vocabulary of aspects, where every aspect is assigned to one concept only and each concept gathers those aspects whose meaning is very similar or indistinguishable.

In order to partition the aspect vocabulary, we need to analyze how people use those aspects in the reviews. Are two aspects often used together in similar sentences? Is the polarity of those two aspect’s judgments correlated over the set of reviews of an entity? Those questions can help elucidate the main concepts referred by people when writing their experiences about digital cameras, in order to identify the most important issues described in people’s experiences.

Challenge 4. *Create the arguments and the bundle of arguments of an entity.*

The vocabulary of concepts partitions aspects into clusters, each of which refers to a specific issue. Moreover, while the judgment analysis detects the aspect involved in a judgment and its polarity, now we can group together the judgments addressing the same concept (i.e. the same issue). Thus, a positive or negative judgment of an individual is now interpreted as referring no longer to an aspect, but as a judgment on a concept.

Since concepts model the important issues found in a set of user experiences, it makes sense to characterize an entity based on the practical knowledge with respect to those concepts. Therefore, we focus first in searching all judgments about a certain concept, for a given entity, in order to obtain a collection of positive and negative judgments created by various individuals. By analyzing this set of user judgments, we will find when there is a recurring pattern among the individual judgments with respect to that concept of that entity.

Arguments are created by aggregating the polarity of user judgments related to an entity's concept in which a recurring pattern is found. We talk about pro arguments when the aggregated polarity of the judgments that form those recurring patterns is positive, and we talk about con arguments if it is negative. The intuition here is that a pro argument suggests than an entity should be preferred over other entities considering that all other arguments are equal, while a con argument suggests the opposite.

Aggregation introduces a new challenge in this monograph: when sound arguments can be created from sets of judgments. For instance, a sound argument cannot be created when the set of user judgments of a concept is too small. Moreover, since experiences may be biased towards individual preferences we may find judgments of opposite polarity on the same concept and entity. The challenge here is that having a "recurring pattern" is not a dichotomy, but rather a matter of degree. For instance, if the positive polarity judgments about an entity's concept outnumber by far the negative polarity judgments, we would create a pro argument by aggregating these judgments — but this argument should have a strength that is lesser than the strength of another argument where the judgments are unanimously positive. Our challenge, therefore, is assessing the *strength* of an argument while aggregating a set of judgments, the intuition being that a high dispersion of polarity values will be reflected in an argument with lower strength than an argument based on a set of judgments whose polarity values are clustered together. Moreover, the more user judgments about a certain concept, the less bias towards the individual user preferences will be present in the acquired knowledge. In Chapter 5 of this monograph we present three different types of arguments created using different aggregation measures over the polarities of user judgments. The three aggregation measures determine the strength of an argument by assessing, in different ways, the dispersion/concentration tradeoff in the distribution of polarity values over a set of judgments. Since the aggregation measures are different, the sets of pro and con arguments for any given entity also vary depending on the argument type.

Now consider all arguments pertaining to an specific entity. We have now a *bundle of arguments* that characterizes that entity by means of the arguments created on the concepts that were found more salient. Therefore, if an argument could be seen as a reason, related to a concept, about why to prefer (or not) an entity, a bundle of arguments can be seen as a collection of reasons, over the set of salient concepts described in user reviews, about why to prefer (or not) an entity over others.

However, arguments have a particular strength, in addition to being pro or con. Therefore, a new challenge is, for any given entity, how to select those arguments that are strong enough to be part of the bundle, in such a way that they constitute an adequate characterization of that entity, as we show in Chapter 5.

The bundles of arguments are the structures that embody the practical knowledge acquired, using our approach, from individual textual expression of their experiences. In order to support the reuse of this practical knowledge, we introduce the notion of *user query*. Moreover, the notion of *query satisfaction* relates a query with a bundle of arguments. A user query expresses the individual preferences of a user, while the degree of query satisfaction estimates to which extent an entity's bundle satisfies the preferences expressed in a query. For instance, considering a user whose preferences are 'HD video' and 'zoom'; then in a given domain with a collection of entities (cameras) characterized by argument bundles, the query ranks the cameras by their degree of query satisfaction, and the one with higher degree is the one that better satisfies the preferences of that user.

1.5 Additional Remarks

In this dissertation, all experiments are performed over three corpora obtained from Amazon.com[5] during April 2014, and a public dataset used to evaluate the unsupervised aspect extraction methods over a set of entity reviews containing manually marked-up entity aspects [Ding et al., 2008; Hu and Liu, 2004a]. The Amazon corpora corresponds to three datasets formed by user reviews about entities of three digital camera categories: Digital Single Lens Reflex (DSLR), Compact (COM), and Point & Shoot cameras (P&S).

These category distribution do not exist anymore in the actual digital photography categorization of Amazon, as Compact and Point & Shoot camera types are joined together under a new category named Mirrorless cameras. During the evaluation sections of the first chapters of this work, we analyze if there exist clear differences in vocabulary among the three camera types (DSLR, Compact, and Point & Shoot) in order to consider them as three separate corpora, or just one single corpus. As we show in Chapter 3, the differences between the aspect vocabularies of the three domains are clear, and as such we consider the three Amazon camera types as three different corpora. Therefore, we talk about DSLR corpus (K_D), Compact corpus (K_C), and Point & Shoot corpus (K_P).

1.6 Structure of the Thesis

The rest of the thesis is organized as follows.

In Chapter 2 we discuss the research work related to this thesis. This thesis is related to information retrieval, aspect extraction and sentiment analysis techniques applied to social networks and the Web of Experience. The literature

[5] https://www.amazon.com/

related to those subjects is large. Therefore, we introduce some of the most relevant and related work in these areas, while describing the differences between their and our work.

Challenge 1 is addressed in Chapter 3, where we analyze user experiences in textual form to identify the salient words used in product reviews to create an *aspect vocabulary*. We present four complementary methods of aspect extraction from a corpus, combining an unsupervised aspect extraction method with the knowledge extraction from two digital camera webresources and *WordNet*[6] in Sections 3.2, 3.3, 3.4, and 3.5. The unsupervised aspect extraction method is evaluated against a set of manually tagged reviews containing manually marked-up product aspects [Ding et al., 2008; Hu and Liu, 2004a] in Section 3.2.2. Each one of these aspect extraction methods generates an *aspect set*, and their union builds the aspect vocabulary for that corpus as described in Section 3.6. In the same section three aspect vocabularies are created from three different corpus formed by user reviews. The three aspect vocabularies are compared and evaluated, to decide whether is better to work with one general photography domain (one general aspect vocabulary), or three domains (three distinct aspect vocabularies). The conclusions of this chapter are presented in Section 3.7.

In Chapter 4 we address **Challenge 2** and **Challenge 3**. We identify the issues addressed in product reviews expressing people's experiences on using those products, and create the *concept vocabulary* of a corpus from the aspect vocabularies created in previous chapter. To do so, we first introduce the notion of *basic level concepts* in Section 4.2. The concepts of the concept vocabulary are created by clustering the aspects of the aspect vocabulary using a bottom-up hierarchical clustering approach. The clustering is based on a similarity measure that considers the way aspects are used by people when expressing their personal experiences in product reviews, explained in Section 4.3. The hierarchical clustering returns a set of possible partitions, and each partition determines the set of concepts that model the collection of issues that are used in a review corpus. The problem of selecting one partition is addressed in Sections 4.4 and 4.5 by choosing, among all possible partitions resulting from the hierarchical clustering, the partition with the highest *sentiment coherence*. The selected partition is considered to describe the set of basic level concepts that form the concept vocabulary of each corpus. Section 4.6 analyzes the concept vocabularies created from the three camera corpus. Finally, Section 4.7 summarizes the conclusions and contributions of the chapter.

Challenge 4 is addressed in Chapter 5, where we introduce the *bundles of arguments* and methods to allow their reuse. Bundles of arguments are knowledge structures created by aggregating the knowledge from user experiences, with respect to the set of concepts of a concept vocabulary. Bundles of arguments are created by analyzing the practical knowledge from user experiences with respect to a product, and are formed by *arguments*, presented in Section 5.2. In the same section, three different types of arguments, created by using different aggregation methods, are presented. Section 5.3 presents the notion

[6] https://wordnet.princeton.edu/

of bundle of arguments of a product, which is the collection of arguments created over the set of concepts defined in the concept vocabulary considering the reviews of a product. To facilitate the reuse of the experiential knowledge contained in the bundles of arguments, we introduce the notion of user query and the query satisfaction based on fuzzy logic in Section 5.4. Section 5.5 presents an evaluation of the quality of the bundles of arguments by comparing the product bundles with the product descriptions and rankings of Dpreview and Amazon. Finally, Section 5.6 summarizes the conclusions.

The conclusions of this dissertation are presented in Chapter 6, where we summarize the work presented in this thesis, the contributions of the thesis, the related publications and a discussion of some future lines of research and final remarks, in Sections 6.1, 6.2, 6.3, and 6.4, respectively.

Appendix A presents the notation and equations used in this dissertation. In Appendix B we present a study about how the vocabulary of aspects used in the individual experiences about digital cameras changes over time. Appendix C presents the aspect vocabulary of the DSLR, Compact and Point & Shoot cameras, created as described in Chapter 3. Appendix D presents the concept vocabularies of the three camera types, created as described in Chapter 4. Appendix E presents a selection of bundles of arguments in detail. Finally, Appendix F presents a comparison between the Gini, agreement and cardinality bundles of arguments of DLSR, Compact and Point & Shoot cameras.

Chapter 2

Background and Related Work

In this chapter, we present some related works on the various tasks faced in this dissertation related with the analysis, extraction and creation of knowledge from experiences expressed in user-generated reviews. This dissertation uses various state of the art techniques related to natural language processing and information extraction. Furthermore, those topics are experiencing great improvements year by year, and the quantity of relevant papers and references keeps growing. In this chapter, we present a collection of the topics and related work most relevant to this monograph.

First, we introduce general background and relevant work about the Web of Experiences in Section 2.1.We review some common natural language processing techniques to extract important words, named features or aspects in the literature, in Section 2.2. Then, in Section 2.3, we summarize some relevant work on sentiment analysis, and present different approaches used to determine the polarity of sentences. We discuss some related work on topic identification and semantic unification in Section 2.4. Finally, Section 2.5 presents some related work concerning arguments and argumentation, argument detection, argumentation schemes, and the creation of pro and con arguments presented in this monograph.

2.1 Web of Experiences

The Web of Experiences (WoE) [Plaza, 2008] is built upon some emerging content on the Web that offers an opportunity and a challenge for the Artificial Intelligence community: people's experiences. These experiences, ranging from client reports on hotels to small explanations on how to do certain things, contain knowledge that is searched for and reused by thousands of people everyday in forums and blogs, normal websites and in specialized services like Question-Answer web pages. The objective of the Web of Experiences is to build systems capable of creating and organizing the knowledge of user experiences, and to facilitate the reuse of this knowledge in order to help other people make more informed decisions.

The Web of Experiences proposed to enlarge the paradigm of Case-based Reasoning (CBR), based on solving new problems by learning from past experiences, to include all forms of experiences expressed in the Web as user-contributed content [Plaza, 2009]. Three workshops were organized in collaboration with the International Conference on Case-based Reasoning ICCBR: the *WebCBR-09* (ICCBR 2009), the *WebCBR-10* (ICCBR 2010), and the *TRUE Workshop* (ICCBR 2012).

As we discussed in Chapter 1, experiences are a very special kind of user-contributed content. They differ from other knowledge sources (e.g. semantic or conceptual knowledge), in that experiences are memories of autobiographical events that can be explicitly stated or conjured up on one's mind.

The notion of episodic memory was introduced by Tulving [Tulving, 1985, 1986], where episodic memory is considered to be the collection of past personal experiences that occurred at a particular time and place. Therefore, human experiences do not merely reflect beliefs or opinions, but describe facts occurred in the real world. Moreover, Tulving introduced another type of memory, the semantic memory [Tulving, 1986]. Semantic memory is a highly structured network of concepts, words and images capable of making inferences and comprehending language. The difference between the two types of memories, episodic and semantic, is similar to the distinction between remembering something that has happened (episodic memory) and knowing something's meaning (semantic memory). As such, we can say that the web of experiences focuses on episodic memory (experiences), while the semantic web focuses on semantic memory (exemplified in the development of organized collection of concepts called 'ontologies').

When an individual describes personal experiences, for instance explaining his interaction with an entity and writing a "review" of those interactions, judging them good or bad, he is able to do that by recalling a set of relevant memories. This recall, this remembering process, is done by means of a human ability named *autonoetic consciousness*. Autonoetic consciousness allows us to mentally place ourselves in the past, and to be able to examine our own thoughts as if we were really at that time frame [Baddeley, 1992]. As such, when a person is writing, for instance, a review about a camera, he is using the autonoetic consciousness to place himself in the exact moments of those episodic memories to remember a set of experiences he had while dealing with that camera.

Experiences, stored in episodic memory, consist a special kind of content that, at the same time, provides a specific form of knowledge, the *experiential knowledge*. The goal of the various implementations under the Web of Experiences framework is to extract this practical knowledge and reuse it to help other people taking more informed decisions.

In that sense, both CBR and the Web of Experiences are based on the same notion of episodic memory and on using knowledge of the world contained in *memories*. Case-based reasoning is a computational model where episodic memory is modeled by the case base, and solves new problems based on the practical knowledge of those cases (past experiences). However, CBR only deals with one individual's memory, contained in the case base. On the other hand, the Web of Experiences enlarges this approach to deal with the experiences of others.

People's experiences can be found and reused in various formats (i.e. textual, video, music), depending on the domain and the reuse task. For instance, Freyne et al. [Freyne and Smyth, 2009] described a novel case-based reasoning application to help users visualize complex datasets utilizing the *Many Eyes* online visualization platform. The application learns from other users's graphs and decisions, called *visualization experiences*, about how to visually represent a given dataset, to recommend visualization configurations for new users with new data sets.

Another example is the Poolcasting system, that focused on creating musical playlists from existing playlists. In this approach, a playlist was interpreted as a human musical experience that put certain songs in a specific playlist because "played well together" [Baccigalupo and Plaza, 2006, 2007; Plaza and Baccigalupo, 2009]. Musical playlists were used to generate new playlists for a set of persons that formed the *audience* of the Poolcasting system. The songs in the final playlist were chosen from songs available in the final user own musical library, but the selection and the ordering of the songs into a new playlist was generated by the system based on the analysis of a set of people's musical experiences, those who were part of the audience of Poolcasting at a specific moment in time.

On the other hand, [Ihle et al., 2009] focused in cooking recipes shared in textual format. They extracted ingredients and comments from online recipes created by people, and analyzed their experiences in order to reuse this knowledge for future recipes. GhostWriter [Bridge and Waugh, 2009] also focused on textual records of experiences. GhostWriter is a case-based reasoning system that makes content authoring suggestions using feature-values extracted from Amazon.com reviews. Later on, GhostWriter-2.0 [Bridge and Healy, 2012; Healy and Bridge, 2010] was an improved version that participated in the WebCBR-10 challenge organized by the ICCBR 2010 conference.

2.2 Aspect Extraction

Aspect extraction is a key task of natural language processing (NLP) that aims to extract salient words from text. Aspect extraction has experienced rapid advances in recent years due to the increase of textual user-generated content in the web, and the interest of the scientific community and retail companies in this domain [Han et al., 2011; Kao and Poteet, 2007].

The salient words obtained when applying aspect extraction techniques are usually entity properties (also called aspects or features) [Andersen, 2007; O'reilly, 2005]. For instance, in sentence "In hotel h the room was alright but the bath was not OK", aspect extraction aims to extract 'room' and 'bath', words that can be considered attributes of the hotel. Aspect extraction is a challenging task, because text data is typically unstructured (in the sense that this information does not usually adhere to any predefined data model). Therefore, the most usual scenario when applying aspect extraction techniques to textual documents is that no common structures can be leveraged apart from the natural structures of the language.

Aspect extraction techniques can be applied to extract information from all kinds of text documents, such as legal documents or stories. Relevant to this monograph are user reviews, a particularly interesting type of text documents where the author expresses his experiences with respect to an entity or product [Petz et al., 2013; Zhang et al., 2010]. User reviews about products represent an extremely valuable source of information for retail companies, and an interesting domain for the scientific community.

In general, text extraction is related to entity identification, in the sense that the set of extracted aspects from the text is usually utilized to identify the entity or categorize the text as belonging to a certain domain [Dumais et al., 1998]. However, this poses no problem for our approach, since using reviews of products the relationship of each review to a specific entity is a given. Thus, our focus is extracting the salient aspects described in user experiences in order to understand what an individual considered important to describe about the entity they are reviewing.

The literature on aspect extraction techniques is extensive, the approaches can be divided into two main classes: unsupervised and supervised aspect extraction.

2.2.1 Unsupervised Aspect Extraction

Unsupervised aspect extraction approaches are typically based on frequency counts of words and in exploiting the semantic structure of sentences to detect a set of candidate aspects. Usually, unsupervised approaches aim at constructing a vocabulary of salient aspects from a given corpus of texts about a domain.

Frequency approaches are based in an observation: a limited set of words is used much more often than the rest of the vocabulary, and these words are more likely to be aspects. Frequency-based aspect extraction approaches usually focus their search in common nouns and compound nouns that are often repeated in a corpus of texts. This is because some studies indicate that 60-70% of the aspects are nouns [Liu, 2007]. Moreover, further studies found that the quantity of noun aspects was almost 90% [Spina et al., 2012] in microblog posts.

This straightforward approach has some downsides, since not all frequent nouns in a corpus of texts can be considered aspects. For instance words like 'daylight' or 'Photoshop' are often mentioned in digital camera reviews, but cannot be considered aspects of the digital camera. Furthermore, other specific aspects not discussed by most people may be missed by frequency-based methods. Nevertheless, frequency methods are simple and they generally achieve good results.

Specific aspects not discussed by most people (and thus with lower frequency) may be identified by using the term frequency - inverse document frequency (tf-idf) approach. The statistical measure tf-idf is used to evaluate how important a word is to a document in a collection or corpus. The importance increases with the word frequency in the document, and is offset by the total word frequency in the corpus. Therefore, specific aspects not discussed by most people can be identified by using tf-idf, if those aspects are characteristic of a small subset of documents and not of the entire corpus [Gupta and Manning, 2011]. The tf-idf technique yields good results if the objective is to determine the most *informative* words of every document with respect to a whole corpus. That is to say, those words that better define a document considering a whole corpus of documents. For this reason, it is not always suited to perform aspect extraction tasks in order to create an aspect vocabulary.

The ealier and more referenced implementations of the frequency-based aspect extraction are [Hu and Liu, 2004a,b], which only consider single nouns and compound nouns as possible aspects. After computing the frequency of the selected aspect candidates, Hu and Liu set up a threshold to filter out those noun combinations that occur in less than 1% of the sentences in the corpus. In order to reduce the quantity of false positives, Hu and Liu introduce two pruning rules in [Hu and Liu, 2004a]: the first rule removes combinations of nouns that do not appear often together in the reviews, and the second rule prunes those single nouns that also form part of a compound noun.

In [Li et al., 2009; Scaffidi et al., 2007], a part-of-speech pattern filter is also applied on top of the frequency filtering in order to only consider as aspects those nouns and compounds followed by an adjective. Furthermore, to remove false positives, they compare the frequency of candidate aspects with a frequency dictionary of words generated from a corpus of 100 million English words. For a word to be considered an aspect, it has to appear more often in the set of reviews than is likely given in the baseline frequency of the dictionary.

Some other relevant work in aspect extraction based on frequency filtering was made by Popescu and Etzioni [Popescu and Etzioni, 2007], Yi et. al. [Yi et al., 2003], Titov et. al. [Titov and McDonald, 2008a] and many others [Blair-Goldensohn et al., 2008; Dong et al., 2013b; Hu and Liu, 2004a; Huang et al., 2012; Moghaddam and Ester, 2010; Ronen et al., 2013; Zha et al., 2014]. Some of the aforementioned work use a combination of different techniques, such as frequency-based and syntax-based techniques, in order to identify words with sentiment associated. The relation between aspect extraction and sentiment analysis is further explained in Section 2.3.

Syntax-based methods, instead of focusing in word frequency to select the set of aspects, find aspects by means of syntactical relations between words [Grishman, 1997]. A very common and effective syntactic relation exploited to extract aspects is the adjectival modifier relation between a sentiment word and an aspect. For instance in the sentence "I think the lens is fantastic", 'fantastic' is an adjective that applies to 'lens'. As such, 'lens' would be considered an aspect. Many other syntactic relations can be defined to extract aspects, such as those between a verb and the subject. In Chapter 3 of this monograph, we explore some of them to create the aspect vocabulary. Using this technique, low frequency aspects can be found. However, many specific syntactical relations need to be described in order to get a good coverage of aspects.

Moghaddam et al. [Moghaddam and Ester, 2012] propose a set of grammatical extraction rules to extract aspects from product reviews. The grammatical extraction rules proposed operate on the dependency relations of the pre-processed and POS tagged sentences. Wu et al. [Wu et al., 2009] extend traditional dependency parsing to phrase level to extract relations between product features and expressions of opinions. Their work is related to sentiment analysis, and aspects are identified by leveraging the syntactical relations between nouns and adjectives, among others. Similarly, [Hai et al., 2011] guides the aspect extraction by means of association rules that consider only those nouns and compounds associated with some predefined sentiment words. On the other hand, [De Marneffe et al., 2006] present a system for extracting typed dependency parses of English sentences by searching for specific patterns applied on phrase structure trees. This system is used in the Stanford parser [Klein and Manning, 2003]. A dependency parse represents dependencies between individual words, while a phrase structure parse represents nesting of multi-word constituents. The selection of grammatical relations included in their work is based on the set of grammatical relations defined in [Carroll et al., 1999] and [King et al., 2003].

In our work, we combine frequency counts with a set of grammatical extraction rules, among other techniques such as frequency filtering, to extract aspects and create an aspect vocabulary in an unsupervised way [Chen et al., 2014; Ferrer et al., 2014]. As we will see in Chapter 3, we further improve the recall of the aspect vocabulary extracted by combining the aspect extraction task with a taxonomy created from two photography web resources.

2.2.2 Supervised Aspect Extraction

Supervised aspect extraction approaches are generally more accurate than unsupervised aspect extraction approaches. However, they need domain specific labeled training data in order to obtain good results. As such, supervised approaches for aspect extraction are less transferable across domains.

Supervised aspect extraction techniques often use conditional random fields (CRF), hidden Markov models (HMM), or support vector machines (SVM), to detect aspects at the phrase level [Sebastiani, 2002].

Conditional random fields (CRF) [Lafferty et al., 2001; Peng and McCallum, 2006] have been successfully applied to information extraction tasks previously. CRF are used in [Jakob and Gurevych, 2010] to label each word of a sentence with a corresponding part-of-speech tag. By considering the context of a word, orthographic features of the sentence, its part-of-speech tag, and dependency relations between the word and the phrase (among other features), they train a CRF model, that is then used to estimate the probability of the various part-of-speech tagging sequences for a new input sentence. [Zhuang et al., 2006] present a supervised algorithm for the extraction of aspects considered opinion targets. They train the system using a combination of dependency rules and part-of-speech paths relevant to opinion targets (aspects on which the reviewers express their opinions), learned from an annotated dataset.

A similar approach is used by [Kessler and Nicolov, 2009], which focus on identifying which opinion expression is linked to which opinion target using support vector machines (SVM). SVM are also used by [Kobayashi et al., 2006] to extract tuples formed by a subject, an aspect, and an evaluation, using an iterative semi-automatic approach which requires human input at every iteration. [Jin et al., 2009] presents a machine learning approach built using hidden Markov models (HMM) to identify opinion expressions. They propose a hybrid approach integrating POS information with the lexicalization technique under the HMM framework, similar to the approaches of [Lee et al., 2000] and [Fu and Luke, 2005].

Other approaches are based on classifying a sentence or paragraph as referring to one or more predefined aspects. [Ganu et al., 2009] trained various support vector machine classifiers over manually labeled data about restaurants. They identify six different categories (food, service, price, ambience, anecdotes, and miscellaneous), and classify each sentence in the reviews as belonging to one of those categories. [Blair-Goldensohn et al., 2008] used maximum entropy models [Berger et al., 1996; Malouf, 2002] to summarize opinions expressed in reviews, by combining a dynamic aspect extractor (where aspects are determined from the text of the review), and a static extractor (where aspects are pre-defined) trained on a set of labeled data.

In this monograph we aim to create a system that is able to work in different domains related with classification tasks. For this reason we do not use a supervised aspect extraction approach, but instead we focus in extracting aspects from text in an unsupervised way. As we present in Chapter 3, our methodology is further enhanced by using domain knowledge extracted from two websites of digital photography.

2.3 Sentiment Analysis

Sentiment analysis is a field of study related with natural language and computational linguistics that addresses the application of techniques to automatically identify and analyze affective states and subjective information in natural language texts. The goal is to determine the (positive, negative or neutral) *polarity* of a sentence, or about a specific topic, as expressed by the author in a text. The subjective information in texts is usually expressed as judgments or evaluations, with different levels of granularity: for instance, sentiment polarity can be determined within individual text passages, or at a higher granularity, like ascertaining the general (positive or negative) tone of a document. Furthermore, sentiment expressions can be associated with different types of semantic categories such as polarity, strength, or type of emotion.

Sentiment analysis is widely applied in reviews and social media, and it is used in a variety of applications, ranging from marketing to customer services. The first works in this direction date back to the 1980s [Wiebe and Rapaport, 1988; Wilks and Bien, 1983], however the prominence of the research problem started to raise a decade ago, with the works of [Pang et al., 2002; Turney, 2002]. Since then, we can observe a steady increase in the number publications year after year. For instance, Google Scholar[1] shows more than 100,000 paper entries related with "sentiment analysis" in 2016. This increase in scientific publications related to sentiment analysis is a direct consequence of the rise of the Web 2.0 and the social media, among other factors such as the improvement in natural language processing techniques that simplify the processing of user judgments. People share their opinions and experiences online more than never, and this increase of sentiment data opens a lot of opportunities for researchers and companies interested in user opinions and experiences. For instance, if a company knows what people disliked about their product, they can easily take countermeasures and improve that specific part of the product to increase sales.

Sentiment analysis is a wide field of study, with various challenging tasks being addressed by the scientific community. Determining if a text is subjective or objective [Jiang et al., 2011; Riloff et al., 2005], or identifying irony and sarcasm in natural language texts [Bosco et al., 2013; Carvalho et al., 2009; Maynard and Greenwood, 2014] are some of those challenges. But the most prominent one and relevant to this dissertation is the sentiment polarity classification task, which consists in determining the general polarity (positive or negative) of a text passage in natural language . This sentiment polarity task may be extended to an ordinal regression problem where the goal is to classify a text according to a rating scale, usually defined from -1 to 1, where -1 expresses the maximum negative polarity and 1 the maximum positive polarity.

Extracting sentiment from natural language passages is a challenge. Sentiment lexicons are often used to ascertain the polarity (positive or negative) and strength of sentiment expressed at word-level [Baccianella et al., 2010; Esuli and Sebastiani, 2006]. Sentiment lexicons are also used to assess the polarity of words for multilingual sentiment analysis [Denecke, 2008]. However sophisticated methods are needed to aggregate these scores at the sentence, paragraph and document level to account for negation and other forms of sentiment modifiers [Muhammad et al., 2013]. Furthermore, sentiment lexicons are usually created for being used in general texts, not taking into account domain-specific nuisances. This methodology has some problems, as some lexicons might assess the polarity of some domain specific texts incorrectly. For instance, adjective 'small' usually carries a negative polarity. However, in the domain of photography, a 'small camera' can carry a positive polarity depending on the context (e.g. if most people prefer a small camera to a bulky one). [Wilson et al., 2005] refer to the polarity of a term in a sentiment lexicon as its prior polarity, suggesting that it may vary with the context. [Ceci et al., 2015] propose a form of contextual polarity that is purely context-dependent by analyzing the text with regular expressions and linguistic patterns, thus adapting the polarity of the sentiment lexicon words based on context.

Since the distribution of a user's sentiment in user-generated content is typi-

[1] https://scholar.google.com/

cally mixed and expressed over the aspects of an entity, the polarity is estimated at the aspect level. For instance, next sentence "I love the *color* but I'm not too fond of the *size*" expresses a positive polarity with respect to aspect 'color', but a negative polarity with respect aspect 'size'. It is clear that the polarity of the whole sentence does not correspond with the specific polarity of the aspects 'color' and 'size'. Therefore, it is necessary to estimate the polarity of the various sentiment targets per separate. This method receives the name of aspect-oriented sentiment analysis (also named aspect-level sentiment analysis or feature-based opinion mining [Ding et al., 2008; Pang and Lee, 2008; Popescu and Etzioni, 2007]), and assesses the polarity of a given aspect in a text by analyzing the surrounding words and natural language structures of the sentences in which the aspect appears [Dong et al., 2014]. Thereafter sentences are grouped by these aspects and sentiment scores assigned to each aspect.

In this work we will not use the term *opinion*, since we consider that the term is not accurate to describe the content of user reviews. From our point of view, user reviews contain descriptions of past experiences, and they may have an implicit polarity or an explicit judgment of some aspects. However, differently from opinions, they are based on facts experiences by users. In that sense, we consider that using the term opinion may lead to confusion, ans we prefer calling them judgments.

Aspect-oriented sentiment analysis is usually combined with aspect extraction techniques to identify aspects in the reviews [Hu and Liu, 2004a; Turney, 2002; Zha et al., 2014]. Afterwards, these aspect sentiment pairs extracted from the reviews can be used to characterize products to create, for instance, a recommender system [Chen et al., 2015a; Dong et al., 2013a,b].

Relevant to this monograph is the work presented in [Dong et al., 2013b; Dong and Smyth, 2016], where the authors extract aspects from `TripAdvisor.com` user-contributed reviews to generate user and product profiles. Product profiles are created by extracting the aspects from the reviews of the product. User profiles are created based on the aspects extracted from that particular user reviews. Aspects are extracted from user reviews by means of two colocation patterns – an adjective followed by a noun (AN) or a noun followed by a noun (NN)– combined with sentiment analysis, eliminating nouns that are rarely associated with sentiment words. Therefore, given a certain hotel, their system returns a set hotels with similar characteristics. That is, if the product profile of a given hotel is characterized by a set of aspects such as 'swimming pool' and 'wi-fi', their system returns other hotels with product profiles characterized by 'swimming pool' and 'wi-fi'. User profiles are used in a similar manner. Given a user profile, also characterized by a set of aspects, their system returns hotels with similar characteristics as the ones of that specific user profile.

Notice that in our work, we do not use sentiment analysis to select the set of aspects, but instead we analyze the context and the dependency structure of the sentences, among other techniques described in Chapter 3, to create the aspect vocabulary. Furthermore, in this monograph, we introduce the notion of query where the user can specify his or her preferences in order to support the reuse of the practical knowledge extracted from user-generated content, instead of recommending similar entities previously liked by the user as presented in [Dong and Smyth, 2016].

Finally, in this work we use the SmartSA [Muhammad et al., 2013] sentiment analysis system to assess the polarity of user reviews and judgments. The SmartSA system obtains the sentiment score of sentiment-bearing words from SentiWordNet [Esuli and Sebastiani, 2006] and uses contextual information to adapt the sentiment score to modify prior polarities of the sentiment lexicon.

2.4 Semantic Unification

In Chapter 4 of this monograph, we focus in identifying the issues used by individuals when describing their experiences. As explained in the Introduction, we aim at creating a concept vocabulary that models the most relevant issues used in a corpus of user-generated reviews. Issues are conceived of as arising from the interaction of an individual person with an entity, and can be expressed by a wide variety of textual descriptions because different lexical items (that have the same semantic content or meaning) can refer to the same issue.

The process of identifying the issues found in user reviews can be viewed as a special case of the general problem of semantic unification: The process where lexically different concept representations are deemed to have the same meaning [Hagoort et al., 2009]. Therefore, we are interested in unifying the set of lexically different aspects that are deemed to have the same semantic content (i.e., meaning), in order to identify the issues addressed by the individuals in the reviews of a corpus.

In this Section we describe two different approximations to solve this problems: 1) probabilistic topic models, and 2) basic level concepts.

Topic models (also referred to as probabilistic topic models) are a type of statistical models used in natural language processing and Artificial Intelligence for discovering the latent semantic structures of an extensive text body, and they are frequently used for text-mining and information extraction. The idea behind topic models is that, given that a document is about a particular topic, one would expect particular words to appear in the document more or less frequently. For instance, 'room' and 'staff' will appear more frequently in documents about hotels, and 'wheel' or 'engine' will appear more frequently in documents about cars. Topic modeling is based in the idea that documents typically concern multiple topics in different proportions. Therefore, a document that is 25% about hotels and 75% about cars will probably contain 3 times more words related to cars than to hotels.

Topic models capture this intuition in a mathematical way, discovering, based on the frequency of the words in each document, what the topics might be and what each document's balance of topics is. In the context of user reviews, the topics ideally cover the mentioned product aspects. Differently from supervised aspect extraction methods, topic models do not rely on training corpora, allowing topic models to be transferable between domains.

Currently, the most prominent approach for topic modeling is the latent Dirichlet allocation (LDA) [Blei et al., 2003]. LDA is similar to probabilistic latent semantic analysis (PLSA) [Hofmann, 1999], and to latent semantic analysis (LSA) [Dumais, 2004; Landauer, 2006], but considers a Dirichlet distribution for the topics instead of a uniform distribution. Central to LDA (and to topic modeling) is the requirement to find a global maximum of a likelihood function characterized by numerous local maxima. LDA assumes that each topic is characterized by the probabilities of usage of a set of words, and that every document in the corpus is generated from a mixture of topics. As such, documents with different topics will have different word probability distributions, because the probability of using a given word depends on the topic.

The objective of the LDA approach is to estimate the topic structure of the documents, which is defined by the probabilities of a topic given a document and the probabilities of a word given a topic. The results obtained from LDA can be difficult to interpret, since the generated topics are unlabeled. Moreover, the aspects included in a topic do not forcibly need to be semantically related, hindering the interpretation of results.

LDA was initially created to operate at a document level, and adapting it to extract aspects is not straightforward, since the topics obtained from applying LDA at a document level are too general. On the other hand, trying to apply LDA solely at a lower level (such as a sentence level) in order to identify more precise aspects does not yield good results, because the bag of words used to define topics results too small [Jin et al., 2011].

To solve this problems, an extension to LDA was introduced: multi-grain LDA (MG-LDA). MG-LDA approaches model topics on two levels, global and local. Global topics are formed by a fixed set of topics usually computed at a document or domain level, while the dynamic local topics are extracted by modeling each document as a set of overlapped sliding windows that cover some adjacent sentences of the document [Lu et al., 2011; Titov and McDonald, 2008b]. Other approaches to topic modeling using LDA include the use of hidden Markov models in order to distinguish between aspect words and background words [Lakkaraju et al., 2011]. This distinction is achieved by analyzing syntactic dependencies between aspects and sentiments. Syntactic dependencies were also used combined with LDA in [Zhan and Li, 2011] to create two vocabularies, one composed by only nouns and the other with noun modifiers such as adjectives and verbs. Their intuition is that adjectives provide useful context for noun features, and that this context information can improve the topic modeling task.

Some research in topic modeling combines the task of aspect extraction with sentiment analysis. Titov et al. (2008a) and Moghaddam et al. (2012), respectively, improve their previous model presented in [Titov and McDonald, 2008b], by including aspect sentiment analysis to determine the polarity of the selected topics. The method assumes to be working upon a collection of user-contributed reviews, where each review has a rating value. They assume that the text and the rating of the reviews are related, in the sense that a low rated review will be accompanied with a text that expresses a negative polarity, and vice versa. Then, a polarity value is aggregated considering the ratings of the aspects grouped in each topic, and a polarity value is assessed for every topic. Although Titov et al. obtain good results, user ratings cannot always be considered a valid ground truth of the user's taste because users are inconsistent in giving their feedback [Amatriain et al., 2009; Ferrer and Plaza, 2016]. Other works related with topic extraction and LDA are [Jo and Oh, 2011; Moghaddam and Ester, 2011; Sauper and Barzilay, 2013; Wang et al., 2011].

Relevant to this monograph is the work in [Lancichinetti et al., 2015], that questions the validity of LDA optimization algorithms for inferring topic models [Wallach et al., 2009]. Lancichinetti et al. demonstrate that the current implementations of latent semantic analysis have low validity to define topic models for documents, because the standard techniques for likelihood optimization are unlikely to infer the best model for the data due to the high degeneracy of the likelihood landscape. Due to this reason, standard techniques for likelihood will likely infer different models for different optimization runs [Blei, 2012; Wallach et al., 2009]. As such, Lancichinetti et al. propose to modify algorithms developed for community detection in networks to define the parameter values in the latent generative model, obtaining high-accuracy for automated topic classification.

A second way of identifying lexically different concept representations that have the same meaning is using the notion of basic level concepts (BLC) from cognitive linguistics presented in [Rosch, 1973; Rosch et al., 2004]. Basic level concepts (BLC) are those that strike a tradeoff between two conflicting principles of conceptualization: inclusiveness and discrimination. Rosch et al. found that there is a level of inclusiveness that is optimal for human beings in terms of providing optimum cognitive economy. This level of inclusiveness is called the basic level, and concept or categories at this level are called basic-level concepts. Basic level concepts are described in detail in Chapter 4, where we model the set of issues used in the user-generated reviews by applying the notion of basic level concepts. Succinctly, we select a set of basic level concepts by evaluating the coherence of the polarity of the judgments whose issues are assembled in a given BLC. The intuition behind this criteria is that if two (or more) aspects are considered to reference the same issue (the same BLC), then the polarity of the judgments about those two aspects over the set of user-generated reviews of a product should be similar [Ferrer and Plaza, 2016].

Basic level concepts were applied to word sense disambiguation by [Izquierdo et al., 2007, 2015]. Izquierdo et al. select, from a set of concepts, a subset that are considered basic level concepts (BLC). They do so by using the structural properties of WordNet (WN), following a bottom-up approach, using first the chain of hypernym relations, and second all types of relations encoded in WN. For each synset in WN, Izquierdo et al. selects as its BLC the first local maximum according to a threshold that specifies the relative number of relations between that synset and the rest of elements in WN. After a certain number of BLC are obtained, the process stops. Afterwards, they set a second threshold parameter to generalize those BLC that do not represent enough concepts; the generalizations of BLC is performed by selecting the next local maximum following the hypernym hierarchy of WordNet. The final set of BLC are concepts that are representative for a set of other concepts, and are used to word sense disambiguation.

In this monograph we unify lexically different aspects by means of a clustering algorithm that considers the semantic, syntactic and lexical similarities between aspects with the objective of identifying the issues used in a corpus of user-generated reviews. We analyze how aspects are used in the reviews of products, and cluster those aspects considered semantically similar into concepts that model the set of issues described in the reviews.

In our work, differently from [Izquierdo et al., 2007], we use the practical knowledge derived from our analysis of recorded experiences to decide what are the most important issues for the persons referring to an entity. We do not select a set of BLC from an already existing ontology such as WordNet, but from those issues more often discussed in the texts that describe people's experiences. This distinction is important: when we create the BLC from our set of aspects, we are not using theoretical knowledge obtained from an external source such as, in the case of cameras, a list of features defined by the manufacturer that are supposedly significant in characterizing that camera. Instead, we focus on the interaction of individual people in the *use of that camera*, and then analyze their explanation of those experiences when dealing with those features, and the issues they choose to mention and judge, positively or negatively. As such, the set of basic level concepts used in this monograph depends on the usage, by the authors of the reviews, of the different words used in the textual reviews of experiences.

2.5 Arguments and Argumentation

There are three areas of research related to our use of judgments and arguments in this monograph: (1) argumentation theory, (2) deductive and inductive arguments, and (3) argumentation frameworks in artificial intelligence. As we shall see, the notion of inductive argument is the most relevant to our approach.

Argumentation theory is the interdisciplinary study of how conclusions can be reached through reasoning, and includes the study of civil debate, dialogue, conversation, and persuasion in humans and in artificial intelligence settings. Argumentation is a key human skill, utilized and studied in various domains and activities: in philosophy, in courtrooms, and even in science by scientists positing new theories.

Stephen E. Toulmin's contributions to Argumentation Theory are the most influential, specially on the study of human use of argumentation in real world settings [Loui, 2005]. Toulmin introduced the concept of argument fields in [Toulmin, 2003], stating that some aspects of arguments vary from field to field, and are hence called *field-dependent*, while other aspects of arguments are the same throughout all fields, and are hence called *field-invariant*. Toulmin attempts throughout his work to develop standards for assessing the worth of ideas based on the combination of field-dependent and field-invariant aspects of arguments.

Furthermore, in [Toulmin, 1972], he proposed an evolutionary model of conceptual change, involving innovation and selection. Toulmin stated that innovation accounts for the appearance of conceptual variations, while selection accounts for the survival and perpetuation of the soundest conceptions.

In philosophy and logic, an *argument* is a set of statements typically used to persuade someone of something or to present reasons for accepting a conclusion [Johnson, 2012]. Arguments are formed by one or more premises (in support of a claim or conclusion), a method of reasoning, and a conclusion. Arguments are structures from where conclusions are derived; therefore, arguments supporting a conclusion need to be consistent and valid, otherwise the conclusion drawn will also be inconsistent.

There are two basic types of arguments in logic: deductive arguments and inductive arguments [Goel et al., 1997]. In a deductive argument, the premises guarantee the truth of the conclusion. That is, a deductive argument asserts that the truth of the conclusion is a logical consequence of the premises. Deductive arguments are either valid or invalid. Therefore, if a set of deductive arguments is true, the conclusion drawn from those arguments will also be true. On the other hand, inductive arguments asserts that the truth of the conclusion is supported to some degree of probability by the premises, but does not entail it. Inductive arguments can be strong or weak. Strong arguments give more support to the conclusion than weak arguments. Therefore, a conclusion supported by strong arguments is more likely to be true than one supported by weak arguments.

As we will see in Chapter 4 of this monograph, we interpret the judgments contained in the experiences of individuals as evidence in order to create inductive arguments, that is, reasons that support (or not) selecting a product over others. Each entity review describes judgments with respect to the various aspects of that entity, based on an individual experience. By analyzing and aggregating those judgments we create inductive arguments that support the decision of selecting this particular entity over others. Furthermore, our arguments also have different strength, depending on the analysis of the data found in the user-generated reviews of a particular entity. Stronger arguments provide stronger support in favor or against selecting an entity over others.

An *Argumentation framework* is a system or methodology that draws conclusions from arguments. Argumentation frameworks can be represented with directed graphs, the nodes being the arguments and the edges between nodes the attack relations. Argumentation frameworks are mostly modeled after the Dung's fundamental model, Dung's abstract formalism for argumentation-based inference, one of the most relevant work that established the basis of the actual AI argumentation theory [Dung, 1995]. Dung studied human argumentation mechanisms in order to explore ways to implement this mechanism on computers. Dung showed that argumentation can be viewed as a special form of logic programming with negation as failure, being a conclusion believable if it can be successfully defended and supported, by means of the arguments, against attacking arguments. Citing Dung:

> " *[...] whether or not a rational agent believes in a conclusion de-*

> *pends on whether or not the argument supporting this statement can be successfully defended against the counterarguments.* ”

Bondarenko et al. [Bondarenko et al., 1997], in collaboration with Dung, extended the work of Dung et al. [Dung, 1995] and introduced the assumption-based framework (ABF), an abstract framework for default reasoning where an assumption can be defeated if the contrary assumption can be proved.

Dung's formalism for argumentation-based inference was an important work that draw near argumentation and AI, leading to a bloom of new subjects that combined logic and deductive reasoning with Artificial Intelligence techniques on argument analysis [Bench-Capon and Dunne, 2007]. Scheuer et al. defines two categories of argumentation analysis approaches in [Scheuer et al., 2010]: argument analysis and discussion analysis. Argument analysis approaches focus on the construction of sound and syntactically valid arguments, usually by means of checking domain-specific structures at a syntactic level [Pinkwart et al., 2006; Suthers, 2001; Thagard, 2006]. On the other hand, discussion analysis approaches are mainly concerned with the social and interaction aspects of discussions, in line with argumentation theories that emphasize the relation between dialog and argumentation [Walton, 2008]. Discussion analysis approaches automatically analyze the textual content to identify the intention and interaction patterns of discussions. Is this contribution arguing for or against and argument? Does this contribution have one or more counterarguments? Some related work was published in [Sierra et al., 1997], where they proposed a general framework for negotiation in which agents exchange proposals backed by arguments in order to persuade other autonomous agents in the context of multi-agent systems.

With the advance of the Web 2.0, there had been attempts to bring together various strands of this work and to produce a standard for representing and exchanging arguments, the Argument Interchange Format (AIF) [Chesñevar et al., 2006; Rahwan et al., 2007]. Arguments can be found in text reviews and in online forums [Janssen and Kies, 2005], where individuals express their arguments and discuss other people's arguments to prove their conclusions.

Another focus is identifying the sets of aspects with higher positive/negative polarity to give insights into the reason why items have been chosen [Muhammad et al., 2015]. Kim et al. [Kim and Hovy, 2006] automatically identify pro and con sentences in online reviews by training a maximum entropy model with a labeled set of pros and cons. Then, each sentence in online reviews is classified either as a pro or a con sentence based on the trained model. The sets of pros and cons of some websites are also used for the task of aspect extraction by Hu an Liu [Liu et al., 2005]. Hu et al. use a supervised form of association mining to extract aspects by targeting pros and cons that are separately specified on some Web sites, because pro and con sentences are known to be rich in aspect descriptions.

In Chapter 5 of this monograph, differently from these websites that explicitly present pro and con reasons introduced by authors, we automatically identify pro and con arguments by leveraging the practical knowledge found in user experiences. We consider the practical knowledge from people's experiences (in plural), as arguments in favor, or contrary, selecting an entity over others. This way, if the overall polarity of user experiences with respect to a feature (or issue) of a camera is positive, we consider this specific feature a positive argument (pro argument) with respect to selecting that product over others. On the other hand, if the overall polarity with respect to another feature is usually negative, we consider it a con argument.

Therefore, by analyzing all arguments found in the set of reviews of an entity, we characterize that entity with a set of pro and con arguments created from the practical experiences of people with the different features of that specific entity.

Chapter 3

Aspect Vocabulary Creation from User Reviews

3.1 Introduction

In this chapter we analyze user experiences in textual form to identify the important issues that define product experiences. Specifically, we are interested in identifying the issues found by users through their own experiences with respect to photographic cameras. The main goal of this section is to discover the vocabulary that people use, which need not be the same as the classical feature list describing the different aspects of a camera (e.g. 12 Megapixels).

In this process, we are not interested in performing so called "opinion mining". Opinions are a views or considerations formed about something not necessarily based on fact or knowledge. Instead, we prefer to talk about finding *user judgments* about product issues, where a judgment is understood as a "considered decision or sensible conclusion". In our approach, judgments expressed by users are based on *facts* encountered while they were actually using a product or a service, since this is the kind of textual content we are interested in analyzing in the web of experiences. In our view, opinion analysis on any kind of text is a more general and less focused task that the one we are interested in: analyzing and discovering the issues and judgments appearing in text expressing actual experiences of users. This chapter addresses the methodology to identify the collection of issues or concepts that the users deem relevant to express in a text describing his or her experiences. The polarity of these judgments (positive or negative sentiment) is not addressed in this chapter but later, when we aggregate large numbers of judgments from a corpus of users experiences to build argument bundles for each product under consideration.

Since we are dealing with textual content, an issue (or concept) can be expressed in a variety of ways in text, and this chapter is focused on finding them. The literature usually calls the issues *aspects*, and the process is called *aspect extraction.* However, in our approach, aspects in themselves are not proper concepts, instead we consider several aspect may refer to a basic level concept. For instance, in the sentence "the video dial, menu button and video shutter are not easily accessible", the user is making a negative judgment over the concept 'button' of the camera, as we will show later in Chapter 4. Thus we take a two phases approach: In this chapter we first analyze and create a vocabulary of aspects appearing in a corpus of user reviews, and then, in Chapter 4 we develop a method to find the matching between aspects and basic level concepts.

We intend to create an *aspect vocabulary*, specifically in the domain of digital photography. This aspect vocabulary will be created using two different sources: a corpus of user product reviews and, ancillary, two professional photography web resources.

User reviews. The main source from where the aspect vocabulary is created are user reviews of products. We are interested in the salient words used by individuals to describe their experiences, and as such we use product reviews as our main source to populate the aspect vocabulary. The unsupervised aspect extraction method is described later in this chapter.

Photography web resources. We use the technical camera descriptions present in two photography web resources to generate a set of possible aspects. These aspects differ from the one extracted from the product reviews in three ways: 1) they tend to be more technical and specialized than the average vocabulary used in user reviews, 2) they do not include misspellings, and 3) they are created by experts in photography.

By combining those two sources of information we are able to extract the salient aspects that people use when they describe their experiences with a product, and create the aspect vocabulary.

To test the approach, we have created a corpora of user reviews, extracted from three different camera types in Amazon.com: Digital Single Lens Reflex, Compact and Point & Shoot. We create an aspect vocabulary for each one of the three corpora, first presenting the methods to extract aspects and select the most relevant aspects, and later analyzing, comparing and evaluating the three aspect vocabularies. Finally, in view of the similarities and differences between the three aspect vocabularies corresponding to the three Amazon camera types, we decide whether we should either consider the three camera types as separate camera (sub)domains, each one with its own vocabulary, or as a unique domain with the a unique and global aspect vocabulary.

The rest of the chapter is organized as follows. We present four different but complementary methods of aspect extraction from a corpus; each one generates an *aspect set* and their union builds the aspect vocabulary for that corpus. Section 3.2 presents an unsupervised aspect extraction method that extracts a set of aspects from a corpus. Section 3.3 explores two photography web resources to extract important camera related aspects and generate a new domain-specific photography taxonomy. Sections 3.4 and 3.5 combine both previous aspect sets, together with the photography taxonomy, to identify more salient aspects from the reviews. The aspect vocabulary creation is described in Section 3.6. In the same section, three aspect vocabularies are created from user reviews of three camera types, experimentally setting threshold values to eliminate less frequent or important aspects. We also analyze, compare, and evaluate, the results depending on the threshold values, the contribution of the 4 aspect extraction methods, and we finally decide on whether is better to work with one general photography domain (one aspect vocabulary) or three domains (three aspect vocabularies, one for each camera type). Finally, Section 3.7 summarizes the contributions and draws the conclusions of this chapter.

3.2 Unsupervised Aspect Extraction

In this section we present an unsupervised method to extract a set of aspects from a corpus of user reviews. We are interested in finding the most salient lexical items used by users to describe their experiences with products.

For this purpose, we present an unsupervised aspect extraction method that combines part of speech *POS* tagging, frequency filtering and WordNet matching, among other techniques. Since user reviews do not need to cope with any format requirements, users can write their experiences the way they want. That means that spelling mistakes, abbreviations, punctuation errors and slang language are commonly found in the reviews. Moreover, a user can use the word 'pic' or 'picture' indistinctly, and our system should be able to identify both of them as aspects (even if they are not accepted dictionary words). As such, the aspect extraction method needs to be robust enough to overcome the misspellings and grammatical errors in order to extract an interesting set of aspects to create an aspect vocabulary.

The aspects extracted using the unsupervised aspect extraction method are the most prominent product features present in the corpus reviews, understanding prominent as commonly and widely used by the authors of the reviews. These form the aspect set $\mathcal{A}_1$, that will later form part of the aspect vocabulary $\mathcal{A}$, together with the 3 other sets of aspects that will be introduced later.

Figure 3.1 presents an overall view of the unsupervised aspect extraction method. To ensure the text consistency, the first step in the unsupervised aspect extraction method is the sentence pre-processing and part of speech tagging presented in Section 3.2.1. In this step, the sentence consistency is ensured by removing special characters, stop words and lemmatizing the sentence before applying the set of grammatical extraction rules that identify aspect candidates. Then we perform an evaluation of the aspects extracted by the grammatical extraction rules with other state of the art aspect extraction techniques in Section 3.2.2. Finally, in Section 3.2.3, aspect candidates are filtered out based on frequency, word lexicons, and WordNet, to ensure that only the most salient aspects form part of the aspect vocabulary.

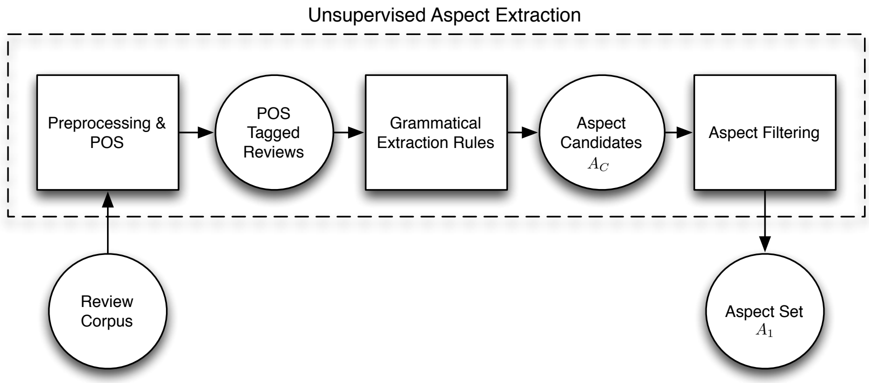


Figure 3.1: Unsupervised aspect extraction process.

3.2.1 Pre-processing and POS tagging sentences

The first step of the unsupervised aspect extraction method is to ensure the text consistency of the sentences by removing entity names (brand, company or product names), stop words and special characters, and lemmatizing the sentence before applying the grammatical extraction rules to extract the candidate aspects.

This step is important since there is no specific format for users to follow when reviewing a product, and as such those reviews contain abuses of exclamation and question marks (such as '??!!!'), special characters (such as '$') or even *smileys* (such as ':-)' or '<3'). Those special characters and stop words can suppose a problem for the part of speech tagging algorithm, and therefore are removed.

The text pre-processing algorithm also removes sentences that do not refer to the camera that is the object of the review. We do so because we are only interested in finding the salient aspects that refer to the products of our corpus, and considering aspects that were meant to describe the author's experiences with other cameras would introduce noise in our data. This behavior is quite common in our corpus of user reviews: authors tend to start their reviews by introducing their experiences with their previous cameras. A name *entity camera lexicon (NE)* was created for this purpose by exploring *Wikipedia*[1] and *Imaging Resource*[2]. We extracted all camera brands and camera models of each brand, and automatically assigned every camera model to a manufacturer. This way, every time we found the word 'D200' (or any other camera model) in the reviews, we knew the author was talking about a Nikon camera.

Figure 3.2 presents an overall view of the text pre-processing process that consists of four steps, applied as a pipeline. The inputs of the process correspond

[1]https://en.wikipedia.org/wiki/List_of_digital_camera_brands
[2]www.imaging-resource.com

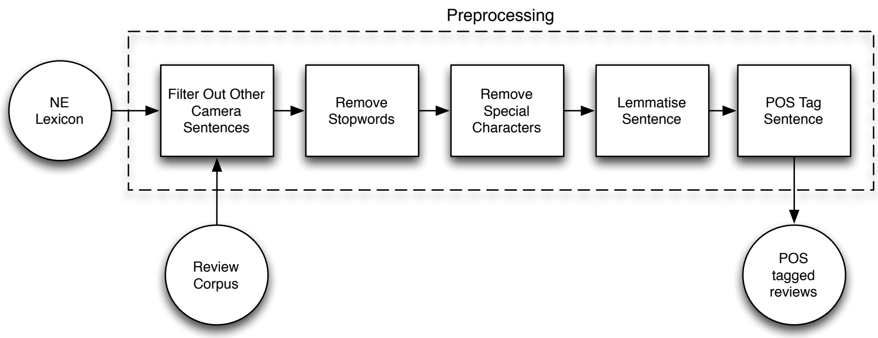


Figure 3.2: Pre-processing process.

to the reviews of the corpus and the camera name entity dictionary. After the pre-processing, we obtain a set of *part of speech* tagged reviews. That is, a set of reviews where each word is identified as a noun, verb, adjective, adverb, pronoun, preposition, conjunction, interjection or determiner. Those POS tagged reviews are then used by the grammatical extraction rules in Section 3.2.2 to select the candidate aspects.

A short description of the pipeline steps is the following:

1. **Filter out sentences that refer to other cameras.** The first step is to filter out the sentences from the reviews of a camera that refer to other cameras. Camera name entities are identified by using the name entity NE lexicon presented before.

2. **Remove stop words.** Such as 'the', 'a', 'or'.

3. **Remove special characters.** Such as symbols '$' or '?' not filtered in previous step. Furthermore, we replace the character '-' with a blank space.

4. **Lemmatization**. Using the Standford parser[3], we find the lemmas of all words in a sentence. This step is performed at a sentence level, sentence by sentence.

5. **POS tagging**. Using the Standford part of speech (POS) tagger[4], we tag every word in each sentence as noun, verb, adjective, adverb, pronoun, preposition, conjunction, interjection or determiner.

At the end of the pipeline, we obtain a group of pre-processed and POS tagged sentences for each camera review in our corpus. Those sentences are the input for the grammatical extraction rules introduced in Section 3.2.2 .

[3]http://nlp.stanford.edu/software/lex-parser.shtml
[4]http://nlp.stanford.edu/software/tagger.shtml

3.2.2 Grammatical Extraction Rules

After pre-processing and POS tagging the reviews, we use the set of grammatical extraction rules presented in [Moghaddam and Ester, 2010] to identify the set of candidate aspects. These rules operate on the dependency relations of the pre-processed and POS tagged sentences. A dependency relation is the notion that words are connected to each other by directed links. The verb is the structural center of the clause structure, and all other words are either directly or indirectly connected to the verb in terms of directed links, named dependencies. In this work we are specially interested in the dependency relations between nouns and adjectives, as we will see later.

$$
\begin{array}{l}
DP\ \{ \\
\quad dp_1 : amod(N, A) \rightarrow \langle N, A \rangle, \\
\quad dp_2 : acomp(V, A) + nsubj(V, N) \rightarrow \langle N, A \rangle, \\
\quad dp_3 : cop(A, V) + nsubj(A, N) \rightarrow \langle N, A \rangle, \\
\quad dp_4 : dobj(V, N) + nsubj(V, N') \rightarrow \langle N, N' \rangle, \\
\quad dp_5 : \langle h, m \rangle + nn(h, N) \rightarrow \langle N + h, m \rangle, \\
\quad dp_6 : \langle h, m \rangle + nn(N, h) \rightarrow \langle h + N, m \rangle \\
\}
\end{array}
$$

Figure 3.3: Grammatical extraction rules. DP is the set of dependency pattern rules used in this work.

Figure 3.3 lists the rules that we have employed in this work, where N is a noun, A an adjective, V a verb, h a noun head term, m an adjective modifier, and $\langle h, m \rangle$ is a candidate phrase. The first dependency rule, dp_1, searches for all adjectives modifying a noun. If any of the pre-processed sentences matches this criteria, the resulting pair $\langle N, A \rangle$ is selected as a candidate phrase and stored for later use. The second rule, dp_2, searches for adjectival complements, and at the same time, looks for the subject of the verb. Similarly, if any of the pre-processed sentences matches the criteria, the adjective and the subject form a new candidate phrase. dp_3 searches for copulas, which are the relation between the complement of a copular verb and the copular verb. At the same time looks for the subject of the sentence, with relation to the verb, to form a candidate phrase. dp_4 searches for nouns performing as direct objects of a verb and for the subject of the same verb. If found, a candidate phrase is formed with the noun of the direct object and the subject of the verb. dp_5 and dp_6 are used to identify noun compound modifiers between the candidate phrases selected with rules dp_1 to dp_4. If any of the pre-processed sentences matches the criteria, the resulting pair, formed by a compound noun and a modifier, is also added as a candidate phrase.

The aspect candidate selection procedure is presented in Algorithm 8. S is the set of pre-processed sentences, and g corresponds to the grammatical relations of the pre-processed sentence s. For every sentence in S, the algorithm

Algorithm 1: Aspect Selection by Dependency Pattern Rules

Input: pre-processed sentences S;
Output: candidate aspects;
$candidatePhrase := \emptyset$;
for $s_j \in S$ **do**
  `// For each sentence in the pre-processed sentences`
  $g := grammaticalRelations(s_j)$;
  for $dp_i \in DP$ *and* $1 \leq i \leq 4$ **do**
    `// If grammatical relations of` s_j `match` dp_1, dp_2, dp_3 `or` dp_4
    if $g.matches(dp_i)$ **then**
      $candidatePhrase.add(dp_i.candidatePhrase)$;
`// Apply` dp_5 `and` dp_6 `to those candidate phrases containing` nn
$candidatePhrase.applyIf(nn, dp_5, dp_6)$;
`// Select nouns (N) and compounds (NN)`
return $candidatePhrase.select(N, NN)$;

first starts obtaining the grammatical relations of the sentence and then searches for matches of the dependency pattern rules from 1 to 4. If any of the rules applies, we obtain a candidate phrase. Next, if the pre-processed sentence where a candidate phrase was found contains a noun compound modifier, then rules 5 and 6 are applied, obtaining a new candidate phrase. Finally, for each candidate phrase, non noun (N) words are eliminated. Thus, the output is a set of candidate aspects $\mathcal{A}_c$ formed by single nouns or compound nouns.

Table 3.1 shows four examples of how these rules apply to several sentences. Consider the first example sentence "The camera lens is good", which according to Algorithm 8 applies to rule two:

$$dp_2 : acomp(is, good) + nsubj(is, lens) \rightarrow \langle lens, good \rangle.$$

Next, if a noun compound modifier (nn) exists in the sentence, rules 5 and 6 apply. In this example rule 5 applies resulting in the following candidate aspects:

$$dp_5 : (lens, good) + nn(lens, camera) \rightarrow \langle camera\ lens, good \rangle.$$

Finally, after removing all non noun words from the candidate phrase ⟨camera lens, good⟩, we are left with the candidate aspect *camera lens*. Figure 3.4 show the dependency parser tree of the previous sentence.

Comparative Study of 4 Aspect Extraction Methods

We turn now to compare the precision and recall of the grammatical extraction rules with three other widely used state of the art aspect extraction methods. Note that here we are not evaluating the precision and recall of the aspect vocabulary, but only the set of extracted aspects using the unsupervised aspect

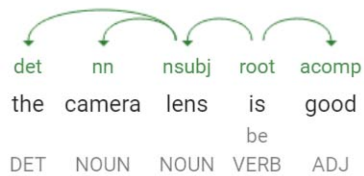


Figure 3.4: Dependency parse tree for the sentence.

Example	Dependencies
The camera lens is good	nsubj(good, lens) acomp(good, is) nn(lens, camera)
The screen has nice colors	amod(colors, nice) nsubj(has, screen) dobj(has, colors)
She looks amazing	acomp(looks, amazing)
I like the image	dobj(like, image) nsubj(like, I)

Table 3.1: Example of dependencies using DP rules.

extraction with grammatical extraction rules. However, it is useful to know the performance of the grammatical extraction rules before continuing with the creation of the aspect vocabulary in next sections.

To evaluate the set of aspects extracted using the grammatical extraction rules, we use a public dataset of product reviews containing manually marked-up product aspects [Ding et al., 2008; Hu and Liu, 2004a], considering only phone category products with at least hundred reviews. Precision and recall is used to compare manually labeled aspects with aspects extracted using the following four extraction algorithms:

- FqItems uses shallow NLP to identify single nouns as candidate aspects that are then pruned using a frequency cut-off threshold [Hu and Liu, 2004a].
- FqPos uses Part-of-Speech(POS) extraction patterns that are then pruned using sentiment informed frequency cut-off threshold [Dong et al., 2013b].
- DPrules uses the grammatical extraction rules in Figure 3.3.
- FqDPrules is the same as DPrules but prunes candidate aspects using a frequency cut-off equal to the 1% of the the total frequency of the most frequent aspect among the reviews [Moghaddam and Ester, 2012].

Precision of the two frequency based extraction approaches are significantly better than DPrules (see Table 3.2). We also confirm that FqPos improves

Approach	Precision	Recall
FqItems	0.60	0.34
FqPos	0.71	0.11
DPrules	0.28	0.66
FqDPrules	0.76	0.25

Table 3.2: Results for aspect extraction pilot study.

over FqItems. As expected, the best precision is that of FqDPrules (that combines deep NLP semantics is combined with frequency pruning). Here we observe a 26% and 7% improvement in precision over FqItems and FqPos respectively. The recall values of FqDPrules are low, so this fact suggests that FqDPrules must have many false negatives and so FqDPrules misses more extraction opportunities than FqItems and DPrules. However, a lower precision is more damaging as it is likely to introduce aspect sparsity problems.

Note that in this evaluation we used a simple frequency filtering to increase the precision of the candidate aspects extracted by means of the FqDPrules rules, obtaining the best precision among the state of the art aspect extraction techniques evaluated. In Section 3.2.3, we present a refined candidate aspect filtering method with the objective of further increasing the precision of the grammatical extraction rules by filtering out more false positives using lexicons, WordNet, and the already used frequency filtering.

3.2.3 Aspect Filtering with Lexicons and WordNet

In this section we increase the precision of the grammatical extraction rules by combining frequency filtering with lexicons and WordNet. The three new filtering steps are motivated by the false positives of the aspect candidate list obtained in the grammatical extraction rules evaluation from previous section. Even though our method scored the best precision over the state of the art aspect extraction algorithms presented, we observed some problems related to the part of speech tagging and name entity recognition of the candidate aspects. Some verbs, adjectives and adverbs were tagged as nouns by the Stanford part of speech tagger and then selected as candidate aspects by the grammatical extraction rules. Those elements were then evaluated as false positives, since they were not phone related aspects. Similar problems were identified when applying the unsupervised aspect extraction method to our camera corpus, indicating that those problems were not domain dependent.

Another problem we found when using the grammatical extraction rules with the reviews in our camera corpus was that not all candidate aspects extracted were related to the photography domain. We found many candidate aspects, such as 'month' or 'time', that are commonly used in the reviews but are not interesting for our purposes, since they are not camera-related terms. While many of them were filtered after applying the frequency cut-off filter, the most

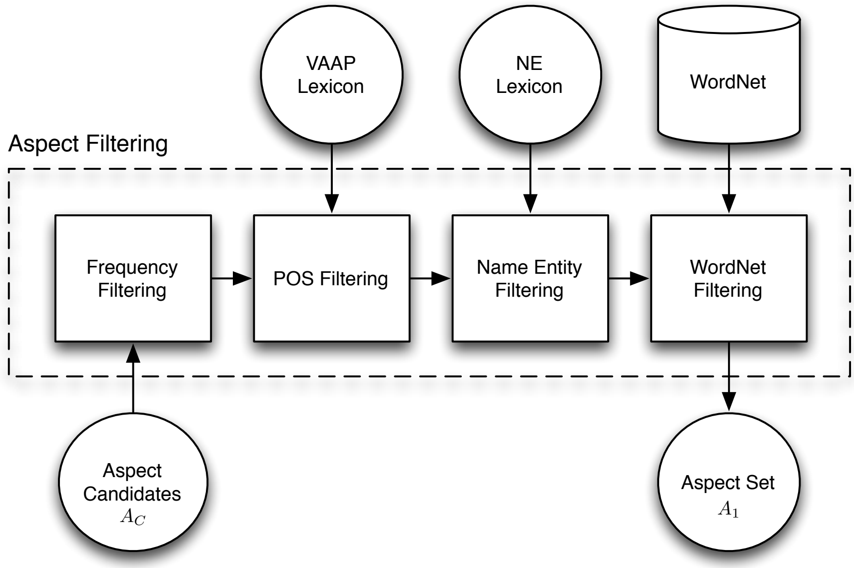


Figure 3.5: Aspect filtering process.

frequent were still selected as candidate aspects.

To address these problems, we now present 3 additional criteria to filter out candidate aspects. Since we increase the filtering process by adding more steps, the process of filtering out candidate aspects becomes more strict. This allows us to have more room to relax the frequency cut-off in the first step. Furthermore, we propose a method that combines the frequency filtering with part of speech filtering and name entity filtering. This way we can remove candidate aspects mistakenly tagged as nouns from our lists of aspects, resulting in an improvement of the precision of our method. To deal with the candidate aspects not related to the photography domain, we use WordNet[5] lexical database. We explore the interlinks that relate all WordNet synsets to determine whether or not a candidate aspect is related to the photography domain in WordNet.

Finally, remember that the aspect vocabulary will be created uniting 4 different aspect sets extracted using different techniques, so even if an important aspect is filtered out in one of the steps described in the filtering process, it can be selected in the next aspect sets presented later. Figure 3.5 presents an overall view of the aspect filtering pipeline process:

1. **Frequency filtering**. Filter out candidate aspects with an occurrence frequency, considering the set of reviews of a corpus, lower than a frequency threshold. If the frequency threshold is very strict, the frequency filtering

[5]https://wordnet.princeton.edu/

could ignore important candidate aspects that are not used by the majority of the reviewers, resulting in a low recall value of the final candidate aspect set. Furthermore, we differentiate two frequency filtering thresholds, one for uni-gram aspects δ_s, and another one for bi-gram aspects δ_c.

2. **POS filtering**. To identify mistakenly tagged nouns between the candidate aspects, we created a lexicon of verbs, adjectives, adverbs and prepositions (the *VAAP lexicon*), using the lists of the most common verbs, adjectives, adverbs and prepositions of *TalkEnglish* website[6]. Using the VAAP lexicon, we filter out those candidate aspects that were mistakenly tagged as nouns by the Standford POS tagger.

3. **Name entity filtering**. Similarly, we remove camera name entities among the candidate aspects tagged as nouns, using the camera NE lexicon introduced in Section 3.2.1.

4. **WordNet filtering**. We use WordNet lexical database to identify interesting aspects related to the photography domain by exploiting the interlinks that relate all WordNet synsets. If the candidate aspect is not related to any of the camera-related terms in WordNet, or it does not exist in WordNet, we remove it. This filter is explained in more detail later in this section.

The first, second and third aspect filtering steps are easily explained: we look for words that we know should not be in the candidate aspects lists, and remove them. However, the WordNet aspect filtering needs further explanation.

WordNet (also referred as *WN* in this work) is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. In this filtering step, we explore the WordNet 3.1 interlinks in order to identify candidate aspects *related* to the photography domain. If a candidate aspect is related to the photography domain, then we keep it in the candidate aspect set. Otherwise, if the candidate aspect is not related to the photography domain, the candidate aspect is removed from the set of candidate aspects.

The relatedness (or similarity) between the synsets of two candidate aspects is estimated by means of WordNet's *path similarity* measure [Meng et al., 2013]. Path similarity measures the semantic relatedness of word senses by counting the number of nodes along the shortest path between the senses in the 'is-a' hierarchy of WordNet. If two WordNet synsets are close in the WordNet hierarchy, it means that they are semantically related.

However, there exist some issues we need to consider before using WordNet. First, WordNet is a general lexical database, and as such it does not contain some domain specific words. This means that some words that are important in the photography domain, such as '*cmos sensor*' or *'macro lens'*, are not present

[6]http://www.talkenglish.com/vocabulary/

in WordNet. As such, since we cannot determine the relatedness of those words with the photography domain domain because they do not exist in WordNet, those words are removed from the aspect candidate set. However, to overcome this shortcoming, these domain-specific aspects will be included in the aspect vocabulary later in this chapter, using the method presented in Section 3.5.

Moreover, there is no direct relation between a word and a WordNet synset. Every word can be related to several synsets, one for every different meaning of the word in WordNet. However, only one of the synsets of a word may be relevant to our photography domain. For instance, consider the word *shutter*. WordNet defines two synsets for shutter: 1) "a mechanical device on a camera that opens and closes to control the time of a photographic exposure" and 2) a "hinged blind for a window". Clearly, the first one is the relevant synset to photography domain.

Identifying the relevant synsets of each aspect candidate is an issue that we need to address before computing WN word similarities, and has been long studied under the topic *word sense disambiguation* [Meng et al., 2013]. In this work we perform a brute force word sense disambiguation by selecting, for each of the top 10 most frequent candidate aspects, the closest synset (using path similarity) to all other synsets of those top 10 candidate aspects. Once the synsets of the top 10 most frequent candidate aspects are determined, we estimate the path similarity between each synset of the rest of the aspect candidates and the top 10 synsets. If the similarity of any of the synsets of an aspect candidate is above a threshold, we consider that it is a related candidate aspect to the photography domain. Otherwise, the candidate aspect is removed. Algorithm 2 shows the procedure for filtering the candidate aspects. $Syns$ are the selected synsets for the top 10 most frequent candidate aspects, and $PathSimAvg$ returns the average path similarity between synset syn and the set of synsets of the top 10 most frequent candidate aspects $Syns$.

At the end of the candidate aspect filtering pipeline, we obtain a set of salient aspects employed by the users when describing their experiences with cameras. The aspect extraction method (text pre-processing, aspect extraction and aspect filtering), is performed in an unsupervised way and can be applied to other domains, since we do not use any domain specific lexicon. This set of unsupervised extracted aspects obtained from the reviews is named $\mathcal{A}_1$, and is a part of the overall aspect vocabulary $\mathcal{A}$ that will be created in this chapter.

3.3 Aspect Set $\mathcal{A}_2$ and PhotoDict

As we noticed in previous Section 3.2.3, WordNet is not rich in content related to photography. Many frequent aspect candidates extracted by the unsupervised aspect extraction method do not exist in WordNet, and as such we are not able to assess their relations with the photography domain. As a result, some interesting camera related aspects were discarded in the process of creating aspect set $\mathcal{A}_1$, because we were not able to determine if they were interesting aspects or noise.

In this section we propose the creation of a new aspect set $\mathcal{A}_2$, created by

Algorithm 2: Candidate Aspect Filtering with WordNet

```
Input: Candidate aspects A_c, Selected synsets set Syns;
Output: A_1;
A_1 := Ø;
for a_c ∈ A_c do
    // Check if a_c exists in WordNet
    if ExistsInWordNet(a_c) then
        // Retrieve all synsets of a_c
        Syn_c := WNSynsets(a_c);
        for syn ∈ Syn_c do
            // Average path similarity between synsets syn and Syns
            pathSim := PathSimAvg(syn, Syns);
            if pathSim ≥ α then
                A_1.add(a_c);
                break;
// Return the aspect set A_1
return A_1;
```

using Dpreview[7] and Snapsort[8] web resources. These two web resources contain camera related words used to describe photographic cameras organized in "categories". Since the aspect set $\mathcal{A}_2$ is created from the photographic content of Dpreview and Snapsort web services, there is no need to apply the WordNet filtering presented in previous section to determine if the selected aspects are related to the photographic domain. Instead, we select interesting aspects by analyzing in the reviews of our corpus the usage of words also occurring in Dpreview and Snapsort. Furthermore, we create a domain taxonomy related to photography, *PhotoDict*, to overcome the domain limitations of WordNet in Section 3.3.3. We do so by exploiting the categories present in Dpreview and Snapsort web resources. PhotoDict taxonomy relates a set of different photographic terms using 'is-a' relations. This way, we know that 'pixel' is related with 'sensor', and 'frame' is usually related with 'sensor size'.

This section explains the process of creating the second aspect set $\mathcal{A}_2$: in Section 3.3.1 we generate a set of candidate words based on the photography keywords found in Dpreview and Snapsort. The candidate words are selected and evaluated against the corpus of reviews in Section 3.3.2. The output is a selected subset of words from Dpreview and Snapsort that constitute the aspect set $\mathcal{A}_2$. Next, using this new aspect set and the category information from Dpreview and Snapsort, we create a photography taxonomy we call PhotoDict. PhotoDict is a taxonomy of camera related uni-grams and bi-grams formed by the aspects in the aspect set $\mathcal{A}_2$. An overall view of the process of the creation

[7] www.dpreview.com
[8] www.snapsort.com

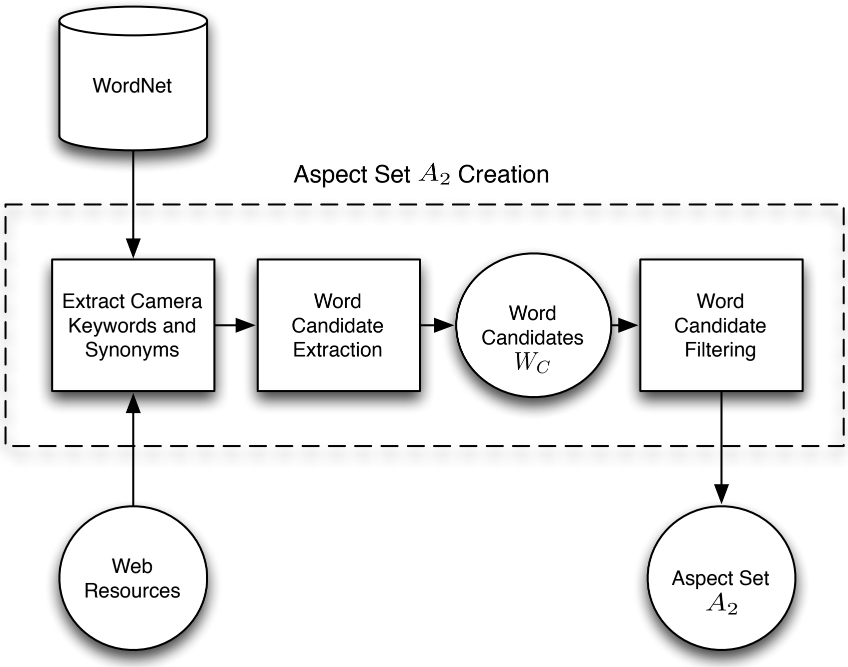


Figure 3.6: $\mathcal{A}_2$ set creation process.

of the aspect set $\mathcal{A}_2$ is shown in Figure 3.6, while Figure 3.7 shows the creation of the PhotoDict taxonomy, detailed later in Section 3.3.3.

3.3.1 Generating Word Candidates W_c

The first step to create $\mathcal{A}_2$ is to generate a list of candidate words using the photography keywords found in the web resources of Dpreview and Snapsort. Figures 3.8 and 3.9 show a small set of the keywords from both web resources. For instance, in Figure 3.8, we observe that 'raw', 'gps' and 'hdr' are grouped in a 'features' category. A similar situation can be observed in Figure 3.9, where 'iso' and 'white balance' are grouped in the 'image' category.

Following the category structure present in both web resources, we define three category levels H, H', and H'' for the Dpreview website, and two category levels H, H' for Snapsort. During the process described next, we will generate two separate lists of candidate words W_c, one for each website, that will then be joined and filtered together to form $\mathcal{A}_2$, and afterward, PhotoDict. Since the process to populate the category levels for the two websites is similar, we will only describe it once.

H is populated with the top category elements $h \in H$, such as 'features'

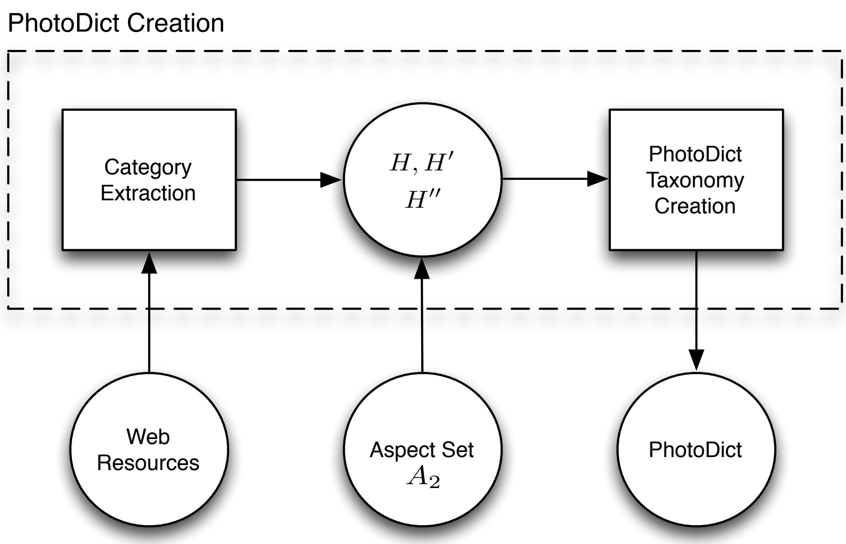


Figure 3.7: PhotoDict creation process.

Low Light	Image Quality	Features	Shutter
Image stabilization	High True Resolution	RAW	Rapid fire
Large sensor	High Megapixels	GPS	No lag
High ISO	Best Overall Quality	HDR	Fast

Figure 3.8: Photography terms from Snapsort.

or 'image'. H' with subcategory elements $h' \in H'$ such as 'hdr' or 'gps', and H'' with third level elements $h'' \in H''$, such as 'cmos' or 'sRGB'. We say that a $h' \in H'$ is a subcategory of $h \in H$ if there exist an *is-a* relation between h and h', represented in the websources by the different nested categories. In Dpreview and Snapsort websites, this is represented by a change in font color and size. Similarly, a $h'' \in H''$ is a subcategory of $h' \in H'$, if there exist an *is-a* relation between h' and h'' elements. Notice however that Snapsort does not have a H'' level. In Figure 3.9, 'cmos' (h'') is a subcategory of 'sensor type' (h'), and 'sensor type' is a subcategory of 'sensor' (h).

To generate all word candidates W_c, we first create the sets H, H' and H'' from Dpreview and Snapsort websites. We further enrich H and H' by including the synonyms of their category and subcategory keywords by using WordNet. To do so, we identify the most relevant WordNet synset for each $h \in H$ with the algorithm introduced in Section 3.2.3. Once the most relevant synset for each h

Sensor	
Max resolution	6000 x 4000
Image ratio w:h	1:1, 3:2, 16:9
Effective pixels	24 megapixels
Sensor size	APS-C (23.6 x 15.6 mm)
Sensor type	CMOS
Processor	X-Processor Pro2
Color space	sRGB, AdobeRGB
Color filter array	X-Trans III CMOS with primary color filter
Image	
ISO	Auto, 200-12800 (expands to 100-51200)
Boosted ISO (minimum)	100
Boosted ISO (maximum)	51200
White balance presets	7
Custom white balance	Yes

Figure 3.9: Photography terms from Dpreview.

is identified, we obtain the synonyms by searching other WordNet synsets with the highest WordNet path similarity. This process is also repeated for all words $h' \in H'$. The synonyms found in WordNet are included in the same set as their respective synonym, H or H'.

After this first step, we lemmatize all words in H, H' and H''. Furthermore, if a word in H or H' contains the character '-' we replace it for an empty space. Similarly, if a word is a compound (e.g. 'white balance'), we create a copy splitting it in two new words, each new word a single noun (e.g. 'white' and 'balance', separately). We also add those new nouns to their respective categories H or H'. Notice that we do not perform any of the previous steps for the words belonging to the lowest level H'' to avoid adding unnecessary noise, since H'' words often contain numbers, sizes or proportions.

After finishing the pre-processing, we generate a list of single and compound word candidates by combining all keywords within a category H with their subcategories H' and H''. The procedure is described in Algorithm 3. The idea behind the algorithm is to generate all possible combinations of keywords considering the category-subcategory relations present in Dpreview and Snapsort websites. For instance, consider the words 'sensor', 'sensor type', 'processor', and 'cmos' in Figure 3.9. The word 'sensor' is a category, 'sensor type' and 'processor' are subcategories of 'sensor', and 'cmos' is a subcategory of 'sensor type'. With this words, Algorithm 3 generates candidate words such as 'sensor', 'cmos sensor', 'sensor type' and 'sensor processor', among many others.

Algorithm 3: Domain Candidate Word Generation

Input: Web resource categories H, H', H'';
Output: Word candidates W_c;
$W_c := \emptyset$;
for $h \in H$ **do**
    `// Include categories and synonyms as candidate words`
    $W_c.add(h)$;
    $W_c.add(Synonyms(h))$;
    for $h' \in H'$ **do**
        `// Include subcategories and synonyms as candidate words`
        $W_c.add(h')$;
        $W_c.add(Synonyms(h'))$;
        `// Combine category names and synonyms to generate new dictionary candidate words`
        W_c.add(Combine(h, h', h''));
return W_c;

In the algorithm, W_c is the list of candidate words generated, and function *Combine* combines the words considering:

- Each subcategory word in H' is split and combined only with its category word H. Using the previous example with $h = \{\text{image}\}$ and $h' = \{\text{iso, white

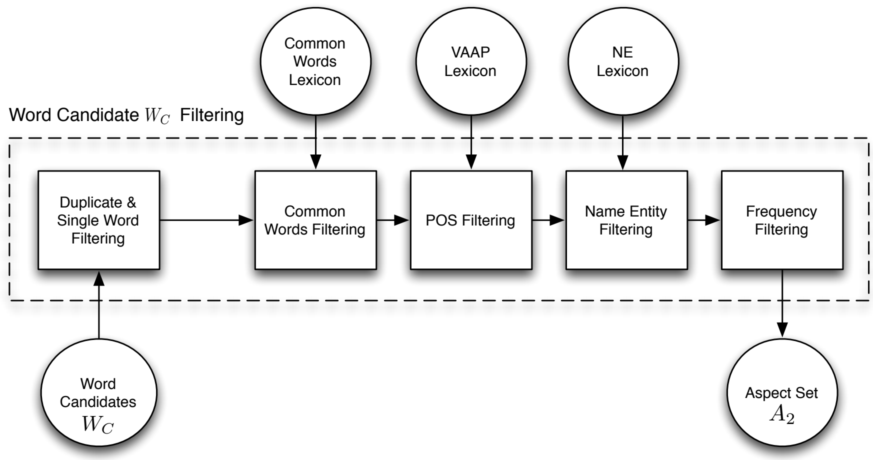


Figure 3.10: Word candidate filtering for aspect set $\mathcal{A}_2$.

balance presets} and supposing those words have no WordNet synonyms, *Combine* will generate bigrams such as: {image iso, image white, image balance, image presets, white image, ...}. *Combine* also reverses the word order of the compound candidate words generated, as in 'image white' and 'white image'.

- A similar approach is taken when combining H' with H''. However, here we ignore all non noun words in H'' before combining them with H' words. This is done to reduce the amount of noise created, since a lot of words in H'' are not interesting to our purposes of creating a camera vocabulary, such as sizes '36 x 24', proportions '3:2', or boolean values. Furthermore, notice that Snapsort does not contain any H'' category.

After applying Algorithm 3 with the photography keywords found in the web resources of Dpreview and Snapsort, we obtain two lists of word candidates that are joined together to form W_c. The resulting list of word candidates W_c is large and contains noise and duplicates. Therefore, this word candidate set is filtered and evaluated against product reviews in Section 3.3.2.

3.3.2 Filtering Word Candidates to Create $\mathcal{A}_2$

In this Section we select those word candidates that will be part of the aspect set $\mathcal{A}_2$. This is no trivial task, since the candidate words generated in previous section contain numerous false positives as the result of the brute-force candidate word generation approach. The filtering process presented in Figure 3.10 consists of 5 steps, each one filtering out word candidates on different criteria.
The first step consists in removing duplicates from the set of candidate words W_c. Then, we remove single or compound candidate words formed by only 1

character. This step removes candidate words such as 'pixels p' or 'lens $' or 'p', generated in previous step by combining category names with subcategories. Next we remove all those candidate words that are among the top 5000 most common English words, in order to remove noise created from the brute force candidate word generation. Most common English words are identified by means of a lexicon created from *WordFrequency*[9] website. Notice that this filter may remove interesting camera related aspects. However, since the possibly removed aspects are among the most frequent English words, those aspects are likely to be present in the first aspect set $\mathcal{A}_1$. Similarly as in previous aspect candidate filtering presented in Section 3.2.3, we remove those word candidates that are uniquely formed by adjectives, prepositions, adverbs, conjunctions or verbs using the VAAP lexicon presented in previous Section 3.2.3. We also filter out candidate aspects formed by company names and camera models with the NE lexicon. As a final step, we evaluate the set of remaining candidate words against our corpus reviews. We are only interested in candidate words used by people in their experiences, because later in this work we need to determine the polarity of the sentiment associated with each aspect and camera model. As such, we count the frequency of occurrence of the candidate words in the reviews in our corpus and filter out those below the uni-gram frequency threshold δ_s, or bi-gram frequency threshold δ_c, as we did in the previous section for $\mathcal{A}_1$. This procedure removes candidate words not often used, together with non occuring words generated by the brute-force approach.

After the word candidate filtering process, we obtain the set $\mathcal{A}_2$, generated using the camera keywords and exploiting the category structures extracted from the two specialized photography websites. This new aspect set differs from aspect set $\mathcal{A}_1$, because it contains specialized camera aspects that were selected without using WordNet as a filtering approach to remove candidate words not related with the photography domain.

3.3.3 Creating the PhotoDict taxonomy

We turn now to the creation of PhotoDict, a photography-related domain taxonomy intended to overcome the inaccuracy and domain limitations of WordNet we have seen in the previous Section 3.2. PhotoDict will be used in next sections to estimate the relations between photography related aspects, and in next chapters of this work to create the concepts and the bundles of arguments.

The PhotoDict taxonomy organizes the large set of aspects related to the photography domain $\mathcal{A}_2$ by creating an aspect hierarchy of categories based on the categories present in Dpreview and Snapsort. This way, using the PhotoDict taxonomy, we know that 'raw' is an aspect related to 'picture', and that 'memory stick', a non-existing compound word in WordNet, is related to 'storage'. PhotoDict taxonomy will be used in Section 3.4 to discover new aspects from user-generated reviews, and later in Chapter 4 to create the concept vocabulary.

To create the PhotoDict taxonomy, we first extract all top category words H

[9] http://www.wordfrequency.info/

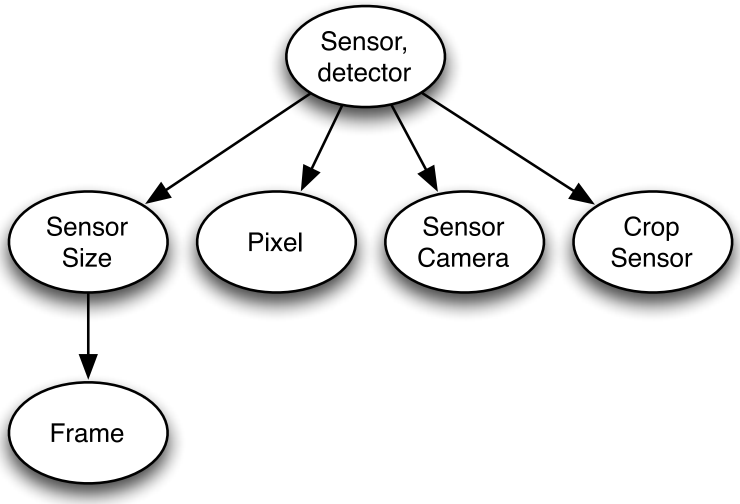


Figure 3.11: PhotoDict taxonomy entry for aspect ‘sensor’. Notice ‘sensor size’ and synonyms, among other aspects, are subcategories of aspect ‘sensor’.

of both Dpreview and Snapsort websites. Since the two websites use a different set of categories to describe the important aspects of a camera, we perform a simple category alignment by comparing the names of the categories of the two websites, together with their WordNet synonyms, using the Jaro-Winkler string distance [Winkler, 1999]. If a category in Dpreview has a similar category name of a category in Snapsort, as happens with ‘Image quality’ and ‘Image’, both categories are mapped together into the same PhotoDict category. This way, all aspects from $\mathcal{A}_2$ generated from ‘Image quality’ Snapsort category and their corresponding subcategories, and all aspects generated from Dpreview ‘Image’ category and its subcategories, will be grouped in the same branch of the PhotoDict taxonomy. On the other hand, if a Dpreview category does not have a Snapsort equivalent category or viceversa, those categories are added to PhotoDict as separate categories. The same procedure is applied to the subcategories H' of Dpreview and Snapsort, and to the H'' sub-categories of Dpreview, until all categories and subcategories are instantiated in the PhotoDict taxonomy. At this point, a manual evaluation is performed to ensure the consistency of the PhotoDict taxonomy category structure, manually modifying the mappings between Dpreview and Snapsort categories if necessary. Table 3.3 shows the mappings created between Dpreview and Snapsort for H categories, and the resulting PhotoDict categories.

Once the PhotoDict structure is created, we populate the PhotoDict taxonomy with all aspects from the aspect set $\mathcal{A}_2$. Remember that every aspect in $\mathcal{A}_2$ was generated in Section 3.3.1 as a result of combining Dpreview and Snapsort

Dpreview H cat.	**Snapsort H cat.**	**PhotoDict**
Viewfinder	-	viewfinder, view finder
Screen	Screen	screen
Focus	-	focus, focal point
Optics	Lens	lens, optics
Connectivity	-	connectivity
Sensor	Low light	sensor
Price	-	price, cost
-	Shutter	shutter
Videography features	Movies	video, movie
Storage	Storage	storage
Physical	-	battery
Photography features, Features	Features	feature, characteristic
Image	Image quality	photograph, image, picture

Table 3.3: Mapping between Dpreview and Snapsort top categories to PhotoDict.

H, H' and H'' category names and synonyms. As such, we add each aspect in $\mathcal{A}_2$ to the corresponding PhotoDict category from where it was generated. Figure 3.11 presents the PhotoDict tree of aspect *sensor*.

The PhotoDict taxonomy contains 182 aspects organized in 13 main categories.

3.4 Aspect Selection Using PhotoDict

In this section, we build a new aspect set $\mathcal{A}_3$ by filtering out the aspect candidate set $\mathcal{A}_c$ using PhotoDict taxonomy instead of WordNet. Remember that $\mathcal{A}_c$ was created with the unsupervised aspect extraction method presented in Section 3.2, and then filtered out using WordNet to create the aspect set $\mathcal{A}_1$. The WordNet filtering removed all candidate aspects that were not found in WordNet, and as a consequence, some interesting camera related aspects were removed. Here we repeat the aspect candidate filtering process to filter the aspect candidates $\mathcal{A}_c$, but using the new PhotoDict taxonomy instead of the WordNet approach used in Section 3.2.

To create the aspect set $\mathcal{A}_3$, we first select a set of salient aspects from A_c using *PhotoDict* taxonomy instead of using the aspect filtering method presented in Section 3.2.3. To do so, we first lemmatize all reviews of our corpus and create the set of lemmatized sentences L_s. Afterwards, we search the set of lemmatized sentences L_s for single words and bi-grams that contain any of the top category aspects H that form the PhotoDict taxonomy. The selected set of nouns and compound nouns is then filtered by frequency based on the frequency threshold δ_s and δ_c, similarly as in previous section, to remove spurious content. However, the quantity of spurious content removed is lower than in section 3.2.3, although the frequency thresholds δ_s and δ_c are the same, because the selected set of nouns and compounds of this section are more frequent. This seems to

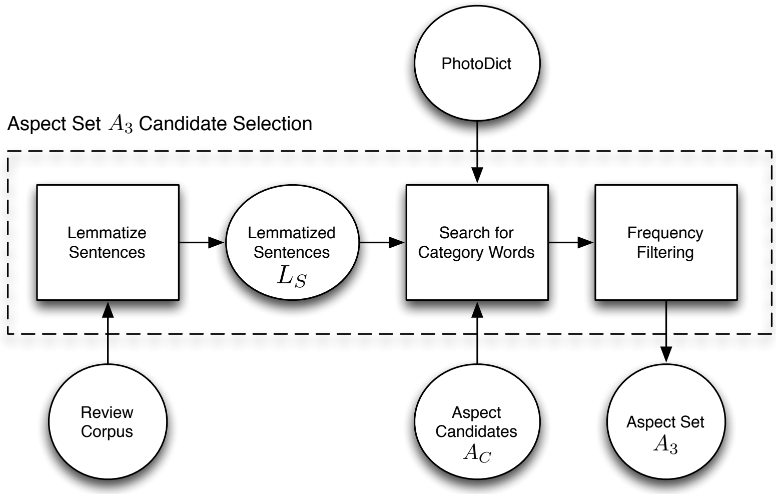


Figure 3.12: Aspect candidate selection using PhotoDict.

be related with the fact that the selected set of nouns and compounds have a common lemma with the top category aspects of PhotoDict, and as such tend to be more frequent.

Notice that, with this method, we do not filter out candidate aspects using WordNet as we did in previous Section 3.2.3, therefore resulting in a different set of selected aspects than $\mathcal{A}_1$. For instance, one of the top level categories in the PhotoDict taxonomy is ‘sensor’. With this approach, we search in all the user reviews from our corpus for single or compound words formed with the word ‘sensor’. By exploring the reviews, we realize that users quite often talk about the ‘ccd sensor’ of certain cameras, and as such ‘ccd sensor’ is selected as a compound aspect. This new compound aspect ‘ccd sensor’ is included into the new aspect set $\mathcal{A}_3$, although it was filtered out in previous aspect set $\mathcal{A}_1$ because it did not exist in WordNet, and was not generated as a candidate word in aspect set $\mathcal{A}_2$ because it was not present in the web resources. Figure 3.12 shows the aspect candidate selection process using PhotoDict, and the creation of the new aspect set $\mathcal{A}_3$.

3.5 Identifying More Compound Aspects

By observing the aspect sets $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ created until this point, we found that the quantity of compound aspects was low with respect to single word aspects in all of the three aspect sets. Analyzing the unsupervised aspect extraction pipeline presented in Section 3.2, we realized that the problem was in the aspect filtering process, specifically with the WordNet filtering. Many of the candidate compound aspects in $\mathcal{A}_c$ were filtered because they were not found in

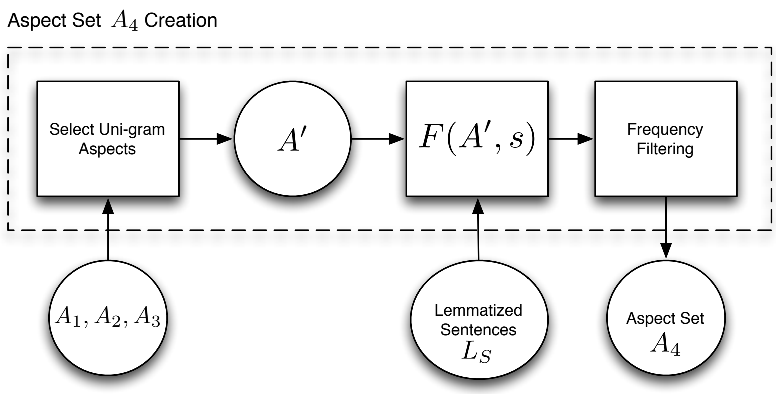


Figure 3.13: Aspect set $\mathcal{A}_4$ creation process.

WordNet. As a result, the final quantity of compound aspects in $\mathcal{A}_1$ is low.

In this section we present an aspect selection method oriented to identify compound aspects bi-grams). Since compound aspects, such as 'zoom lens', are usually formed by joining two single aspects (uni-grams), we search for all uni-gram aspects selected until now in the three aspect sets of the set of lemmatized sentences L_s introduced in previous Section 3.4. Once all uni-gram aspects of a sentence are identified, we create compound aspects by joining them in pairs, finally only keeping those above a certain frequency threshold. The process is presented in Figure 3.13. Let s be a sentence that belongs to the list of lemmatized sentences L_s, A' the union between all single aspects from the aspect sets $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$, and $F(A', s)$ a function that identifies all aspects from aspect set A' found in sentence s. Then, $\forall a_i, a_j \in F(A', s)$, we create all pairs (a_i, a_j) and (a_j, a_i), forming compound aspect candidates. Those compound aspect candidates are then evaluated against the user reviews of our corpus, similarly as we did in previous section, only keeping those above a frequency cut-off δ_c. The resulting set of compound aspects forms a new aspect set of compound aspects named $\mathcal{A}_4$.

3.6 Aspect Vocabulary Creation

The creation of the aspect vocabulary of a corpus is the union of the four aspects sets we have developed so far:

$$\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4 \tag{3.1}$$

Figure 3.14 shows the creation of the aspect vocabulary from the aspect sets. Notice that all four aspect sets were created after applying two frequency filters: δ_s for the uni-gram aspects, and δ_c for the bi-gram aspects or compounds. The objective of the frequency filters is to filter out spurious content and aspects not

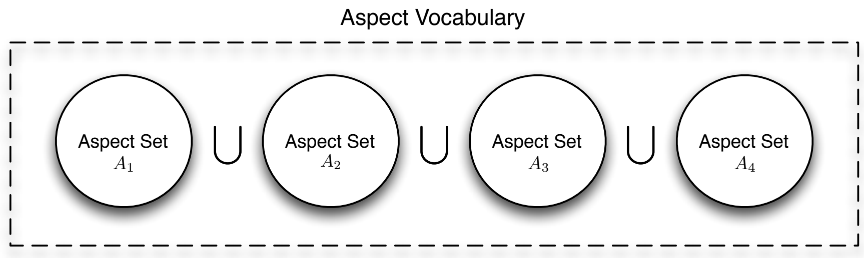


Figure 3.14: Creation of the Aspect Vocabulary AV.

frequently used in the reviews. The reason of using separate frequency thresholds for single and compound aspects is because single aspects are more frequent in the reviews than compound aspects. If we were to use only one frequency threshold for both uni-gram and bi-gram aspects, the resulting aspect vocabulary would almost only contain uni-gram aspects. Furthermore, when we say we apply the two frequency thresholds to generate different aspect vocabularies, we are in fact applying both frequency thresholds to all four aspect sets. In this section we will also analyze the effect of different values of frequency thresholds, in order to select the best δ_s and δ_c, which will be then used in creating the final aspect vocabularies for the three camera types that will be used in the rest of this work.

An aspect vocabulary contains all identified salient lexical items (uni-grams and bi-grams) used by people when expressing their experiences about a product in textual form. All words present in an aspect vocabulary were obtained from and evaluated against the set of user reviews that form one review corpus. Furthermore, all four sets of aspects were selected using automatic and unsupervised methods, with the exception of the PhotoDict taxonomy (although we also show how can it be generated automatically). This means that the methods presented in this chapter are transferable to other product-related domains, as we present in the evaluation performed over phone reviews in Section 3.2.2. The only domain dependent part is the PhotoDict taxonomy, which obviously cannot be used to assess the relations between aspects not related to photographic cameras. However, the creation of the PhotDict taxonomy is detailed in Section 3.3, and is easily replicable to other domains with publicly available resources.

In order to create our review corpora, we extracted camera reviews from Amazon.com belonging to three different camera types. The camera types defined by Amazon are: Digital single lens reflex or DSLR, Compact system cameras or Compact, and Point & Shoot or P&S. The reviews were grouped in 3 corpora, one corpus for each camera type. A brief description of these camera types is the following:

- **Digital Single Lens Reflex cameras or DSLR**: Digital cameras that combine the optics and the mechanisms of a single-lens reflex camera with a digital imaging sensor. The reflex design scheme is the primary difference between a DSLR and other digital cameras.

- **Point & Shoot cameras or P&S**: Point & Shoot cameras are designed primarily for simple operation. Most use focus free lenses or autofocus for focusing, automatic systems for setting the exposure options, and have flash units built in. Point-and-shoots are by far the best selling type of separate camera. They are popular with people who do not consider themselves photographers but want easy-to-use cameras for snapshots of vacations, parties, reunions and other events.

- **Compact cameras or COM**: Compact cameras, also named *bridge* cameras, are a trade-off between the features of DSLR and Point & Shoot cameras. They are usually smaller in size than the DSLR cameras and may have interchangeable lenses, allowing more user customization and specific configurations that are not possible with the simple configurations permitted with a Point & Shoot camera.

From the extracted reviews organized in three corpora, and after evaluating the frequency thresholds, we create three camera aspect vocabularies, one for each camera type. We analyze the quantity of uni-gram and bi-gram aspects that form each aspect vocabulary in order to understand the boundaries that define each camera type. We are also keen to study the contribution of the four aspect sets to the aspect vocabulary of each of the three camera types. Specifically, we want to compare the set of aspects extracted using the unsupervised aspect extraction method $\mathcal{A}_1$ with the rest of aspect sets $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$.

This section is structured as follows. We first describe our corpora of user reviews in Section 3.6.1. Section 3.6.2 studies the impact that different frequency thresholds have in the creation of an aspect vocabulary; the comparison is made against a a "ground truth" formed by manually tagged user reviews. Then, after selecting the frequency thresholds that perform best, we describe in detail the aspect vocabularies of the three camera categories, analyzing the contribution of the 4 aspects sets that form each vocabulary in Section 3.6.3. We finally compare the aspect vocabularies of the three Amazon camera categories DSLR, Point and Shoot and Compact, presenting the most interesting aspect differences between camera types.

3.6.1 Review Corpora

During September 2015, we extracted more than 100,000 reviews of 2,264 Amazon digital cameras from three different camera Amazon.com categories: Digital SLR, Compact System Cameras, and Point & Shoot. We filtered out those products that were older than 1st January 2010 and had less than 15 different user-generated reviews. Then, we united all synonymous products leaving us data for 102 products in the DSLR category, 95 in Compact category, and 599 products in Point & Shoot category. Finally, we grouped the resulting products in three corpus: K_D for DSLR category, K_C for Compact category, and K_P for Point & Shoot category. Each corpus is formed by a set of product-reviews pairs

Category	$\mathbf{K}_D$	$\mathbf{K}_C$	$\mathbf{K}_P$
No. of Products	102	95	599
No. of Reviews	7,552	6,334	84,138

Table 3.4: DSLR, Compact and PAS Camera Corpus.

$\{(p_i, Rev(p_i)\}$, where p_i is a digital camera and $Rev(p_i)$ is the set of reviews about camera p_i.

Table 3.4 shows the quantity of products and reviews that form each corpus. Notice that the set of products in K_P is much larger than K_D and K_C.

3.6.2 Frequency Threshold Evaluation

Before creating the aspect vocabularies that will be used in next chapters, we need to select the frequency thresholds to filter the spurious content. To do so, we take the aspect vocabularies $\mathcal{A}$ created by the methodology presented in Section 3.6 and we apply different frequency thresholds δ_s and δ_c. These results in a number of filtered aspect vocabularies, that we will note $\mathcal{A}_{\delta_s}^{\delta_c}$. Then, we will evaluate each $\mathcal{A}_{\delta_s}^{\delta_c}$ using the measures of precision, recall and f-score. The objective is to identify which are the δ_s and δ_c that create the best aspect vocabulary and to assess how useful the aspect vocabulary is to our purposes of reusing other people's experiences. The evaluation is performed against a manually tagged subset of reviews from K_D, that we consider the "ground truth" corpus against which the measures of precision, recall and f-score are applied. Every review in the ground truth corpus is tagged by three different annotators. The annotators searched and marked aspects that are interpreted to be related to the photographic domain. Then, for different δ_s and δ_c frequency threshold values, we compare how many of the manually tagged aspects are found in an aspect vocabulary $\mathcal{A}$, and how many aspects of a vocabulary aspect $\mathcal{A}$ are not manually tagged as aspects by the annotators.

The frequency thresholds for uni-grams δ_s and for bi-grams δ_c play a very important role in the creation of the final aspect vocabulary. Depending on those two frequency thresholds, the resulting vocabularies may drastically vary in size, an as a consequence, the precision and recall obtained in the evaluation may be affected. Figure 3.15 shows the frequency distribution of the top 1000 most frequent aspects of the three aspect vocabularies. The x-axis corresponds to the aspects in a vocabulary, ordered by frequency, and the y-axis corresponds to the number of occurrences of an aspect in the reviews (presented using a logarithmic scale).

Notice that, among the top 1000 aspects, most of them have less than 1000 occurrences. Moreover, the frequency of the aspects decreases exponentially: the accumulated occurrences of the top thirty aspects in each corpus is greater than all occurrences of the rest of aspects added together. That explains why small changes on the frequency threshold may result in very different aspect

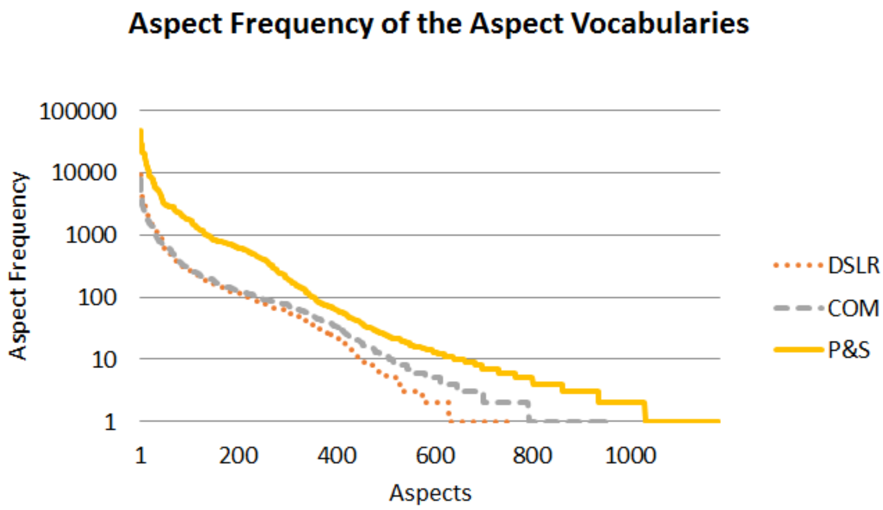


Figure 3.15: Aspect frequency of the top 1000 most frequent aspects for the three aspect vocabularies, ordered by frequency.

vocabularies, and the significance of analyzing the various filtered vocabularies $\mathcal{A}_{\delta_s}^{\delta_c}$.

Remember that δ_s and δ_c are frequency thresholds relative to the most frequent aspect in an aspect vocabulary. For this reason, we write δ_s and δ_c as percentages. For instance, suppose $\delta_s = 0.1\%$. Applying δ_s as the frequency threshold to create $\mathcal{A}_D$, considering that the most frequent aspect in $\mathcal{A}_D$ has a frequency of $7,770$, would be equivalent to filter out all aspects that are used less than 7 times in the reviews of K_D. In fact, since an aspect vocabulary is the union of the four aspect sets, a filtered vocabulary $\mathcal{A}_{\delta_s}^{\delta_c}$ is created as the union of the four aspect sets ($\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$, $\mathcal{A}_4$) filtered using the frequency thresholds δ_s and δ_c.

We proceed as follows: first we present the selection and annotation of the ground truth reviews, and then we analyze the filtered aspect vocabularies generated with different frequency thresholds δ_s and δ_c.

Manual Tagging of Reviews

The first step is to select a subset of reviews from the DSLR corpus K_D that will be tagged by human annotators, since manually tagging 7,552 reviews is not feasible. The annotators, as instructed, will read the subset of reviews; for each review, they will tag the lexical items they consider are aspects relevant to the photographic domain.

The subset of K_D reviews should contain as many different aspects belonging to $\mathcal{A}_D$ as possible. Moreover, the subset of reviews should to be rather small to

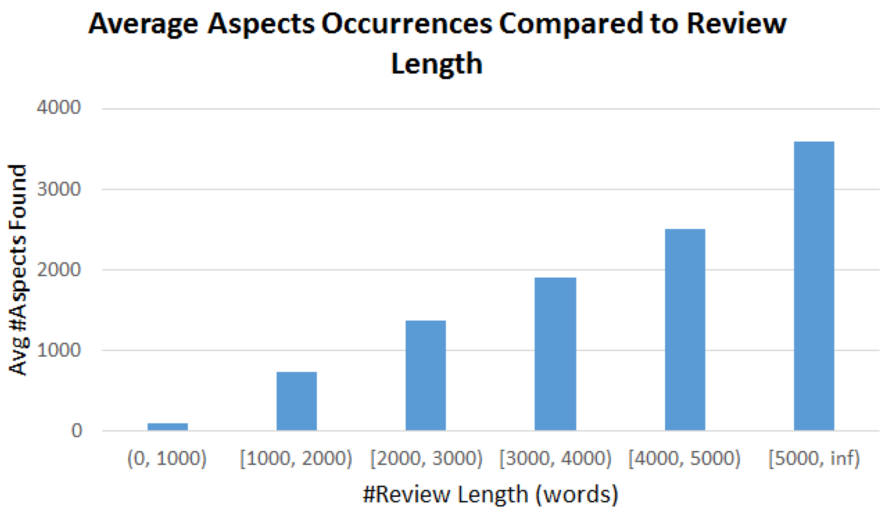


Figure 3.16: Average number of aspect occurrences compared to review length in words.

facilitate the tagging task to the annotators. Thus, we analyzed which reviews contained more aspects, and found a strong correlation between the quantity of words of a review and the average quantity of aspects it contains, presented in Figure 3.16. We see that, on average, the reviews that have between 1 and 1000 words contain a small quantity of aspect occurrences, while the reviews that have 5,000 or more words contain 3,500 aspect occurrences on average.

Since longer reviews contain more aspects, we sorted all K_D reviews from longest to shortest review length. First, we selected the 5 longest reviews to be in the ground truth corpus. Then, iteratively until 70 reviews were selected, we added to the ground truth corpus those review that included more new aspects considering the aspects already present in the selected reviews. Finally, we randomly added 10 more reviews from K_D to complete the ground truth corpus.

Thus, the ground truth corpus has 80 reviews of the DSLR corpus K_D, and contains more than 65,000 words, an average of approximately 900 words per review. Our aspect extraction approach approach identified 1,484 different aspects in the ground truth corpus, an average of 18 unique aspects per review. Among those 1,484 identified aspects, 94 belong to the top 100 most frequent words of the DSLR corpus K_D, and 184 to the top 200. This means that, within the 80 reviews that will be tagged by the three human annotators, we can find a representative set of the most frequent aspects of the DSLR corpus K_D.

Finally, this set of 80 reviews is given to three annotators, with instructions to tag all existing uni-gram or bi-gram aspects related with the domain of digital photography. To facilitate the annotation task, we developed a supporting program that allows the annotator to double click the text to tag or remove

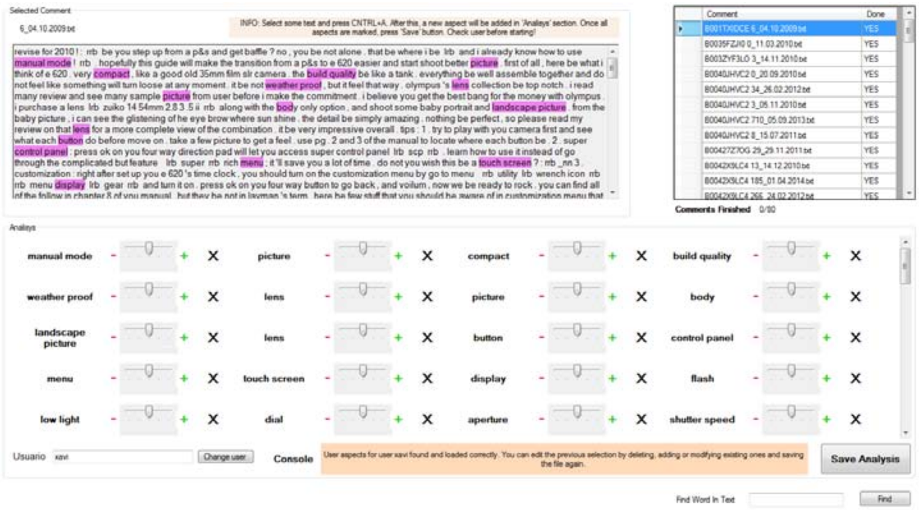


Figure 3.17: Tagging helper and a tagged comment.

an already tagged aspect, and save the review tagging progress so it can be resumed later. The annotator assistant also allow annotators to assign a polarity, positive or negative, to each selected aspect. Although we did not ask the annotators to assess the polarity of the selected aspects, this tool can also facilitate the creation of a sentiment vocabulary in order to evaluate the performance of sentiment analysis systems. Figure 3.17 presents an screenshot of this program.

The three annotators produced 240 annotated reviews, that we will use in next section to evaluate the different aspect vocabularies. The three annotators, Ann_1, Ann_2 and Ann_3, tagged 1,275 aspects in the 80 selected reviews. Table 3.5 presents the number of different aspects tagged. The column *Top 100* shows how many of the tagged aspects belong to the top 100 most frequent aspects in $\mathcal{A}_D$. The row *Two annotators* shows the number of different aspects tagged by at least two different annotators, and *Three annotators* shows the number of different aspects tagged by the three annotators. Notice that at least two annotators agreed in tagging 325 aspects, including 71 of the 100 most frequent aspects in this corpus, while 226 words were considered aspects by the three annotators. Although we want to keep the aspect set ground truth corpus as big as possible, we are not prepared to accept as a valid aspect in our ground truth a aspect only considered by one annotator. Therefore, we have decided to consider valid aspects for our ground truth those aspects tagged by at least two annotators; we will note this aspect set as $\mathcal{A}_M$.

Table 3.6 shows the Jaccard similarity between the aspect sets tagged by the annotators. The Jaccard similarity between the tagged sets of aspects of Ann_1 and Ann_2 is strong, considering that Ann_1 tagged only 323 different aspects compared with the 447 tagged by Ann_2. The Jaccard similarity between Ann_3

	#Different Aspects Total	Top 100
Ann_1	323	75
Ann_2	447	70
Ann_3	505	67
Two annotators	325	71
Three annotators	226	62

Table 3.5: Ground truth annotation summary.

	Ann_1	Ann_2	Ann_3
Ann_1	1	0.524	0.425
Ann_2		1	0.382
Ann_3			1

Table 3.6: Jaccard similarity between sets of annotated aspects.

and the rest is a somewhat lower, mainly because of the large number of words Ann_3 considered to be aspects (Ann_3 tagged 100 aspects more than Ann_1).

Analysis of Frequency Threshold

In this section we use the apect set $\mathcal{A}_M$ to evaluate the precision and recall of several aspect vocabularies $\mathcal{A}_{\delta_s}^{\delta_c}$, created using different frequency thresholds δ_s and δ_c. Recall that $\mathcal{A}_M$ is set of aspects tagged by at least two annotators from the selected DSLR partition of 80 reviews. The goal is to identify which frequency thresholds values for δ_s and δ_c create a better aspect vocabulary for a camera type; therefore our goal is to determine the pair of values δ_s and δ_c for each one of the three camera types. To this end, we selected 8 different value pairs δ_s and δ_c, from which we constructed 8 filtered aspect vocabularies, namely

$$\mathcal{A}_0^0, \mathcal{A}_{0.2}^{0.05}, \mathcal{A}_{0.4}^{0.1}, \mathcal{A}_{0.6}^{0.15}, \mathcal{A}_{0.8}^{0.2}, \mathcal{A}_1^{0.25}, \mathcal{A}_2^{0.5}, \text{ and } \mathcal{A}_5^{1.25}$$

$\mathcal{A}_{0.1}^{0.03}$ means that we used a frequency threshold for uni-gram aspects $\delta_s = 0.1\%$ of the most frequent aspect in $\mathcal{A}$, and a frequency threshold for bi-gram aspects of $\delta_c = 0.03\%$. $\mathcal{A}_0^0$ means that we did not use any frequency threshold to filter spurious content. Since uni-gram words were approximately 4 times more frequent than bi-gram words in the corpora reviews, we decided that the threshold for bi-gram aspects would be set to $\delta_c = \frac{1}{4}\delta_s$. Moreover, recall that this process is performed for each one of the three camera types.

We will now compare each aspect vocabulary $\mathcal{A}_{\delta_s}^{\delta_c}$ against $\mathcal{A}_M$. We consider an aspect vocabulary to be good to the degree that is close to the aspects in $\mathcal{A}_M$. Notice that here we are only interested in finding the frequency thresholds δ_s and δ_c to get rid of the spurious content for the three camera categories. For this reason, we consider that the tagged ground truth created from K_D reviews suffices for the Compact and P&S camera types, as we will see later in this section. Recall that, to overcome the annotator's subjective bias when tagging the ground truth reviews, we decided to include in $\mathcal{A}_M$ only those aspects tagged by at least two annotators. By doing this, we also remove any tagging mistakes the annotators could have made.

Comparing the different aspect vocabularies $\mathcal{A}_{\delta_s}^{\delta_c}$ with $\mathcal{A}_M$ we can obtain the precision, recall and f-score measures. The precision value is the fraction of retrieved instances that are relevant, or in other words, the ratio of how many of the aspects that are in the vocabulary $\mathcal{A}_{\delta_s}^{\delta_c}$ are also in $\mathcal{A}_M$. The recall value is the fraction of relevant instances that are retrieved, that is, the ratio of how many aspects in $\mathcal{A}_M$ are also in $\mathcal{A}_{\delta_s}^{\delta_c}$. The f-score F is a a k-parametric measure that combines precision and recall:

$$F_k = (1 + k^2)\frac{precision \cdot recall}{(k^2 \cdot precision) + recall}$$

Notice that when $k > 1$ we weight recall higher than precision, and when $k < 1$ we weight precision higher than recall. In our experiments we used $k = 1$, since we consider precision and recall equally important.

Tables 3.7, 3.8 and 3.9 show the precision, recall and f-score of 8 generated vocabularies using the DSLR (K_D), the Compact (K_C) and P&S (K_P) corpora. Notice that, in Table 3.7, the recall value keeps decreasing from 0.822 in $\mathcal{A}_0^0$ to 0.221 in $\mathcal{A}_5^{1.25}$, while the precision increases from 0.180 in $\mathcal{A}_0^0$ to 0.866 in $\mathcal{A}_5^{1.25}$. The same behavior is observed in Tables 3.8 and 3.9 for Compact and P&S types. This is because of the frequency filter, that removes less frequent aspects from the aspect vocabulary. The majority of those aspects are spurious content, and as such the precision of the aspect vocabulary increases when those aspects are removed. However, the recall value, decreases indicating that there are also camera related aspects that were tagged, and then removed by the frequency threshold. Notice that without applying any frequency filter the system identifies an 82% of all the words considered aspects by two or more annotators. That means that there is a 18% of words considered aspects by the annotators not detected as such by our system. The f-score weights equally both measures, and the best f-scores are obtained when $\delta_s = 0.8$ and $\delta_c = 0.2$ (with a $F_1 = 0.595$) for DSLR, and $\delta_s = 0.6$ and $\delta_c = 0.15$ for both Compact and P&S.

Notably, the best f-scores for the vocabularies of the three camera types are obtained with $\mathcal{A}_{0.6}^{0.15}$ and $\mathcal{A}_{0.8}^{0.2}$, that are contiguous. Therefore, we may sensibly conclude that the best frequency thresholds are obtained when $0.6 \leq \delta_s \leq 0.8$ and $0.15 \leq \delta_c \leq 0.2$. Since the ground truth corpus was formed by annotated reviews from the DSLR K_D, and the f-score differences between $\mathcal{A}_{0.6}^{0.15}$ and $\mathcal{A}_{0.8}^{0.2}$ are not remarkable, we decided that, from this point onwards, we will use $\delta_s = 0.8$ and $\delta_c = 0.2$ as the frequency filtering values to create the three aspect vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$ and $\mathcal{A}_P$ used in later chapters of this work.

3.6.3 Creating Three Photography Aspect Vocabularies

In this section we create the three aspect vocabularies, $\mathcal{A}_D$, $\mathcal{A}_C$ and $\mathcal{A}_P$, from the three corpus K_D, K_C, and K_P, using the frequency filters selected in previous section. We explore the quantity of uni-gram and bi-gram aspects that form each aspect vocabulary and we analyze the contribution of the four aspect sets to the aspect vocabulary of each of the three camera categories.

In the previous section we determined to use $\delta_s = 0.8\%$ and $\delta_c = 0.2\%$. For the DSLR camera type, this means that the δ_s threshold is set to 60, approximately the 0.8% of 7,770, the number of occurrences of the most frequent aspect ('lens') in $\mathcal{A}_D$, and the δ_c is set to 15. Table 3.10 presents the resulting

DSLR	$\mathcal{A}_0^0$	$\mathcal{A}_{0.2}^{0.05}$	$\mathcal{A}_{0.4}^{0.1}$	$\mathcal{A}_{0.6}^{0.15}$	$\mathcal{A}_{0.8}^{0.2}$	$\mathcal{A}_1^{0.25}$	$\mathcal{A}_2^{0.5}$	$\mathcal{A}_5^{1.25}$
Precision	0.179	0.369	0.491	0.622	0.713	0.745	0.788	**0.866**
Recall	**0.837**	0.680	0.610	0.523	0.511	0.455	0.371	0.221
F_1	0.295	0.478	0.544	0.568	**0.595**	0.565	0.504	0.352

Table 3.7: Precision, recall and f-score of 8 DSLR aspect vocabularies generated using different δ_s and δ_c.

COM	$\mathcal{A}_0^0$	$\mathcal{A}_{0.2}^{0.05}$	$\mathcal{A}_{0.4}^{0.1}$	$\mathcal{A}_{0.6}^{0.15}$	$\mathcal{A}_{0.8}^{0.2}$	$\mathcal{A}_1^{0.25}$	$\mathcal{A}_2^{0.5}$	$\mathcal{A}_5^{1.25}$
Precision	0.222	0.389	0.510	0.586	0.641	0.693	0.775	**0.878**
Recall	**0.738**	0.623	0.539	0.511	0.461	0.430	0.333	0.224
F_1	0.341	0.479	0.524	**0.546**	0.538	0.531	0.466	0.357

Table 3.8: Precision, recall and f-score of 8 COM aspect vocabularies generated using different δ_s and δ_c.

P&S	$\mathcal{A}_0^0$	$\mathcal{A}_{0.2}^{0.05}$	$\mathcal{A}_{0.4}^{0.1}$	$\mathcal{A}_{0.6}^{0.15}$	$\mathcal{A}_{0.8}^{0.2}$	$\mathcal{A}_1^{0.25}$	$\mathcal{A}_2^{0.5}$	$\mathcal{A}_5^{1.25}$
Precision	0.231	0.393	0.537	0.586	0.616	0.697	0.761	**0.852**
Recall	**0.713**	0.619	0.520	0.493	0.454	0.386	0.277	0.215
F_1	0.349	0.479	0.528	**0.535**	0.523	0.496	0.406	0.343

Table 3.9: Precision, recall and f-score of 8 P&S aspect vocabularies generated using different δ_s and δ_c.

	$\delta_s = 0.8\%$	$\delta_c = 0.2\%$
$\mathcal{A}_D$	60	15
$\mathcal{A}_C$	70	20
$\mathcal{A}_P$	280	80

Table 3.10: Resulting frequency thresholds δ_s and δ_c for the three camera categories.

frequency thresholds of the three camera categories. Figure 3.18 shows the δ_s and δ_c thresholds for the DSLR aspects vocabulary.

Table 3.11 shows the sizes of the aspect vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$ and $\mathcal{A}_P$, before (top) and after (bottom) applying the frequency filtering with $\delta_s = 0.8\%$ and $\delta_c = 0.2\%$. As we can observe, by applying a small frequency filtering we remove a large quantity of aspects. We noticed that the size of the three filtered aspect vocabularies is similar: the three vocabularies contain between 190 and 250 aspects. Moreover, the ratio between uni-gram and bi-gram aspects is almost equal in the three vocabularies, showing that the two frequency thresholds δ_s and δ_c are balanced with respect to uni-grams and bi-grams for the three camera types. The three aspect vocabularies, $\mathcal{A}_D$, $\mathcal{A}_C$, and $\mathcal{A}_P$, are presented in Appendix C.

Aspect Set Contribution

Now we will compare the total amount of aspects contributed by every aspect set $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$ for the three aspect vocabularies in Figure 3.19. Keep

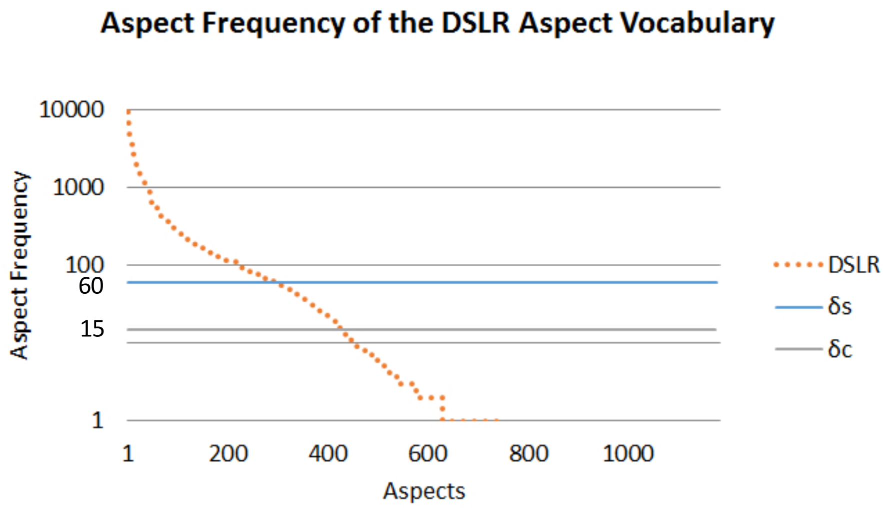


Figure 3.18: DSLR aspect vocabulary selected frequency thresholds δ_s and δ_c.

Before Frequency Filtering			
Corpus	Uni-gram aspects	Bi-gram aspects	Total
$\mathcal{A}_D$	8,153	5,526	13,679
$\mathcal{A}_C$	14,062	5,869	19,931
$\mathcal{A}_P$	13,763	5,360	19,123
After Frequency Filtering			
Corpus	Uni-gram aspects	Bi-gram aspects	Total
$\mathcal{A}_D$	128	138	266
$\mathcal{A}_C$	149	119	268
$\mathcal{A}_P$	110	89	199

Table 3.11: Aspect vocabularies for K_D, K_C, and K_P.

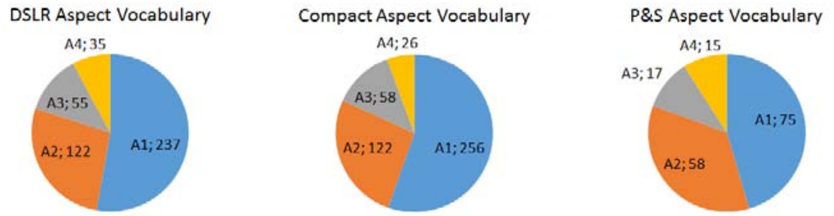
	$\mathcal{A}_D$		$\mathcal{A}_C$		$\mathcal{A}_P$	
	U-gram	B-gram	U-gram	B-gram	U-gram	B-gram
Unique in $\mathcal{A}_1$	44	21	65	21	18	2
Unique in $\mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$	9	31	4	30	21	42

Table 3.12: Unique aspect sets of the aspect vocabularies K_D, K_C, and K_P, where 'U-gram' means uni-gram and 'B-gram' means bi-gram.

in mind that an aspect may belong to more than one aspect set, since the same aspect may have been selected by more than one method presented in this chapter. For instance, this is the case of aspect 'lens', present in both aspect sets $\mathcal{A}^1_D$ and $\mathcal{A}^2_D$, that is $\mathcal{A}_1$ and $\mathcal{A}_2$ of the DSLR aspect vocabulary $\mathcal{A}_D$. As such, aspect *lens* is represented in the figure as belonging to both aspect sets $\mathcal{A}^1_D$ and $\mathcal{A}^2_D$. The fact that an aspect forms part of more than one aspect group can be used as a measure of confidence about how *relevant* that aspect is. In fact, 9 of the top 10 most frequent aspects of $\mathcal{A}_D$ are found in more than one aspect set. This aspect confidence measure could be then used to further prune those aspects that, for instance, were not selected in two or more aspect sets. This would improve the precision, but at the expense of drastically reducing the recall of the aspect vocabulary. Therefore, we chose not to implement this measure in our work.

Next, Table 3.12 compares the quantity of aspects found exclusively by the unsupervised aspect extraction method in $\mathcal{A}_1$, or exclusively with the other aspect methods that created the other aspects sets $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$. This way we can assess how useful $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$ are to the final vocabulary, and compare it with $\mathcal{A}_1$, the aspect set that contributes with most aspects to the final vocabulary as shown in Figure 3.19. As expected, $\mathcal{A}_1$ is the aspect set that adds more unique content to the aspect vocabulary. If the aspect vocabulary $\mathcal{A}_D$ were to be formed by only aspects sets $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$, it would contain 44 uni-grams and 21 bi-grams less. On the other hand, we observe how the aspects sets $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$ are also important to the final vocabulary, in the sense that they also provide an important quantity of uni-grams and bi-grams that were not extracted by the unsupervised aspect extraction techniques that created $\mathcal{A}_1$. Some aspects belonging to $\mathcal{A}_2$, $\mathcal{A}_3$ and $\mathcal{A}_4$ but not to $\mathcal{A}_1$ are 'external microphone', 'screen resolution' and 'exposure compensation'. Some aspects belonging exclusively to $\mathcal{A}_1$ are 'lcd display', 'lag' and 'iso setting'.

Figure 3.19 shows that all three aspect vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$ and $\mathcal{A}_P$ have a similar distribution of aspects over the 4 aspect sets, even considering the difference in the quantity of reviews between P&S and DSLR/Compact types. This means that all four methods used to construct the four aspect sets are stable across the three vocabularies.



(a) DSLR category $\mathcal{A}_D$ (b) Compact category $\mathcal{A}_C$ (c) P&S category $\mathcal{A}_P$

Figure 3.19: Size of the aspect sets in the three aspect vocabularies.

Comparison between $\mathcal{A}_D$, $\mathcal{A}_C$ and $\mathcal{A}_P$

We will now focus on the aspects that form the three aspects vocabularies. Table 3.13 presents a list of the top 10 most frequent aspects of the three aspect vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$ and $\mathcal{A}_P$. Notice that, even if the aspect vocabularies are created from three different camera corpora, the top 10 most frequent aspects of the three types of cameras are similar. We can observe that aspect 'lens', 'video' and 'picture' are between the top 10 most frequent aspects of each aspect vocabulary. Those aspects seem to be important for users of all kinds of cameras, no matter their category. Aspect 'lens' is ranked 1st for DSLR and Compact categories, as expected, since both categories possess interchangeable lenses. However it is ranked 5th in Point & Shoot. Furthermore, seems that the 'price' of the camera is a more important factor for those people who bought a Point & Shoot cameras, while the camera 'body' is important for DSLR buyers. We also observe how 'battery' (ranked 4th in PAS) is an important aspect for Point & Shoot cameras. Finally, aspect 'zoom' occupies the 2nd place in the ranking of most frequent aspects for Point & Shoot cameras, but it is not present in the top 10 most frequent aspects of the other two camera categories. This can be surprising, however it can be explained because P&S cameras do not usually have interchangeable lenses. As such, the optical and digital zoom of the Point & Shoot cameras are usually commented in the camera reviews.

Next, we are interested in analyzing the similarity between the aspects contained in the three aspect vocabularies. To this end, we use the Jaccard set similarity coefficient (see Equation 3.2) to perform a pairwise comparison of vocabularies:

$$J(\mathcal{A}_i, \mathcal{A}_j) = \frac{|\mathcal{A}_i \cap \mathcal{A}_j|}{|\mathcal{A}_i \cup \mathcal{A}_j|} \tag{3.2}$$

where $\mathcal{A}_i$ and $\mathcal{A}_j$ are two different aspect vocabularies.

The results in Table 3.14 show that the three camera types have clearly different aspect vocabularies. The camera types that share more aspects are DSLR and Compact, with a Jaccard similarity of 0.6262. Point & Shoot obtains a lower similarity when compared with both DSLR and Compact aspect vocabularies (0.5349 and 0.5172 respectively), indicating that Point & Shoot

Top 10 aspects $\mathcal{A}_D$	Top 10 aspects $\mathcal{A}_C$	Top 10 aspects $\mathcal{A}_P$
lens	lens	picture
picture	shoot	zoom
shoot	picture	photo
video	focus	battery
shot	photo	lens
focus	shot	video
photo	flash	shot
feature	image	feature
image	feature	quality
body	video	price

Table 3.13: Top 10 most frequent aspects of the aspect vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$, and $\mathcal{A}_P$.

	$\mathcal{A}_D$	$\mathcal{A}_C$	$\mathcal{A}_P$
$\mathcal{A}_D$	1	0.6262	0.5349
$\mathcal{A}_C$		1	0.5172
$\mathcal{A}_P$			1

Table 3.14: Jaccard similarity between the three aspect vocabularies.

Top 50	$\mathcal{A}_D$	$\mathcal{A}_C$	$\mathcal{A}_P$
$\mathcal{A}_D$	1	0.7544	0.6129
$\mathcal{A}_C$		1	0.5625
$\mathcal{A}_P$			1

Table 3.15: Jaccard similarity between the top 50 most frequent aspects of the three aspect vocabularies.

category has a slightly more different aspect vocabulary than those of the other two camera types.

Similarly, Table 3.15 shows the Jaccard similarity between the top 50 most frequent aspects of the three aspect vocabularies. As expected, the similarity between the aspect vocabularies of the different camera types increases for the top 50 most frequent aspects. Notice that DSLR and Compact have a Jaccard similarity of 0.75 (0.13 above the similarity when considering both complete aspect vocabularies). Also, the similarities between DSLR and Point & Shoot, and Compact and Point & Shoot increase between 0.05 to 0.09 points with respect to the results of Table 3.14. We can draw two conclusions from these results: 1) the most frequently used aspects of the three different camera types are more similar than less frequent aspects, which tend to vary between camera types, and 2) the Point & Shoot aspect vocabulary is the most different compared with the other two.

At this point, we know that the aspect vocabularies are different, specially considering the top 50 most frequent aspects of the aspect vocabulary of DSLR and Compact types. Now we want to study how the relative frequency of occurrence of those aspects is related to the camera type. To do so, we order the aspects of each aspect vocabulary by frequency and create three aspect frequency rankings r_D, r_C and r_P, one for each camera type. This will allow us to

	r_D	r_C	r_P
r_D	1	0.8525	0.6374
r_C		1	0.5590
r_P			1

Table 3.16: Spearman rank correlation between the three frequency ranked vocabularies.

Top 50	r_D	r_C	r_P
r_D	1	0.8595	0.5614
r_C		1	0.6323
r_P			1

Table 3.17: Spearman rank correlation between the top 50 most frequent aspects of the three frequency ranked vocabularies.

compare pairs of frequency rankings using the Spearman rank correlation:

$$Sp(r_1, r_2) = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{3.3}$$

where r_1 and r_2 are two frequency ordered aspect vocabularies, $d = r_1(a_i) - r_2(a_i)$ is the difference between the two ranks of aspect a_i in r_1 and r_2, and n is the number of common aspects between the two aspect vocabularies.

Tables 3.16 and 3.17 show the Spearman rank correlation between the commons aspects shared by the three different aspect vocabularies: Table 3.16 considering the whole aspect vocabulary, and Table 3.17 considering only the top 50 most frequent aspects. It it interesting to point the high correlations between r_D and r_C (Spearman correlation is 0.8525), and the correlations of both r_D and r_C with r_P. This result should be considered in the context of the previous results presented in Tables 3.14 and 3.15. Namely, that the relative importance of the common aspects shared between the frequency ranked vocabularies of the three camera categories is similar: 0.8525 Spearman rank correlation between DSLR and Compact, 0.6374 between DSLR and Point & Shoot, and 0.5590 between Compact and Point & Shoot. The Spearman rank correlations between camera types obtain similar values when only considering the top 50 most frequent aspects of the vocabularies.

Next we explore the three frequency ranked aspect vocabularies r_D, r_C, and r_P, and search for those aspects found between the top most frequent aspects of a camera type, but not between the top most frequent aspects of other types. Our goal is to find which aspects better define the boundaries between the aspect vocabularies of the three camera types. Table 3.18 shows that, 'live view' is an aspect very frequent in the reviews of K_D (the 44th most frequent aspect in DSLR), however 'live view' is not that frequent for Compact and Point & Shoot types (ranked 139th and 159th, respectively). Similarly, we observe that almost all aspects related to 'lens' are deemed important for DSLR and Compact since they are usually found between the top 50 more frequent aspects in both camera categories. As expected, those aspects are not very frequent in Point & Shoot reviews. Finally, among others, we may observe how aspects 'touch screen', 'charge' and 'wifi' are more frequently found in the reviews of Compact and Point & Shoot cameras than in DSLR. This result was also to be expected, since battery and charging speed, together with connectivity, are highly desired features for this set of all-around cameras. Analyzing the aspects to create this

r_D	r_C	r_P
live view (44)	139	159
evf (79)	36	-
hd video (80)	152	67
kit lens (33)	31	-

r_C	r_D	r_P
touch screen (45)	99	63
wifi (61)	115	56
kit lens (31)	33	-

r_P	r_D	r_C
wifi (56)	115	61
charger (41)	127	92
memory (69)	105	147
gps (97)	162	-
battery (4)	15	18
waterproof (50)	-	-

Table 3.18: Difference in aspect ranking for DSLR, Compact and PAS.

table, we found that aspect ‘waterproof’ was the 50th most frequent aspect in Point & Shoot cameras, but it was filtered out by the frequency filter in both DSLR and Compact aspect vocabularies. This shows that the interests of the users that buy a Point & Shoot camera are more oriented to use them in different types of activities that require an easy-to-use camera, whilst users that buy a DSLR or a Compact are more interested on the quality of image and lenses.

In this section we have seen that the aspect vocabularies used by people when describing their experiences about cameras are unique and depend on the camera type the experience is describing. the three aspect vocabularies, $\mathcal{A}_D$, $\mathcal{A}_C$, and $\mathcal{A}_P$, contain a different set of aspects. Furthermore, the importance of the common aspects of each camera vocabulary changes between camera types. For instance, aspects such as ‘live view’ or ‘lens’ are very frequently used by people that define product experiences with DSLR cameras. However, those aspects are rarely used when describing Point and Shoot product experiences. On the other hand, ‘price’ and ‘battery’ are important aspects related with Point and Shoot cameras but not with DSLR or Compact. Moreover, some aspects only occur in the reviews of certain camera types. That is the case of, for instance, aspect ‘waterproof’ of $\mathcal{A}_P$. For a Point and Shoot, being waterproof is an interesting feature, and as such the P&S experiences frequently contain a reference to the aspect. However, the same aspect does not frequently occur in the reviews of DSLR and Compact cameras, because it is not a desired feature in those cameras.

Because of these differences, we have decided it is better to work with three different aspect vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$ and $\mathcal{A}_P$, each for one camera type, than to fuse them in one general digital camera dictionary. The res of the Chapters will thus work with each one of the aspect vocabularies separately.

3.7 Conclusions

The aspect vocabulary presented in this monograph is created mainly analyzing user textual experiences expressed as reviews and, ancillary, by analyzing two web resources about photography. Furthermore, differently from other opinion mining approaches, we create the aspect vocabulary without taking into consideration (yet) the polarity of judgments, but by exploiting the syntactic structure of the sentence and the domain knowledge available in two photography websites.

An important factor when creating an aspect vocabulary from user-generated product descriptions is that it may change over time, because what makes a product feature interesting now may become the accepted standard in the future. Therefore, it is important to assess how time influences the aspect vocabularies used by people when describing products in reviews. Although in this monograph we considered a fixed time window comprised between 2010 and 2015 to perform the experiments, we studied the temporal dynamics of the aspect vocabularies and user preferences of DSLR cameras in Appendix B. The results presented in Appendix B suggest that better aspect vocabularies are created when considering the temporal dynamics to model aspect vocabularies.

In this monograph, we present a combination of four methods that create the vocabulary used by people when explaining important issues that define product experiences. The four methods used to create the aspect vocabulary are complementary, in the sense that they discover different set of aspects from user reviews by using different techniques. First, we use grammatical extraction rules, combined with part of speech tagging and frequency filtering to select a set of aspects from user reviews in an unsupervised way. This set of aspects proved to obtain good precision when compared with other state of the art techniques, but obtained a low recall due to the frequency filtering applied to remove spurious content. To improve the frequency filter, we used WordNet to asses the relation of every aspect with the photography domain, removing those that were not related *enough*. This filtering method correctly filtered out words not related with the photography domain, increasing the recall. However we realized that some interesting aspects were filtered out by this approach, because WordNet was missing a lot of content related to photography. That is, some interesting aspects (mostly compound), did not exist in WordNet, and as such we were discarding them.

Motivated by the problems found with WordNet, we decided to create our own photographic taxonomy, *PhotoDict*. To do so we explored two web resources, Dpreview and Snapsort, and analyzed the categories and vocabulary they use when describing cameras. From those resources, and after evaluating their vocabulary against our set of user reviews, we created another aspect set formed by more technical and specialized vocabulary.

Later we used the PhotoDict taxonomy as a way to filter out the spurious content extracted by the unsupervised aspect extraction method. That is to say, instead of assessing the relatedness of the candidate aspects with the photography domain using WordNet, we used PhotoDict. As a result, we created the third aspect set, which included new candidate aspects not found in WordNet. We need to clarify that this method is not meant to replace the WordNet filtering, but to be used as a complement. PhotoDict content is specialized in the photographic domain, and as such, it does not contain the necessary general vocabulary to asses the majority of word relations found in a free-text environment like user reviews.

The fourth aspect set was created with the goal of finding more compounds (bi-grams) aspects than what we had acquired with the three previous aspect sets. The approach is a kind of exhaustive generation and later filtering of possible bi-grams at a sentence level. Thus, we first generated all combinations from the uni-gram aspects selected in the previous three aspect sets and then we checked how many occurrences of them were found in a review corpus. This created a new aspect set, uniquely formed by compound aspects.

The result of aggregating those aspect sets is a rich aspect vocabulary, created by the combination of the four methods described, and similar to what can be created by a human annotator. Using this methods we created the three aspect vocabularies, one for each of the three corpora extracted from Amazon (K_D, K_C, and K_P), that will be used in next chapters.

We noticed that an important step when creating an aspect vocabulary is the selection of a frequency filter that will remove spurious content and false aspects. The frequency filter should be carefully selected, otherwise it will also remove interesting camera related aspects. In this chapter, we found that the optimal aspect frequency filtering values for the vocabularies of the three camera types, are found between the 0.6% and 0.8% of the top most frequent aspect for uni-grams, and the 0.15% and 0.20% of the top most frequent aspect for bi-grams. Furthermore, filtering the less frequent aspects is also necessary for our work, since in later chapters we will be performing aspect polarity aggregation to create the basic concepts that will define the bundles of arguments. If an aspect does not occur often enough in the reviews, we will not be able to assess its sentiment polarity with enough confidence.

Furthermore, we observed differences between the aspect vocabularies of three Amazon camera types, DSLR, Compact and P&S. The reviews of each camera type contain a different set of aspects, and users define their product experiences differently. For instance, 'price' and 'battery' occur more frequently in the reviews of Point and Shoot cameras than in the reviews of the other two camera types, while 'lens' is of prior interest for users who bought a DSLR or a Compact camera. Since the difference between camera types is clear when comparing their aspect vocabularies, we decided to work with three corpora and three vocabularies instead of conflating them into one general corpus and vocabulary for all types of cameras. This decision is important, since it will affect the results of future chapters of this monograph.

Chapter 4

Concept Discovery

4.1 Introduction

In this chapter, we are interested in identifying the issues addressed in product reviews expressing people's experiences on using those products, in the domain of digital cameras. An issue can be expressed by a wide variety of words in text. The previous chapter has addressed the extraction of *aspects* from a corpus, however in our approach each aspect need not be an individual issue. Instead, we consider that several aspects may refer to the same issue. This is the case of synonyms: two aspects can be synonymous, and therefore refer to the same issue. However, there may be aspects that are not strictly synonymous in the dictionary sense, but may refer to the same issue by the way those aspects are used in the text. Furthermore, issues are part of user judgments found in the product reviews, and judgments have a polarity based on the experiences of people with a product. The coherence of the sentiment polarity of the judgments present in user reviews can help elucidate those aspects that refer to the same issue.

Consider, for instance, these three aspects: 'picture', 'pic' and 'JPEG'. One may surmise that people using those aspects in reviews are in fact referring to the same issue, i.e. *the picture obtained by my digital camera*, because they have the same intended meaning. However, although 'picture' and 'pic' may be considered synonyms, not so 'JPEG' (that is a type of digital format for images). So, strictly speaking 'JPEG' refers to a different concept (digital formats), but in the text of the reviews it is often used to refer to *the picture obtained by my digital camera*. The purpose of this chapter is to find these more subtle relations in the usage of words, that do not follow the standard definitions in a dictionary. That is to say, our analysis may decide that in our corpus 'JPEG' does not refer to the concept of digital formats but to the concept *the picture obtained by my digital camera*. In practice, this implies that we need to find groups of aspects, beyond strict synonymy, that belong to the same concept (they refer to same issue in a review corpus).

Clustering similar aspects in groups that are considered to belong to the same concept opens a problem of granularity: how many groups partition the set of aspects into concepts? On one side we can have a few large groups, or on the other side, a large number of small groups, or something in between. This decision on granularity is a tradeoff between having a small number of general concepts or a large number of specific concepts.

To solve this decision on granularity we use the notion of *basic level concepts* (BLC) from cognitive linguistics, explained in detain in Section 4.2. In our example above, an example of a basic level concept for the digital camera domain is *the picture obtained by my digital camera*, rather than 'picture' or 'photo' in abstract. Even if we later use one of the aspects to name this concept, e.g. calling it the 'photo' concept, we view this concept 'photo' as a *model of the issue* addressed by the users writing reviews, that is to say *the picture obtained by my digital camera* (even this phrase is never used in practice). In other words, when we create a group of aspects, we will consider it a concept, but this concept is considered by us a model of the issue being discussed, instead of a representation of an abstract concept (e.g. 'photo').

Thus, using the notion of basic level concepts we will be able to identify an adequate granularity for clustering an aspect vocabulary $\mathcal{A}$ into a set of groups. The set of basic level concepts (groups of aspects) will form the concept vocabulary $\mathcal{C}$ for a given corpus. The concept vocabulary models the important issues used by people when expressing their experiences in product reviews, and will be used in Chapter 5 to define the bundles of arguments of each product.

This chapter is organized as follows. We first introduce the cognitive linguistics notion of basic level concepts in Section 4.2. In Section 4.3 we present a hierarchical clustering approach that groups aspects based on the aspect usage in the user reviews. The hierarchical clustering returns a set of possible partitions, formed by aspects grouped at different levels of granularity. Each partition determines the set of concepts that model the collection of issues that are used in judgments in a review corpus.

Section 4.5 addresses the problem of selecting a single partition of the set of aspects, and therefore committing to a set of concepts that will be our vocabulary for the rest of our research work. In order to decide which partition is selected, we define a measure over partitions that explore the degree of coherence of the aspects in a group with respect to the judgments' sentiment polarity. The analysis of judgments' sentiment polarity is explained in Section 4.4. The selected partition is considered to describe the basic level concepts (BLC) of digital cameras based on the user-reported experiences in a corpus, and will form the concept vocabulary $\mathcal{C}$ that we will use for such a corpus. Then, Section 4.6 analyzes the three concept vocabularies created for the three digital camera types. Finally, Section 4.7 summarizes the approach and contributions of this chapter.

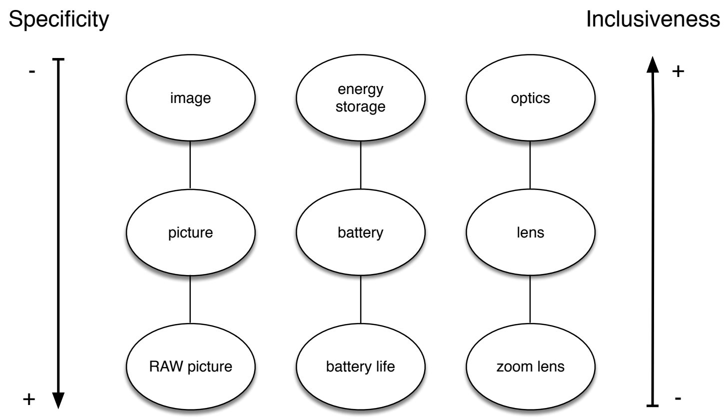


Figure 4.1: Examples of several levels of categorization on digital cameras.

4.2 Basic Level Concepts

In this section we introduce the notion of basic level concepts (BLC) and its relation with the basic level categories of cognitive linguistics. Furthermore, we explain how we use this notion to create adequate groupings of aspects.

In cognitive linguistics, Rosch et al. [Rosch et al., 2004] define two basic principles that guide the human categorization system: (1) the principle of cognitive economy, and (2) the principle of perceived world structure. These basic principles state that humans group similar stimuli into categories to maintain economy in cognitive representation, and that humans rely on the correlational structure between categories in order to form them. For this reason, Rosch et al. present the human categorization as having a vertical and horizontal dimension. The principle of cognitive economy, represented in the vertical dimension, has implications for the level of inclusiveness with which categories are formed. The higher up the category is in the vertical axis, the more inclusive it is. On the other hand, the principle of correlational structure is represented as the horizontal dimension, and has implications for the representativeness of the categories formed. The horizontal dimension relates to the category distinctions at the same level of inclusiveness.

Figure 4.1 presents various categories related with the camera domain. For instance consider the category 'picture'. Relative to this category, we find 'image' higher up the vertical axis. Therefore the category 'image' is more inclusive than the category 'picture'. Similarly, the category 'RAW picture' is more specific than 'picture' and 'image'. On the other hand, 'RAW picture' and 'battery life' are distinct categories that occupy the same position in the vertical axis. Hence, according to Rosch et al., they operate at the same level of detail.

According to Rosch et al. [Rosch, 1973, 1977; Rosch and Lloyd, 1978], information-rich bundles of perceptual and functional attributes occur that form natural discontinuities, and basic cuts in categorization are made in such discontinuities. Following this idea, categories within taxonomies are structured such as there is one level of abstraction at which the most basic category cuts can be made; the greater the inclusiveness of a category within a taxonomy, the higher the level of abstraction. We are interested in identifying the natural discontinuities between the sets of aspects that form the aspect vocabulary, to obtain the precise level of abstraction that will allow us to reuse people experiences. With that level of abstraction, we will create the set of basic level concepts, and the concept vocabulary $\mathcal{C}$.

Basic level concepts (BLC) are those that strike a tradeoff between two conflicting principles of conceptualization: inclusiveness and discrimination. Rosch et al. found that there is a level of inclusiveness that is optimal for human beings in terms of providing optimum cognitive economy. This level of inclusiveness is called the basic level, and concept or categories at this level are called basic-level concepts. More inclusive concepts, that is, those concepts that give less details and are found higher up in the vertical axis, are called superordinate concepts. Concepts lower down the vertical axis, providing more detail, are called subordinate concepts.

In Figure 4.1, the basic level is found at the middle level of the hierarchy, and the concepts 'picture', 'battery' and 'lens' are basic level concepts. Concept 'image' is the superordinate concept for 'picture' (as are 'energy storage' and 'optics' to their respective BLCs), and it is more general: A 'picture' is considered here an *image made using a camera*, and used in common parlance more often than the technical term 'photograph'; on the other hand, 'image' is considered an abstract term, a *physical likeness or representation of a person, animal, or thing*, including painted canvases, and sculptures. On the other hand 'RAW picture' is a subordinate concept of the BLC 'picture', and it is more specific. Notice that the BLC concept subordinates are usually formed by combining the BLC concept name with another word that specifies the nature of the given BLC. For instance, 'RAW picture', or other concepts such as 'gray-scale picture' and 'color picture' are also subordinate concepts of the BLC 'picture', because they are more specific and can be abstracted to have the same intended meaning as 'picture'.

In our approach, we create a collection of basic level concepts in an unsupervised way from the review corpus, where each BLC assembles a set of aspects that, according to our analysis, are used in a similar way by the reviewers. As we show in section 4.3, we estimate this similarity by taking into account the semantic and lexical similarities of the aspects, and analyzing its usage in the reviews. Furthermore, we evaluate the coherence/incoherence of the polarity of the judgments whose issues are assembled in a given BLC. The first step towards the creation of the set of BLCs is to create the categories of aspects included in the aspect vocabulary. We do so by using a hierarchical clustering algorithm, described in Section 4.3.

4.3 Aspects Hierarchical Clustering

The objective of this chapter is to obtain the set of basic level concepts users deem relevant to express in a text describing his or her experiences. Since in previous chapter we identified all salient aspects used by people and created the aspect vocabulary $\mathcal{A}$, here we need to assess the similarity between those aspects and group them to select the set of basic level concepts we will use in later sections. We say that two aspects are similar if they are used in similar ways in the reviews of a product. This approach is detailed later in this section.

To group the aspects in concepts, we use a hierarchical clustering algorithm. This process returns a dendrogram: a tree diagram used to illustrate the arrangement of the clusters produced by the clustering algorithm. Exploring this dendrogram, we are able to select the set of basic level concepts, and define our concept vocabulary $\mathcal{C}$.

We divide the selection of basic level concepts in two parts: First, in this section we create a dendrogram, using a hierarchy clustering algorithm, where we cluster the set of aspects in aspect groups by analyzing the aspect usage among the reviews of the corpus and exploring the semantic and lexical similarities between them. Once the hierarchical clustering is finished, in Section 4.5 we explore the resulting dendrogram to select the set of basic level concepts that will form our basic level concept vocabulary $\mathcal{C}$.

The first step to create the clustering dendrogram, is to identify the synonyms of the aspects that form the aspect vocabulary. To this end we use WordNet, a lexical database of English. Every aspect a in the aspect vocabulary $\mathcal{A}$ is mapped to the corresponding WordNet synset with the same noun word form, if it exists, and is disambiguated by identifying the synset with the shortest aggregated WordNet *Path Distance* [Meng et al., 2013] to a set of manually selected WordNet synsets formed by the top 5 most frequent aspects of the aspect vocabulary. The aspects that have a synonymy relation among them are grouped together into *aspect groups* G_j. Aspects without synonyms form a group of cardinality 1. This collection of aspect groups G is the input of the aspect hierarchical clustering that we present in this section.

Next, we iteratively cluster the most similar groups of aspects and create a dendrogram from where the set of basic level concepts will be selected. To group the aspects we use an unsupervised bottom-up hierarchical clustering algorithm that takes the most similar pair of groups at each stage and joins them together in a higher level aspect group. Joining two aspect groups G_i and G_j results in a new aspect group G_k formed by uniting the aspects of both G_i and G_j. For instance suppose the next two aspect groups $G_i = \{\text{battery, battery pack}\}$ and $G_j = \{\text{battery life}\}$ are to be joined. The resulting aspect group G_k will be formed by the three aspects {battery, battery pack, battery life}. Algorithm 4 presents the complete aspect hierarchical clustering algorithm.

We will define now similarity measures over aspects and over groups, used to select the most similar aspect groups in every iteration of the hierarchical clustering. The similarity between two groups of aspects G_i and G_j is defined

Algorithm 4: Aspect Hierarchical Clustering Algorithm

```
Input: Aspect Groups G;
Output: dendrogram H;
repeat
    // Find the two most similar aspect groups in G
    (G_i, G_k) = FindMostSimilarGroups(G);
    // Join them together
    G_k = Join(G_i, G_j);
    // Remove old aspect groups G_i, G_j, and add new G_k
    G.remove(G_i, G_j);
    G.add(G_k);
    // Update dendrogram
    H = UpdateDendrogram(G);
until |G| := 1;
return H;
```

as:

$$Sim(G_i, G_j) = \frac{1}{|G_i||G_j|} \sum_{n=1}^{|G_i|} \sum_{m=1}^{|G_j|} SimA(a_n, a_m)$$

where a_n and a_m are aspects from the aspect groups G_i and G_j respectively. The similarity measure between two aspects is:

$$SimA(a_i, a_j) = \alpha \cdot \Gamma(a_i, a_j) + \beta \cdot \Delta(a_i, a_j) + \gamma \cdot \Lambda(a_i, a_j)$$

where α, β and γ are weighting parameters in $[0, 1]$ such that $\alpha+\beta+\gamma = 1$. The values of *SimA* are in $[0, 1]$. Functions $\Gamma(a_i, a_j)$, $\Delta(a_i, a_j)$ and $\Lambda(a_i, a_j)$ estimate the aspect similarity between aspects a_i and a_j in three different dimensions: Semantic similarity, string similarity and PhotoDict similarity. We introduce them next:

Semantic Similarity (Γ) Compares two aspect co-occurrence vectors to estimate the similarity between aspects [Sani et al., 2011; Wiratunga et al., 2006]. The co-occurrents of an aspect are the other aspects that have a first order co-occurrence with it within a sentence window. By passing this window over the entire corpus we obtain, for each aspect, a list of its co-occurrent aspects. The lists of co-occurrent aspects represent the global contexts of words, and we use it to estimate the semantic word similarity between aspects. That is to say, we consider that two aspects are semantically close if the co-occurrence vectors of both aspects with respect to all other aspects in the aspect vocabulary are similar.

Figure 4.2 shows the co-occurrent relations between some of the top most frequent aspects in the aspect vocabulary $\mathcal{A}_D$. Notice that the figure does not show the similarity between the co-occurrence vectors, but only the first order co-occurrences between the most frequent aspects of $\mathcal{A}_D$. The size of the nodes represent the frequency of occurrence of the aspects over the reviews of K_D, the most frequent aspects are shown in bigger nodes, and the edges represent the strength of the co-occurrence between aspects, the wider the edge the more times those aspects co-occur in the same sentence. Notice that, for instance, aspect 'lens' is frequently used in the same sentence as aspect 'zoom', aspect 'shoot' frequently co-occurs with 'point', and aspect 'iso' is frequently used with 'noise'.

For every aspect a, we represent the set of first order co-occurrences between a and the rest of the aspects in $\mathcal{A}$ in a vector space, where the dimensions of this space are the set of aspects in the aspect vocabulary, and the values that form the vector are the total of times those aspects co-occur in a sentence in the reviews. The co-occurrence values that form the vector are then normalized to facilitate the comparison between aspects with different frequency usage in the reviews. The vectors are then compared using the cosine similarity, the higher the cosine similarity the more semantically equivalent are the two aspects.

Table 4.1 shows the co-occurrence vectors for aspects 'picture', 'photo' and 'lens', considering the co-occurrence frequency between the aspects (top), and the resulting normalized co-occurrence vectors (bottom). Remember that the co-occurrence vectors contain a co-occurrence value for all aspects of $\mathcal{A}$. In this example, we reduced the dimensionality of the co-occurrence vectors to only a subset of 8 aspects as an example. At the top of Figure 4.1 we show the amount of times the two aspects co-occur together in the same sentence considering the set of reviews in corpus K_D. For instance, aspect 'picture' was used in the same sentence as aspect 'image' 140 times, and aspect 'lens' was used together aspect 'zoom' 529 times. Notice that 'lens' and 'lcd' were never used together in the reviews. This sparseness problem is well known and serious in the co-occurrence statistics. However it is not very common in our corpus, because the aspects that form the aspect vocabulary were selected after a frequency filter, assuring that all aspects occur at least enough number of times in the reviews.

The bottom of Figure 4.1 shows the normalized co-occurrence vectors of the three example aspects. The normalization is done with respect to the total frequency of the aspect in the reviews of a corpus. If we were to add all co-occurrence values from the real co-occurrence vectors of one aspect, it would add to 1. Notice that the co-occurrence vectors of 'picture' and 'photo' are very similar. In fact, the semantic similarity between them Γ('picture','photo') = 0.891. On the other hand, the co-semantic similarity between the co-occurrence vectors of 'lens' and 'picture' is Γ('lens','picture') = 0.810. That means that the aspect 'photo' is semantically more similar to 'picture' than the aspect 'lens'.

String Similarity (Δ) Estimates the similarity of two aspects by using the Jaro-Winkler string similarity. Jaro-Winkler string similarity is a variant of the Jaro similarity metric that compares the characters of two strings. The higher the Jaro-Winkler similarity, the more similar the strings are. Differently from the Jaro similarity metric, the Jaro-Winkler similarity gives more importance to

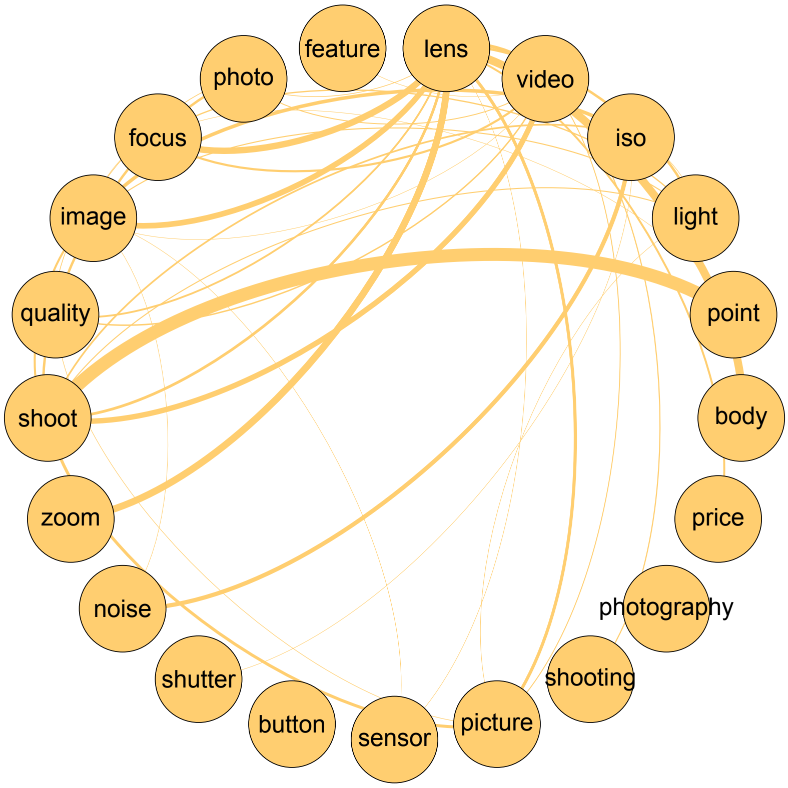


Figure 4.2: Co-occurrences between pairs of the most frequent aspects of the aspect vocabulary $\mathcal{A}_D$.

Co-occurrence frequency vectors									
	zoom	video	image	shoot	lcd	noise	shutter	screen	...
picture	105	248	140	345	71	86	116	98	...
photo	57	249	162	319	54	49	81	91	...
lens	529	353	469	333	0	104	129	0	...
Co-occurrence vectors (normalized)									
picture	0.017	0.041	0.023	0.057	0.011	0.014	0.019	0.016	...
photo	0.012	0.053	0.034	0.068	0.011	0.010	0.017	0.019	...
lens	0.039	0.026	0.034	0.024	0	0.007	0.009	0	...

Table 4.1: Co-occurrence vectors for aspects 'picture', 'photo' and 'lens', with co-occurrent frequency counts (top), and normalized co-occurrence vectors (bottom).

the left-most characters of the aspects to reward aspects with similar lemmas φ, in this work we use $\varphi = 3$. We use 0.1 as the constant scaling factor for the prefix φ. The Jaro-Winkler similarity between two aspects is defined as:

$$\Delta(a_i, a_j) = (1 - J) + (0.1 \cdot \varphi(1 - J))$$

where J is the Jaro distance between two strings. The Jaro distance J is defined as:

$$J(a_i, a_j) = \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3}(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}) & \text{otherwise} \end{cases}$$

where m is the number of *matching characters* between the strings, and t is the number of *transpositions* needed to form the same word considering only the matching characters between the two strings. Two characters from a_1 and a_2 match, if they are the same character and not farther than $\frac{max(|a_1|,|a_2|)}{2} - 1$ in their respective aspect strings. Table 4.2 shows the Jaro-Winkler similarity between aspects, using $\varphi = 4$ and 0.1 as the constant scaling factor prefix.

a_1	a_2	$\Delta(a_1, a_2)$
photo	photography	0.89
pic	picture	0.87
lens	zoom lens	0
lens	lens zoom	0.88
video	focus	0

Table 4.2: Jaro-Winkler similarity between pairs for aspects using $\varphi = 4$.

PhotoDict similarity (Λ) Estimates the similarity between two aspects by exploring the shortest path between the two terms in the *PhotoDict* taxonomy. PhotoDict is a small taxonomy of camera-related terms created in previous Chapter 3. If two aspects are close in the PhotoDict taxonomy, that means that they are related. For two aspects a_i and a_j in the aspect vocabulary, the length of the shortest path between them can be determined from one of those three cases:

- a_i and a_j are the same aspect. The path length between a_i and a_j, $len(a_i, a_j) = 0$, thus the path similarity between them is 1.
- a_i and a_j are different nodes, and a_i is related in the taxonomy with a_j. The similarity between a_i and a_j is estimated as $\frac{1}{len(a_i,a_j)+1}$.
- Either a_i and a_j do not exist in the taxonomy. The similarity between a_i and a_j is 0.

Table 4.3 shows some examples using path similarity in PhotoDict taxonomy. Notice that 'focus' and 'auto focus' are close in the PhotoDict taxonomy, as well as 'lens' with 'zoom', or 'sensor' with 'image sensor'. 'Image sensor' and 'resolution' are also related in the PhotoDict taxonomy, but the distance between the two aspects is longer than in the previous examples. Finally, the similarity between 'sensor' and 'sensor' is 1, since they are the same aspect.

a_1	a_2	$\Lambda(a_1, a_2)$
focus	auto focus	0.5
lens	zoom	0.5
image sensor	resolution	0.33
sensor	image sensor	0.5
sensor	sensor	1

Table 4.3: PhotoDict taxonomy similarity between pairs of aspects.

There is a special treatment of compound nouns in the presented clustering. Since compounds are formed by two or more words (e.g. *image quality*), we group them with the most frequent aspect among the compound forming words.

Figure 4.3 shows a small part of the resulting dendrogram for DSLR, considering only a representative small subset of uni-gram and bi-grams aspects. Since hierarchical clustering gives multiple partitions (clusterings) at different levels, we have to select one single partition to create our concept vocabulary.

The result of the hierarchical clustering is a dendrogram (or clustering tree) of aspects, where each aspect is grouped with its most similar aspect group. This way, and according to our semantic and lexical user review analysis, all similar aspects are grouped together in aspect groups, formed at different levels of the dendrogram.

Those different aspect groups have different inclusiveness and discrimination, and it is up to us to find the select the partition of basic level concepts that will form the concept vocabulary. To do so, we introduce the polarity of judgments and the concept of sentiment coherence in Section 4.4. Those techniques are then used in Section 4.5 to select the set of basic level concepts of our concept vocabulary $\mathcal{C}$.

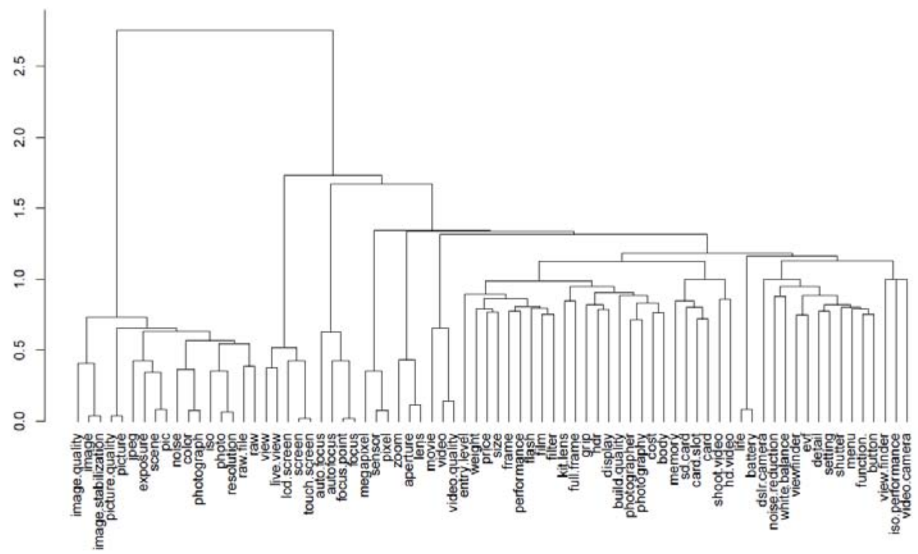


Figure 4.3: Representative portion of the DSLR resulting dendrogram.

4.4 Analyzing Judgment Polarity

In this section we present a method to identify the polarity of judgments found in user reviews. Judgments expressed by users are based on issues they encountered while using a product or a service, and have a positive, negative or neutral polarity. Here we show how, by assessing the polarity of the judgments, we can leverage information about the similarity of usage of those aspects, and thus select a good partition of basic level concepts.

Assessing the polarity of these judgments is important to determine if the experiences of the users with respect to the judgment's issue were positive or negative, and to determine whether two or more aspects are related with the same issue, as we will see later in this chapter. The analysis of judgment polarity is used in Section 4.5 to select partition of basic level concepts, and in Chapter 5 to create the bundles of arguments for each camera under consideration.

For example, people usually complain about the issue 'battery' of the camera, and sentences such "this camera has a short battery life" are commonly found in camera user reviews. We call such sentences judgments, because they express facts, from the user point of view, encountered by users while they were actually using a product or a service. Polarity can be extracted from those facts: As we can understand from previous sentence, the author of the reviews did not have a pleasant experience with the battery of the camera.

Assessing the polarity of a judgment falls within a broad task known in the literature as *sentiment analysis*. Due to its nature, sentiment analysis is widely applied to reviews and social media for a variety of applications, such as marketing, estimating sales or customer service. There exists numerous approaches to extract the subjective sentiment of words in a sentence, mostly of them by combining text analysis and natural language processing techniques. We describe some of them later in this section.

For instance, in the previous sentence "this camera has a short battery life", the adjective *short* has a negative connotation in the context of the judgment, and since it appears near to 'battery' (the issue of the judgment), we infer that this judgment shows a negative sentiment polarity over the issue 'battery'. Now consider the next judgment, found in a user review for the Nikon D7100 camera: "The camera auto-focus works well [...]". The user judgment over the issue ('auto-focus') is clearly positive, and the author expresses a positive experience about the auto-focus of the camera. The perceived polarity of the judgment could have been more intense if the sentence were to contain the adverb *very*, as in "The camera auto-focus works *very* well". Note how the adverb *very* increases the polarity strength of the sentiment word 'good'. Adverbs can also diminish the sentiment strength of a judgment, as in "The camera auto-focus works *fairly* well". Finally, suppose the judgment "The camera auto-focus does *not* work good". With the addition of the negation 'not', we completely switched the polarity of the judgment over the issue auto-focus, from positive to negative.

All those considerations must be taken into account when analyzing the sentiment or polarity of the judgments by using sentiment analysis methods. Since extracting sentiment from natural language constructs is a challenge, lexicons are often used to ascertain the polarity (positive, negative or neutral) and strength of sentiment expressed at word-level. SentiWordNet [Esuli and Sebastiani, 2006] is an example of a sentiment lexicon, where every WordNet synset has three sentiment scores associated: positivity, objectivity and negativity. Table 4.4 shows the positive (*PosScore*), objective (*ObjScore*), and negative (*NegScore*) scores given by SentiWordNet for some example words. Notice how 'well' and 'love', as expected, are words that usually carry a strong positive polarity. On the other hand, 'short' and 'awful' carry a predominant negative sentiment, being 'awful' (with a negative score of 0.875) considered more negative than 'short' by SentiWordNet (0.5). The word 'short' also has a positive score (0.375), but it is not as strong as the negative (0.5). Finally, notice how 'banana' is not considered a positive or negative word by SentiWordNet, but purely objective.

Word	PosScore	ObjScore	NegScore
well	0.75	0.25	0
love	0.5	0.5	0
short	0.375	0.125	0.5
awful	0	0.125	0.875
banana	0	1	0

Table 4.4: SentiWordNet polarity scores of example words.

By combining those three sentiment scores, we can assess the sentiment value of a word. The resulting sentiment value is normally expressed by a normalized score [-1,1], being 1 the most positive sentiment score, and -1 the most negative sentiment score.

Using those word sentiment vocabularies, and analyzing the judgments with natural language techniques, we can adjust the sentiment of a given term relative to its environment based on the way sentiment words relate. Words that intensify, relax or negate the sentiment expressed by the concept, will affect the sentiment score of the judgment. Returning to previous user judgment "The camera auto-focus works *very* well", 'well' has a positive sentiment associated that is intensified by the adverb 'very'. As such, the overall sentiment score of the judgment is clearly positive.

As we have said, those sentiment dictionaries are useful to determine the polarity of words. However, the subjectivity of words and phrases sometimes depend on their contexts and the domain of the text. For instance, the judgment "the Canon PowerShot G3 is small", has a positive polarity, because a point-and-shoot camera should be easy to carry, and as such, the smaller the size the better. However, if we say "the television is small" we might be stating a problem, because, usually, the bigger the television the better. As such, the sentiment dictionaries like SentiWordNet should be used carefully, with a previous evaluation on how correctly the sentiment dictionary applies to our domain. If necessary, the sentiment lexicon can be adapted to our domain using contextual semantics [Saif et al., 2014].

In this work, we use a sentiment analysis system named SmartSA [Muhammad et al., 2013], to assess the polarity of user judgments. The SmartSA system obtains the sentiment score of sentiment-bearing words from SentiWordNet [Esuli and Sebastiani, 2006]. The score will be modified to take into consideration negation terms and lexical valence shifters such as intensifier and diminish terms. Furthermore, SmartSA uses contextual information to improve the sentiment scores by first identifying the correct senses when extracting scores from SentiWordNet, and second by adjusting the sentiment scores on the basis of contextual analysis to modify prior polarities of documents' terms. The negative and positive strength is expressed as a value in the range [-1,1].

Let a be an aspect occurring in sentence x, we define $s(a, x) \in [-1, 1]$ as the sentiment value, estimated using the SmartSA sentiment analysis system, of aspect a in sentence x. Table 4.5 shows some example judgments whose polarity is assessed by the SmartSA system. Notice that the SmartSA system correctly identifies the positive polarity of the first judgment "The shutter speed is impressive" (with a sentiment score of 0.250). The polarity of previous sentence is increased when we add the intensifier 'very' (sentiment score of 0.375). The last judgment is correctly identified as having a negative polarity (-0.542) as well.

Using the SmartSA system, we assess the polarity of the judgments presented by people when describing their experiences with digital cameras. In Section 4.5, we introduce a method to assess the sentiment coherence between aspects by analyzing the average polarity of the sentiments of those aspects among the corpus reviews. This method is used in the same section to discover the basic level concepts contained in the aspect vocabularies created in the previous chapter.

Judgment	$s(a, x)$
"The shutter speed is impressive"	+0.250
"The shutter speed is very impressive"	+0.375
"All telephoto lens are awful"	-0.542

Table 4.5: Example judgments with SmartSA sentiment scores.

4.5 Concept Vocabulary Creation

From the concept dendrogram created in Section 4.3, here we are interested in selecting a partition that is able to describe the basic level concepts of digital cameras based on the usage of the aspects occurring in the experiences of our corpus. The groups of aspects of the dendrogram can be considered concepts at different levels of granularity, grouped together in Section 4.3 by analyzing the semantic and lexical information from user reviews. Those issues are found in user judgments expressed in user experiences. As such, by exploring the user generated reviews, we can assess the sentiment polarities related to those judgments, and extract the polarity of the experiences users had with the different granularity issues from the dendrogram. We say that a good level of granularity for an aspect group, and consequently for a partition, would be that in which the average polarity of the judgments of all aspects in an aspect group cohere with respect to the products in our corpus. For instance, let 'battery' and 'battery life' be two aspects grouped in the same issue. We consider that the judgments related to those two aspects over the reviews of a product should have a similar polarity. It is unlikely that some reviewer would judge 'battery' as a positive feature of a certain camera, and 'battery life' as a negative feature of the same camera. Since, always according to our analysis, both of them are grouped in the same issue from the dendrogram, the polarities of the judgments related to those two aspects should cohere.

In this section, we leverage the aspect sentiment information from those judgments related to the dendrogram issues, to define a sentiment coherence measure between pairs of aspects of an aspect group. Using this measure, we estimate the sentiment coherence of a partition by aggregating the sentiment coherence values of all aspect groups in that partition, and finally select the partition with the highest aggregated sentiment coherence to form the concept vocabulary $\mathcal{C}$.

Before continuing, we introduce some notation. Let $\mathcal{G}$ be the set of aspect groups that form the hierarchical clustering dendrogram H, created in previous Section 4.3. $\mathcal{K} = \{G_1 \ldots G_n\}$ is the set of possible partitions, formed by aspect groups $G_i \in \mathcal{G}$, from the dendrogram H such that each partition $K_i \in \mathcal{K}$ is pairwise disjoint $G_i \cap G_j = \emptyset$ for $i \neq j$, and exhaustive $\bigcup_{i=1 \ldots n} G_i = \mathcal{A}$. Section 4.5.1 describes a method to assess the sentiment coherence between aspects and aspect groups, and later, in Section 4.5.2 we select the best partition to create the concept vocabulary.

4.5.1 Aspect Sentiment Coherence

We have seen how to assess the polarity of judgments present in user reviews in the previous Section 4.4. Now, let us consider $Occ(p, a)$, all judgments concerning aspect a in reviews about a product p. Given two aspects a_1 and a_2, we are interested in estimating the degree of coherence between their sets of judgments $Occ(p, a_1)$ and $Occ(p, a_2)$, related with a product p. By averaging the polarity of the judgments in $Occ(p, a_1)$ and $Occ(p, a_2)$, we can compare the polarity coherence between the experiences of people with respect to aspects a_1 and a_2 of product p. We repeat this process for all products in the corpus, with the intuition that two aspects will have a high degree of coherence if their polarity of judgments are highly correlated over a set of products $\forall p \in \mathcal{P}$.

In our approach, judgment sentiment plays an important role in the creation of BLCs. We say that the average of the polarity judgments related to the aspects grouped in the same BLC should correlate when considering the same product. Consequently, since all aspects grouped in a given BLC are used in similar contexts representing the same concept, the polarity of the judgments related to the aspects assembled in the same BLC should be similar over the set of products of the corpus.

We do so in the following way. Let $p \in \mathcal{P}$ be a product, $a \in \mathcal{A}$ an aspect and $S_{av}(p, a)$ the average sentiment, of the set of sentences from the reviews of product p related with aspect a:

$$S_{av}(p, a) = \frac{1}{M} \sum_{x \in Occ(p,a)} s(x, a)$$

where x is a sentence, $Occ(p, a)$ is the set of sentences from the reviews of product p in which aspect a occurs, and $M = |Occ(p, a)|$. Remember that $s(a, x)$ assesses the sentiment of aspect a in sentence x using the SmartSA sentiment analysis system. $S_{av}(p, a)$ always returns a sentiment value between $[-1, 1]$.

For each aspect $a \in \mathcal{A}$, we create a vector $D(a)$ formed by the normalized sentiment averages of aspect a over the set of products p in the corpus:

$$D(a) = (S'_{av}(p_i, a))_{i \in 1 \dots |\mathcal{P}|}$$

where, $S'_{av}(p_i, a)$ is in [0,1] and is estimated over the set of products in $\mathcal{P}$. By comparing two aspect vectors $D(a_i)$ and $D(a_j)$, we can assess the polarity coherence between two aspects a_i and a_j. The aspect vector D is the *polarity profile* of aspect a: it contains the average polarity of aspect a, considering all user judgments from user experiences, over the set of products in the corpus.

Figure 4.4 shows the judgment average polarity distribution of three aspects 'battery', 'battery life' and 'lens', over a set of products. According to our analysis, aspects 'battery' and 'battery life' belong to the same BLC, whilst 'lens' does not. As can be observed, the judgment average polarity distribution of the two aspects grouped in the same BLC, 'battery' and 'battery life', is similar over the set of selected products. For instance, the average polarity of the judgments related to 'battery' aspect is slightly negative (-0.31 polarity

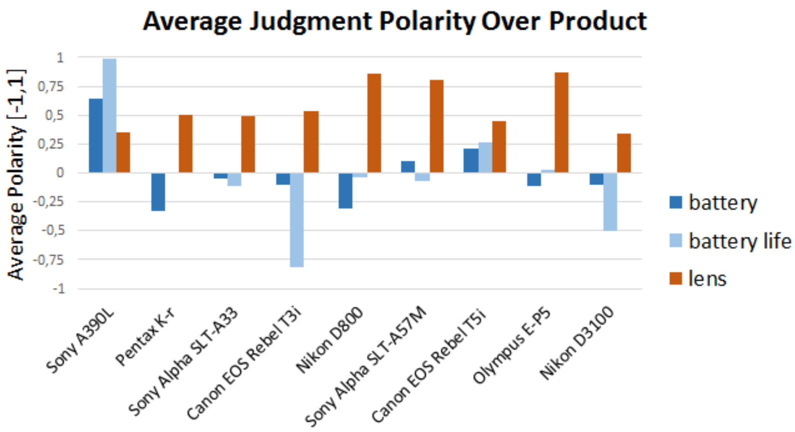


Figure 4.4: Judgment polarity distribution over a set of 9 products for aspects 'battery', 'battery life' and 'lens'.

score) for the Nikon D800. Similarly, the average polarity of the aspect 'battery' over the set of reviews of Nikon D800 is also negative (-0.04). That means that the judgments related to those two aspects found in the reviews of the Nikon D800 and grouped into the same BLC are negative on average. On the other hand, the average polarity of the judgments related to the aspect 'lens' is very positive (+0.86) for the same camera. People is talking about a different concept when using the aspect 'lens' with respect to the other two aspects, but they talk about the same concept when using 'battery' and 'battery life'.

Next, we define the polarity similarity measure between two aspect polarity profiles D_i for aspect a_i, and D_j for aspect a_j. To do so, we compare the two vectors using the cosine similarity. Since the vector values are normalized, the cosine similarity returns a similarity value between $[0, 1]$, 1 meaning maximum orientation similarity, and 0 no similarity:

$$Sim(D_i, D_j) = cos(D_i, D_j) \tag{4.1}$$

Table 4.6 shows the similarity Sim between the polarity profiles of aspects 'battery', 'battery life' and 'lens' from previous example. Notice that the similarity between the polarity profiles of 'battery' and 'battery life' is high (0.896), while it is not between 'battery' and 'lens' (0.423). The polarity similarity measure between two aspect polarity profiles is used in next section to define a coherence measure for aspect groups, and then extended to estimate the polarity coherence of a partition.

4.5.2 Concept Vocabulary

Using the aspect polarity profiles and the similarity Sim defined in the previous section, here we define a sentiment coherence measure for aspect groups, by

a_i	a_j	$Sim(D_i, D_j)$
battery	battery life	0.896
battery	lens	0.423
battery life	lens	0.594

Table 4.6: Similarity Sim between polarity profiles.

comparing the polarity profiles of all aspects grouped together. Furthermore, by aggregating the sentiment coherence of the aspect groups that form a partition K, we obtain a partition score. The partition with highest partition score is selected, and its aspects groups form the concept vocabulary $\mathcal{C}$.

For instance, let *picture*, *photo* and *image* be three aspects in an aspect group. If those three aspects are used by people to refer to the same concept ("picture obtained by my digital camera"), then the sentiment values of those aspects with respect to the reviews of each product should have a high coherence degree.

We define the average sentiment similarity *IS* of a group of aspects G as the average Sim among the polarity profiles of all pairs of aspects in G:

$$IS(G) = \frac{1}{|G| \cdot (|G| - 1)} \sum_{i=1}^{|G|} \sum_{j=1, j \neq i}^{|G|} Sim(D_i, D_j)$$

where D_i and D_j are the polarity profiles of aspect a_i and a_j respectively, and $Sim(D_i, D_j)$ computes the cosine similarity between aspect polarity profiles as introduced in Section 4.5.1. Aspect groups formed by only one aspect will have an IS value of 0.

Next, the *Partition Ranking score* $R(K)$ of a partition K is estimated as follows:

$$R(K) = \frac{1}{|K|} \sum_{i=1}^{|K|} IS(G_i)$$

where $|K|$ is the size of the set of aspect groups that form partition K. The higher *R(K)*, the better the partition K.

Algorithm 5 presents the method to select the partition of basic level concepts from all possible dendrogram partitions $\mathcal{K}$. Notice that the input variables max_c and min_c correspond to the maximum and minimum accepted size for a partition. This way we can adapt the partition selection strategy to only estimate the partition score of partitions with a certain granularity.

The partition K with the highest partition ranking score $R(K)$ is selected. Now, let us consider $Occ(\mathcal{P}, G)$, all judgments concerning the aspects in the aspect group G in the reviews about all products $\mathcal{P}$ in a given corpus. All aspect groups $G_i \in K$ with $|Occ(\mathcal{P}, G_i)| \geq 100$, form the concept vocabulary $\mathcal{C}$. This selection is performed to ignore small aspect groups with little similarity to other aspect groups in the dendrogram, as those are usually formed by spurious

Algorithm 5: Concept vocabulary partition selection

Input: dendrogram Partitions $\mathcal{K}$, **int** max_s, min_s
Output: Concept Vocabulary $\mathcal{C}$;
// Initialize the partition-score list
$Ks\langle K, score\rangle := \emptyset$;
// Select all partitions $K_i \in \mathcal{K}$ of size $min_c \leq |K| \leq max_c$
for $K_i \leftarrow \mathcal{K}$ **where** $min_c \leq |K_i| \leq max_c$ **do**
  // Estimate the partition ranking score $R(K_i)$
  $score_i := R(K_i)$;
  // Add the partition-score pair to the list
  $Ks.add(\langle K_i, score_i\rangle)$;
// Return the partition K with highest score
return $Ks\langle R, score\rangle.selectHigestScore$;

Concept Name	Aspects in Concept
Wireless	wireless, wifi
Card	card slot, cf card, sd card, card, card memory
Battery	life battery, battery life, ion battery, charger, battery, life
Button	button, shutter button, speed shutter, shutter speed, shutter, button layout, release shutter, button shutter

Table 4.7: Three of the basic level concepts in $\mathcal{C}_D$ and their aspects.

content that overcame the aspect filtering from previous chapter. In Section 4.6 we give some examples about the aspect groups filtered.

Each $c \in \mathcal{C}$ is identified with the name of the most frequent aspect grouped within the concept. Table 4.7 presents a small subset of the concepts obtained in this section when considering the the aspect vocabulary $\mathcal{A}_D$ extracted from the DSLR corpus.

In Section 4.6 we create three concept vocabularies, one for each Amazon camera type, using the aspects from $\mathcal{A}_D$, $\mathcal{A}_C$, and $\mathcal{A}_P$ respectively. We compare the selected partitions of the resulting hierarchical clustering dendrogram for each camera type, and the aspects that form the concepts in the concept vocabularies. The concept vocabularies will be used in Chapter 5 to create the bundles or arguments that will allow the reuse of people's experiences.

4.6 Analysis of Concept Vocabularies

In this section we create the concept vocabularies for the three camera types: DSLR ($\mathcal{C}_D$), Compact ($\mathcal{C}_C$), and Point & Shoot ($\mathcal{C}_P$). We do so by following the methodology presented in this chapter. Using the aspect vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$ and $\mathcal{A}_P$ discovered in Chapter 3, we create three hierarchical clustering dendrograms (one for each camera type). Afterwards, we select the partition

Concept Vocabularies	$\lvert\mathcal{C}\rvert$	$\sum_{i...\lvert\mathcal{C}\rvert} \lvert c_i\rvert$
$\mathcal{C}_D$	41	225
$\mathcal{C}_C$	39	197
$\mathcal{C}_P$	39	179

Table 4.8: Concept vocabularies for DSLR, Compact and Point & Shoot categories.

with with highest $R(K)$ from each dendrogram, removing those groups $G_i \in K$ whose $|Occ(\mathcal{P}, G_i)| \leq 100$, and only considering partitions with 35 to 45 groups, a reasonable concept vocabulary size for our purposes. From the selected partitions, one for each camera category, we create the concept vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$ and $\mathcal{C}_P$, that determine the set of concepts that model the collection of issues that are used in the judgments of the reviews from the tree corpus of cameras. The concept vocabularies for the DSLR ($\mathcal{C}_D$), Compact ($\mathcal{C}_C$), and Point & Shoot (C_P) cameras, are presented in Appendix D.

Table 4.8 shows the quantity of concepts and aspects that form the three selected concept vocabularies $\mathcal{C}_D$ (formed by 41 concepts), $\mathcal{C}_C$ (formed by 39 concepts), and $\mathcal{C}_P$ (formed by 40 concepts). Notice that the quantity of aspects that form the concept vocabularies of the three camera categories is smaller than their respective aspect vocabularies. That is because, before creating the concept vocabulary, we discarded those aspect groups from the selected partition that were found less than 100 times as review judgments. We interpret the low occurrence of the elements within those concepts as that they are not deemed important by people when describing their experiences with digital cameras. Some of the discarded aspects for DSLR were 'burst mode', 'aspect ratio' and 'chromatic aberration' together with 'manual' and 'auto'. For Compact, some of the discarded aspects were 'bargain', 'shoot', and 'histogram'. On the other hand, for Point & Shoot, we discarded 'preset', 'speed', and 'manual control', among others.

Figure 4.5 shows the most frequent aspects that form the top 10 concepts with more occurrences in DSLR concept vocabulary $\mathcal{C}_D$. Each concept has a different color assigned: Concept 'picture' (blue), 'iso' (yellow), 'focus' (pink), 'viewfinder' (light pink), 'screen' (teal), 'sensor' (brown), 'resolution' (dark green), 'video' (green), 'battery' (light blue) and 'lens' (orange). The closer the elements in the graph, the more similar they are considering the similarity equations presented in Section 4.3. By observing the figure, we obtain an idea of how the aspects are used in the reviews of the DSLR corpus. For instance, we observe concept 'iso', according to our analysis, is very similar to 'picture'. Furthermore, 'iso' occurs often in the reviews, and is selected to form a separate concept from 'picture'. Another interesting aspect is 'image sensor', found between the concepts 'picture' (blue) and 'sensor' (brown). In the figure, 'image sensor' appears closer to concept 'sensor' than concept 'picture', however our analysis considered it should be clustered in concept 'picture'. We also observe that 'focus', 'lcd' and 'screen' are similar concepts, as they appear close in the figure. On the other hand 'lens',

Top 10 concepts $\mathcal{C}_D$	Top 10 concepts $\mathcal{C}_C$	Top 10 concepts $\mathcal{C}_P$
picture	image	picture
lens	lens	zoom
video	price	video
screen	video	price
focus	focus	battery
button	zoom	screen
iso	button	price
photography	screen	focus
price	battery	button
battery	sensor	flash

Table 4.9: Top 10 most frequent concepts of $\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$.

'video' and 'battery' form very well defined concepts.

Now, we will compare some concepts of the three concept vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$ and $\mathcal{C}_P$. Do people talk about different issues depending on the camera type? Does the usage of the same aspects vary between the reviews of different camera types, and as a consequence, the resulting concepts vary between categories? Table 4.9 shows the top 10 most frequent concepts in $\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$. The name of a concept corresponds to the most frequent aspect grouped in that concept. We observe how the most important issues expressed by users in their experiences are similar for the three camera categories. For instance, concept 'picture' (named 'image' in $\mathcal{C}_C$) is found between the top 5 most frequent concepts of the three camera types, together with 'video' and 'lens' (named 'zoom' in $\mathcal{C}_P$). Also notice that 'price' is a common concept for the three camera types, ranked 3rd and 4th for Compact and Point & Shoot cameras respectively, but not deemed that important for DSLR (ranked 9th).

Let us focus on the aspects that form the concept 'button' of the three vocabularies in Table 4.10. In this table, we considered 'shutter button' and 'button shutter' as the same aspect. As such, the table only presents the aspect 'shutter button'. Similarly, we considered 'shutter speed' and 'speed shutter' as the same aspect. Notice that, the description of concept 'button' has some common aspects between $\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$, however the quantity of aspects grouped in $\mathcal{C}_P$ is smaller compared to $\mathcal{C}_D$ and $\mathcal{C}_C$. Furthermore, it is interesting to point that only the description of 'button' in $\mathcal{C}_C$ includes aspects such as 'lag' or 'shutter lag', not present in concept 'button' of $\mathcal{C}_D$ nor $\mathcal{C}_P$. This happens because in both $\mathcal{C}_D$ and $\mathcal{C}_P$, 'shutter lag' (or 'lag') is considered a concept by itself.

We need to define a measure to assess the similarity between concepts from different concept vocabularies, and we do so using the Jaccard similarity introduced in Equation 3.2 of previous Chapter 3. Next Tables 4.11, 4.12, 4.13 and 4.14 show the Jaccard similarity between four of the most frequent concepts of the three concept vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$ and $\mathcal{C}_P$: 'picture', 'video', 'button' and 'price'. We compare if the aspects that form those concepts are similar or vary

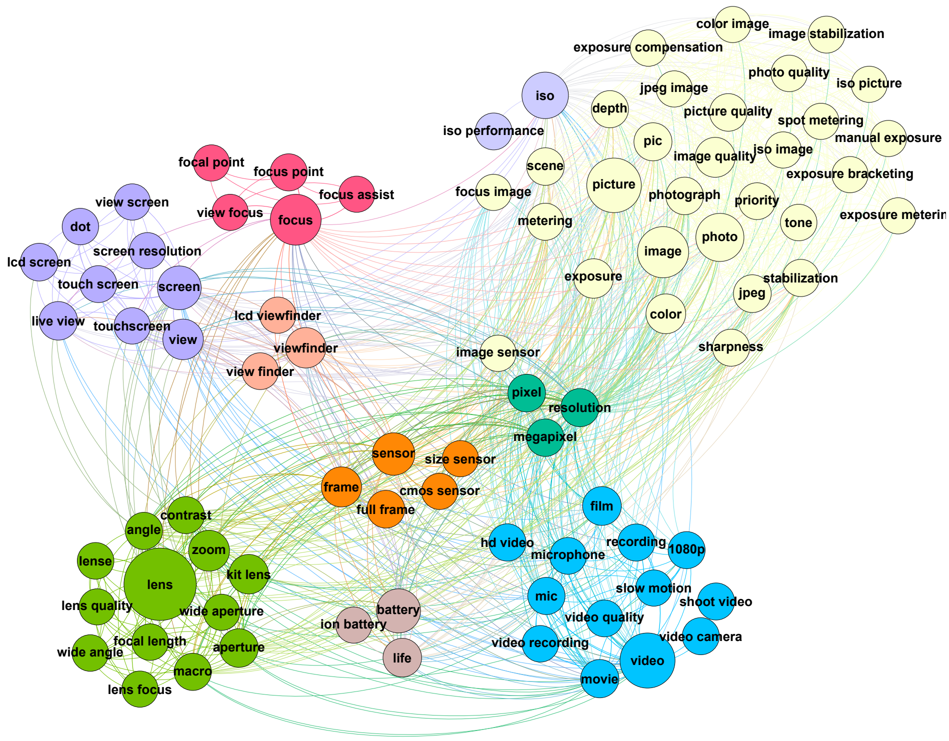


Figure 4.5: Top 10 most important concepts in $\mathcal{C}_D$ and its aspects, related by edges.

	$\mathcal{C}_D$	$\mathcal{C}_C$	$\mathcal{C}_P$
button	button shutter button speed shutter button layout release shutter shutter	button shutter button speed shutter menu button lag shutter lag shutter	button shutter button shutter

Table 4.10: Aspects grouped in 'button' concepts for $\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$.

$c(picture) \in$	$\mathcal{C}_D$	$\mathcal{C}_C$	$\mathcal{C}_P$
$\mathcal{C}_D$	1	0.6842	0.5833
$\mathcal{C}_C$		1	0.6666
$\mathcal{C}_P$			1

Table 4.11: Jaccard similarity between concept 'picture' of the three concept vocabularies.

$c(video) \in$	$\mathcal{C}_D$	$\mathcal{C}_C$	$\mathcal{C}_P$
$\mathcal{C}_D$	1	0.8500	0.6363
$\mathcal{C}_C$		1	0.7500
$\mathcal{C}_P$			1

Table 4.12: Jaccard similarity between concept 'video' of the three concept vocabularies.

$c(button) \in$	$\mathcal{C}_D$	$\mathcal{C}_C$	$\mathcal{C}_P$
$\mathcal{C}_D$	1	0.4444	0.5000
$\mathcal{C}_C$		1	0.4285
$\mathcal{C}_P$			1

Table 4.13: Jaccard similarity between concept 'button' of the three concept vocabularies.

$c(price) \in$	$\mathcal{C}_D$	$\mathcal{C}_C$	$\mathcal{C}_P$
$\mathcal{C}_D$	1	0.0800	0.6666
$\mathcal{C}_C$		1	0.0833
$\mathcal{C}_P$			1

Table 4.14: Jaccard similarity between concept 'price' of the three concept vocabularies.

between camera types. It is interesting to notice the similarity between the selected concepts of $\mathcal{C}_D$ and $\mathcal{C}_C$ (except for concept 'price'): 0.68 between concept picture of $\mathcal{C}_D$ and $\mathcal{C}_C$, or 0.85 between concept 'video'. Similarly as what we have seen in previous chapter, Point & Shoot also has a slightly different concept vocabulary compared with DSLR and Compact. Finally, notice the high Jaccard similarity obtained when comparing the concept 'video' between the three concept vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$ and $\mathcal{C}_P$. Concept 'video' is one of the largest concepts in all three camera concept vocabularies, and the Jaccard similarities obtained tell that it is also consistent in all three concept vocabularies.

At this point, we are interested in comparing the concept vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$ and $\mathcal{C}_P$. Equation 4.2 presents a measure that estimates the similarity between two concept vocabularies $\mathcal{C}_i$ and $\mathcal{C}_j$. Notice that the measure is not symmetric, and aggregates the best Jaccard similarity obtained between the elements of two concept vocabularies.

$$SimC(\mathcal{C}_i, \mathcal{C}_j) = \frac{1}{|\mathcal{C}_i|} \sum_{m=0}^{|\mathcal{C}_i|} \max_{\forall c \in \mathcal{C}_j} \left(J(c_m, c)\right) \tag{4.2}$$

	$\mathcal{C}_D$	$\mathcal{C}_C$	$\mathcal{C}_P$
$\mathcal{C}_D$	1	0.3885	0.3993
$\mathcal{C}_C$	0.4400	1	0.3304
$\mathcal{C}_P$	0.4054	0.2798	1

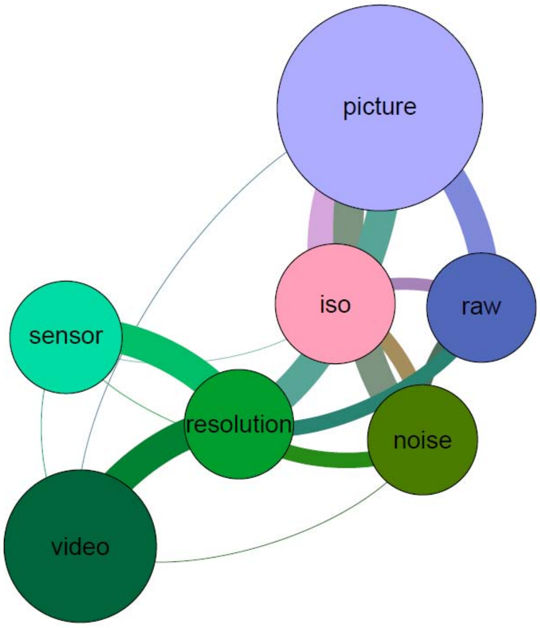
Table 4.15: Concept vocabularies for DSLR $\mathcal{C}_D$, Compact $\mathcal{C}_C$ and Point & Shoot $\mathcal{C}_P$ categories.

where concept $c_m \in \mathcal{C}_i$, concept $c \in \mathcal{C}_j$, $J(c_m, c)$ is the Jaccard similarity between concepts c_m and c, and max returns the maximum Jaccard similarity considering concept c_m and all concepts $\forall c \in \mathcal{C}_j$. Table 4.15 shows the results of estimating the similarity between all three concept vocabularies. We observe that all three concept vocabularies are noticeably different from each other. Since the concept vocabularies represent the important issues found in user experiences, we can say that people are interested in different issues depending on the camera type. The most similar concept vocabularies are $\mathcal{C}_D$ and $\mathcal{C}_C$, with an average $SimC$ of 0.4142, and the less similar vocabularies are $\mathcal{C}_C$ and $\mathcal{C}_P$ ($SimC$ of 0.3051). This results are similar to the ones obtained when comparing aspect vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$ and $\mathcal{A}_P$, where $\mathcal{A}_D$ and $\mathcal{A}_C$ had a higher similarity than with $\mathcal{A}_P$.

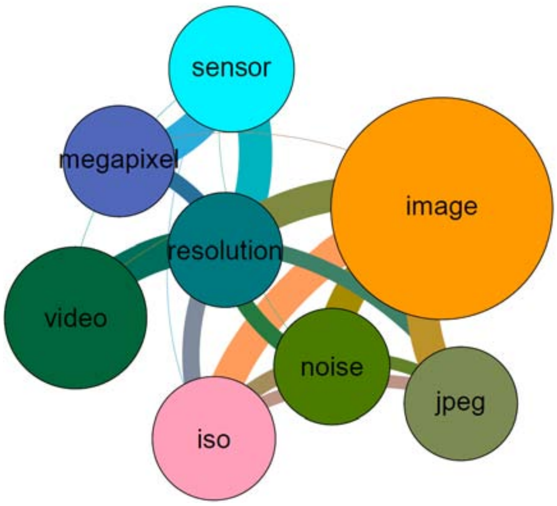
We further analyze these vocabulary differences by comparing the most similar concepts related to concept picture of the three concept vocabularies in Figure 4.6. We observe how DSLR and Compact differentiate other picture related concepts such as ‘iso’, ‘noise’ or ‘resolution’ from ‘picture’. Our analysis showed that those concepts were popular issues in the reviews of DSLR and Compact, and as such were selected as concepts. On the other hand, Point & Shoot reviews do not deem ‘iso’, ‘noise’ or ‘resolution’ that important, and as a consequence they are grouped in concept ‘picture’.

Table 4.16 shows the concepts where the most frequent aspects where clustered for the three concept vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$ and $\mathcal{C}_P$. Notice that the most frequent aspects are normally grouped in similar concepts without depending on the camera type concept vocabulary. However, it is interesting to point out that aspect ‘lens’ is clustered in concept ‘lens’ in $\mathcal{C}_D$ and $\mathcal{C}_C$, however it is clustered in concept ‘zoom’ for $\mathcal{C}_P$. That happens because in $\mathcal{C}_P$, ‘zoom’ is more frequent than ‘lens’, and as such the concept receives the name ‘zoom’ in $\mathcal{C}_P$. However, it is clear that Point & Shoot users talk more about issues related to ‘zoom’ than ‘lens’, contrary as we can see in DSLR and Compact camera types. We can also observe that ‘iso’ is a concept by itself in $\mathcal{C}_D$ and $\mathcal{C}_C$, however it is clustered inside concept ‘picture’ in $\mathcal{C}_P$. Finally, we observe a strange fact that is ‘photography’ being clustered in concept ‘price’ for Compact concept vocabulary $\mathcal{C}_C$.

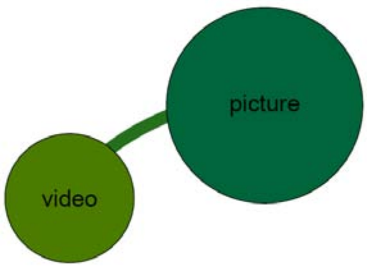
The evaluation of the concept vocabularies created for the three camera types cannot be performed in this chapter, since the concepts are not correct or wrong by themselves but need be useful for the reuse of experiences. In the next chapter, we analyze the bundles of arguments that define the products of each camera type using the concept vocabularies created here. There, we will be able to evaluate if the concept vocabularies correctly represent the most important issues from the



a) DSLR $\mathcal{C}_D$.



b) Compact $\mathcal{C}_C$.



c) Point & Shoot $\mathcal{C}_P$.

Figure 4.6: Concepts similar to 'picture' for the three concept vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$ i $\mathcal{C}_P$

	Clustered in concept		
aspect	c_D	c_C	c_P
lens	lens	lens	zoom
picture	picture	image	picture
video	video	video	video
focus	focus	focus	focus
photo	picture	image	picture
image	picture	image	picture
body	body	body	body
iso	iso	iso	picture
price	price	price	price
battery	battery	battery	battery
setting	setting	setting	setting
photography	photography	price	photographer
flash	flash	flash	flash
button	button	button	button
shutter	button	button	button
sensor	sensor	sensor	sensor
screen	screen	screen	screen

Table 4.16: Concepts were aspects are clustered in $\mathcal{C}_D$, $\mathcal{C}_C$ and $\mathcal{C}_P$.

user point of view found in their product experiences, and facilitate the reuse of their experiences.

4.7 Conclusions

We present a methodology to identify the judgment issues found in reviews of user experiences about products. We do so in two steps: First, we group the semantically and syntactically similar aspects from an aspect vocabulary using a hierarchical clustering to create a dendrogram, where each cluster in the dendrogram represents a judgment issue found in the reviews of products with a certain level of granularity. Second, we select a partition of issues from this dendrogram, to create the concept vocabulary. The partition is selected using the cognitive linguistics notion of basic level concepts, and finding a good level of concept granularity, in terms of providing good cognitive economy, by estimating the sentiment coherence between aspects grouped in the same cluster. Then, we select the partition with higher sentiment coherence from the possible partitions of the dendrogram. The resulting selected partition forms the concept vocabulary, which models the important issues used by people when expressing their experiences in product reviews.

Using these techniques and the aspect vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$ and $\mathcal{A}_P$ created in previous chapter, we create three concept vocabularies, one for each camera type $\mathcal{C}_D$, $\mathcal{C}_C$ and $\mathcal{C}_P$ (see Appendix D). We observed that the methods presented in Section 4.3 to create the hierarchical clustering dendrogram are effective to cluster the most similar aspects in issues, in the sense that similar aspects are grouped in similar issues. Furthermore, we consider that using the sentiment coherence/incoherence of the aspect groups proved to be a good technique to identify a good level of granularity for the partition, since it leverages the usage of the sentiment polarity with respect to the aspects of a set of reviews to decide whether two or more aspects should be grouped together in the same issue. In both cases, to create the dendrogram and to select a partition, we could not implement an evaluation because aspects are grouped depending on their usage in our corpus of user reviews, and as such, the resulting concepts may be different from the usual concepts found in a more general corpus and clustering approach. We will evaluate the concept vocabularies in next chapter, by observing if they are useful to facilitate the reuse of people's experiences. However, we can see that the resulting concepts are coherent with what could be expected from a general camera category classification: all aspects related to 'picture' are included in similar concepts named 'picture', concept 'zoom' is similar to concept 'lens', and the same for the rest of selected concepts (see Figure 4.5).

By analyzing the resulting concept vocabularies, we noticed that the usage of aspects and the important issues over the reviews of the three corpus is different depending on each camera type. The highest similarity between two concept vocabularies $SimC$ (as described in Equation 4.2) is obtained between the concept vocabularies $\mathcal{C}_D$ and $\mathcal{C}_C$ (0.4400), while the lower $SimC$ is obtained between $\mathcal{C}_C$ and $\mathcal{C}_P$ (0.2798). These values show that the concept vocabularies of the three camera types are very different.

That means that the important issues expressed in the reviews of a user that bought a DSLR are different than the issues considered important in the reviews of another user that bought a Compact or a Point & Shoot camera. For instance, in $\mathcal{C}_P$ we have concept 'waterproof'. This concept is not present neither in $\mathcal{C}_D$ nor $\mathcal{C}_C$, because our algorithm did not consider 'waterproof' and important issues when analyzing the user reviews of our DSLR and Compact corpus. On the other hand, Figure 4.6 shows the difference between concepts related with concept 'picture', in all three concept vocabularies. Our approach considered that for $\mathcal{C}_D$ and $\mathcal{C}_C$, the concept 'picture' was too general to express the important issues present in the reviews of DSLR and Compact. As such, we observe some other 'picture' related concepts, such as concept 'iso', 'noise' or 'resolution' in $\mathcal{C}_D$ and $\mathcal{C}_C$. On the other hand, all those concepts are grouped in concept 'picture' of $\mathcal{C}_D$. Another interesting difference are concepts 'gps', 'hdr', 'usb' and 'wireless'. Our system did not find strong sentiment coherence between those aspects, and decided to select them as separate concepts of the concept vocabulary. On the other hand, the polarity of other aspects such as 'wireless' or 'wifi' cohered among the judgments of the products, and the system grouped them into the same concept in $\mathcal{C}_D$.

From this point onwards, we use those three concept vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$ and $\mathcal{C}_P$ to characterize the set of products of our corpus for each camera type, and create the argument bundles. Those argument bundles will allow us to reuse people's experiences in Chapter 5.

Chapter 5

Bundle of Arguments

5.1 Introduction

In this chapter we are interested in representing the experiential information found in user reviews as some knowledge structures that are amenable to reuse. We leverage the experiential knowledge from user's individual judgments about some product domain (digital cameras) in order to create new knowledge structures; these knowledge structures should be in a form that facilitates other users to make more informed decisions when buying, consulting or searching information about those products.

To do so, we use the three concept vocabularies created in previous chapter, that define the important issues found in the reviews that express the experiences of users in the three corpora. Then, for each review of each product in a corpus K, we identify the specific polarity value in $[-1, 1]$ of every sentence related to one of the concepts in the corpus vocabulary $\mathcal{C}$; we call a *judgment* the pairing (concept, polarity) — the concept occurring in a sentence in K with the corresponding polarity value.

By aggregating the polarity values of judgments related to a specific concept $C \in \mathcal{C}$ in the set of reviews $Rev(p)$ of a specific product p, we can assess whether the user experiences with respect to that concept C are positive or negative regarding that product p. Repeating the same process over the set of concepts of $\mathcal{C}$ over $Rev(p)$, we can obtain a characterization of that product p based on the individual experiences of people expressed in the terms of the concepts in $\mathcal{C}$.

We call an *argument* the result of aggregating the polarity values of user judgments over a specific concept C over a product p; this way we determine an overall polarity value for C on p. Arguments can be understood as reasons in favor (or against) buying a product p. If the overall polarity of the judgments related to a concept over the reviews $Rev(p)$ is positive, we consider there is a *pro argument* for the product p regarding that concept C, in the sense that C plays a positive role in the decision about why choosing this product over another one. On the other hand, if the overall polarity of the judgments related to concept is negative, we consider there is a *con argument* for the product p regarding that concept C. Finally, if the overall polarity of judgments of a concept over the reviews $Rev(p)$ is not clearly positive or negative, we consider that there is no conclusive argument in favor or against for that product in relation to the concept at hand, and we say it is a *moot argument* on the product.

By considering the pro, con and moot arguments of a product p over every concept in a vocabulary $\mathcal{C}$, we obtain a characterization about what people like or dislike about p. We call the *bundle of arguments* of a product the collection of pro, con, and moot arguments with respect to all concepts in a concept vocabulary $\mathcal{C}$. That is to say, a collection of pro and con reasons that have been generated by extracting and analyzing the practical knowledge of user's experiences from the reviews of a product $Rev(p)$; moreover, the moot arguments represent the concepts whose judgments are inconclusive and therefore have no influence on selecting that product over another one.

Argument bundles allow us to reuse the experiential knowledge of past users for new users (that may have different individual preferences from the past users). The reason is that a bundle of arguments abstracts the knowledge from the experiences of multiple users —whose individual preferences also vary— and they embody the overlapping consensus over specific concepts of the corresponding product. Furthermore, to support this reuse, we will introduce later the notions of *user query* and *query satisfaction by a bundle of arguments*. A user query expresses the individual preferences of a user. The degree of satisfaction of a query by a bundle of arguments is modeled using fuzzy logic and estimates to which extent a product satisfies the preferences expressed in a query.

To test the approach, we first create the bundles of arguments for all the cameras belonging to the three corpora, Digital Single Lens Reflex cameras (K_D), Compact cameras (K_C), and Point & Shoot cameras (K_P). Since we are interested in observing the differences between the bundles of arguments of the three different camera types, we compare the sets of pros and cons of the DSLR, Compact and Point & Shoot argument bundles. Then, to evaluate the quality of the product bundles, we compare the bundles of arguments of the three camera types with the camera descriptions of Dpreview.com, a renowned website specialized in digital cameras. Specifically, we compare if the lists of pros and cons created by Dpreview experts agree with the pro and con arguments of the bundles of products. Finally, we present a ranking measure for bundles. We use this measure to rank the three sets of products (DSLR, Compact, and Point & shoot) based on their respective bundles of arguments. Afterward, we compare the resulting rankings of bundles with the product rankings of Dpreview and Amazon, analyzing if there is any correlation between the different rankings.

This chapter is organized as follows: Section 5.2 introduces the notion of argument, and presents three different argument types based on different polarity aggregation methods. Then, Section 5.3 presents the notion of bundle of arguments for a product. Next, Section 5.4 introduces the user query and the semantics of query satisfaction by bundles based on fuzzy logic. Bundles of arguments are evaluated against Dpreview experts in Section 5.5, and Section 5.6

summarizes the approach and the contributions of this chapter.

5.2 From Individual Judgments to Arguments

In this section we analyze judgments occurring in the user reviews of digital cameras in order to create arguments. Arguments may be understood ae reasons about why to buy (or not) a camera.

Arguments are created by aggregating the polarities of user judgments over each specific concept $C \in \mathcal{C}$ over the set of reviews $Rev(p)$ corresponding to a product p. With the aggregation of user judgments regarding one concept C, we are able to exploit the experiential knowledge and create an argument about C. For instance, if the vast majority of users complain about the 'grip' of a certain camera, then we have evidence to believe that the camera's grip is bad. Similarly, if everyone praises a camera's 'flash' light, then we know the manufacturers did a good job with that specific part of the camera.

Let $Occ(p, C)$ be the set of sentences in which the judgments related to concept C occurring in the set of reviews of product $Rev(p)$. That is to say, $Occ(p, C)$ denotes all sentences from $Rev(p)$ where a user judgment about any of the aspects clustered in concept C occurs. The polarity of each judgment is obtained by applying the SmartSA sentiment analysis system presented in Section 4.4. We define the vector $V(p, C)$ as the collection of sentiment values of the judgments in $Occ(p, C)$:

$$V(p, C) = \{s(C, x) | x \in Occ(p, C)\} \tag{5.1}$$

where $s(C, x)$ returns the sentiment value in $[-1, 1]$ of sentence x using the SmartSA sentiment analysis system, and $|V(p, C)| = |Occ(p, C)|$.

By aggregating the polarity values of the vector of user judgments $V(p, C)$, we obtain an overall argument sentiment value $s = AGG(V(p, C))$, where AGG is some aggregation measure (we will presently discuss three of them). In other words, the argument sentiment value s is the aggregated sentiment polarity about a concept C based on the particular experiences of users in $Rev(p)$.

Thus, we define an argument as a tuple formed by a product $p \in \mathcal{P}$, a concept $C \in \mathcal{C}$, and an aggregated sentiment value s:

$$Arg = \langle p, C, s \rangle$$

We will use the dot notation $Arg.p$, $Arg.C$, and $Arg.s$ to refer to the three components of an argument Arg.

If the aggregation of the judgment sentiment values in V is clearly positive, we will create a pro argument. That means that according to our analysis, user judgments with respect to that specific concept of a product are overall positive. On the other hand, if the the aggregation of the sentiment values of V is clearly negative, we will create a con argument. Finally, if according to our analysis the aggregation of the sentiment values of V is neither clearly positive or negative, we will create a moot argument. Moot arguments are arguments of little to no practical interest in our characterization of the product, since they are considered inconclusive arguments —the reason being that there was not enough positive or negative evidence in user experiences to consider them pro or con arguments.

Therefore, depending on the sentiment of the argument $Arg.s$, we define pro arguments (Arg^{+}), con arguments (Arg^{-}) and moot arguments ($Arg^{\emptyset}$), as follows:

$$Arg = \begin{cases} Arg^{+} & \text{if } Arg.s > \delta \ \} \\ Arg^{-} & \text{if } Arg.s < -\delta \ \} \\ Arg^{\emptyset} & \text{if } -\delta \leq \ Arg.s \leq \delta\} \end{cases} \tag{5.2}$$

where $\delta \in [0, 1]$ is a parameter that defines what we consider "clearly" positive or negative, and thus a pro, or a con argument —and (also implicitely) a moot argument. Higher values of δ result in more moot arguments, since the criteria for considering pro and con arguments becomes more strict. On the other hand, low values of δ result in more pro and con arguments and, consequently, less moot arguments.

Moreover, sound arguments cannot be created from a small set of user judgments, because when a user expresses an experience with respect to a concept of a product the user is performing an individual judgment based on her individual experience, and that experience can be biased towards individual preferences. To overcome this problem, arguments need to be created from practical knowledge, obtained by analyzing a set of experiences from a sizable number users that may have, in principle, different individual preferences. For this reason, the quality of an argument is directly related to the size of $Occ(p, C)$ —i.e. the quantity of user judgments about a concept C occurring in the reviews of that product p. The more user judgments about a certain concept, the less bias towards the individual user preferences in the acquired knowledge, and the better the quality of the argument.

For instance, suppose a user who likes interchangeable lenses, and used to own a DSLR camera. Chances are that the judgments given by this user with respect to concept 'lens' of a Panasonic Lumix ZS50 (a Point & Shoot camera) will probably be negative, since P&S do not possess good lenses, a big drawback when comparing this particular concept against her previous DSLR camera. The practical knowledge we will be able to acquire about this user's judgments with respect to concept 'lens' will be biased. The only way to overcome individual bias is to have enough number of judgments that insure enough variety in the pool of judgments, so that there is a consensual polarity value (or at least a clear majority regarding the polarity).

In this work we consider three different types of arguments, each one created using a different aggregation measure over judgment sentiment polarities: the Gini argument Arg_G, agreement argument Arg_σ, and cardinality argument Arg_F. Moreover, they share a parameter Δ that works as a threshold value: those arguments with a set of judgments $Occ(p, C)$ that is smaller in size that Δ are considered moot arguments. Since the sentiment aggregation measures used to create each one of those types of arguments are different, pro, con and moot arguments for any given product may also vary depending on the argument type (Arg_G, Arg_σ, Arg_F).

We will now introduce the three types of arguments Arg_G, Arg_σ and Arg_F, defined using different aggregation measures over polarity values of judgments.

5.2.1 Gini Arguments

A Gini argument (Arg_G) is created by combining the average judgment sentiment value with respect to a product p concept C with the *Gini coefficient* [Yitzhaki, 1979], that measures to the inequality of the distribution of these judgments sentiment values.

The Gini coefficient $Gini(p, C)$ is usually used in economics to measure the income inequality in an economy. It ranges from 0 (perfect income equality) to 1 (perfect income inequality), and is calculated from the associated Lorenz curve, being equal to the area between that curve and the line of perfect income equality. The Gini coefficient for a set of similar incomes will be close to 0, since the area between the Lorenz curve generated from this set of incomes and the perfect income equality will be very small.

In this work, we adapt the Gini coefficient measure to evaluate the sentiment dispersion (i.e. inequality) over a set of sentiment values instead of monetary incomes. By estimating the Gini coefficient between the set of sentiment values, we obtain a value that indicates how similar (or dissimilar) the sentiments of a set of judgments are. If the estimated Gini coefficient among the judgments of a concept C in the reviews of product p is close to 0, we know that people judged concept C similarly (either in a positive, negative or neutral way). On the other hand, if the Gini coefficient is close to 1, we know that people judged concept C very differently.

A Gini argument (Arg_G) has the form:

$$Arg_G = \langle p, C, S_G(p, C) \rangle$$

where S_G aggregates the sentiment values of concept's C related judgments using $S_{av}(p, C)$ and then we use the Gini Coefficient to penalize the average judgment's sentiment according to the degree of dispersion of the polarity values of the judgments of concept C over product p. This way, given a clearly positive or negative $S_{av}(p, C)$ sentiment value, the higher the degree of dispersion of the polarity values of the judgments of concept C, the closer it will be shifted to the neutral sentiment value 0. If there is a low degree of dispersion between the polarity values of the judgments, S_G will be similar to the average sentiment value $S_{av}(p, C)$.

We define the sentiment average $S_{av}(p, C)$ of concept's C related judgments with respect to the reviews of product p as:

$$S_{av}(p, C) = \frac{1}{M} \sum_{x \in Occ(p,C)} s(x, C)$$

where x is a judgment, $Occ(p, C)$ is the set of judgments from the reviews of product p in which any aspect of concept C occurs, $M = |Occ(p, C)|$, and $s(C, x)$ assesses the sentiment value of concept C in sentence x using the SmartSA sentiment analysis system (see Section 4.4). $S_{av}(p, C)$ always returns a sentiment value between $[-1, 1]$.

The aggregated Gini argument sentiment $S_G(p, C)$ is estimated as follows:

$$S_G(p, C) = \begin{cases} 0, & \text{if } |Occ(p, C)| < \Delta \\ & \text{or } \; -\delta_G < S(p, C) < \delta_G \\ S(p, C) & \text{otherwise} \end{cases}$$

where $S(p, C)$ measures the argument's sentiment using the Gini coefficient:

$$S(p, C) = S_{av}(p, C)(1 - Gini(p, C))$$

Notice that when $|Occ(p, C)| < \Delta$ we consider that we don't have enough judgments of product p and concept C; thus, we assign the neutral sentiment value 0 to the argument. Similarly, when $-\delta_G < S(p, C) < \delta_G$ we consider that the judgments average polarity is not strong enough to define an argument as a pro or a con, and we assign the neutral sentiment value to the argument.

Now we will define $Gini(p, C)$. However, since Gini coefficient can only be computed with positive values, we need to translate the argument sentiment values from [-1,1] to [0,1], as we will show later. This is only necessary to compute the Gini coefficient, the sentiments of the argument bundles are not modified. Let $V(p, C) = \{s(C, Occ(p, C))\}$ (Equation 5.1), be an ordered vector formed by the sentiments of the judgments related with concept C found in the reviews of product p, such that $V_i \leq V_{i+1}$. The size of the vector is $n = |V(p, C)| = |Occ(p, C)|$. We define $\vec{V} = \{v_1, \ldots, v_n\}$ as another vector formed by mapping, from [-1,1] to [0,1], the sentiment values of $V(p, C)$, using the function $f(s) = \frac{s+1}{2}$, such that $v_i = f(V_i)$ for $i = 1 \ldots n$.

The Gini coefficient $Gini(p, C)$ between the judgments of concept C of product p returns a value in $[0, 1]$, and is defined as follows:

$$Gini(p, C) = \frac{1}{n}\left(n + 1 - 2\frac{\sum_{i=1}^{n}(n + 1 - i)v_i)}{\sum_{i=1}^{n} v_i}\right)$$

where $v_i \in \vec{V}$, and $n = |V(p, C)|$.

Finally, recall that the parameter δ_G determines when the argument is considered a pro, con or moot (see Equation 5.2). We use $\delta_G = 0.1$ in the experiments presented in Section 5.5.

5.2.2 Agreement Arguments

An agreement argument (Arg_σ) is created by estimating, over a number of judgments' polarity values, whether these values agree (they are clustered closely) or not (they are scattered).

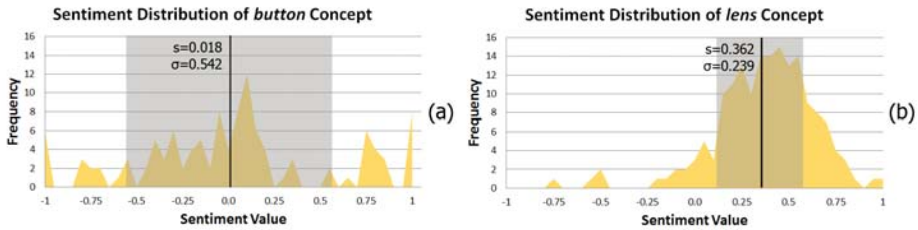


Figure 5.1: Judgment sentiment value distribution with respect to (a) *button* concept and (b) *lens* concept for Pentax K-5. Notice that values have a higher degree of dispersion in (a) than in (b).

Let $Dev(p, C)$ be the standard deviation of the sentiment values of the judgments in $Occ(p, C)$. The agreement sentiment measure $S_\sigma(p, C)$ is the sentiment average S_{av} of the sentiment values of the judgments in $Occ(p, C)$, but only for those concepts whose $Dev(p, C) < \delta_{max}$ (i.e. when values are reasonably clustered together and thus their standard deviation is low).

This measure uses two threshold parameters δ_{max} and δ_σ. First, δ_{max} specifies the maximum acceptable standard deviation over the distribution of judgment sentiment values in $Occ(p, C)$: when $Dev(p, C) > \delta_{max}$ we consider that we have no grounds for an informed decision on the overall polarity of C with respect to product p (the values are too scattered). Second, δ_σ specifies the threshold for an argument sentiment value to be considered a pro, a con, or a moot argument. An agreement argument has the form:

$$Arg_\sigma = \langle p, C, S_\sigma(p, C) \rangle$$

where S_σ is defined as follows:

$$S_\sigma(p, C) = \begin{cases} 0, & \text{if } Dev(p, C) > \delta_{max} \\ & \text{or } |Occ(p, C)| < \Delta \\ S_{av}(p, C), & \text{otherwise} \end{cases}$$

We use $\delta_\sigma = 0.1$ (see Equation 5.2) in the experiments presented in Section 5.5.

Figure 5.1 presents the sentiment value distribution of two agreement arguments of Pentax K-5, *button (a)* and *lens (b)*. The *button* argument of the Pentax K-5 has a sentiment value deviation $\sigma = 0.542$, showing a high dispersion of sentiment values among the judgments of concept *button* among the reviews of Pentax K-5. Since the deviation of the sentiment values of *button* is higher than δ_{max}, we have no clear overall polarity. On the other hand, the deviation of the sentiment values of *lens* is lower than the threshold and has a positive average sentiment $(0.235 > \delta_\sigma)$. Therefore, argument *lens* is considered a pro argument with respect to Pentax K-5.

5.2.3 Cardinality Arguments

A cardinality argument (Arg_F) is created by comparing the number of positive judgments versus the number of negative judgments of a concept C over a product p in a set of judgments $Occ(p, C)$. The number of positive (O^+) and negative (O^-) occurrences of a concept C in the reviews of a product p are defined as:

$$O^+(p, C) = |\{x \in Occ(p, C) | s(C, x) > 0\}|$$
$$O^-(p, C) = |\{x \in Occ(p, C) | s(C, x) < 0\}|$$

where $s(C, x)$ is the sentiment value in $[-1, 1]$ of concept C in judgment $x \in Occ(p, C)$. The function that compares the number of positive judgments versus the number of negative judgments of concept C in the reviews of product p is:

$$O(p, C) = \left(2 \cdot \frac{O^+}{O^+ + O^-}\right) - 1$$

where $O^+ = O^+(p, C)$ and $O^- = O^-(p, C)$. A cardinality argument has the form:

$$Arg_F = \langle p, C, S_F(p, C)\rangle$$

where S_F is defined as:

$$S_F(p, C) = \begin{cases} 0, & \text{if } O(p, C) = 0 \\ & \text{or } |Occ(p, C)| < \Delta \\ O(p, C), & \text{otherwise} \end{cases}$$

Notice that $S_F(p, C)$ takes values on $(0, 1]$ if $O^+ > O^-$, and in $[-1, 0)$ if $O^+ < O^-$. Also, when $|Occ(p, C)| < \Delta$ we consider that we don't have enough judgments of product's p concept C to make an informed decision, and we assign a neutral value to the sentiment of the argument. In the experiments presented on Section 5.5 we use $\delta_F = 0$ (see Equation 5.2) as the threshold parameter.

In this section we explored how to create types of arguments by leveraging the polarities of the judgments, in user experiences about a product, related to specific concepts of the concept vocabulary. We presented three types of arguments, Arg_G, Arg_σ, and Arg_F, each one created using a different judgment sentiment aggregation measure. In next section, we turn to build the arguments for all concepts relevant to a product p (a bundle of arguments for p), using the three different argument types presented.

5.3 Bundle of Arguments

In this section we characterize a product p by the collection of arguments relative to p, which we call bundle. The union of arguments with respect to the set of concepts in $\mathcal{C}$ over a product $p \in \mathcal{P}$ forms the *bundle of arguments* of that product $B(p)$:

$$B(p) = \bigcup_{C \in \mathcal{C}} Arg\,\langle p, C, s\rangle$$

Notice that thebundle of arguments $B(p)$ is defined over all concepts of the concept vocabulary $\mathcal{C}$.

As we introduced in previous sections, arguments can be pros, cons or moots. Now that we have all arguments with respect to a product p grouped together in the argument bundle $B(p)$, we obtain a characterization about what people like or dislike of that product. Therefore, we can distinguish three sets of arguments *Pros*, *Cons* and *Moots* by grouping the pro, con and moot arguments of a bundle $B(p)$:

$$Pros(p) = \{Arg^{+} \in B(p)\}$$
$$Cons(p) = \{Arg^{-} \in B(p)\}$$
$$Moots(p) = \{Arg^{\emptyset} \in B(p)\}$$

The set $Pros(p)$ contains all favorable arguments acquired from user experiences related to a product p (a digital camera in our experiments). That is, all possible positive reasons that play a favorable role in the task of choosing this camera over another from viewpoint of the users' experiences. Moreover, the set $Cons(p)$ contains all negative arguments, acquired from user experiences, that play a negative role in the task of choosing this camera over another. The set $Moots(p)$ contains the arguments of a product that are inconclusive — those that could not be considered clearly neither pros nor cons, or those that did not have a sufficient number of judgments about that particular argument's concept.

Now, since we have three argument types (Arg_G, Arg_σ and Arg_F), for any product p we can build three types of argument bundles: the Gini bundle $B_G(p)$, the agreement bundle $B_\sigma(p)$ and the cardinality bundle $B_F(p)$. A Gini bundle B_G only contains Gini arguments Arg_G, an agreement bundle B_σ only contains agreement arguments Arg_σ, and a cardinality bundle B_F only contains cardinality arguments Arg_F. Clearly, a bundle of arguments can only be formed by arguments of the same type, otherwise the bundle would not be consistent —since it would contain pro, con and moot arguments selected using different criteria.

Figures 5.2 and 5.3 shows the pros and cons of the Gini bundle of arguments (B_G) for Pentax K-r, and the pros and cons of the agreement bundle of arguments (B_σ) for Nikon D5300. Note that moot arguments are not represented in the figures. Each word represents an argument, painted green if it is considered a pro, and painted red if it is considered a con. The size of the arguments is proportional to the number of occurrences of that argument in the reviews of that product. for instance, in Figure 5.2, we observe how arguments 'picture', 'resolution', 'display' and 'lens' are considered pro arguments (painted in green), and are bigger in size compared to the other arguments (meaning that are frequent in the reviews of the Pentax K-r). The Gini bundle of arguments for Pentax K-r only has two arguments considered cons: 'button' and 'flash'. That means that the experiences of people with respect to those two arguments were mostly negative, as assessed by out system. Furthermore, we can observe a difference in size between both bundles in Figures 5.2 and 5.3, namely that the agreement bundle of the Nikon D5300 has a larger number of arguments than that of Pentax

K-r. As we will see later in Section 5.5, this is due to the nature of the bundles. Gini pro and con argument sets are, usually, smaller in size than the pro and con arguments sets of the agreement and cardinality bundles.

Figure 5.2: Gini bundle of arguments (B_G) for Pentax K-r.

Figure 5.3: Agreement bundle of arguments (B_σ) for Nikon D5300.

As a final step to create the bundles, for a given set of products $\mathcal{P}$ of a corpus K, and a concept vocabulary $\mathcal{C}$, we normalize the sentiment values of the arguments of the collection of product bundles in K. Let $Args = \bigcup_{\forall p \in \mathcal{P}} B(p)$ be the union of all arguments that form the bundles of arguments of the set of products $\mathcal{P}$ of a corpus K. The normalization considers the maximum and minimum sentiment value of the arguments in $Args$ with respect to a concept $C \in \mathcal{C}$. To do so, we first identify those arguments in $Args$ that refer to concept C. The normalization shifts the sentiment values of the selected arguments such that the argument with most negative sentiment value with respect to C is normalized to -1, and the argument with most positive sentiment value with respect to concept C is normalized to -1. The rest of the arguments of concept C are normalized proportionally to the most positive sentiment value (if the argument is considered a pro), or to the most negative sentiment value (if it is considered a con). This process is repeated for all $C \in \mathcal{C}$, until all bundles are normalized for all set of concepts in $\mathcal{C}$.

This way, after applying the argument sentiment value normalization, the product with the most positive sentiment value for argument 'lens' (for instance), over the set of product bundles of a corpus $B(p)\ \forall p \in \mathcal{P}$, will have a normalized sentiment value of 1. This step is necessary for the next sections (as we will see later in Section 5.4), where we use the arguments' sentiment values to define the concept of query satisfaction degree.

To normalize an argument, the function $s_{max}(Args, C)$ returns the highest positive sentiment value of all arguments in $Args$ related with concept C. Similarly, $s_{min}(Args, C)$ returns the highest negative sentiment value of all arguments in *Args* related with concept C. A normalized argument $\overline{Arg}$ is defined next:

$$\overline{Arg}\,\langle p, C, \overline{s}\rangle = Normalize(Arg\,\langle p, C, s\rangle)$$

where $\overline{s} = f'(s,\ s_{min}(C, Args),\ s_{max}(C, Args))$ as presented in next equation:

$$f'(s, min, max) = \begin{cases} \frac{s}{max} & \text{if } s > 0 \\ -\frac{s}{min} & \text{if } s < 0 \\ 0 & otherwise \end{cases} \tag{5.3}$$

Finally, the normalized bundle of arguments of a product $\overline{B}(p)$ is formed by the set of normalized arguments of the bundle of arguments $B(p)$:

$$\overline{B}(p) = \bigcup_{Arg \in B(p)} Normalize(Arg)$$

For example in Figure 5.4, concerning the concept 'zoom' in the DSLR domain, the product that has the highest positive sentiment value for this concept is the Leica D-LUX5 (with a sentiment value of 0.78). As such, after normalizing the bundles of arguments of all DSLR products, the sentiment value of the normalized argument for concept 'lens' of $\overline{B}$(D-LUX5) is 1. The positive sentiments of the other normalized bundles are rescaled accordingly to Equation 5.3. For instance, the sentiment of the argument 'lens' of the bundle of arguments of the Canon Rebel T5i was 0.30. To normalize it, we apply Equation 5.3, $f'(s, min, max)$, where $s = 0.30$, $s_{min} = -0.34$ (FinePix HS50 argument sentiment for concept 'lens'), and $s_{max} = 0.78$ (Pentax K-5 argument sentiment for concept 'lens'). Thus the sentiment value for the Canon Rebel T5i normalized concept 'lens' argument is: $f'(0.30, -0.34, 0.78) = 0.38$. A similar situation can be observed for FinePix HS50, the only product in the example with a negative argument sentiment value for 'zoom'. Since it has the most negative sentiment value for concept 'zoom', the normalized argument is rescaled to -1.

Finally, the argument bundle characterization of a product facilitates the reuse of product experiences. Since arguments aggregate the polarity of user experiences with respect to the product features defined in the concept vocabulary, a bundle of arguments is a good compact representation of what people experienced with respect to that product. As we will see in next section, this representation can be easily adapted and compared against a set of user preferences, over the concepts of the concept vocabulary.

5.4 User Query over Argument Bundles

A user query defines the preferences of a user expressed using the concept vocabulary $\mathcal{C}$. Since not all preferences are equally important for the user, every preference over a concept has a utility value. Given a set of products $\mathcal{P}$ characterized with their corresponding normalized argument bundles $\overline{B}(p)$, we can decide which is the product that has a higher level of satisfaction for a particular query.

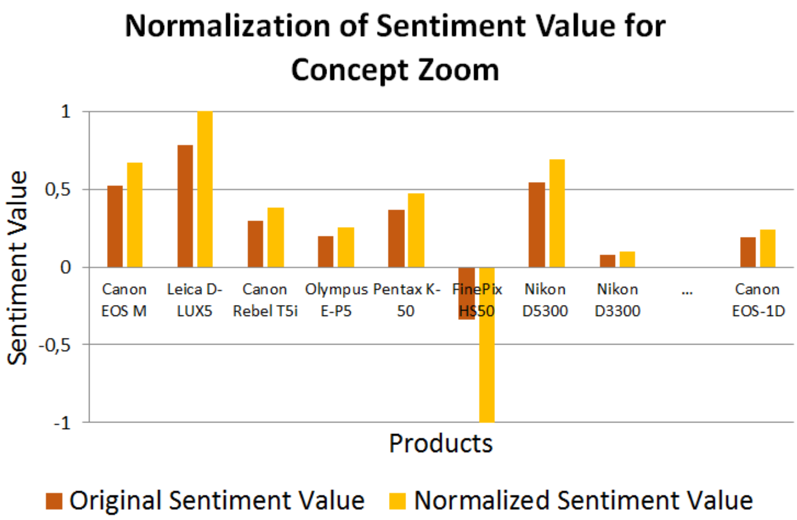


Figure 5.4: Sentiment normalization of concept 'zoom' example for DSLR products.

We define a user query as a set of concept utility pairs:

$$Q = \{(C_j, U(C_j))\}_{j=1,\ldots,k}$$

where $k \leq |\mathcal{C}|$. Notice that the there is no need to define a user preference for every concept in the concept vocabulary to create a user query, but only to for those concepts deemed important for the user. Each concept utility pair $(C_j, U(C_j))$ expresses a preference from the user over concept C_j with a strength $U(C_j) \in [0.5, 1]$. For instance in a query $Q = \{(lens, 0.9), (video, 0.6)\}$, the user prefers a good lens and good video, although, a good lens is more important than good video. Furthermore, a good lens or video are more preferred for the user than any other feature the camera could possess.

We will now define the *Degree of Satisfaction* for a query, that determines the degree in which a bundle satisfies a user query using the notion of fuzzy logic. Since t-norms and implications in fuzzy logic are defined in the interval $[0, 1]$, we need to rescale the sentiment values of all arguments that form all normalized product bundles from [-1,1] to [0,1] by applying the linear mapping $\widehat{s} = \frac{s+1}{2}$. For example, consider a normalized argument $\langle p,\ lens,\ 0.83\rangle \in \overline{B}(p)$, the sentiment of the rescaled argument will be $\widehat{s} = 0.915$, and the resulting rescaled argument is $\langle p,\ lens,\ 0.915\rangle \in \widehat{B}(p)$. Notice that the neutral value 0 in [-1,1] is mapped to the neutral value 0.5 in [0,1].

The degree of query satisfaction is defined by aggregating the degrees in which an argument, with respect to concept C, satisfies a user preference with respect to the same concept C. Therefore, we first define a concept-wise satisfaction degree, using the notion of fuzzy implication associated to the t-norm product

Q_1 Preferences	(C_1, 0.7)	(C_2, 0.6)	DS
$\widehat{B}_F$(D7100)	0.75	1.00	
$\widehat{B}_F$(EOS70D)	0.97	0.50	
$U(C_j)$ for $\widehat{B}_F$(D7100)	1.00	1.00	**1.00**
$U(C_j)$ for $\widehat{B}_F$(EOS70D)	1.00	0.83	0.83

Q_2 Preferences	(C_1, 0.7)	(C_2, 0.6)	(C_3, 0.9)	DS
$\widehat{B}_F$(D7100)	0.75	1.00	0.64	
$\widehat{B}_F$(EOS70D)	0.97	0.50	1.00	
$U(C_j)$ for $\widehat{B}_F$(D7100)	1.00	1.00	0.72	0.72
$U(C_j)$ for $\widehat{B}_F$(EOS70D)	1.00	0.83	1.00	**0.83**

Table 5.1: Degree of satisfaction of two cameras for each requirement and the overall DS for the query Q_1 and the Q_2, where C_1 = *picture*, C_2 = *resolution*, C_3 = *video*, and DS is the degree of Query Satisfaction.

($\Rightarrow_\otimes$). The fuzzy implication models this notion of degree of satisfaction: if the sentiment of an argument related with concept C, is higher than the user preference with respect to C, we consider that the user preference is fulfilled. On the other hand, if the user preference with respect to a concept C is higher than the sentiment provided by an argument, then we consider that the user preference is not fulfilled:

$$U(C_j) \Rightarrow_\otimes \widehat{s_j} = \begin{cases} 1, & \text{if } U(C_j) \leq \widehat{s_j} \\ \frac{\widehat{s_j}}{U(C_j)} & \text{otherwise} \end{cases}$$

where $\widehat{s_j}$ is the rescaled sentiment value of argument $\langle p, C_j, \widehat{s_j} \rangle$.

We need now to aggregate these k concept-wise satisfaction degrees into an overall degree of bundle satisfaction (DS) of a query Q. For this purpose, we do the conjunction of the resulting concept wise satisfaction degrees between the arguments of a bundle and the k query preferences. We use the t-norm product to do so, since conjunctions in fuzzy logic are represented using t-norms:

$$DS(Q, \widehat{B}(p)) = \prod_{j=1}^{k} (U(C_j) \Rightarrow_\otimes \widehat{s_j})$$

where $\widehat{s_j}$ is the rescaled sentiment value of argument $\langle p, C_j, \widehat{s_j} \rangle$ of the argument bundle $\widehat{B}(p)$, and $\widehat{B}$ is a rescaled argument bundle (either $\widehat{B}_G$, $\widehat{B}_\sigma$ or $\widehat{B}_F$).

Table 5.1 shows the degree of satisfaction of two user queries Q_1 and Q_2 against the cardinality bundles of two cameras: Nikon D7100 and Canon EOS70D (sentiment values are rescaled). The first query is created by a user who likes to go hiking and that is looking for a camera to capture landscape and nature while valuing fine detail. Assume her query is $Q_1 = \{(\textit{picture}, 0.7), (\textit{resolution}, 0.6)\}$ because she wants a camera with good image quality and resolution. Table 5.1 shows on the first two rows the argument sentiment values

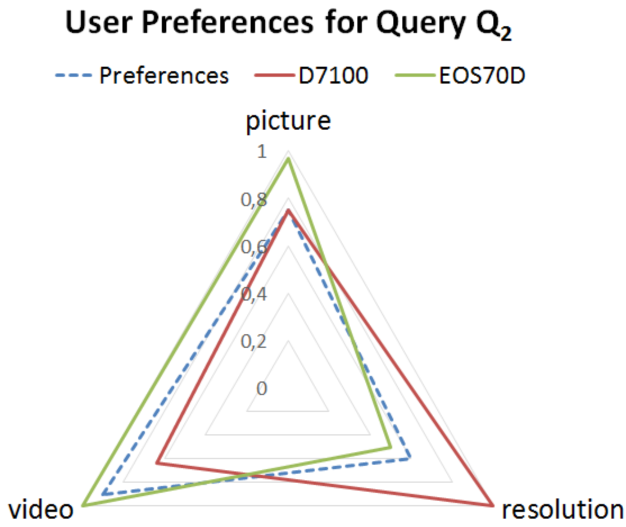


Figure 5.5: User preferences for concepts 'picture', 'resolution' and 'video' compared with the bundles $\widehat{B}_F$(EOS70D) and $\widehat{B}_F$(D7100).

of the two cameras corresponding to the concepts appearing in the query. The second two rows show the satisfaction degree of the two cameras for each preference and the overall DS for the query. Notice that satisfaction is 1 when the argument sentiment value is higher than the query's utility value for a concept.

The second example is query $Q_2 = \{(picture, 0.7), (resolution, 0.6), (video, 0.9)\}$ (second half of Table 5.1). It is created by a user that, besides hiking, also loves recording video. Figure 5.5 shows a comparison between the user preferences regarding concepts 'picture', 'video' and 'resolution', and the sentiment of the corresponding arguments of the two camera bundles. Now, according to user reviews, Canon EOS70D has an outstanding video quality (1.0), while Nikon D7100 has an average quality video (0.64). Because of this newly added preference now the higher ranking camera for Q_2 is Canon EOS70D instead of Nikon D7100, the best ranking camera for Q_1. This is clear when we observe Figure 5.5, where EOS70D better fits the user preferences.

5.5 Evaluation

In this section we compare and evaluate the different types of bundles of arguments over the three camera types (DSLR, Compact, and Point & Shoot); we compare the three types of bundles with each other and also with the characterization of some cameras written by professional reviewers from *Dpreview.com*, a renowned website specialized in digital cameras.

We are keen to study the differences between the sets of pros, cons and moots

Category	$\mathbf{K}_D$	$\mathbf{K}_C$	$\mathbf{K}_P$
No. of Products	102	95	599
No. of Reviews	7,552	6,334	84,138
Avg. Reviews/Product	74.03	66.67	140.46

Table 5.2: DSLR, Compact and PAS Camera Corpus.

between the three types of bundles of arguments, B_G, B_σ and B_F, while assessing the impact that the number of reviews of a product has over the quality of the bundle of arguments. Moreover, we are interested in observing the differences between the bundles of arguments of the three camera types (DSLR, Compact, and Point & Shoot). In addition, we also want to ascertain whether the three different types of argument bundles (B_G, B_σ, and B_F) are consistent among the product domains of DSLR, Compact and Point & Shoot cameras.

For this purpose we evaluate the precision and recall of the bundles of arguments by comparing them with the expert evaluations of products presented in Dpreview. Finally, we present a ranking strategy for bundles and compare the rankings of products obtained with the three bundle types B_G, B_σ, B_F, with two ranking of the products (those of Dpreview and Amazon), for DSLR, Compact, and Point & Shoot cameras.

We have three corpora: the Digital SLR camera corpus K_D formed by 102 products, the Compact camera corpus K_C formed by 95 products, and the Point & Shoot camera corpus K_P formed by 599 products (see Table 5.2). For every product in a corpus, we create three types of bundles of arguments as described in Section 5.3, B_G, B_σ, and B_F, using the corresponding concept vocabulary of that corpus. This way, the bundles of arguments of the K_D products are created using the corresponding concept vocabulary $\mathcal{C}_D$, the bundles of arguments of the K_C products are created using the corresponding concept vocabulary $\mathcal{C}_C$, and the bundles of arguments of the K_P products are created using the corresponding concept vocabulary $\mathcal{C}_P$.

Notice in Table 5.2 that Point & Shoot camera corpus K_P has a larger number of products and of reviews than the other two corpora.

5.5.1 Comparison between Bundle Types B_G, B_σ and B_F.

Here we study the differences between the the three bundle types B_G, B_σ and B_F for the products of each camera corpus K_D, K_C, and K_P. Since the criteria to establish an argument as pro, con, or moot varies between the three bundle types, the quantity of pros, cons, and moot arguments obtained by each bundle type may differ. Table 5.3 presents a comparison between the average quantity of pros, cons and moot arguments of each bundle type for DSLR cameras K_D, Compact cameras K_C, and Point & Shoot cameras K_P. Notice that the values presented in the table are averages between all products of the same corpus. As we will see later, products with more reviews usually have more pro and con

	Gini Bundle B_G	Agreement Bundle B_σ	Cardinality Bundle B_F
DSLR (K_D)			
Avg # pros	11.82	15.61	16.06
Avg # cons	0.59	4.08	3.19
Avg # moots	28.59	21.31	21.75
Compact (K_C)			
Avg # pros	9.42	12.70	13.16
Avg # cons	0.47	3.66	2.90
Avg # moots	29.11	22.64	22.94
Point & Shoot (K_P)			
Avg # pros	9.64	18.49	19.01
Avg # cons	0.82	6.59	5.58
Avg # moots	28.54	13.92	14.41

Table 5.3: Average number of pros, cons and moot arguments for the three bundle types B_G, B_σ, and B_F for K_D, K_C, and K_P.

arguments —and much less moot arguments.

In Table 5.3, the Agreement and Cardinality bundles have a similar average number of pros and cons, while Gini bundles are slightly smaller for the three camera corpora (DSLR, Compact and Point & Shoot). The Gini average tends to move the argument sentiment value towards 0 when there is dispersion in the distribution of sentiment values, and thus more arguments tend to be moots. The highest average quantity of pro arguments in a bundle is obtained with the cardinality bundle B_F of the Point & Shoot corpus K_P, whilst the lower quantity of pro arguments is obtained by the Gini bundle B_G of the Compact camera corpus K_C.

On the other hand, the quantity of cons is lower than the quantity of pros for all bundles B_G, B_σ, and B_F, for the three camera corpora. Either the SmartSA sentiment analysis system is biased towards positive sentiments, or the reviews of our three corpora contain more positive than negative judgments. Notice that the quantity of pro and con arguments is directly related with the average quantity of reviews per product, presented in Table 5.2. For instance, the bundles of arguments of the Point & Shoot category K_P, with an average quantity of reviews per product of 140.46, contain more pro and con arguments than the bundles of the other two camera corpora K_C and K_P, with an average quantity of reviews per product of 74.03 and 66.67, correspondingly.

Next we analyze which concepts are considered pros in the three bundle types. Figure 5.6 focuses on the the top 50 products with more reviews from the DSLR corpus K_D on the x axis. The y axis shows the quantity of pros shared by two or three bundle types for each product. The analogous figures for the Compact corpus and Point & Shoot corpus are presented in Appendix F. Here "shared" means the intersection, i.e. those pro concepts that appear in the three bundle types. For instance, concept 'lens' is considered a pro argument in the

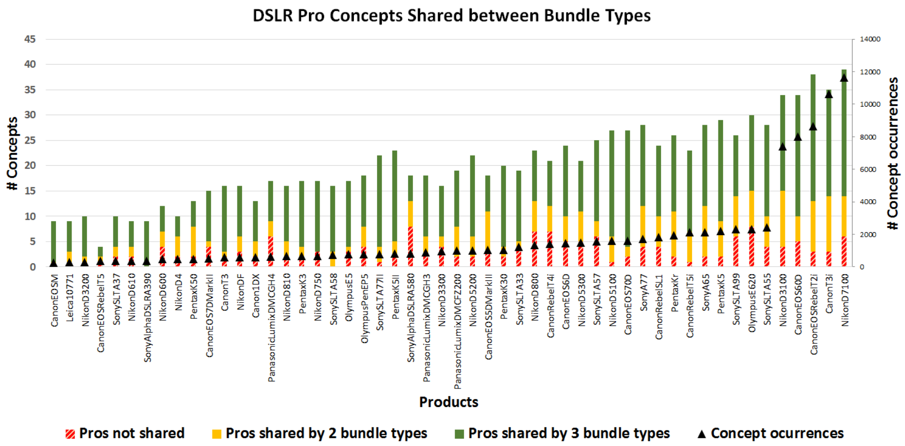


Figure 5.6: Quantity of pros shared between the three bundles of arguments B_G, B_σ and B_F of the top 50 products with more reviews of DSLR corpus K_D, together with the number of occurrences of the pro concepts in the reviews of the product.

three bundles types (B_G, B_σ and B_F) of the Nikon D7100 DSLR camera, thus is *shared* by the three types of bundle of arguments of that product.

The results in Figure 5.6 show that most pro arguments (almost 8 out of 10) are shared between two or three bundle types of a product, a good indicator of the consistency of our approach. This means that a concept of a pro argument in a B_G is also likely be present in B_σ pros and in B_F pros. Furthermore, the number of pros (and also cons, not included in this figure because results are similar) of a bundle is directly related with the quantity of judgments in the reviews of that product: the more judgments the richer the bundles are. Notice that we are only analyzing whether a concept is part of a pro argument in more than one of bundle types of a product; we are not comparing the concrete positive sentiment values of the arguments.

5.5.2 Argument Bundles Evaluation

To evaluate the quality of bundles, we compared the bundles of arguments of the 15 products with more reviews from the DSLR cameras K_D with the product pros and cons textual descriptions we found in the Dpreview website. The Dpreview pros and cons of a product are separately formed by lists of judgements such as 'good detail and color in JPEGs at base ISO (pro)' or 'buggy Live View / Movie Mode (con)'. In order to compare the Dpreview pro and con items with our bundles of arguments, we first manually identify the issues referenced in each item text and interpret that issue as included in one of the concepts in our DSLR vocabulary $\mathcal{C}_D$, if it exists. For instance, we consider that the Dpreview sentence 'good detail and color in JPEGs at base ISO' refers positively to the

		Precision	Recall	F_2-score	*Contradictions*
	B_G	**0.567**	0.644	0.627	**0.004**
Pros	B_σ	0.506	0.761	0.691	0.135
	B_F	0.513	**0.822**	**0.733**	0.065
	B_G	0.333	0.046	0.056	0.046
Cons	B_σ	0.285	0.558	0.468	0.132
	B_F	0.388	0.488	0.464	0.165

Table 5.4: Measure on precision, recall, F_2-score and contradictions between pros and cons of bundles B_G, B_σ and B_F, of DSLR products from corpus K_D, with respect to Dpreview pros and cons.

concepts *jpeg*, *color* and *picture*, whilst 'buggy Live View / Movie Mode' refers negatively to concepts *live view* and *video*. Those sentences from Dpreview that did not clearly refer to a concept in $\mathcal{C}_D$ were ignored. By grouping the vocabulary concepts present in the DPreview pro and con items of a product, we create the sets of Dpreview pros and cons ($Pros_{dp}$ and $Cons_{dp}$), but without associating any numeric value of sentiment —only the positive or negative nature of the polarity expressed in the text. We compare those Dpreview $Pros_{dp}$ and $Cons_{dp}$ sets with the pros and cons of the three different bundles of arguments of each product —without taking into account the sentiment values, only whether the concept is selected as a pro or as a con.

Table 5.4 presents the average precision, recall and F_2-score between the sets of pros and cons of the three bundle types and those of Dpreview. We use the F_2-score to weight recall higher than precision, since we are keen to study whether the three different bundle types identify as pros and cons the same concepts listed in Dpreview. Furthermore, we analyze the percentage of *contradictions*, which are those concepts selected as pros in our bundles of arguments but considered cons in Dpreview and vice versa. A low rate of contradictions is a good indicator of the quality of the bundles if we take Dpreview as a standard for comparison.

The bundle of arguments that performs best for the pro arguments of the selected DSLR products is the cardinality bundle B_F, with an average recall of 0.822 and an F_2-score of 0.733. This means that the 82.2% of the concepts listed as pros of product a p in Dpreview also form part of the pros of the cardinality bundle $B_F(p)$. On the other hand, the sets of cons of all three bundles of arguments perform poorly. The reason is the granularity of the sentences, which is different between our concept vocabulary and that used in Dpreview. For us the granularity level is given by our concept vocabulary, while Dpreview sentences address issues that are at different levels of granularity. Furthermore, the granularity of Dpreview sentences varies whether the sentence is a pro or a con. Dpreview pro sentences tend to be more general: 'camera buttons and dials are useful and easily configurable', while con sentences tend to be more specific: 'the video dial is not easily accessible'. Although for us both sentences reference concept *button*, in this example the Dpreview pro sentence addresses a more general view of the buttons of the camera than the second one. Table

5.5 shows some examples of Dpreview sentences in which the granularity varies between pro and con arguments.

Camera	Concept	Type	Sentence
Canon EOS Rebel t2i	Resolution	*Pro*	Excellent detail and resolution
	Resolution	*Con*	Chroma noise takes a big chunk out of resolution
Sony SLT A-99	Video	*Pro*	Good video specifications
	Video	*Con*	Magnified live view not available in video mode
Canon Rebel SL1	Autofocus	*Pro*	AF noticeably improved
	Autofocus	*Con*	AF illuminator integrated into flash

Table 5.5: Granularity differences in Dpreview pro and con sentences.

Furthermore, notice that the precision values of all bundles are lower than 0.6, suggesting that the sets of pros of the bundles of arguments are richer in concepts compared to those of Dpreview summaries. This is not surprising, since the sets of Dpreview pros and cons are not exhaustive but a short list of the concepts that stand out from their point of view. The average size of the pros in a bundle is 12-14 arguments, while the average pro set size of Dpreview identified issues is 7-9. Finally, notice the number of contradictions between the bundles of arguments and the Dpreview sets is low. Nevertheless, we are interested in studying which concepts incur more often in contradictions.

The most common contradictions between the bundles and the set of pro and con extracted from Dpreview for the 15 selected products are: battery (10), viewfinder (5), recording (5) and button (3). In Dpreview battery is often selected as a pro, however it is usually selected as a con in the bundles of arguments. That is because in the reviews people usually complain about the battery of a camera, while they do not seem to express positive opinions on cameras with a good battery (it would seem it is taken as a given). Other frequent contradictions are *viewfinder*, *recording* and *button*. This is because in Dpreview those are commonly selected as cons for having not optimal behavior in certain types of situations (e.g. 'the video dial is not easily accessible') while the overall opinions about the rest of the buttons are positive. Therefore, our bundles will capture this average higher granularity sentiment of *button*. Similar situations are observed for *recording* and *jpeg* concepts.

5.5.3 Bundle Ranking Evaluation

In this section we are interested in comparing the bundles of arguments with the camera descriptions of Dpreview photography website. We have seen that the sets of pro and con arguments between the arguments of the DSLR products and Dpreview characterizations are similar in Table 5.4. Now we want to evaluate how good our experiential characterizations of products are, compared to those created by the photography experts of Dpreview.

To do so, we define function $\Phi(\overline{B}(p_i), \overline{B}(p_j))$ that estimates the degree in which a normalized bundle $\overline{B}(p_i)$ is better or superior to another normalized bundle $\overline{B}(p_j)$.

$$\Phi(\overline{B}(p_i), \overline{B}(p_j)) = \frac{1}{2|\mathcal{C}|} \sum_{k=1}^{|\mathcal{C}|} s_k^i - s_k^j$$

where s_k^i and s_k^j are the sentiment values of respective arguments $\langle p_i, C_k, s_k^i \rangle$ and $\langle p_j, C_k, s_k^j \rangle$ in the normalized bundles of p_i and p_j. Φ is the average of these differences, a value in $[-1, 1]$. If the value of $\Phi(\overline{B}(p_i), \overline{B}(p_j))$ is in $(0, 1]$, it means that $\overline{B}(p_i)$ is superior than $\overline{B}(p_j)$. If the value of $\Phi(\overline{B}(p_i), \overline{B}(p_j))$ is in $[-1, 0)$, it means that $\overline{B}(p_i)$ is worse than $\overline{B}(p_j)$.

Using Φ, we create five rankings with the products of each camera corpus: one for each normalized bundle type $\overline{B}_G, \overline{B}_\sigma$ and $\overline{B}_F$, a Dpreview ranking based on the Dpreview overall score, and an Amazon ranking based on each product star rating. At the end, we have five different rankings for DSLR cameras ($\overline{B}_G$ ranking, $\overline{B}_\sigma$ ranking, $\overline{B}_F$ ranking, Dpreview ranking and Amazon ranking), five camera rankings for Compact cameras, and five camera rankings for Point & Shoot cameras.

In case two or more products of the same camera type had the same Dpreview score, such as Olympus E620 and Nikon D3100, both DSLR cameras with a score of 72 out of 100, we only kept the product with most reviews, in this example the Nikon D3100. This left us with 10 different DSLR cameras, 10 Compact cameras, and 10 Point & Shoot cameras for the rankings.

Let us now compare the 5 DSLR rankings. The top 3 products for the $\overline{B}_G$ ranking are Nikon D7100, Pentax K-5 and SonySLT A-55. The top 3 products for $\overline{B}_\sigma$ are Nikon D7100, SonySLT A-99 and SonySLT A-55, and the top 3 ranked products for $\overline{B}_F$ are Nikon D7100, SonySLT A-99 and Pentax K-5. Notice that Nikon D7100 is the top product in all three bundles, and it is also ranked 1st (with a score of 85 points) in the Dpreview ranking, followed by SonySLT A-99 and Pentax K-5. The top products of Compact and Point & Shoot camera types are also similar between their respective set of bundles and Dpreview.

Table 5.6 shows the Spearman Rank Correlation (Equation 3.3) of the 3 rankings of bundles with the Dpreview score ranking and the Amazon star ranking, for DSLR, Compact and Point & Shoot cameras. We added a random ranking strategy to facilitate comparison. The random ranking correlation values were obtained by averaging the Spearman correlations of 10.000 randomly generated product rankings with Dpreview ranking and Amazon ranking. The 'Avg. Dpreview Ranking' of Table 5.6 shows the average Spearman rank correlation between the Dpreview ranking and the Amazon ranking of the three camera types DSLR, Compact, and Point & Shoot.

The results show that, for the DSLR camera type, the $\overline{B}_G$ ranking has the highest Spearman correlation with Dpreview ranking (correlation of 0.80), followed by the cardinality ranking $\overline{B}_F$ (correlation with Dpreview of 0.76). For Compact cameras, $\overline{B}_F$ ranking obtained the highest Spearman correlation with Dpreview ranking (correlation of 0.68). Finally, for Point & Shoot cameras, the bundle ranking that obtained the best correlation with Dpreview ranking was $\overline{B}_\sigma$ (correlation of 0.86), closely followed by $\overline{B}_F$ (correlation of 0.83). These

	Dpreview Ranking	Amazon Ranking
s		
$\overline{B}_G$ Ranking	**0.80**	0.39
$\overline{B}_\sigma$ Ranking	0.70	0.62
$\overline{B}_F$ Ranking	0.76	0.61
Compact Spearman Rank Correlation		
$\overline{B}_G$ Ranking	0.65	0.54
$\overline{B}_\sigma$ Ranking	0.62	0.55
$\overline{B}_F$ Ranking	**0.68**	0.57
Point & Shoot Spearman Rank Correlation		
$\overline{B}_G$ Ranking	0.57	0.33
$\overline{B}_\sigma$ Ranking	**0.86**	0.32
$\overline{B}_F$ Ranking	0.83	0.16
Avg. Random Ranking	0.27	0.27
Avg. Dpreview Ranking	1	0.23

Table 5.6: Spearman rank correlation between the top 10 DSLR, Compact and Point & Shoot bundle rankings with Dpreview product ranking and Amazon star ratings ranking.

values tell us that there is a very strong correlation between Dpreview and the rankings of our three bundle types, a good indicator of the high quality of the bundles. This is specially true with the rankings created from the cardinality bundles $\overline{B}_F$, which obtained an average Spearman correlation of 0.76 between the DSLR, Compact and Point & Shoot cameras and the corresponding Dpreview rankings. The correlations for $\overline{B}_\sigma$ and $\overline{B}_G$ rankings are also strong, being notably higher than the random ranking correlations.

On the other hand, notice that the ranking correlations between the bundle rankings and Amazon star rankings are in comparison lower than the average Spearman correlation between the bundle rankings and the Dpreview ranking. In fact, the average Spearman correlation between the three bundle rankings and the Amazon ranking is around 0.40, showing a not very strong similarity between the star-rating ranking and the rankings of the bundles acquired from the reviews. Furthermore, the Amazon star-based ranking does not correlate with the Dpreview score ranking either (Spearman correlation of 0.23), indicating that there exists a notable difference between the star rating ranking of Amazon and the Dpreview ranking, to the point that the random ranking has a higher average correlation with Amazon ranking (0.27) than with Point & Shoot cardinality bundles (0.16). These results may be understandable, since two people expressing similar arguments about a product can give different star-rating values (as an overall score). Nevertheless, that fact is that Amazon's star rating seems unsuitable to be used as ground truth to test the quality of the argument bundles.

The last experiment consists on evaluating the Spearman rank correlation between rankings formed by a larger set of products, to prove that the bundles of arguments correctly elucidate the pro and con arguments not only for top rated products but for lower rated products as well. We would have liked to perform this evaluation for all three camera types (DSLR, Compact, and Point & Shoot), but Dpreview does not have many DSLR and Compact cameras scored, and without Dpreview scores it is impossible to create the Dpreview rankings. However, we were able to obtain 30 Point & Shoot Dpreview scores from among the top 100 cameras that form K_P. The Dpreview scores of those cameras range from 67 (Canon PowerShot SX150) to 83 (Sony Cyber-shot DSC-RX100 V). As such, we are performing this evaluation with those 30 K_P products.

The experiment is similar to the one presented before in Table 5.6: we have 5 camera rankings, 3 bundle rankings ($\overline{B}_G$, $\overline{B}_\sigma$, and $\overline{B}_P$), one Dpreview score ranking, and one Amazon star ranking, each of those rankings for the same 30 products. However, this time we assign fractional ranks to those products with equal Dpreview scores. In fractional ranking items that compare equal receive the same ranking number, which is the mean of their ordinal ranking positions. Notice that this is different from the ordinal rankings of Table 5.6, where all items receive distinct ordinal numbers (including items that compare equal) that represent their ranking positions. For instance, suppose product p_1 and product p_2 are rated with a 79 Dpreview score, and p_3 with a 85 Dpreview score. If we create a Dpreview ordinal ranking from higher to lower score, p_3 will be ranked 1st (because it has the highest score), and p_1 and p_2 will follow with 2nd and 3rd position indistinctly, because both products have a Dpreview score of 79. In a fractional ranking, p_3 will be equally ranked 1st, but p_1 and p_2 will be both ranked in position 2.5 (the average of p_1 and p_2 ranking positions considering the ordinal ranking). When using fractional rankings, the Spearman rank correlation is also computed with Equation 3.3. The only difference being that the ranks of products in fractional rankings can be rational numbers $\mathbb{Q}$ (as in previous example, where p_1 and p_2 were both ranked in position 2.5), instead of natural numbers $\mathbb{N}$ (for the ordinal rankings).

Using fractional ranking is equivalent to averaging the ranking correlations over all possible ranking permutations of elements with the same score [Basler, 1988]. At this point, we could have used Kendall τ-b rank correlation [Abdi, 2007] to manage Dpreview score ties, but we decided to use Spearman's rank correlation instead so we can compare the results with the ones presented in Table 5.6.

Table 5.7 shows the Spearman rank correlation between the 5 rankings of 30 Point & Shoot cameras. The correlation between the bundle rankings and the Dpreview ranking is strong —the cardinality bundle $\overline{B}_F$ ranking has the highest correlation with the Dpreview ranking (with a 0.77 Spearman rank correlation). This shows that the cardinality bundles of arguments are able to correctly capture the information expressed in user experiences with respect to a set of products, and that the resulting bundle of arguments used to characterize products correlate with the products scores given by the experts of Dpreview. $\overline{B}_G$ (Spearman correlation of 0.57) and $\overline{B}_\sigma$ (Spearman correlation of 0.75) also

	Dpreview Ranking	Amazon Ranking
Point & Shoot Spearman Rank Correlation		
$\overline{B}_G$ Ranking	0.57	0.68
$\overline{B}_\sigma$ Ranking	0.75	0.48
$\overline{B}_F$ Ranking	**0.77**	0.43
Avg. Random Ranking	0.061	0.061
Dpreview Ranking	1	0.57

Table 5.7: Spearman rank correlation between bundle rankings formed by 30 Point & Shoot cameras, with Dpreview product ranking and Amazon star ratings ranking.

correlate stongly with Dpreview ranking, being $\overline{B}_\sigma$ very close to the Spearman correlation obtained by the agreement bundle ranking $\overline{B}_F$, and the Spearman correlations obtained in Table 5.6.

A noteworthy result is the change of the correlation degree between the rankings of bundles and the Amazon star rating presented in Table 5.7. Seems that, since the quality difference between the top 100 Point & Shoot products is more differentiated than between the top 10 products, Amazon star rating performs slightly better than in Table 5.6. The same situation can be observed between Dpreview ranking and Amazon ranking, now having a stronger Spearman correlation of 0.57. Finally, notice that, since the size of the rankings increased to 30 ranked products, the random ranking strategy performs 5 times worst than in Table 5.6, with an average Spearman's rank correlation over the 10.000 randomly generated product rankings of 0.061.

5.6 Conclusions

In this chapter, we present a method that reuses the experiences of other people with products in order to help others make more informed decisions. We achieve this goal by means of acquiring the practical knowledge from a set of individual user judgments, to create pro, con and moot arguments that characterize the set of products in our corpus. Every argument is related with a certain issue of a product, being the issues defined in the concept vocabulary created in previous chapter. An argument can be seen as a reason about why to buy or not to buy a product, obtained from analyzing a set of user experiences.

We define three different types of arguments, each one created using a different sentiment aggregation measure over judgment polarities: Gini arguments, agreement arguments, and cardinality arguments. Gini arguments are created by combining the average sentiment values of user judgments with respect to a product's concept with the Gini coefficient. Agreement arguments are created by estimating, over a number of judgment's polarity values related to a concept, whether these values agree in the polarity of the sentiment or they are scattered. Finally, cardinality arguments are created by comparing the number of positive judgments versus the number of negative judgments of a concept over the reviews of a product.

Then, a bundle of arguments $B(p)$ is defined as the grouping of all arguments of a product p. The bundles of arguments are compact characterizations of products based on the set of concepts in a vocabulary $\mathcal{C}$. A bundle is a knowledge structure created from user experiences, that aggregate and organize the arguments, created from the experiential knowledge found in user reviews about a given product. A bundle of arguments is formed by pro, con, and moot arguments, and as such, can be seen as the aggregation of reasons about why to buy, or not, a given product. Since there exist three types of arguments (the Gini, cardinality, and agreement arguments), we also define three different types of bundles depending on the argument types that form the bundle. If the bundle is formed by Gini arguments, we have a Gini bundle B_G. If the bundle is formed by agreement arguments, we have an agreement bundle B_σ and, if the bundle is formed by cardinality arguments, we have a cardinality bundle B_F.

Argument bundles allow us to reuse the knowledge for other users with other individual preferences, because they abstract the knowledge from multiple user experiences. To do so, in this chapter we introduce the notion of query and query satisfaction by a bundle of arguments. A query expresses the individual preferences of a user, and by means of the fuzzy implication associated to the t-norm product, we determine the bundles of arguments of products that better satisfy the user preferences. Notice that, when we query a set of argument bundles for the product that better satisfies our preferences, we are *effectively reusing* the experiential knowledge acquired by our methodology from other people's experiences. Notice that the user query cannot be evaluated *per se*, since it depends on the preferences of the user. The user determines the preference value of the different concepts that form the query. Furthermore, the degree of query satisfaction estimates how much a bundles satisfies the preferences of a query. It is only a way to model how fulfilled are the preferences of a user query by a bundle of arguments. Although the degree of query satisfaction can be modeled using different operators, the elements that need to be evaluated are the bundles of arguments. If the bundles of arguments are accurate with respect to the experiences of the reviews, the bundle with a higher degree of query satisfaction will be the bundle that better satisfies the preferences of the user query considering the user experiences of the reviews of that product.

In this chapter, an evaluation of the three types of argument bundles is performed and compared with the expert descriptions of the Dpreview website for each of the three camera types DSLR, Compact and Point & Shoot. We show that the quantity of pro and con arguments of the bundle of arguments is directly correlated with the average number of reviews per product in Table 5.3. The more reviews of a product, the more rich are the bundles of arguments for that product. Furthermore, in Figure 5.6, we show that the three types of DSLR bundles B_G, B_σ, and B_F share a big quantity of pro arguments, a good indicator of the consistency of our approach. Similar results are obtained in for Compact and Point & Shoot cameras, presented in the Appendix F.

We show that the bundles of arguments are very close to the pro and con features listed in Dpreview, with the cardinality bundles B_F obtaining a recall of 0.822, and F_2-*score* of 0.733 (Table 5.4) when comparing the sets of pros of the bundles with those of Dpreview. Moreover, the three types of bundles obtained a high correlation with the overall Dpreview score ranking over the subset of the most frequent products for DSLR, Compact, and Point & Shoot, a good indicator of the quality of the bundles and its consistency among the three camera categories. Specifically, the cardinality bundle B_F for Point & Shoot cameras obtained a 0.77 Spearman rank correlation (Table 5.7) with the Dpreview score ranking formed by 30 products, 13 times better than the average random ranking strategy. Amazon star ranking presented small correlations with bundle rankings and Dpreview rankings in all the experiments, sometimes slightly better (and once worst) than the random ranking strategy.

Chapter 6

Conclusions and Future Work

In this chapter we summarize and discuss the research work presented in this monograph. We first present a summary of this work in Section 6.1, discussing how we have addressed and solved the challenges stated in the Introduction. Then we present and discuss the main research contributions of this doctoral thesis in Section 6.2, and the publications related to those contributions in Section 6.3. Finally, we present and discuss several challenging lines for future work in Section 6.4.

6.1 Summary

The work presented in this monograph tackles the challenge of analyzing and representing experiences from user-generated content in order to facilitate their reuse, within the framework of the Web of Experiences. A first main goal in this research work has been to analyze and discover the practical knowledge contained in the *textual* records of experiences of a (large) number of people, experiences that imply people interacting with entities in the real world. Moreover, a second main goal has been to acquire and represent that practical knowledge about a domain in such a way that it is amenable to be reused in helping other people to make more informed decision dealing with entities of the same domain.

In this section we will draw conclusions by analyzing and discussing how we have addressed and solved the challenges we set up to address in the Introduction (Chapter 1). The first challenge we posited is this:

Challenge 1. Identify the set of salient terms, in a given domain, used by people while writing about their experiences with a concrete entity of that domain.

The first challenge of our work has been to analyze the user experiences to identify the set of salient terms used by people while writing about their experiences about digital cameras. This challenge has been addressed in Chapter

3, by creating three aspect vocabularies ($\mathcal{A}_D$, $\mathcal{A}_C$, and $\mathcal{A}_P$), one for each corpus of digital cameras: DSLR K_D, Compact K_C, and Point & Shoot K_P. The three aspect vocabularies model the uni-gram and bi-gram words used by people when expressing their experiences with the three types of digital cameras, and was created using two different sources: a corpus of user-generated reviews and, additionally, two professional photography web resources. User reviews provided the main source for populating the aspect vocabularies, whereas the two professional web resources were used to refine it by improving the aspect selection process with specialized domain knowledge related to digital cameras.

Each aspect vocabulary was created by joining four aspects sets generated by combining various natural language processing techniques. Moreover, the four methods used to create the aspect vocabulary are complementary, in the sense that they discover different set of aspects from user reviews by using different techniques. Grammatical extraction rules, combined with part of speech tagging and frequency filtering were used to select a set of aspects from the user reviews of the corpora in an unsupervised way. This unsupervised aspect extraction approach was compared against other state of the art aspect extraction techniques with a manually marked know dataset for testing aspect extraction in Section 3.2.2, and obtained a good precision. Furthermore, we improved the recall of the aspect vocabularies by using WordNet and *PhotoDict*, our own photographic taxonomy created from Dpreview and Snapsort web resources. The resulting three aspect vocabularies ($\mathcal{A}_D$, $\mathcal{A}_C$, and $\mathcal{A}_P$) succeed in capturing the most salient lexical items found in the reviews of our three corpora (K_D, Compact K_C, and Point & Shoot K_P).

During the process of creating an aspect vocabulary using unsupervised aspect extraction techniques, we noticed that an important step is the selection of an accurate frequency filter that removes spurious content and false aspects, while retaining interesting domain related aspects. Furthermore, we observed differences between the aspect vocabularies of three Amazon camera types, $\mathcal{A}_D$, $\mathcal{A}_C$, and $\mathcal{A}_P$: the reviews of the three corpus contain different aspects, and the individuals define their experiences differently depending on the corpus. Therefore, from that chapter onwards we decided to work with three separate corpora (K_D, K_C, K_P) and three aspect vocabularies ($\mathcal{A}_D$, $\mathcal{A}_C$, and $\mathcal{A}_P$).

The decision of considering the three camera types as different domains was important in order to improve the accuracy of the bundles of arguments of DSLR, Compact and Point & Shoot cameras. If, instead of considering three separate domains, we considered all cameras to belong to the same domain, the resulting concept vocabularies and argument bundles would be different and less accurate, because they would not include specific concepts and arguments belonging uniquely to DSLR, Compact, or Point & Shoot cameras. This is an important insight of this monograph: given a corpus about a domain (here digital cameras) it is not self-evident that it should be *one application domain* for the Web of Expertise approach. Therefore, we consider that the task of determining whether a corpus constitutes one or several domains is important —specially in the context of this monograph, where user-contributed content is the basis by which the vocabularies and the bundles of arguments are built. The important problem of identifying there are different (sub)domains of a given corpus is further discussed in the future work presented in Section 6.4.5.

It is interesting to point the limitations of the methods presented to create the aspect vocabulary from a corpus of user-generated reviews. First of all, we focused in extracting uni-gram and bi-gram lexical items using the grammatical extraction rules to create the aspect vocabulary. Our techniques are not fit to extract n-grams from user-generated reviews, and although n-gram aspects are not as frequent as uni-gram and bi-gram aspects (and therefore they might end up being filtered by the frequency filter of the aspect vocabulary), we are missing some aspects that might be interesting to include in our aspect vocabulary. In order to extract n-gram aspects, two methods could be implemented on top of our aspect extraction system: 1) add a new grammatical extraction rules to facilitate the extraction of n-grams, and 2) modify the methodology used to create $\mathcal{A}_4$ to extract n-grams, instead of focusing only in uni-gram and bi-grams.

Moreover, notice that, differently from opinion mining techniques that exploit the structure of sentences that contain sentiment bearing words in order to extract aspects, in our approach aspects are extracted without applying any sentiment analysis techniques. We create the aspect vocabulary without taking into consideration the polarity of user judgments, but by exploiting the syntactic structure of the sentence and the domain knowledge available in two photography websites. Sentiment analysis is applied later, after the creation of the aspect vocabularies, to assess the polarity of the judgments in order to create the concept vocabulary.

The second challenge was to be able to identify and evaluate user *judgments* and its polarity, since user judgments contain practical knowledge with respect to an entity's issues:

Challenge 2. Identify user judgments and their polarity.

This challenge has been addressed in Chapter 4, by identifying those sentences, from the three corpus of digital camera (K_D, K_C, and K_P), that refer to one (or more) of the salient words of the corresponding aspect vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$, and $\mathcal{A}_P$. That is to say, to identify user judgments in the reviews of a DSLR camera, we search for those sentences in the DSLR camera corpus K_D that contain an aspect (or more) of the DSLR aspect vocabulary $\mathcal{A}_D$.

User judgments are "considered decisions or sensible conclusions" about certain issues of an entity, and that have a positive, negative or neutral polarity. However, we do not have explicitly defined the entity's issues in the text, we just have the aspects that are related to each entity: those aspects from the aspect vocabularies. Therefore, in order to identify user judgments, we search, in the user reviews for sentences that refer to one or more aspects from the aspect vocabulary.

The polarity of the judgments is estimated using the SmartSA sentiment analysis system (see Section 4.4). The SmartSA system obtains the sentiment score of sentiment-bearing words from SentiWordNet, taking into consideration negation terms and lexical valence shifters such as intensifiers. Therefore, by applying the SmartSA system we can determine the polarity of user judgments with respect to the various aspects referenced in the judgments, belonging to the three aspect vocabularies.

Although the SmartSA system proved to be effective when estimating the sentiment polarity of judgments found in user-generated reviews of our corpus, we found some exceptions in where the SmartSA system did not assess the polarity correctly. Those exceptions were basically caused by the prior sentiments assigned by the SmartSA system to some polarity words specific of the domain of digital cameras. For instance, adjectives such as 'small' or 'RAW' digital format where considered to carry a negative polarity by the SmartSA system, although in some domains (such as in the domain of digital cameras), they might be considered to carry a positive or even a neutral polarity since a 'small' camera is easier to carry. Therefore, there is room for improving the sentiment detection by, for instance, adapting the SmartSA sentiment vocabulary to the domain of digital cameras, as we discuss later in future work lines on Section 6.4.2.

The third challenge of this dissertation was to identify what were the *issues* described in the user experiences about an entity:

Challenge 3. Discover the main issues addressed by people when writing their experiences and create a concept vocabulary.

This challenge has been addressed in Chapter 4 of this monograph, and has been overcome by creating three vocabularies of concepts, $\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$, one for each camera corpus K_D, K_C, and K_P, where each concept in those concept vocabularies models an issue addressed in the textual descriptions of experiences.

To create a concept vocabulary for a given corpus K, we first identify those aspects from the corpus aspect vocabulary $\mathcal{A}$ that are treated as indistinguishable in the user judgments found in the reviews of corpus K. That is to say, to create the concept vocabulary for DSLR cameras ($\mathcal{C}_D$), we identify those aspects from the DSLR aspect vocabulary ($\mathcal{A}_D$) that were used similarly in the judgments found in reviews of the DSLR corpus (K_D).

Therefore, we defined a similarity measure between aspects that takes into account the semantic and lexical similarities of the aspects of an aspect vocabulary $\mathcal{A}$, as used in a corpus of camera reviews K. Using that similarity measure, we created an unsupervised bottom-up hierarchical clustering algorithm that groups the most similar aspects from an aspect vocabulary $\mathcal{A}$. The result of that clustering process is a dendrogram where each cluster of aspects in the dendrogram represents a concept found in the reviews of a corpus K, but with different levels of granularity. The set of aspects clustered in a concept are deemed to refer to the same issue of the camera, because, according to our analysis, they have the same (or a similar) intended meaning when used by individuals in the reviews.

From a computational point of view, creating a vocabulary of concepts means determining a partition of the vocabulary of aspects, where every aspect is assigned to one concept only and each concept gathers those aspects whose meaning is very similar or indistinguishable. Therefore, from that dendrogram, we selected a partition of concepts with a good level of granularity by using the cognitive linguistics notion of basic level concepts (see Section 4.2), and by estimating the sentiment coherence among aspects grouped in the same cluster. The selected partition of concepts forms the *concept vocabulary* $\mathcal{C}$ of a corpus K, where each concept models a relevant issue used by people when expressing their experiences in the reviews.

We repeated this process for each corpus K_D, K_C, and K_P, resulting in three concept vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$, one for each camera type. By analyzing the resulting concept vocabularies, we noticed that the concept vocabularies of the three camera types are quite different: not only the set of concepts varies between the concept vocabularies of the different camera types, but also the usage of the aspects and the issues involved in the judgments of the reviews of the three camera corpus is different. This result supports the decision of considering the three camera types (DSLR, Compact, and Point & Shoot) as separate domains, since the set of concepts (and their clustered aspects) used in the reviews of an entity depends on the domain of the entity.

Notice that, in order to correctly identify those aspects from the aspect vocabulary that are treated as indistinguishable aspects in the reviews of a corpus, a lot of reviews are necessary. The similarity between two aspects is more accurately estimated if those two aspects are frequently judged in the reviews. Similarly, the more user judgments with respect to the issues represented in a concept vocabulary, the more accurate the polarity profiles of the aspects over the set of products in a corpus, resulting in a better partition selection from the clustering dendrogram. Therefore, the quantity of reviews and data is important in the process of creating a concept vocabulary, since we cannot expect to determine the similarity between two aspects or an accurate average sentiment of user judgments if those aspects are merely used and judged by the individuals. This problem is an important caveat of our methodology to create a concept vocabulary, and should be kept in mind when using smaller corpora.

Knowing the set of issues (modeled as concepts) deemed more important by users when expressing their experiences in entity reviews, and the aspects that refer to each issue, the next challenge was to study how this knowledge could be used to characterize a set of entities based on the practical knowledge found in the experiences from individuals.

Challenge 4. Create the arguments and the bundle of arguments of an entity.

This challenge has been addresses in Chapter 5, where we introduce a method that leverages the practical knowledge found in the experiences of individuals with respect to an entity, over the set of relevant issues of a concept vocabulary.

Considering a corpus of digital cameras K, we aggregated the individual judgments over the issues described in a concept vocabulary $\mathcal{C}$, in order to identify recurring patterns to create *arguments*. Arguments are knowledge structures that aggregate the practical knowledge of user experiences with respect to the different issues of an entity, and can be understood as reasons in favor or against buying or recommending the entity. Therefore, every argument is related with a certain issue of an entity, being the issues defined in a concept vocabulary $\mathcal{C}$.

Depending on the aggregated polarity of the argument, we find pro, con and moot arguments. Pro arguments are those in which the aggregation of the polarities of user judgments with respect to an entity's issue is positive. On the contrary, con arguments are those in which the aggregation of the polarities of user judgments with respect to an entity's issue is negative. Moot arguments are considered inconclusive: arguments in which there is not enough positive or negative evidence in the user experiences to consider them pro or con.

Furthermore, we defined three different ways of aggregating the judgment polarities of an entity with respect to an issue, creating three different types of arguments: the Gini argument Arg_G, the agreement argument Arg_σ, and the cardinality argument Arg_F.

Gini arguments (Arg_G) are created by combining the Gini coefficient with the average sentiment values of user judgments with respect to an entity's issue. If the Gini coefficient is low, meaning that the polarities of the user judgments with respect to an entity's issue are similar, the argument average sentiment is not weakened, while if the Gini coefficient is high then the argument average sentiment is weakened. Thus, Gini arguments penalize the aggregated sentiment value if the judgment's polarities with respect to an entity issue are very different among the individuals.

Agreement arguments (Arg_σ) are created by estimating, over a number of judgment's polarity values related to a concept, whether these values agree in the polarity of the sentiment or they are scattered. The aggregation process is similar to the one of the Gini arguments, however agreement arguments are defined by a threshold δ_{max} that specifies the maximum acceptable standard deviation over the distribution of judgment sentiment values of a concept in the reviews of a product. If the standard deviation of the polarities of set of user judgment's is bigger than δ_{max}, the agreement argument is considered a moot. Otherwise, the arguments are considered pros or cons normally, and the sentiment strength of the arguments is directly related to the average polarity of the user judgments.

Finally, cardinality arguments (Arg_F) are created by comparing the number of positive judgments versus the number of negative judgments of a concept over the reviews of a product. This type of arguments do not use the polarity value of judgments as assessed by the SmartSA sentiment analysis system. Instead, only the quantity of positive user judgments versus negative user judgments is compared in order to estimate whether the resulting argument is positive or negative, and to which degree. A strong cardinality bundle will be one in which the judgment polarities with respect to an entity's issue are mostly either positive or negative.

By considering the set of pro, con and moot arguments over the set of relevant issues of the concept vocabulary $\mathcal{C}$ with respect to an entity, we obtain the *bundle of arguments* B of that entity: that is, a characterization of that entity based on people's experiences over the relevant issues described in the individual judgments. Depending on the type of the arguments (Gini, cardinality, or agreement) that form the bundle, we can have a Gini bundle of arguments B_G, an agreement bundle of arguments B_σ, or a cardinality bundle of arguments B_F. Since the aggregation of the polarity of the judgments is different for the three types of arguments, the resulting bundles of arguments are also, in principle, different.

Notice that, since arguments are the result of aggregating user judgments with respect to an issue found in the reviews of an entity, the quality of an argument is directly related to the quantity of user judgments with respect to that issue found in the reviews of an entity. The more user judgments, the more information to identify more accurate recurring patterns in order to create better arguments. This is a limitation of our approach, since products with few reviews will most likely have less judgments, resulting in less accurate pro and con arguments, and more moot arguments. For this reason, our approach can only be applied meaningfully to domains where the corpus has enough number of reviews for the entities of that domain, in order to insure the adequate creation of arguments. Notice that, in Section 3.6.1 where we presented our corpora, we initially discarded those products with less than 15 reviews. Moreover, to evaluate the bundles of arguments against the pro and con lists of Dpreview in Section 5.5.2, we only considered the 15 products of the DSLR corpus K_D with more reviews.

In order to support the reuse of this practical knowledge, we introduce the notion of *user query* and *degree of query satisfaction*. The degree in which a query is satisfied by a bundle of arguments is defined using fuzzy logic (specifically, using a fuzzy implication operation). A user query expresses the individual preferences of a user over the set of relevant issues defined in the concept vocabulary, and the degree of query satisfaction by an argument bundle estimates to which extent a bundle of arguments satisfies the preferences expressed in the user query. By estimating the degree of query satisfaction over a set of bundles of arguments, we can assess which product better satisfies the preferences of a user on the basis of the practical knowledge we have acquired from a corpus of reviews.

Moreover, since all bundles of arguments are defined over the set of relevant issues of the concept vocabulary $\mathcal{C}$, we can estimate the degree in which a bundle is better or superior to another bundle by comparing its arguments over the set of issues of the concept vocabulary.

We created and evaluated the three types of bundles of arguments (B_G, B_σ, and B_F) for each of the digital cameras in the three corpus (K_D, K_C, and K_P), using their corresponding concept vocabularies ($\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$). We noticed that the quantity of pro and con arguments of the bundle of arguments is directly correlated with the average number of reviews of an entity: the more reviews of an entity, the more rich are the bundles of arguments for that entity. Finally, we compared the bundles of arguments we created from the user experiences in the reviews with the camera characterizations of Dpreview, a renowned website specialized in digital cameras. The results shows that the sets of con and pro argument of our cameras are very close to the pro and con features listed in Dpreview.

Moreover, we defined a function to compare two bundles of arguments, and using it we ranked a subset of the cameras with more reviews for the three corpus,

K_D, K_C, and K_P, obtaining a high correlation with the overall Dpreview score ranking. These results are a good indicator of the quality of the bundles and its consistency among the three camera types.

Finally, notice that the approach presented in this thesis integrates various state-of-the art tasks related to natural language processing, such aspect extraction and sentiment analysis, in order to extract knowledge from the experiences of the individuals to create the bundles of arguments of an entity. Consequently, the quality of the bundles of arguments of an entity is directly related to the accuracy of those tasks. Therefore, the approach presented in this monograph will be able to incorporate these new methods and improve the vocabularies and bundles of arguments, with these advances that we will surely surface in upcoming years.

Next section describes and discusses the main contributions of this monograph relevant to aspect extraction, semantic unification and concept creation, and how to acquire and represent practical knowledge in the framework of the Web of Experience.

6.2 Contributions

The first contribution of this monograph is related to aspect extraction and the creation of an aspect vocabulary.

In this monograph we presented an unsupervised aspect extraction method to create a vocabulary of aspects formed by a set of salient words from a corpus of user-generated entity reviews. The method used to create the aspect vocabulary presented in this monograph is domain independent; it does not use domain specific aspect extraction methods, and all aspects are extracted by analyzing the content of the user-generated reviews of a corpus. Furthermore, all aspects in the aspect vocabulary are lexical units frequently used in the reviews of the corpus, since we were only interested in the salient lexical units used by the individuals when describing their experiences with entities. Therefore, the methods presented in this monograph can be applied, given a corpus of user-generated reviews of entities, to create an aspect vocabulary independently of the domain of the corpus (e.g. hotels, laptops, cars, etc). Finally, as we show in Section 3.6, the resulting aspect vocabulary of a corpus is a rich aspect vocabulary, similar to what can be created by a human annotator.

The aspect vocabulary $\mathcal{A}$ of a corpus K is created with the union of four complementary aspect sets, $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$, and $\mathcal{A}_4$, that combine different natural language processing techniques to select different sets of aspects from the reviews of a corpus. For instance the method used to create aspect set $\mathcal{A}_4$ (see Section 3.5) focuses on extracting compound aspects that were not extracted by the techniques used to create aspect sets $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$. Although the best aspect vocabulary is created when joining the four aspects sets, notice that $\mathcal{A}_1$ and $\mathcal{A}_2$ are the aspects sets that contribute most to the final aspect vocabulary (see Section 3.6.3).

The methodology presented to create an aspect vocabulary is domain independent, albeit it is worth noticing the impact of the domain specific lexicon used to create aspect sets $\mathcal{A}_2$ and $\mathcal{A}_3$. However, as we show in Section 3.3, the creation of the lexicon can also be automated. For instance, in the experiments presented in Chapter 3, we used the PhotoDict taxonomy, a taxonomy we created from two websites specialized in digital cameras, to identify aspects such as 'external microphone' and 'exposure compensation' discarded by the grammatical extraction rules applied while creating $\mathcal{A}_1$.

It is important to notice that in this monograph, differently from opinion mining approaches such [Dong and Smyth, 2016], the aspect vocabulary $\mathcal{A}$ is created without leveraging the sentiment information of sentences and words, but only by exploiting the semantic structure of the sentences found in user-generated reviews.

The second main contribution of this monograph is related to solve the problem of semantic unification of aspects in the context of corpus reviews. Our approach is to create a concept vocabulary that intends to model the set of entity issues considered relevant by the individuals that shared their experiences in textual reviews. Given an aspect vocabulary and a corpus of reviews, a concept groups a subset of aspects that, according to our analysis, are used in an indistinguishable way (to a certain degree) in that corpus. The concept vocabulary induces a partition over the aspect vocabulary.

The creation of a concept vocabulary has two main stages. The first stage performs a hierarchical clustering process over the aspect vocabulary $\mathcal{A}$; the clustering is based on a semantic and lexical similarity measure. The second stage uses sentiment polarity analysis of the clusters, and selects the best set of clusters that (1) create a partition over $\mathcal{A}$ and (2) maximizes sentiment polarity coherence (Section 4.5.1) in the selected clusters.

In the first stage, aspects were grouped by a bottom-up hierarchical clustering with a similarity function that combined their semantic and lexical similarities between aspects with the parent-child relations of the PhotoDict taxonomy in a weighted function. This similarity function is specially interesting because it captures the way users use the different aspects of the aspect vocabulary in their reviews. For instance, suppose that a group of individuals frequently used the word 'pic' to describe the pictures of a camera, while others used 'image' for the same purpose. By means of this similarity function, we are able to assess that, in this particular corpus, the words 'pic' and 'image' are related, because those two words are used in similar contexts and similar sentences by different individuals.

The second stage is based on the idea of sentiment coherence of a concept (defined in Section 4.5.1). Based on the notion of basic level concepts (BLC), our goal is to select the partition of concepts with an adequate specificity from the clustering dendrogram, in order to model the set of entity issues described in the user experiences of a corpus.

Our assumption is that, if two aspects are referring to the same issue (and should therefore be part of the same concept), the polarities of the judgments related to those two aspects should have a high correlation (should cohere) over all judgments with respect to each product of a corpus. Therefore, we can identify the partition that better defines the set of issues addressed by the individuals in the reviews of a corpus by selecting the partition of concepts with a higher sentiment coherence from the clustering dendrogram. The selected partition forms the concept vocabulary.

Notice that in the process of creating a concept vocabulary we are combining the semantic, lexical and polarity information found in a set of textual reviews. That is, we are analyzing in various ways how the individuals use the set of aspects of the aspect vocabulary in order to group them accordingly. We are first using the semantic and lexical information to create a dendrogram of concepts, grouping aspects that are used similarly in the reviews. This clustering process could lead to clusters with very different granularity: we can have small and very specific clusters of similar aspects, and on the other hand bigger general clusters of aspects. Then, in order to select the groups with an accurate level of granularity, we utilize the sentiment coherence of the concepts from that dendrogram to select the best partition of concepts.

Moreover, the method used to create the concept vocabulary is fully parametrized. For instance, the similarity measure between aspects is defined by three weighting parameters (see Equation 8), corresponding to the semantic, lexical and PhotoDict similarity between two aspects. Therefore, we can decide which of these three similarities has more weight when assessing the similarity between aspects only by changing the corresponding weighting factor. Changing the similarity measure will lead to different clustering dendrograms, and therefore to different concept vocabularies. Our approach also allows the possibility to specify the minimum and maximum size of a partition.

Notice that in our work, differently from [Izquierdo et al., 2007], we do not select a set of concepts from an already existing ontology such as WordNet, but it is an unsupervised, bottom up approach that is aimed at discovering those issues more often discussed in the texts that describe people's experiences. When we create a concept from our set of aspects, we are not using theoretical knowledge obtained from an external source such as, in the case of cameras, a list of features defined by the manufacturer that are supposedly significant in characterizing that camera. Instead, we focus on the interaction of individual people in the use of an entity, and then analyze their description of those experiences when dealing with that entity, and the issues they choose to mention and judge, to create the concept vocabulary.

The third main contribution of this monograph is concerned on how to acquire and represent practical knowledge from the experiences with entities of a large group of individuals that have different preferences and biases.

In this monograph we present two related notations, the arguments and the bundles of arguments of an entity. Arguments are constructs formed with the practical knowledge present in the reviews of individuals, with respect to an issue of an entity. The problem to address here is that this practical knowledge expressed in judgments may vary from individual to individual according to their (in principle different) preferences and biases. Arguments are introduced to address this problem. Arguments are clear and consolidated collective judgments with respect to an issue of an entity, when they exist. When a collective judgment is inconclusive, we cannot acquire practical knowledge for that issue and we consider it a moot argument. Otherwise, we are able define pro and con arguments depending on the polarity of the collective judgments with respect to an issue of an entity: if the overall polarity of the individual judgments with respect to an entity issue is positive, we consider it a pro argument, if it is negative, we consider it a con argument.

The next contribution of this monograph are the definition of three methods to identify clear and consolidated collective judgments considering the biases and preferences of the individuals. These methods consist in three different ways of aggregating the polarity of individual judgments and determining when such aggregation is clear and consistent or not. Each method is used to create a different type of argument: Gini arguments Arg_G, agreement arguments Arg_σ, and cardinality arguments Arg_F.

A bundle of arguments for an entity is the collection of arguments, over the set of issues discussed by the individuals in the reviews of a corpus, with respect to the reviews of that entity. Thus, a bundle of arguments is a *characterization* of an entity based on people's experiences over the relevant issues described in the individual judgments, and a *representation* of the practical knowledge discovered from the individual experiences expressed in those judgments.

Moreover, argument bundles allow us to define the notions of user query and the satisfaction degree of a bundle by a user query. We show how, using standard notions from fuzzy logic, we can relate a query to a bundle by a measure of satisfaction degree. This shows that argument bundles are not only capable of representing practical knowledge but they are useful to perform inference: given a set of user preferences specified in a query, the fuzzy measure yields a degree of satisfaction for each entity's bundle of arguments. Consequently, we can give a personalized ranking of all entities with respect to a user query.

Finally, this monograph presents a proposal that includes all aspects of the Web of Experience framework, focused in textual user-contributed content and analytical (classification) tasks. From a set of individual user experiences with a class entities, expressed in textual form, we extracted practical knowledge and created knowledge structures to characterize each entity. This process was divided in three main parts, corresponding to the three chapters of this monograph: 1) the creation of an aspect vocabulary, formed by the set of salient lexical units used by the individuals to describe their experiences with digital cameras, 2) the creation of a concept vocabulary, built from the aspect vocabulary, that we use to model the set of issues described in the reviews of a corpus, and 3) the creation of the arguments and the bundles of arguments of entities, knowledge structures that aggregate the practical knowledge extracted from the individual's experiences, and the reuse of that knowledge via fuzzy queries.

Therefore, the results of this monograph, given a corpus of user-generated reviews of entities, are the aspect vocabulary used by the individuals in the reviews of the corpus, the concept vocabulary that describes the issues described in the set of reviews of the corpus, and the arguments and the bundles of arguments of the entities of a corpus, knowledge structures that facilitate organize the extracted knowledge and facilitate the selection of the entity that better fits a set of user preferences.

6.3 Publications

Our first publication focuses in analyzing and identifying the salient words used by people when defining their own experiences in textual reviews about products, in the context of social recommender systems and case based reasoning in collaboration with the Robert Gordon University of Aberdeen. In [Chen et al., 2014]:

- Chen, Y., Ferrer, X., Wiratunga, N., and Plaza, E. (2014). Sentiment and preference guided social recommendation. In *Proceedings of the 22nd International Conference on Case-Based Reasoning (ICCBR'14)*, pages 79–94. Springer.

we focus on two knowledge sources, user generated reviews of a product and preferences from purchase summary statistics, to create a formalism by which we could recommend products.

To do so, we define an unsupervised aspect extraction method to obtain salient aspects from product reviews, and compare it to three state of the art aspect extraction techniques, using the manually tagged mobile phone corpus presented in [Ding et al., 2008; Hu and Liu, 2004a]. The unsupervised aspect extraction algorithm used in the paper corresponds to one of the first versions of the algorithm presented in Section 3.2. The salient words identified in the product reviews are aspects: uni-gram and bi-grams words used by people to describe their experiences with cameras. Similarly as we do in this thesis, the polarity of those sentences related to each aspect from the reviews of a product is estimated using the SmartSA sentiment analysis system presented in Section 4.4, and used to characterize a product. Furthermore, by exploiting the product preference graph from Amazon, we identify those aspects mostly possessed by the most preferred products and not by others. We assigned weighting factors to every aspect accordingly, in a way that the most preferred aspects, those aspects belonging to preferred products, had more importance when recommending a product.

By combining and weighting the aspect sentiment information of every product together with a preference rank value of a product estimated using the PageRank, we obtained a product score for every product. This score was then used to recommend better products based on a query product, and compare the results with the list of top sellers from Amazon. The results showed that higher precision in aspect extraction from user reviews was achieved with the grammatical extraction rules, and that aspect weighting and preference knowledge can be conveniently exploited to recommend based on user experiences.

In [Ferrer et al., 2014], also in collaboration with the Robert Gordon University of Aberdeen, we used a similar approach to create a set of aspect-sentiment pairs to characterize the products of our DSLR corpus, analyzing how users' preferences and interests change over time, and tracing aspect importance trends:

- Ferrer, X., Chen, Y., Wiratunga, N., and Plaza, E. (2014). Preference and sentiment guided social recommendations with temporal dynamics. In *Research and Development in Intelligent Systems XXXI*, pages 101–116. Springer.

Tracking users preference over time raises unique challenges for recommendation systems, since every product potentially goes through a series of changes which typically involves functional improvements. What makes a product feature interesting now may become the accepted standard in the future. Since the corpus we are using contains products from 2008 to 2014, we decided to study how time affects the recommendations and characterizations of products, and formalized an aspect-based sentiment ranking that utilizes both preference and time contexts.

The benefits were demonstrated in a realistic recommendation setting using benchmarks generated from Amazon and Cnet[1]. We show that monitoring aspect frequency in user generated reviews allows to capture changes to aspect importance over time. For instance, Table 6.1 shows the most frequent aspects used to describe DSLR camera experiences from 2008 to 2012. We observe how aspect 'resolution' was the most used aspect in 2008, however, this aspect's importance diminished during the following years finally disappearing from the top 10 in 2010. However, other aspects such as 'battery', 'feature' and 'picture' maintain their positions in the ranking for years. Finally, in this paper we confirm that time context can be conveniently exploited by using the recent time frame to improve recommendations.

In [Chen et al., 2015b], we extended our previous work on aspect extraction and social recommender systems to harness knowledge from product reviews, and explored the utility of frequency based approach and supervised Information Gain (IG) to rank and select the most useful aspects for recommendation:

- Chen, Y., Ferrer, X., Wiratunga, N., and Plaza, E. (2015b). Aspect selection for social recommender systems. In *Proceedings of the 23rd International Conference on Case-Based Reasoning (ICCBR'15)*, pages 60–72. Springer.

In previous papers we found that the quantity of aspects used to characterize products was very large, and some of the aspects that formed the aspect vocabulary were not related to the photography domain. Furthermore, synonyms and acronyms were common among the aspects of the aspect vocabulary. Thus there was a need to further explore the aspect selection algorithms to improve the aspect vocabularies we used to characterize products.

[1] https://www.cnet.com/topics/cameras/products/

2008	2009	2010	2011	2012
resolution (0.011497558)	photographer (0.010162692)	photographer (0.0102211265)	photographer (0.010785269)	photographer (0.009730392)
photographer (0.010942243)	***picture (0.009932901)***	***picture (0.010030088)***	***picture (0.010662347)***	***picture (0.009580281)***
software (0.010741603)	feature (0.009924242)	feature (0.010021289)	feature (0.010643691)	feature (0.009560834)
feature (0.010699623)	battery (0.009656516)	setting (0.009714422)	battery (0.010373606)	battery (0.009321224)
picture (0.010689487)	setting (0.009596125)	battery (0.009521719)	setting (0.010341785)	setting (0.009276917)
battery (0.01046086)	reason (0.009533786)	result (0.009065794)	result (0.009634106)	result (0.008670973)
setting (0.010346161)	**resolution (0.009430549)**	photography (0.008940687)	photography (0.009578535)	photography (0.008653971)
reason (0.010275534)	result (0.009028885)	reason (0.008864645)	video (0.009363669)	video (0.008407041)
noise (0.01002432)	photography (0.008967144)	image (0.008774484)	capability (0.009352991)	image (0.008382303)
result (0.00971449)	software (0.008902968)	video (0.008729932)	image (0.009333819)	time (0.008291209)

Table 6.1: Top 10 most used aspects in DSLR reviews, ordered by relative frequency between 2009 to 2013.

In this paper we decided to explore different ways of reducing the dimensionality of the aspect vocabulary, by using information gain and other techniques to rank and decide which are the most salient aspects. After creating different aspect vocabularies with the different aspect ranking strategies, products were recommended, and recommendations evaluated similarly as in previous papers. In the experiments, we confirmed that aspect selection using feature selection techniques improved product recommendations, specially when using a small subset of features selected with the semi-supervised information gain algorithm. Finally, in this paper we introduced the three distinct camera types that form part of our actual corpora: DSLR, Compact and Point & Shoot, and evaluated the similarity between the aspect vocabularies of the three camera types.

The aforementioned papers [Chen et al., 2014, 2015b; Ferrer et al., 2014] partially cover the Challenges 1 and 2 ("Identify the set of salient terms, in a given domain, used by people while writing about their experiences with a concrete entity of that domain." and "Identify user judgments and their polarity.", respectively). However, the unsupervised aspect extraction and the creation of the aspect vocabulary from user reviews were improved in Chapter 3 of this monograph by exploiting domain knowledge available in two photography websites in order to improve the precision and recall of the extracted aspects.

In [Ferrer and Plaza, 2016], we present a complete account of the Web of Experiences approach, which is based on solving new problems by learning from past experiences. In this paper, we extend our previous work on aspect extraction and sentiment analysis and propose a novel approach to create a vocabulary of basic level concepts with the appropriate granularity to characterize a set of products. This vocabulary is created from analyzing the usage of the aspects over product reviews (the experiences of users), finding those product features with a clear positive or negative polarity to create bundles of arguments. The bundles of arguments are collections of pro and con reasons that have been generated by obtaining and analyzing the practical knowledge of user's experiences from the reviews of a product; and allow us to reuse past experiences of users by means of a user query and a degree of query satisfaction, defined in fuzzy logic.

This work is presented in the following publication:

- Ferrer, X. and Plaza, E. (2016). Concept discovery and argument bundles in the Experience Web. In *Proceedings of the 24th International Conference on Case-Based Reasoning (ICCBR'16)*, pages 108–123. Springer.

The aspect vocabulary used in this work is improved from previous publications, and is the one described in Chapter 3. Furthermore, in this paper we introduce the idea of concept vocabulary, which models the RELEVANT issues used by people when expressing their experiences in product reviews. As such, it partially covers Challenge 3 ("Discover the main issues addressed by people when writing their experiences and create a concept vocabulary."). See chapter 4 for a detailed description of the concept vocabulary creation.

Furthermore, we present the arguments, and the bundles of arguments, which are the knowledge structures created from the user experiences with respect to a product. As such, this paper also partially covers Challenge 4 ("Create the arguments and the bundle of arguments of an entity."). This paper also introduces the concept of user query and the semantics of query satisfaction by the argument bundles presented in this monograph, based on fuzzy logic, and performs an evaluation of the resulting knowledge structures.

6.4 Open Issues and Future Work

In this section we describe several open issues that raise from the research presented in this monograph, and the corresponding lines of future work that could address them. The future work lines presented in this section range from improving the aspect extraction methodology and the detection of judgment polarity, to study the temporal dynamics of the bundles of arguments.

6.4.1 Improving Aspect and Judgment Detection

In this monograph, the detection of user judgments and the creation of the aspect vocabulary play an important role (see Chapters 3 and 4 for reference). Aspects are identified by means of grammatical extraction rules presented in [Moghaddam and Ester, 2010], and, although improved to fit the task of extracting important aspects from user reviews, they are limited to uni-grams and bi-grams. Improving the aspect extraction rules by considering n-grams will result in a more rich vocabulary, and would improve the argument bundles of those products with less reviews (since actually we are missing some aspects such as 'electronic viewfinder speed' and similar).

Furthermore, we consider that we can improve the detection of user judgments by leveraging the particular structures of natural language commonly found in user reviews [Zhang et al., 2010]. In this dissertation, the techniques used to identify user judgments are not domain specific, in the sense that can be used to identify user judgments in a very wide range of texts. However, we think that better results in the identification of user judgments and the polarity of those judgments can be achieved by adapting this techniques to our domain, by, for instance, using a domain specific sentiment vocabulary to determine the polarity of the judgments (see Section 6.4.2).

6.4.2 Improving Domain Specific Sentiment Analysis

In Chapter 4 we analyze the polarity of user's judgments in reviews. We use the SmartSA sentiment analysis system to assess the polarity. SmartSA uses SentiWordNet, a general sentiment lexicon where every WordNet synset has three sentiment scores associated, positivity, objectivity, and negativity, to assess the polarity of words. The positive, negative, and objective sentiment values of each word in SentiWordNet are related with WordNet synsets. The problem occurs when we are trying to assess the polarity of a word whose synset is not found in WordNet, or changes it's polarity depending on the domain. This problem is often solved by using the sentiment value associated with the most common WordNet synset for that word. However, this sentiment value might fail to correctly identify the polarity of the domain specific word, since we are, in fact, talking about two different meanings (associated with two different domains) of the same word.

We studied the impact of this problem in our camera domain by analyzing if the polarity values given by the SmartSA system correspond with the real polarity, assessed manually, for a small set of domain specific camera related words. We found that there were not much camera specific words whose polarities were incorrectly assessed by the SmartSA system. However, some of the incorrectly assessed polarities were related with important words of the camera domain.

Table 6.2 presents some of those words whose polarity was incorrectly assessed by the SmartSA system. The symbol '+' represents a positive polarity, '0' a neutral polarity, and '-' a negative polarity. For instance, aspect 'RAW' is considered a word with a negative polarity by the SmartSA system, but should be considered neutral (or maybe even positive) when talking about the RAW file format of digital cameras. Another example is the word 'small'. Smaller is a positive factor in almost all digital cameras, since it allows for easier transportation, however it is considered a negative factor by the SmartSA system probably because small has a commonly negative connotation in most domains. The last three rows of Table 6.2 present sentences with a clearly positive polarity, assessed as neutral by the SmartSA sentiment analysis system.

In this dissertation we decided to not create a domain specific sentiment dictionary because we observed that the impact of the incorrectly guessed polarities was not important. Furthermore, sometimes it was not clear if the mistakes in assessing the polarity of some words and sentences were due to the prior sen-

Word	SmartSA polarity	Expected domain polarity
Small	-	0 or +
RAW	-	0 or +
GPS	-	0 or +
Cheaper	0	+
Less shutter lag	-	+
Faster max shutter speed	0	+
Faster autofocus	0	+

Table 6.2: Comparison between SmartSA polarities and the expected domain specific polarities of some words considering the domain of digital cameras.

timent information associated to the SentiWordNet sentiment lexicon, or some problems related with how the SmartSA system analyzes the natural language structure of the sentences. And moreover, we decided that the creation of a completely new sentiment analysis system was out of the scope of this dissertation because of the time constraints we were facing.

However, we consider that the improvement of the sentiment analysis techniques with respect to our particular domain of digital cameras would improve numerous parts of this dissertation: the aspect vocabulary, the concept vocabulary, and the bundles of arguments. A possible solution would be to create a domain specific sentiment dictionary related with digital cameras, although further analysis on the impact of those incorrectly assessed word polarities would be necessary.

6.4.3 Creating a More Complex Concept Vocabulary

In this work, we created the concept vocabulary in Chapter 4 by analyzing the way words are used in the reviews of products and leveraging the idea of basic level concepts used in cognitive linguistics. However, right now, all concepts of the concept vocabularies are found at a certain level of granularity. That means that when a concept is created (by selecting a partition from the hierarchical clustering dendrogram), the level of granularity of this concept is decided, and does not change unless a new concept vocabulary is created.

We consider that we can explore the creation of more complex concept vocabularies to better define the bundles of arguments of each camera. For instance, by combining vocabularies with different granularity when creating the bundles of arguments.

6.4.4 Modeling Query Satisfaction Using Different Operators

In Chapter 5 we presented the user query and the concept of query satisfaction by a bundle of arguments. The degree of satisfaction of a query by a bundle of arguments is modeled using fuzzy logic and estimates to which extent a product satisfies the preferences expressed in a query.

In this dissertation we used the fuzzy implication associated to the t-norm product ($\Rightarrow_\otimes$), together with the t-norm product, to model this degree of satisfaction. However, the degree of query satisfaction could also be modeled using other implications associated to other t-norms, such as the Lukasiewicz t-norm or the minimum t-norm. The difference between modeling the degree of query satisfaction by using different t-norms is that they estimate the degree of satisfaction differently. For instance, the minimum t-norm is less strict than the Lukasiewicz t-norm. Consequently, different bundles of arguments could be retrieved by utilizing different implications associated to other t-norms.

6.4.5 Determining the Domain(s) of a Corpus

Determining if a corpus is formed by one unique domain, or can be split in several domains, is important in order to create accurate and representative vocabularies, as we have shown in this monograph. However, it is not a straightforward task: How can you assess the different domains, if more than one, of the entities in a corpus? Do we need to analyze the entities, review per review, in order to identify similarities between entities and define different domains?

In this monograph, we analyzed and compared the aspect and concept vocabularies of three camera types (that are commonly used in digital photography terminology, and not just Amazon). We found that they were noticeably different, and therefore we created three different corpora (K_D, K_C, and K_P). However, in other domains the typing of entities may not be clear or consensual in their terminology. Therefore, we consider that further research needs to be done in this field, considering the impact it might have in the resulting vocabularies created from user-generated reviews of entities.

6.4.6 Recommending with Hard and Soft Constraints

With our definition of a query and the query satisfaction model presented in Chapter 5, we cannot define queries such as “I do not want a camera with gps” or “I want only cameras that cost less than 500 euros”. This caveat makes impossible the creation of a recommendation system that fully leverages the knowledge of the bundles of arguments, because a simple query such as “I want a black camera” cannot be modeled using our query system.

The inclusion of hard constraints in the user query could overcome the actual limitations of the query system. This way, requirements such as “camera without gps” could be modeled, discarding automatically those bundles of arguments that contain the undesired feature when estimating the degree of query satisfaction.

Furthermore, including hard constraints will be the first step to create a recommendation system. With that being said, it is important to point that the creation of a recommender system was not the goal of this dissertation, but we acknowledge that it may be an interesting way to allow the reuse of user experiences and the experiential knowledge expressed with the bundles of arguments.

6.4.7 Bundle Visualization

In this monograph we extracted and modeled practical knowledge from user-generated reviews about digital cameras to create knowledge structures able to represent this information. These structures can then be reused in numerous ways and to various ends by other people interested in the digital camera domain.

However, we think it would be interesting to create an online platform to present, in a user-friendly way, the bundles of arguments of the different cameras used in this monograph. By creating a bundle visualization tool, we would facilitate the reuse of the practical knowledge contained in the bundles of arguments, as people will be able to easily understand the positive and negative arguments associated to every camera and the arguments in each bundle. Furthermore, an online bundle visualization platform will also help disseminating the research presented in this monograph.

6.4.8 Temporal Dynamics of the Bundles of Arguments

Time is an important variable to consider when working with user reviews, as it has a remarkable impact on people's opinions and language: The vocabulary used in user-generated content and the sentiment of people with respect to the different aspects of the entities described in user reviews changes over time.

We studied the temporal dynamics of an aspect vocabulary in [Ferrer et al., 2014]. In Appendix B of this monograph, we present a summary with the most relevant findings. Based on the results obtained, we think it would be interesting to further explore temporal dynamics with respect to the argument structures presented in this monograph. What effect has time on the bundles of arguments? Do *old* arguments become invalid to characterize digital cameras in the present? Or can their knowledge be reused?

Appendices

Appendix A

Notation

A.1 Corpora and Products

- K a corpus of reviews; can be one of K_D, K_C or K_P, that correspond to the three camera types DSLR, COMPACT and PAS respectively.
- $p \in \mathcal{P}$. A product p belongs to the set of products $\mathcal{P}$ in a K.
- $Rev(p)$. The set of reviews of a product p.
- $K = \bigcup_{i=1...|\mathcal{P}|}\{(p_i, Rev(p_i))\}$. Each corpus is formed by a set of product-reviews pairs, where p_i is a digital camera in $\mathcal{P}$, and $Rev(p_i)$ is the set of reviews about camera p_i.

A.2 Aspects

- $\mathcal{A}$ an aspect vocabulary; can be one of $\mathcal{A}_D$, $\mathcal{A}_C$ or $\mathcal{A}_P$, that correspond to the three camera types DSLR, COMPACT and PAS respectively.
- $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$. The aspect vocabulary is the union of the four aspect sets.
- $a \in \mathcal{A}$. An aspect a belongs to the aspect vocabulary a in a $\mathcal{A}$.
- $Occ(p, a)$. Occurrences of aspect a in reviews of product p, i.e. the sentences from the reviews $Rev(p)$ in which aspect a occurs.
- $s(x, a) \in [-1, 1]$. The sentiment value, ranging from -1 to +1, for aspect a occurring in sentence x.
- $S_{av}(p, a) = \frac{1}{M} \cdot \sum_{x \in Occ(p,a)} s(x, a)$, where $S_{av}(p, a) \in [-1, 1]$ and $M = |Occ(p, a)|$. Average sentiment of an aspect a in the reviews of product p.

- $Dev(p,a) = \sqrt{\frac{1}{M} \cdot \sum_{x \in Occ(p,a)} (S_{av}(p,a) - s(x,a))^2}$ where $M = |Occ(p,a)|$. Standard deviation of the sentiment values for aspect a in the reviews of product p.

A.3 Partitions and Aspect Groups

- $G = \{a_1, \ldots, a_n\}$. An aspect group G is formed by a set of aspects $a \in \mathcal{A}$.

- $K = \{G_1, \ldots, G_n\}$. A partition K is formed by a set of aspect groups G. The partition with higher $R(K)$ will form a concept vocabulary $\mathcal{C}$.

- $Sim(G_i, G_j) = \frac{1}{|G_i||G_j|} \sum_{n=1}^{|G_i|} \sum_{m=1}^{|G_j|} SimA(a_n, a_m)$. The similarity between two groups of aspects G_i and G_j is defined as the average similarity between all combinations of aspects of the two groups, where a_n and a_m are aspects from the aspect groups G_i and G_j respectively.

- $SimA(a_i, a_j) = \alpha \cdot \Gamma(a_i, a_j) + \beta \cdot \Delta(a_i, a_j) + \gamma \cdot \Lambda(a_i, a_j)$. The similarity between two aspects a_i and a_j is defined as a weighted addition of the semantic similarity, string similarity and PhotoDict similarity between two aspects, where α, β and γ are weighting parameters in $[0,1]$ such that $\alpha + \beta + \gamma = 1$, and the values of $SimA$ are in $[0,1]$

- $Occ(\mathcal{P}, G)$. Is the set of occurrences of all aspects forming the aspect group G considering all products in $\mathcal{P}$.

- $R(K) = \frac{1}{|K|} \sum_{i=1}^{|K|} IS(G_i)$. The Partition Ranking score $R(K)$ of a partition K is the average sentiment similarity of all aspect groups G in K.

- $IS(G) = \frac{1}{|G| \cdot (|G|-1)} \sum_{i=1}^{|G|} \sum_{j=1, j \neq i}^{|G|} Sim(D_i, D_j)$. The average sentiment similarity IS of a group of aspects G as the average Sim among the polarity profiles of all pairs of aspects in G, where D_i and D_j are the polarity profiles of aspect a_i and a_j respectively, and $Sim(D_i, D_j)$ computes the cosine similarity between aspect polarity profiles.

- $D(a) = (S'_{av}(p_i, a))_{i \in 1 \ldots |\mathcal{P}|}$ is the polarity profile of aspect a. D is a vector of normalized sentiment averages of aspect a over the set of products $\mathcal{P}$, where $S'_{av}(p_i, a)$ is in [0,1].

A.4 Concepts

- $\mathcal{C}$ a concept vocabulary; can be one of $\mathcal{C}_D$, $\mathcal{C}_C$ or $\mathcal{C}_P$, that correspond to the three camera types DSLR, COMPACT and PAS respectively.

- $C \in \mathcal{C}$. A concept C belongs to the concept vocabulary $\mathcal{C}$.

- $Occ(p, C)$. Occurrences of concept C in reviews of product p, i.e. the sentences from the reviews $Rev(p)$ in which any aspect of concept C occurs $Occ(p, C) = \bigcup_{a \in C} Occ(p, a)$.

- $s(x, C) \in [-1, 1]$. The sentiment value for the aspects of C occurring in sentence x. $s(x, C) = \frac{1}{\sum_{a \in C} |s(x,a)|} \sum_{a \in C} s(x, a)$.

- $S_{av}(p, C) = \frac{1}{M} \cdot \sum_{x \in Occ(p,C)} s(x, C)$, where $S_{av}(p, C) \in [-1, 1]$ and $M = |Occ(p, C)|$. Average sentiment of a concept C in the reviews of product p.

- $Dev(p, C) = \sqrt{\frac{1}{M} \cdot \sum_{x \in Occ(p,C)} (S_{av}(p, C) - s(x, C))^2}$ where $M = |Occ(p, C)|$. Standard deviation of the sentiment value for a concept C in the reviews of product p.

A.5 Arguments and Bundles

- $V(p, C) = \{s(x, C)|x \in Occ(p, C)\}$. Collection of sentiment values ranging from [-1,1] of the judgments found in Occ(p,C) sentences, and $|V(p, C)| = |Occ(p, C)|$.

- $Arg = \langle p, C, s \rangle$. An argument is a tuple formed by a product $p \in \mathcal{P}$, a concept $C \in \mathcal{C}$, and an aggregated sentiment value s.

- An Arg can be one of Arg_G, Arg_σ, Arg_F, that corresponds to the three types of arguments Gini argument, Agreement argument and Cardinality argument, respectively.

- Arg^+ if $Arg.s > \delta$. An argument is considered a pro argument if the aggregated sentiment value is bigger than a threshold parameter δ.

- Arg^- if $Arg.s < -\delta$. An argument is considered a con argument if the aggregated sentiment value is smaller than a threshold parameter $-\delta$.

- $Arg^\emptyset$ if $-\delta \leq Arg.s \leq \delta$. An argument is considered a moot argument if the aggregated sentiment is between $-\delta \leq Arg.s \leq \delta$.

- $Gini(p, C) = \frac{1}{n} \left(n + 1 - 2 \frac{\sum_{i=1}^{n} (n+1-i) v_i)}{\sum_{i=1}^{n} v_i} \right)$. The Gini coefficient between the judgments of concept C of product p returns a value in $[0, 1]$, where $v_i \in \vec{V}$, $\vec{V}$ is the polarity vector $V(p, C)$ ordered from lower to higher sentiment value, and $n = |V(p, C)|$.

- $O^+(p, C) = |\{x \in Occ(p, C)|s(C, x) > 0\}|$. The quantity of positive occurrences of a concept C in the reviews of a product p.

- $O^-(p, C) = |\{x \in Occ(p, C)|s(C, x) < 0\}|$. The quantity of negative occurrences of a concept C in the reviews of a product p.

- $B(p) = \bigcup_{C \in \mathcal{C}} Arg\langle p, C, s\rangle$. The bundle of arguments of a product $p \in \mathcal{P}$ is the union of arguments with respect to the set of concepts in $\mathcal{C}$.

- $Pros(p) = \{Arg^{+} \in B(p)\}$. The set of pro arguments of product p of the bundle of arguments $B(p)$.

- $Cons(p) = \{Arg^{-} \in B(p)\}$. The set of con arguments of product p of the bundle of arguments $B(p)$.

- $Moots(p) = \{Arg^{\emptyset} \in B(p)$. The set of moot arguments of product p of the bundle of arguments $B(p)$.

- $\overline{B}(p) = \bigcup_{Arg \in B(p)} Normalize(Arg)$. A normalized bundle of arguments of a product $\overline{B}(p)$ is formed by the set of normalized arguments of the bundle of arguments $B(p)$.

- $Normalize(Arg\langle p, C, s\rangle) = \overline{Arg}\langle p, C, \overline{s}\rangle$, where $\overline{s} = f'(s,\ s_{min}(C, Args), s_{max}(C, Args))$. A normalized argument $\overline{Arg}$ is created by normalizing the sentiment s of an argument Arg with f':

$$f'(s, min, max) = \begin{cases} \frac{s}{max} & \text{if } s > 0 \\ -\frac{s}{min} & \text{if } s < 0 \\ 0 & otherwise \end{cases} \tag{A.1}$$

- $\Phi(\overline{B}(p_i), \overline{B}(p_j)) = \frac{1}{2|\mathcal{C}|} \sum_{k=1}^{|\mathcal{C}|} s_k^i - s_k^j$. Degree in which a normalized bundle $\overline{B}(p_i)$ is better or superior to another normalized $\overline{B}(p_j)$, where s_k^i and s_k^j are the sentiment values of respective arguments $\langle p_i, C_k, s_k^i\rangle$ and $\langle p_j, C_k, s_k^j\rangle$ in the normalized bundles of p_i and p_j.

- $\widehat{B}(p) = \bigcup_{Arg \in B(p)} Rescale(Arg)$. A rescaled bundle of arguments of a product $\widehat{B}(p)$ is formed by the set of rescaled arguments of the bundle of arguments $B(p)$. Normalized bundles $\overline{B}(p)$ can also be rescaled.

- $Rescale(Arg\langle p, C, s\rangle) = \widehat{Arg}\langle p, C, \widehat{s}\rangle$, where $\widehat{s} = \frac{s+1}{2}$. A rescaled argument $\widehat{Arg}$ is created by rescaling the sentiment s of an argument Arg. Normalized bundles $\overline{B}(p)$ can also be rescaled.

- $Sp(r_1, r_2) = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$ is the Spearman rank correlation between two rankings r_1 and r_2, $d = r_1(a_i) - r_2(a_i)$ is the difference between the two ranks of the same element a_i in r_1 and r_2, and n is the quantity of common elements between the two aspect vocabularies.

A.6 User Query and Degree of Satisfaction

- $Q = \{(C_j, U(C_j))\}_{j=1,\ldots,k}$ where $k \leq |\mathcal{C}|$. A query Q is formed by a set of concept-utility pairs $(C_j, U(C_j))$ that express the user preferences over concept C_j, and where $C \in \mathcal{C}$.
- $U(C) \in [0,1]$. A utility function U assigns a user preference value to a concept $C \in \mathcal{C}$.
- The Satisfaction degree in which an argument with respect to concept C_j satisfies a user preference related with C_J:

$$U(C_j) \Rightarrow_{\otimes} \widehat{s_j} = \begin{cases} 1, & \text{if } U(C_j) \leq \widehat{s_j} \\ \frac{\widehat{s_j}}{U(C_j)} & \text{otherwise} \end{cases}$$

  where $\widehat{s_j}$, is the rescaled sentiment value of argument $\langle p, C_j, \widehat{s_j} \rangle$.

- $DS(Q, \widehat{B}(p)) = \prod_{j=1}^{k} (U(C_j) \Rightarrow_{\otimes} \widehat{s_j})$. Degree of bundle satisfaction (DS) of a query Q over a rescaled bundle $\widehat{B}(p)$. DS is the conjunction of the resulting concept wise satisfaction degrees between the arguments of a rescaled bundle and the query preferences of a user, using the t-norm product.

Appendix B

Aspects Over Time

In this appendix we present an evaluation about how the aspect vocabulary and user preferences change over time in the context of user-generated reviews and temporal dynamics [Ferrer et al., 2014]. Temporal dynamics is a crucial dimension for both content and collaborative approaches, because what makes a product feature interesting now may become the accepted standard in the future. The vocabulary used in user-generated reviews of products, and more importantly, the individual criteria when evaluating or describing the different aspects of a product in a review, may change influenced by the temporal dimension. Therefore, temporal dynamics are of great interest in the context of the Web of Experiences and those systems that leverage the information from user-generated content.

In this work, we create a social recommender system using sentiment rich user generated product reviews and temporal information. Specifically we integrate these two resources to formalize a novel aspect-based sentiment ranking that captures temporal distribution of aspect sentiments and so the preferences of the users over time. We demonstrate the utility of our proposed model by conducting a comparative analysis on data extracted from `Amazon.com` and `Cnet`. We show that considering the temporal preferences of users leads to better recommendation, and that preferences and the vocabulary of users change over time.

B.1 Social Recommendations with Temporal Dynamics

Recommender systems traditionally provide users with a list of recommended items based on users preferences. The huge success of these systems in the retail sector demands innovative and improved recommendation algorithms. The dawn of the social web has created opportunities for new recommendation algorithms to utilize knowledge from such resources and so the emergence of social recommender system. These systems harness knowledge from user generated reviews to generate better recommendations by incorporating sentiment expressed in opinions to bias the recommendation list [Dong et al., 2013b]. Similarly preference knowledge and temporal dynamics have also separately been applied to influence recommendations [Hong et al., 2012; Vasudevan and Chakraborti, 2014].

Purchase choices are based on comparison of artifacts; which implicitly or explicitly involve comparison of characteristics or aspects of these artifacts. In particular a user's purchase decision hints at the aspects that are likely to have influenced their decision and as such be deemed more important. Additionally it is also not unusual to expect that the criteria used for this comparison may also change with time. For example, in the domain of Cameras, the LCD display may have been an important aspect users were interested in the past but now this is given in almost every camera and so is likely to be an aspect of contention.

In recent work [Chen et al., 2014] we explored how preference knowledge can be captured and exploited within a social recommendation application. Our findings suggested that preference knowledge allows us to extract important aspects from reviews, in terms of those that are likely to have influenced the users' purchase decision. However, would recency of reviews have an impact on aspect weights? How far back in time must we go before extracted weights improve the recommendations? Our main focus in this paper is to study temporal and preference context for social recommendations with a view to integrate these contexts with aspect-based sentiment analysis.

Our contribution is three-fold: firstly we demonstrate how sentiment distribution analysis can impact the quality of recommendations; and secondly show how a preference-based algorithm can be incorporated to derive rankings on the basis of preference relationships; and finally provide a formalism to combine sentiment and temporal information. Our results confirm that incorporating temporal information in aspect-based sentiment analysis is comparable to preference knowledge.

The rest of the paper is organized as follows: In Section B.2 we present the background research related to this work. Next in Section B.3 we describe aspect preference over time and how preference graphs can be generated by using a case study from Amazon.com. The process of aspect extraction and weight learning for sentiment analysis is presented in Section B.4. Finally, our evaluation results are presented in Section B.5 followed by conclusions in Section B.6.

B.2 Related Work

Social recommender systems recognize the important role of sentiment analysis of user reviews [Dong et al., 2013b]. Instead of relying on user logs and sessions to model user preference [McCarthy et al., 2010], in this paper we infer aspect preferences from comparing the sentiment-rich content generated by users. However, extracting sentiment from natural language constructs is a challenge. Lexicons are often used to ascertain the polarity (positive or negative) and strength of sentiment expressed at word-level (e.g. SentiWordNet [Esuli and

Sebastiani, 2006]). However sophisticated methods are needed to aggregate these scores at the sentence, paragraph and document level to account for negation and other forms of sentiment modifiers [Muhammad et al., 2013]. Increasingly aggregation is organized at the aspect level, since the distribution of a user's sentiment is typically mixed and expressed over the aspects of the artifact (e.g. I love the *color* but not too keen on *size*). Hu and Liu [Hu and Liu, 2004a] propose an association mining driven approach to identify frequent nouns or noun phrases as aspects. Thereafter sentences are grouped by these aspects and sentiment scores assigned to each aspect group [Moghaddam and Ester, 2010]. Whilst there are many other statistical approaches to frequent noun extraction [Popescu and Etzioni, 2007]; others argue that identifying semantic relationship in text provides significant improvements in aspect extraction [Moghaddam and Ester, 2012]. Here we explore how semantic based extraction can be augmented by frequency counts.

Temporal dynamics is a crucial dimension for both content and collaborative approaches. Initial work on concept drift was applied to classification tasks with focus on optimising the learning time window [Kolter and Maloof, 2003]. More recently with recommender systems, temporal knowledge in the form of rules was used to predict purchase behavior over time [Cho et al., 2005]. Similarly the temporal influence on changes in user ratings has been observed on the Netflix movies [Koren, 2010], and short and long-term preference changes on products [Xiang et al., 2010]. The association of importance weights to aspects according time is not new [Ding and Li, 2005]. Here they generate aspect weights that are a function of time. Whilst our work also acknowledges the need for time-aware aspect weight learning, we exploit knowledge from both user review histories and preferences for this task.

B.3 Social Recommendation Model

An overview of our proposed process appears in Figure B.1. The final outcome is a recommendation of products (hitherto referred to as artifacts) that are retrieved and ranked, with respect to a given query product. Central to this ranking are aspect weights, which are derived from two knowledge sources: sentiment rich user generated product reviews and preferences from purchased summary statistics. Generally preference knowledge is captured in a graph according to purchase behavior and reviews depending on recency will influence both weight extraction and ranking algorithms. Here, we are interested in exploring how aspects can be weighted. Accordingly alternative aspect weight functions, AW_i, will be explored by taking into account the time and preference contexts of aspects. These in turn will influence the sentiment scores assigned to each extracted aspect, AS_i. Therefore the final ranking of products is based on an aspect weighted sentiment score aggregation.

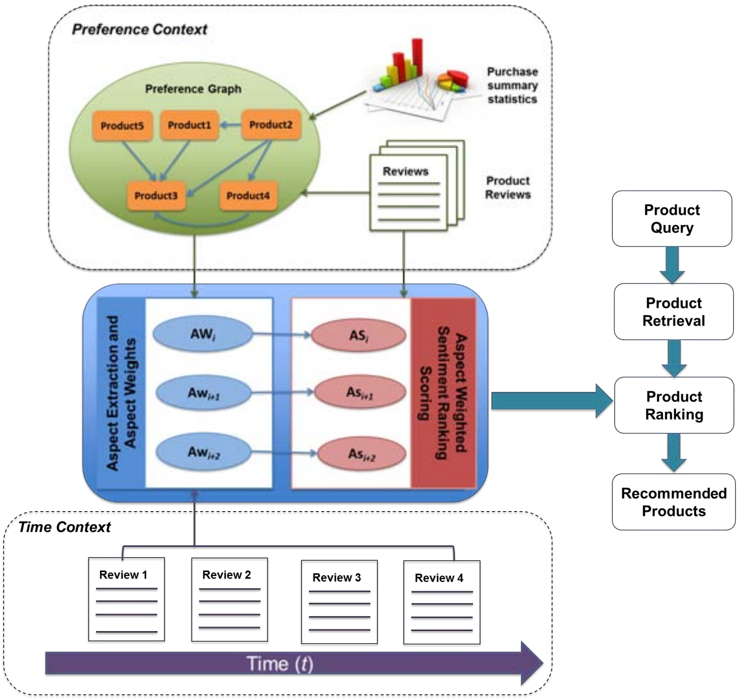


Figure B.1: Social recommendations with temporal dynamics

B.3.1 Time context

New and improved product aspects grow over time. While there are product aspects that are continuously improving, others stabilize when a majority of the products possess them. We can observe such trends on a sample of data collected from Amazon between 2008 and April 2014 (see Table B.1). Here we summarize the statistics of aspect frequency and show the top 10 most frequently mentioned aspects. In 2008, *resolution* was most frequent, however, this aspect's importance diminished during the following years finally disappearing from the top 10 in 2010. On the other hand, aspects like *picture* or *battery* remain in the top 10 list. *Resolution* might have been an important aspect for a camera in 2008, but it is no longer a determinant as the majority of the SLR cameras are now equipped with high resolution. On the contrary, users keep seeking for better pictures or longer battery life in SLR cameras.

Another interesting observation is that the number of aspects grows with time (see Table B.1). This is not surprising as manufacturers introduce new product features every year (e.g. HD video, creative image function etc.) to attract customers. This situation also explains the top aspect weights presented in Table B.1 decreasing over time: a higher number of aspects per year means a lower aspect average frequency. Accordingly such situations challenge existing recommendation systems, calling for adaptive extraction and weighting algorithms that can leverage temporal information from product reviews for product recommendation.

Top 10 aspects						
2008	2009	2010	2011	2012	2013	2014 until April
resolution	photographer	photographer	photographer	photographer	photographer	*picture*
photographer	*picture*	*picture*	*picture*	*picture*	*picture*	feature
software	feature	feature	feature	feature	feature	photographer
feature	battery	setting	battery	battery	setting	battery
picture	setting	battery	setting	setting	battery	setting
battery	reason	result	result	result	result	photography
setting	*resolution*	photography	photography	photography	photography	image
reason	result	reason	video	video	image	result
noise	photography	image	capability	image	video	time
result	software	video	image	time	time	quality
# Aspects						
315	557	672	721	866	934	896

Table B.1: Top 10 aspects and total of aspects (#aspects) ordered by frequency between years 2008 until April 2014

B.3.2 Preference Context

Like time, the product preference behavior of users also paints a picture about what aspects are likely to be important when making purchase decisions. For instance if we know which product was preferred over which other product then by comparing the product aspect differences we can infer a degree of aspect importance. To do this we need to establish preference knowledge and thereafter quantify product aspect differences on the basis of sentiment.

We acquire preference knowledge from preference graph generated from viewed and purchased product pairs. The weight of an aspect is derived by comparing the sentiment difference between node pairs in the graph. A preference relation between a pair of products denotes the preference of one product over the other through the analysis of viewed and purchased product relationships. A preference graph, $G = (\mathcal{P}, \mathcal{E})$, is generated from such product pairs (see Figure B.2). The set of nodes, $p_i \in \mathcal{P}$, represent products, and the set of directed edges, $\mathcal{E}$, are preference relations, $p_j \succ p_i$, such that a directed edge from product p_i to p_j with $i \neq j$ represents that, for some users, p_j is preferred over product p_i. For any p_i, we use $\mathcal{E}^i$ to denote in-coming product sets, and $\mathcal{E}_i$ for outgoing product sets.

Figure B.2 illustrates a preference graph generated from a sample of Amazon data on *Digital SLR Camera*. The number of reviews/questions for a product is shown below each product node. It is not surprising that such products appear in Amazon's *Best Seller* ranking (e.g. *B003ZYF3LO* is amongst Amazon's top 10 list). In our recent work [Chen et al., 2014], we observed that the higher the number of incoming edges (quantity) from preferred products (quality), the more preferred is a product. However we also observed that while our assumption is true with most studied products, it is not always the case that a product with higher number of incoming edges will also have a higher rank in Amazon's *Best Seller*. This motivates the need to leverage further dimensions of knowledge sources such as sentiment from product reviews.

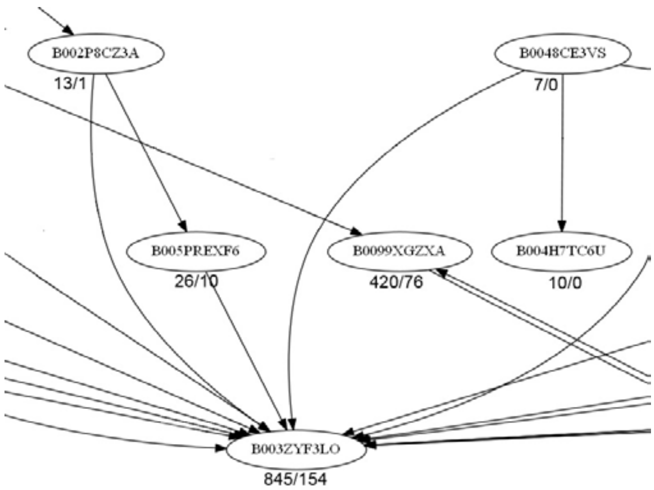


Figure B.2: Preference graph for *Amazon DSLR Cameras*, showing the products ids (i.e. B003ZYF3LO) and the quantity of incomeing and outgoing edges for each product (i.e. 420/76)

B.4 Aspect Weighted Sentiment-based Ranking

User generated product reviews contain user opinion in the form of positive and negative sentiment. Strength of sentiment expresses the intensity with which an opinion is stated with reference to a product [Turney, 2002]. We exploit this information as a means to rank our products, such that products ranked higher denote higher positive sentiment. The product sentiment ($ProdSenti$) of a camera p, given a set of related reviews, is a weighted summation of sentiment expressed at the aspect level:

$$ProdSenti(p_i) = \frac{\sum_{j=1}^{|\mathcal{A}^i|} AW(a_j, t) * AS(p_i, a_j)}{|\mathcal{A}^i|} \tag{B.1}$$

$$AS(p, a) = \frac{\sum_{r \in Occ(p,a)} s(r, a)}{|Occ(p, a)|} * (1 - Gini) \tag{B.2}$$

Where $Occ(p, a)$ is the set of sentences from the reviews of product p related to aspect a, AW is a function of a's weight over time t, and $\mathcal{A}^i$ is the set of unique aspects extracted from the reviews of product p_i. $s(r, a)$ is the sentiment of aspect a in sentence r, a value between [-1,1] generated by the Smart sentiment analysis system [Muhammad et al., 2013]. The negative and positive strength is expressed as a value in the range [-1:1]. It allows the sentiment of product, p, to be associated with individual aspects $a \in \mathcal{A}^i$ where $\mathcal{A}^i \subseteq \mathcal{A}$, being $\mathcal{A}$ the set of unique aspects extracted from the reviews of all products in $\mathcal{P}$.

These sentiment instantiated values for product aspects allow the comparison of product pairs in terms of sentiment. We exploit this further to derive aspects weights based on preference and sentiment knowledge. We use Gini index [Yitzhaki, 1979] to acknowledge higher sentiment scores to an aspect when there is consensus on the distribution of the sentiment and otherwise penalized accordingly.

Aspects of products are extracted by using Algorithm 8 presented in Section 3.2 of this monograph. Aspects are extracted by applying a set of grammatical extraction rules [Moghaddam and Ester, 2012] that operate on dependency relations in parsed sentences parsed using the Stanford Dependency parser [De Marneffe et al., 2006]. Rules involving negations are not included because the SmartSA system already takes this into consideration when generating sentiment scores.

B.4.1 Time-dependent Aspect Weight

The first of our weighting schemes assumes that an aspect is deemed important when it is frequently mentioned by the authors of the reviews. This allows us to monitor aspect trends in straight forward manner as reviews can be fairly easily grouped into seasonal time windows. Based on this principle, an aspect weight is derived by the relative aspect frequency at time window t.

$$AW(a_j, t) = \frac{Freq(a_j, t)}{\sum\limits_{a_i \in \mathcal{A}} Freq(a_i, t)} \tag{B.3}$$

Where $Freq$ returns the frequency of an aspect a in a given time window t into which reviews have been grouped. Frequency here is the number of times a term appears in a specified group of reviews. Table B.2 shows an example of aspect weight calculations in two different time windows i.e. 2008 and 2011. We observe that although the aspect frequency of 'megapixel' has increased overtime, its relative weight compared to all other aspects decreased significantly. Therefore, we suggest that the importance of the aspect 'megapixel' has dropped over the years. Whilst frequency of aspects over time allows us to infer global trends about aspect usage, it does so without considering the relationship between aspects from preferred products. Therefore an alternative approach is to compare aspects shared between preferred products.

Year	***Freq***$(a, Year)$	$\sum_{a \in \mathcal{A}} Freq(a, Year)$	***AW***$(a, Year)$
2008	434	$1.4 \cdot 10^6$	$3 \cdot 10^{-5}$
2011	504	$4.5 \cdot 10^6$	$1.1 \cdot 10^{-5}$

Table B.2: Time-dependent *megapixel* aspect weight in 2008 - 2011

B.4.2 Preference Aspect Weight

A product purchase choice is a preference made on the basis of one or more aspects. The notion of aspect importance arises when the same set of aspects contribute to similar purchase decisions. Using this principle, aspects weights are derived by comparing the aspect sentiment score differences between purchased and viewed product pairs (p_x, p_y), where $p_x \neq p_y$.

$$AW(a_j) = \frac{\sum_{x=1}^{|\mathcal{P}|} \sum_{y=1}^{|\mathcal{P}|} \delta(a_j, p_x, p_y)}{|d|} \tag{B.4}$$

where either $p_x \succ p_y$, or $p_y \succ p_x$, or both, and $d \in \mathcal{E}$ is the set of product preference pairs whose products have aspect a_j. The preference difference between any pairs of products is computed as:

$$\delta(a_j, p_x, p_y) = |L_{min}(\mathcal{A}, \mathcal{E})| + \delta'(a_j, p_x, p_y) \tag{B.5}$$

$$\delta'(a_j, p_x, p_y) = AS(a_j, p_x) - AS(a_j, p_y) \tag{B.6}$$

Here $|L_{min}(\mathcal{A}, \mathcal{E})|$ is the least minimum preference difference scores obtained over all aspects and product preference pairs, and $AS(a_j, p_x)$ is the sentiment associated to aspect a_j of product p_x.

Therefore, to compute the preference aspect weight (AW) of an aspect a, we search all those product preference pairs (p_x, p_y) whose products possess aspect a and estimate $\delta(a_j, p_x, p_y)$. Higher values of δ for an aspect contribute to higher aspect weights AW for that aspect. This indicates that aspect a may have played an important role in the selection of a preferred product. indicating that one of the reasons product p_i was preferred was because of aspect a, .

Since δ' computation can result in negative scores, we use $|L_{min}(\mathcal{A}, \mathcal{E})|$ to bound the lowest value to zero. Thereafter we normalize $\delta(a_j, p_x, p_y)$ such that it is in range [0,1].

B.5 Evaluation

In this section we evaluate our proposed integration of time and preference guided aspect weight extraction applied to product recommendation. we evaluate how well the recommendation system works in practice on Amazon and Cnet data using two derived benchmark rankings.

For the experiments, we use the DSLR corpus of digital cameras (K_D) introduced in this monograph in Section 3.6.1. Since we are not focusing on the cold-start problem, newer products and those without many user reviews are removed

The aspect extraction algorithm that uses dependency extraction rules (FqDPrules) extracted 981 unique aspects, 128 aspects on average per product. Importantly more than 50% of the products shared at least 70 aspects, while 30% shared more than 90 aspects on average. The fact that there are many shared aspects is reassuring for product comparison when applying Equation B.1.

B.5.1 Ranking Strategies

The retrieval set of a query product consists of products that share a similar number of aspects. This retrieval set is ranked using the following sentiment-based recommendation strategies considering only the k top shared aspects between the retrieved and the query product. The comparative weighting schemes used in our experiments are as follows:

- Base: recommend using aspect sentiment analysis without considering aspect weights presented in Equation B.2;
- $PrefW$: same as Base but with the additional preference aspect weighting component from Equation B.4;
- $TimeW_t$: same as $PrefW$ but considering the time context for aspect weighting component (instead of preference) presented in Equation B.3.

It is worth noting that Amazon only provides the current snap shot of preferences. Therefore we are unable to explore the impact of time on preference-based aspect weight extraction. We will present the $TimeW_t$ strategy considering all reviews created between three different time windows: 2008-2014, 2011-2014 and 2014.

B.5.2 Evaluation Metrics

In the absence of ground truth data and individual user specific purchase trails, we generated two different benchmark lists according to the following dimensions:

- *Popular*: Derived from Amazon's reviews, questions and timeline data. Products are ranked based on its popularity computed by means of Equation B.7.

$$Popular(p) = \frac{nReviews + nQuestions}{days\_online} \tag{B.7}$$

  where *nReviews* and *nQuestions* refer to the number of reviews and questions of a product respectively, and *days_online* is the number of days the product has been on Amazon's website. We found that this formula has some correlation with the top 100 Amazon Best Seller ranking (*Spearman* *correlation* of -0.4381). Unlike Amazon's benchmark this allows us to experiment with query products that may not necessarily be in the Amazon top 100[1]. Using a *leave-one-out* methodology, the average gain in rank position of recommended products over the left-out query product is computed relative to a benchmark product ranking.

$$\%RankGain = \frac{\sum_{i=1}^{n=3} benchmark(P_q) - benchmark(P_i)}{n * |\mathcal{P} - 1|} \quad \text{(B.8)}$$

where n is the size of the retrieved set and *benchmark* returns the position on the benchmark list. The greater the gain over the query product the better.

- *Precision*: Derived from Cnet.com expert recommendations for DSLRs cameras[2]. This ground-truth is divided in three subcategories (*entry-level DSLRs*, *midrange DSLRs*, and *professional DSLRs*), each containing a list of cameras recommended by Cnet experts.

$$Precision = \sum_{i=1}^{n} \frac{TopN_{cat} \cap Cnet_{cat}}{n} \quad \text{(B.9)}$$

In the absence of a defined Cnet ranking per category, we use a leave-one-out methodology to evaluate the capacity of our strategies to recommend the expert-selected cameras in each category. We compute the precision by means of Equation B.9, where $TopN_{cat}$ is the list of the top n recommended products for category *cat*, and $Cnet_{cat}$ is the list of Cnet expert recommended cameras for that category.

B.5.3 Results - Amazon

Here we present results from our exploration of aspects trends in terms of weights over time followed by a comparison of the two weighting schemes.

Importance of Time on Aspect Weighting

Figure B.3 shows the weight of the aspects 'megapixel', 'autofocus' and 'battery' computed by using strategy $TimeW_t$ for years between 2008 and 2014. We observe that 'megapixel' was an important aspect in *2008* with a frequency weight of close to 0.009. However, its importance decreased dramatically during the following years, reducing its weight up to five times in *2011*. In contrast, with 'autofocus', we see an increasing trend. A different trend can be observed for 'battery' in Figure B.3. Here it is interesting to note that the aspect weight is maintained over the years. Whilst there is a negligible loss in the raw score this is explained by the difference in number of unique aspects in the time period.

[1]http://www.amazon.co.uk/Best-Sellers-Electronics-Digital-SLR-Cameras
[2]http://www.cnet.com/uk/topics/cameras/best-digital-cameras/

For example, in *2008* we found approximately 250 aspects whilst in *2014* this had increased to 900.

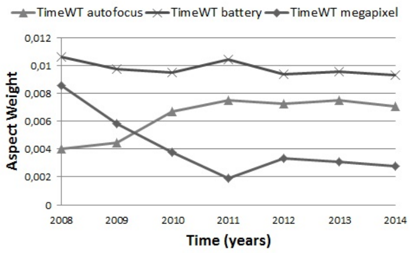


Figure B.3: Aspects weight over time (in years)

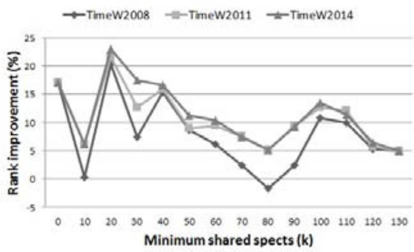


Figure B.4: *ProdSenti* on *Popular* benchmark

In Figure B.4, we use $TimeW_t$ with $t =$ *2008-2014*, *2011-2014* and *2014* to rank the recommendations for increasing number of shared aspects *(k)* on benchmark *Popular*. In general, we observe that weights learned using $TimeW_{2014}$ perform up to a 15% better for k = *30* when recommending actual products. $TimeW_{2011}$ falls close to the recommendations made by $TimeW_{2014}$, being the weights learned by $TimeW_{2008}$ the ones that perform worst. These results indicate that considering the most recent time frame for computing the aspect weights improves the recommendations made by the system, and that aspect frequency over time is a good indicator of what users care most when considering cameras.

Time vs Preference Weighting

In Figure B.5 we compare the three strategies, $TimeW_{2014}$, $PrefW$ and Base using the *Popular* benchmark. We include the strategy agnostic of aspects weights, Base, in order to compare the impact that weights have on the recommendations while considering preferences and time weighted aspects.

As we can observe, $TimeW_{2014}$ and $PrefW$ strategies outperform Base by more than 10% on average. On the other hand, we observe that $PrefW$ outperforms $TimeW_{2014}$ for all values of k comprised between 30 and 100. This suggests that preference weights used by $PrefW$ are able to recommend better products for the *Popular* benchmark since they represent the most recent snapshot of the current users preferences. We also observe that there seems to be a sweet spot in terms of the number of shared aspects (k), with $PrefW$ achieving best results with k=30, and a gradual decrease thereafter.

However, the rank improvement obtained by considering time in $TimeW_{2014}$ should not be ignored as it performs 15% better than $PrefW$ with smaller subsets of shared aspects (e.g. $k = 20$) and obtains a similar rank improvement with increased numbers of shared aspects (e.g. $k \geq 100$). Close examination of $TimeW_{2014}$'s performance suggests that the retrieval set consisted of a high number of products (85%) with k = *20* shared aspects. This is in contrast

Top 10 aspects for $TimeW_{2014}$		Top 10 aspects for $PrefW$	
Aspect	**Weight**	**Aspect**	**Weight**
picture	.00982	shutter	.00229
feature	.00974	photography	.00181
photographer	.00956	point	.00179
battery	.00948	system	.00176
setting	.00942	video	.00166
photography	.00877	setting	.00165
image	.00857	picture	.00158
result	.00847	advantage	.00152
time	.00843	sensor	.00150
quality	.00842	manual	.00150

Table B.3: Top 10 aspects for $TimeW_{2014}$ and $PrefW$

to its poorer performance with higher values of k. This suggests that with less frequently occurring aspects the frequency based weight computation of $TimeW_{2014}$ is likely to be less reliable compared to $PrefW$.

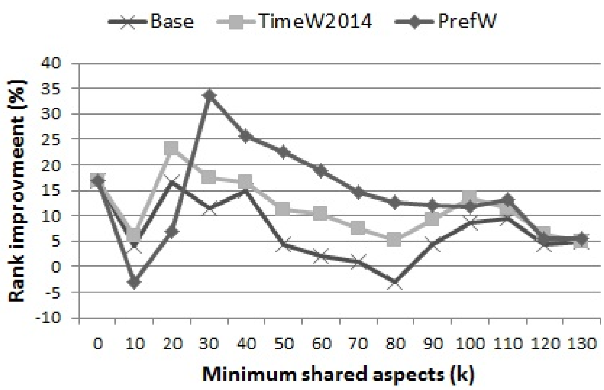


Figure B.5: Comparison of different strategies

Table B.3 presents the top 10 aspects extracted by means of $TimeW_{2014}$ and $PrefW$ and its correspondent weight. Here the lists and weights of the top aspects obtained by the two strategies are different except for aspects 'picture', 'setting' and 'photography'. Although occupy different ranking positions and are weighted differently, both strategies seem to agree on their importance. Also, the weight distribution of each strategies is different. For example, $TimeW_{2014}$ gives almost five times more weight to 'picture' (0.00982) than $PrefW$ does to its top aspect 'shutter' (0.00229). We also notice that there are several semantically related aspects that appear in the top 10: 'image' represents a similar concept to 'picture' and 'photography'; similarly 'system' and 'settings'. It is likely that mapping such related aspects to a common concept is able to generate more accurate aspect weights.

B.5.4 Results - Cnet

Next we compare our recommendations using weights extracted from Amazon sources against expert advice. For this purpose we use Cnet expert advice on DSLR camera recommendations. We divided all cameras from our dataset in three subsets, corresponding to *entry-level*, *midrange* and *professional* categories presented on the website, based on price (see Table B.4). Table B.4 also shows the number of products included in every subset and the number of Cnet products used as the gold standard.

	Entry-level	**Mid-range**	**Professional**
Price range ($)	0-1k	1k-2.2k	2.2k-7k
$\#products_{cat}$	60	14	6
$\#products_{cat} \cap Cnet_{cat}$	7	8	3

Table B.4: Cnet dataset classification for Entry-level, Mid-range and Professional

Table B.5 shows the average precision of $TimeW_{2014}$, $PrefW$ and BASE for every Cnet category and different recommendation set sizes (n=1 and n=3) computed by means of Equation B.9. We included a strategy that randomly recommends products, *Random*, to facilitate understanding of the results.

As we observe in Table B.5, strategies that are aware of aspect sentiments are able to improve precision in every category. In *Popular* dataset, $PrefW$ is the strategy that performs better: for the *Cnet Entry-level* subset, it is able to recommend a Cnet top product 37% of the time for *n=1* and a 24% for *n=3* on average. These results are promising considering that the probability of recommending a top Cnet product amongst the entire subset in this category is 20% and 7% respectively. Precision results for all three strategies are much higher when applied to smaller Cnet datasets; *Mid-range* and *Professional*, achieving a precision of close to 85% and 80% for *n=1* respectively and doubling the precision of the random recommender. Overall, since our system recommendations closely match with Cnet expert's recommendations, we conclude that the aspect weights learned from Amazon are likely to correspond with criteria that the Cnet expert's might have implicitly used. However, we cannot accurately verify this without manually demanding user trial. Nevertheless it is interesting that consensus knowledge discovered from social media seem to closely echo views of the domain experts.

Strategy	**Entry-level**		**Mid-range**		**Professional**	
	n=1	*n=3*	*n=1*	*n=3*	*n=1*	*n=3*
Random	0.203	0.077	0.461	0.376	0.2	0.277
BASE	0.220	0.146	0.769	0.423	0.4	0.377
$TimeW_{2014}$	0.254	0.128	**0.846**	0.444	**0.8**	0.4
$PrefW$	**0.372**	**0.243**	**0.846**	**0.461**	0.6	**0.422**

Table B.5: Precision for different retrieved set sizes n in Cnet.

B.6 Conclusions

Tracking users preference over time raises unique challenges for recommendation systems. Every product potentially goes through a series of changes which typically involves functional improvements resulting in a broader range of aspects which in turn will be echoed in changes to user preferences. Our previous findings suggested that preference knowledge allows us to identify aspects that are important to users but lacks the capability to trace aspect importance trends.

In this paper, we formalize a novel aspect-based sentiment ranking that utilize both time and preference contexts. The benefits are demonstrated in a realistic recommendation setting using benchmarks generated from Amazon and Cnet. We show that monitoring aspect frequency in product reviews allows to capture changes to aspect importance over time. Importantly, we confirm that time context can be conveniently exploited by using the recent time frame to improve recommendations.

We compare the performance of aspect-based sentiment ranking in the context of time and preference. We observed that both contexts perform well in different number of shared aspects, further work is needed to study the benefit of integrating both contexts in ranking algorithm. Moreover, our results show that similar aspects were mentioned using different terms and further work is needed to study how sparsity problems might impact the difference calculations. Finally, it would be interesting to integrate trending information within the aspect weight computation to infer their importance.

Appendix C

Aspect Vocabularies $\mathcal{A}_D$, $\mathcal{A}_C$, and $\mathcal{A}_P$

In this appendix we create the aspect vocabularies for DSLR cameras ($\mathcal{A}_D$), Compact cameras ($\mathcal{A}_C$), and Point & Shoot cameras ($\mathcal{A}_P$), from the three camera corpora K_D, K_C, and K_P. The three aspect vocabularies were created as described in Chapter 3, using the frequency threshold determined in Section 3.6.2.

Table C.1 contains the 266 aspects that form the DSLR aspect vocabulary $\mathcal{A}_D$. The Compact aspect vocabulary is formed by 268 aspects and is presented in Table C.2. Finally, the Point & Shoot aspect vocabulary is presented in Table C.3, and is formed by 199 aspects. The aspects are ordered in frequency in the tables from left to right and top to bottom. That means that aspect 'lens' is more frequent than aspect 'picture', and 'picture' more frequent than 'view', in the reviews of the DSLR corpus K_D presented in Table C.1.

Notice that, event if the aspect vocabularies are created from three different corpus, the most frequent aspects of the three vocabularies are similar. Aspects such as 'lens', 'picture', and 'battery' are found between the most frequent aspects in the reviews of all three camera corpora K_D, K_C and K_P, indicating that those aspects are similarly used in the reviews of the three types of cameras. On the other hand, the differences between the three aspect vocabularies are clear when not considering the top 10 most frequent aspects. The occurrences of aspects such as 'live view' or 'price' clearly vary between the reviews of different camera corpus. For instance, aspect 'waterproof' is only found in the Point & Shoot cameras aspect vocabulary $\mathcal{A}_P$.

DSLR aspect vocabulary $\mathcal{A}_D$			
lens	card memory	light situation	user interface
picture	view	distortion	focus manual
shoot	megapixel	tone	shutter lag
video	photograph	stabilization	body camera
shot	scene	photo quality	dslr user
focus	white balance	layout	lens selection
photo	compact	iso setting	slow motion
feature	hdr	focus issue	button shutter
image	evf	cam	iso level
body	hd video	stereo	shutter speed
point	focus point	dot	lcd viewfinder
iso	slr camera	menu system	cmos sensor
price	depth	lag	flash picture
quality	landscape	crop sensor	lens aperture
battery	audio	focus lens	view focus
setting	firmware	light capability	iso image
photography	contrast	variety	lens quality
shooting	macro	raw image	exposure bracketing
flash	accessory	composition	camera system
button	video camera	detection	iso light
auto	shoot video	thrill	flash control
manual	mount	gps	manual exposure
shutter	stick	touchscreen	lens telephoto
sensor	sharpness	liveview	proof weather
screen	build quality	video feature	lcd panel
photographer	1080p	lens work	ratio aspect
zoom	spec	timer	camera quality
noise	image stabilization	mic external	focus image
speed	touch screen	daylight	speed write
viewfinder	angle	magnification	speed shutter
bit	ergonomics	sync	wide aperture
image quality	mic	lcd display	hdr feature
kit lens	iso performance	array	ion battery
menu	metering	slr cameras	jpeg image
color	memory	storage	flash exposure
size	multiple	iso range	menu setting
autofocus	shooter	zoom lens	chromatic aberration
function	camcorder	videographer	iso picture
level	video quality	length	image light
card	priority	lens mount	lens point
system	wireless	exposure compensation	macro lens
raw	view finder	image sensor	focus movie
lcd	telephoto	cf card	quality video
live view	raw file	frame sensor	exposure metering
frame	wifi	focus motor	raw processing
exposure	adapter	mode burst	screen resolution
grip	enthusiast	button layout	lens macro
weight	noise reduction	lens zoom	focus assist
detail	card slot	image iso	flash photo
aperture	light performance	iso button	focus autofocus
dial	custom	battery life	release shutter
film	functionality	lens canon	mount lens
cost	jpg	remote control	focus picture
life	price range	raw format	spot metering
performance	video capability	manual focus	pixel sensor
full frame	wide angle	view screen	color image
movie	charger	lens focus	
resolution	recording	grip hand	
pic	interface	lens nikon	
lcd screen	lense	external microphone	
lighting	frame camera	zoom range	
sd card	focal length	video shoot	
pixel	manual control	quality lens	
filter	hobby	size sensor	
display	film camera	life battery	
entry level	video recording	crop factor	
picture quality	mirrorless	iso capability	
jpeg	usb	focus screen	
value	microphone	lcd monitor	
dslr camera	focus system	focal point	

Table C.1: DSLR aspect vocabulary $\mathcal{A}_D$.

Compact Aspect vocabulary $\mathcal{A}_C$			
lens	interchangeable	portability	usb cable
shoot	lcd screen	lens zoom	lens pancake
picture	pancake	film camera	video button
focus	value	photo quality	frame sensor
photo	lighting	iso performance	focus system
shot	depth	spec	quality lens
flash	build quality	memory	view screen
image	card	camera sensor	raw format
feature	touchscreen	daylight	lens focus
video	accessory	mft	shutter lag
point	scene	rangefinder	life battery
quality	view finder	hd video	mode burst
price	image stabilization	tone	lens option
button	hdr	distortion	dslr user
size	video camera	bokeh	mic external
sensor	macro	lens mount	panel control
iso	contrast	manual focus	menu button
battery	pixel	dof	screen protector
body	photograph	recording	button shutter
setting	stick	camcorder	end dslr
zoom	sd card	noise reduction	light capability
shutter	charger	port	iso image
image quality	white balance	raw image	end point
manual	landscape	layout	nex lens
viewfinder	menu system	variety	jpeg engine
auto	wide angle	focus manual	lens manual
shooting	priority	ovf	lens aperture
screen	video quality	focus lens	slow motion
menu	lug	video recording	grip hand
photography	download	user interface	sync speed
kit lens	card memory	iso setting	card slot
bit	ergonomics	iauto	flash picture
speed	shooter	camera system	focal point
color	sharpness	size camera	macro lens
system	dslr camera	lens selection	lcd viewfinder
evf	light performance	brightness	menu setting
photographer	lens adapter	length	hood lens
dial	jpg	histogram	upgrade firmware
noise	custom	hotshoe	shutter speed
function	focal length	battery charger	jpeg image
lens camera	multiple	slr camera	lens quality
weight	interface	stereo	chromatic aberration
aperture	manual control	kit zoom	focus picture
lcd	focus point	bargain	flash photography
touch screen	stabilization	timer	focus autofocus
autofocus	cam	microphone	manual exposure
level	raw file	beauty	external microphone
mirrorless	angle	metering	speed shutter
mount	thirds	nfc	system lens
adapter	wireless	ipad	lens interchangeable
firmware	lag	dot	ratio aspect
raw	telephoto	image sensor	focus image
grip	megapixel	customization	video shoot
performance	smartphone	internet	size sensor
cost	oly	finder	best quality
life	detection	large sensor	aperture lens
exposure	zoom lens	facebook	lens macro
detail	focus speed	panny	
picture quality	body camera	quirk	
full frame	functionality	lense	
wifi	audio	composition	
display	usb	versatility	
frame	exposure compensation	hood	
resolution	sensor size	frame camera	
film	shoot video	lens system	
pic	handling	video capability	
view	1080p	focus assist	
jpeg	mic	mount lens	
movie	live view	remote control	
filter	ibis	crop factor	

Table C.2: Aspect vocabulary $\mathcal{A}_C$ of Compact camera corpus K_C.

Point & Shoot aspect vocabulary $\mathcal{A}_P$		
picture	stick	focus point
zoom	pixel	minimum
photo	lcd screen	sync
battery	megapixel	raw image
lens	video quality	raw format
video	image stabilization	cmos sensor
shot	photograph	focus picture
feature	grip	size sensor
quality	recording	button shutter
price	autofocus	ratio aspect
point	multiple	gps feature
flash	landscape	image sensor
setting	photo quality	manual focus
focus	view finder	quality picture
image	lag	flash photo
manual	jpeg	focus manual
size	manual control	ion battery
button	depth	large sensor
auto	white balance	picture flash
pic	video camera	live view
color	1080p	color picture
screen	sharpness	best quality
shutter	priority	large screen
bit	contrast	pic quality
shooting	hdr	zoom feature
picture quality	build quality	picture focus
function	gps	lens focus
iso	touchscreen	remote control
speed	mount	quality video
menu	port	feature zoom
image quality	focal length	speed write
photographer	freeze	color photo
sensor	angle	card slot
photography	stabilization	photo flash
card	wireless	picture color
viewfinder	custom	quality photo
noise	length	focus image
movie	mode burst	pic flash
card memory	variety	mic external
macro	view screen	best photo
charger	lense	chromatic aberration
performance	description	grip hand
cost	timer	focus photo
lcd	full frame	manual exposure
resolution	detection	lens point
detail	raw file	overall quality
body	shutter lag	iso image
powershot	battery life	lithium battery
scene	zoom lens	lens mount
waterproof	life battery	focus assist
display	flash picture	color quality
view	mirrorless	point price
sd card	lithium	file format
lighting	microphone	jpeg image
weight	capacity	video recording
wifi	optics	overall picture
aperture	mic	focus lens
usb	iso setting	battery lithium
system	storage	lcd viewfinder
level	exposure compensation	
film	preset	
exposure	interchangeable	
touch screen	tone	
raw	slow motion	
coolpix	saturation	
filter	lens zoom	
hd video	dot	
wide angle	shoot video	
memory	taking	
value	video button	

Table C.3: Point & Shoot aspect vocabulary $\mathcal{A}_P$.

Appendix D

Concept Vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$

In this appendix we show the resulting concept vocabularies for DSLR ($\mathcal{C}_D$), Compact ($\mathcal{C}_C$), and Point & Shoot cameras ($\mathcal{C}_P$). The three concept vocabularies were created as described in Chapter 4, using the DSLR aspect vocabulary $\mathcal{A}_D$ and the DSLR corpus K_D to create the DSLR concept vocabulary ($\mathcal{C}_D$), the Compact aspect vocabulary $\mathcal{A}_C$ and the Compact corpus K_C to create the Compact concept vocabulary ($\mathcal{C}_C$), and the Point & Shoot aspect vocabulary $\mathcal{A}_P$ together with the Point & Shoot corpus K_P to create the Point & Shoot concept vocabulary ($\mathcal{C}_P$).

Table D.1 shows the complete DSLR concept vocabulary $\mathcal{C}_D$. It is formed by 41 concepts that contain 225 different aspects (considering that aspects such as 'shutter speed' and 'speed shutter' are the same aspect, thus not counting them twice). The Compact concept vocabulary $\mathcal{C}_C$ is presented in Table D.2, and it is formed by 39 concepts that contain 197 different aspects. Finally, Table D.3 shows the complete Point & Shoot concept vocabulary $\mathcal{C}_P$, formed by 39 concepts and 179 different aspects.

Notice that the concept vocabularies of the tree camera types ($\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$), are different. Since the concept vocabularies model the important issues used by people when expressing their experiences in product reviews, that means that a user that bought a DSLR camera considers important different issues than another user that bought a Compact or a Point & Shoot. For instance, in $\mathcal{C}_P$ we have concept 'waterproof', but this concept does not exist in any of the other two vocabularies $\mathcal{C}_D$ or $\mathcal{C}_C$. However, aspects such as 'picture', 'iso' and 'noise' form different concepts in the concept vocabularies $\mathcal{C}_D$ and $\mathcal{C}_C$, but are grouped in the same concept in the Point & Shoot concept vocabulary $\mathcal{C}_P$. The system considered that "picture" was too general to express the important issues present in the reviews of DSLR and Compact cameras, but not for Point & Shoot.

Concept	**Aspects in Concept**
autofocus	auto focus, autofocus, focus auto
battery	life battery, battery life, ion battery, charger, battery life
body	body camera, body
build quality	build quality
button	button, shutter button, speed shutter, shutter speed, button shutter shutter, button layout, release shutter
card	card slot, cf card, sd card, card, card memory
detail	detection, detail
display	lcd display, display
dslr camera	entry level, dslr camera
enthusiast	enthusiast, hobby
filter	filter, adapter, telephoto, accessory, distortion
flash	flash control, flash exposure, flash
focus	manual focus, focus autofocus, screen focus, focus screen, focus view focus, focus assist, focus issue, focal point, focus point focus manual, autofocus focus, focus movie, focus system, focus motor
function	layout, magnification, composition, custom, function functionality, camcorder, audio, sync, cam
grip	grip hand, grip
hdr	hdr feature, hdr, feature hdr
iso	iso level, iso performance, iso button, iso, iso capability iso range, iso setting
jpg	jpg
lag	lag, daylight, shutter lag
lens	aperture lens, lense, mount lens, lens, lens nikon lens mount, lens selection, zoom range, lens macro, angle focal length, macro, lens focus, wide angle, zoom lens length, lens canon, lens work, focus lens, lens zoom zoom, aperture, lens quality, contrast, wide aperture kit lens, lens aperture, macro lens, quality lens
light situation	light capability, light situation
memory	memory, storage
menu	menu, menu system
mount	mount, mirrorless
noise	noise, noise reduction
performance	light performance, performance, evf, interface, liveview user interface
photography	variety, array, videographer, photographer, photography
picture	photo quality, picture focus, flash picture, tone, jpeg image stabilization, color, raw image, focus picture, image focus image jpeg, focus image, pic, picture quality, iso picture scene, image iso, quality image, photo flash, metering exposure metering, jpeg, sharpness, quality photo, exposure bracketing image, exposure compensation, iso image, image sensor, spot metering photo, image stabilization, priority, flash photo, quality picture picture, picture iso, image color, image quality, photograph depth, exposure, manual exposure, color image, image light
price	cost, price, price range
raw	raw file, raw processing, raw, raw format
resolution	megapixel, pixel, resolution
screen	view, live view, view screen, dot, screen lcd screen, touch screen, touchscreen, screen resolution, screen view
sensor	crop sensor, frame camera, frame sensor, sensor, full frame pixel sensor, cmos sensor, size sensor, sensor size, frame
setting	menu setting, setting
size	size, weight
stereo	timer, stereo
usb	usb
video	hd video, recording video, video recording, movie, shoot video mic, video camera, recording, video quality, video shoot video capability, film, video, 1080p, video feature feature video, external microphone, quality video, microphone, slow motion film camera
viewfinder	lcd viewfinder, view finder, viewfinder, viewfinder lcd
white balance	white balance
wireless	wifi, wireless

Table D.1: DSLR concept vocabulary $\mathcal{C}_D$.

Concept	Aspects in Concept
autofocus	auto focus, autofocus, focus auto
battery	life battery, battery life, battery charger, battery, life
body	portability, body, body camera
bokeh	bokeh
button	button, speed shutter, shutter button, lag, shutter speed button shutter, shutter, menu button, shutter lag
cam	camcorder, cam
card	card slot, sd card, card
custom	custom
dot	dot
flash	flash photography, flash
focus	picture focus, manual focus, focus autofocus, focus picture, focus focus assist, focal point, speed focus, focus point, focus manual autofocus focus, focus speed, focus system
hdr	hdr
image	flash picture, photo quality, stabilization, tone, jpeg image color, raw image, image focus, focus image, image jpeg pic, picture quality, scene, image iso, quality image metering, raw format, sharpness, quality photo, image exposure compensation, iso image, sensor image, image sensor, raw photo, image stabilization, priority, quality picture, picture image quality, raw file, photograph, depth, exposure manual exposure, picture flash
interface	interface, user interface
iso	iso performance, iso, iso setting
jpeg	jpeg engine, jpeg
lcd	touch screen, touchscreen, lcd
lens	lens camera, lens option, aperture lens, lense, mount lens system lens, nex lens, lens, lens mount, lens selection lens macro, manual lens, hood lens, zoom lens, lens focus lens adapter, lens zoom, focus lens, lens hood, lens quality lens system, kit lens, lens aperture, lens manual, quality lens macro lens
megapixel	megapixel
memory	memory
menu	menu, menu system
noise	noise reduction, noise
port	port
price	detection, grip, cost, adapter, telephoto display, functionality, mount, light performance, filter thirds, performance, charger, evf, variety price, function, handling, photographer, detail accessory, photography, mirrorless, build quality
resolution	pixel, resolution
screen	view screen, screen protector, screen, lcd screen, screen lcd screen view
sensor	frame sensor, sensor, large sensor, camera sensor, sensor size size sensor
setting	daylight, menu setting, setting
size	ibis, size, size camera
smartphone	smartphone
usb	usb, usb cable
video	hd video, recording video, video recording, movie, shoot video mic, recording, video quality, video camera, video shoot film, video capability, video, 1080p, quality video external microphone, microphone, video button, film camera, slow motion
view	view, live view
viewfinder	lcd viewfinder, view finder, viewfinder, finder
weight	weight
white balance	white balance
wifi	wifi
wireless	wireless
zoom	macro, aperture, wide angle, contrast, kit zoom length, angle, zoom, focal length, interchangeable

Table D.2: Compact concept vocabulary $\mathcal{C}_C$.

Concept	**Aspects in Concept**
battery	life battery, ion battery, lithium, lithium battery, battery lithium battery
body	filter, coolpix, body, mount, mirrorless
build quality	build quality
button	button, shutter button, button shutter, shutter
card	card slot, sd card, card, card memory
charger	charger, freeze
chromatic aberration	chromatic aberration
description	description
detail	timer, display, detail, taking
detection	detection
dot	dot
flash	pic flash, flash
focus	manual focus, auto focus, focus point, focus manual, focus focus assist, autofocus, focus auto
full frame	full frame
function	function
gps	feature gps, gps feature, gps
grip	grip, grip hand
lag	lag, shutter lag
lcd	lcd viewfinder, view finder, viewfinder, viewfinder lcd, lcd
memory	memory, storage, capacity
menu	menu
minimum	minimum
performance	performance
photographer	photographer, photography
picture	photo quality, flash picture, picture focus, iso, stabilization raw image, focus picture, image focus, image jpeg, pic picture quality, scene, saturation, raw format, sharpness image, iso image, image sensor, raw, flash photo priority, quality picture, stabilization image, exposure manual, image quality photo focus, exposure, color quality, color, jpeg image tone, overall picture, focus image, picture color, image iso focus photo, quality image, photo flash, noise, jpeg photo color, quality photo, exposure compensation, quality pic, sensor image photo, image stabilization, picture, photograph, raw file color photo, depth, pic quality, color picture, manual exposure best photo, picture flash
powershot	powershot, weight
price	cost, price
remote control	remote control
screen	view, live view, view screen, screen, lcd screen touchscreen, touch screen, large screen, screen lcd, screen view
sensor	megapixel, sensor, large sensor, cmos sensor, pixel
setting	iso setting, setting
size	size, size sensor, sensor size
sync	sync
usb	usb, port
video	video recording, recording video, hd video, movie, shoot video mic, video quality, video camera, recording, pic film, video, resolution, 1080p, quality video microphone, video button, slow motion
waterproof	waterproof
white balance	white balance
wifi	wifi, hdr, wireless
zoom	macro, zoom lens, wide angle, lens focus, feature zoom lense, length, lens zoom, focus lens, zoom mount lens, lens point, optics, lens, interchangeable lens mount, aperture, contrast, angle, focal length zoom feature

Table D.3: Point & Shoot concept vocabulary $\mathcal{C}_P$.

Appendix E

Bundle Examples

In this appendix we show the bundles of arguments of various digital cameras from the three corpora K_D, K_C, and K_P, created by following the methodology presented in this dissertation. The objective is to show the reader the pro and con arguments of the different types of bundles of arguments B_G, B_σ, and B_F, for a selection of cameras, in order to help understand the tasks faced in this monograph and the results obtained.

This appendix is structured as follows: Every section presents the bundles of arguments, together with other information, of a digital camera belonging to one of the three corpora K_D, K_C, and K_P. We start every section by introducing some data about the camera: when did the camera come out to the market, Amazon.com star rating, and the Dpreview score (if any). We also show, for every concept in the corresponding concept vocabulary $\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$, how many concept occurrences were found in the reviews of that specific camera. Notice that cameras belong to one of the three corpora K_D, K_C, and K_P, and the concept vocabularies $\mathcal{C}_D$, $\mathcal{C}_C$, and $\mathcal{C}_P$, vary accordingly. Moreover, since the set of camera reviews varies between cameras, the quantity of concept occurrences may also vary between digital cameras. Furthermore, some features may be preferred over others depending on the digital camera, and consequently it may be more likely to find those features occurring in the reviews of certain products. Concept occurrences are also presented in a visual way, for every product, in a word-cloud like figure. More frequent concepts are larger, while less frequent are smaller.

Finally, we present the three bundles of arguments of the corresponding product B_G, B_σ, and B_F, together with a visual representation for each of the bundles of arguments. Bundles are created using $\delta = 0.1$ to select the pro and con arguments (see Equation 5.2). In the figures persenting the bundles, pro arguments are painted green, while con arguments are red. The size of the arguments corresponds to the strength of the argument sentiment. For pro arguments, a larger size means a more positive polarity, while for con arguments, a larger size means a more negative polarity. Moot arguments are not shown in those figures.

E.1 Nikon D7100 (DSLR)

Figure E.1: Nikon D7100 DSLR camera.

Name:	Nikon D7100
Type:	Digital SLR
Date Online:	13/09/2010
Reviews:	908
Amazon Score :	4.5/5
Dpreview Score:	85/100

Table E.1: Information about Nikon D7100 DSLR camera.

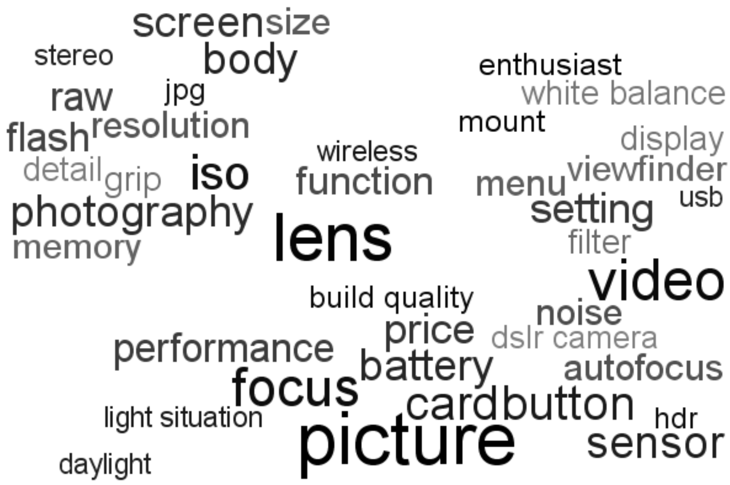


Figure E.2: Concept occurrences found in the reviews of camera presented in Table E.1 (larger means more frequent).

Concept	Occurrences	Concept	Occurrences
picture	2107	menu	143
lens	1510	memory	101
video	951	viewfinder	101
focus	772	filter	94
iso	504	grip	78
card	498	display	60
button	460	white balance	59
sensor	412	detail	56
body	357	dslr camera	53
battery	353	enthusiast	37
photography	348	mount	29
screen	327	build quality	28
price	315	hdr	20
setting	285	jpg	19
performance	210	wireless	16
raw	208	light situation	16
flash	207	daylight	15
function	201	stereo	14
noise	174	usb	11
autofocus	171		
resolution	170		
size	162		

Table E.2: Quantity of concept occurrences in the reviews of camera in Table E.1.

Figure E.3: Gini Argument Bundle B_G of camera in Table E.1. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Gini Argument Bundle B_G					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
build quality	0,4947	pro	filter	0,1425	pro
price	0,2982	pro	photography	0,1415	pro
performance	0,2699	pro	mount	0,1397	pro
white balance	0,2352	pro	body	0,1346	pro
size	0,2193	pro	focus	0,1337	pro
noise	0,2187	pro	enthusiast	0,1224	pro
lens	0,2030	pro	grip	0,1214	pro
memory	0,2008	pro	sensor	0,1192	pro
card	0,1837	pro	dslr camera	0,1148	pro
autofocus	0,1811	pro	setting	0,1113	pro
display	0,1695	pro	picture	0,1095	pro
iso	0,1641	pro	function	0,1053	pro
jpg	0,1641	pro	usb	-0,1865	con
flash	0,1488	pro			

Table E.3: Gini Argument Bundle B_G of camera in Table E.1 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

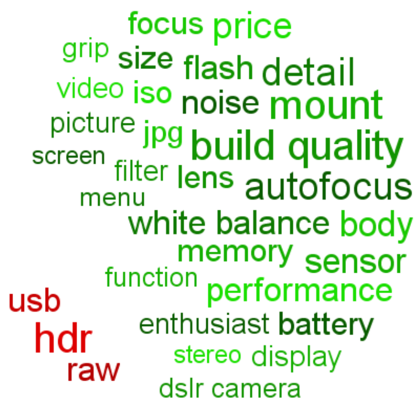


Figure E.4: Gini Argument Bundle B_G of camera in Table E.1. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Agreement Argument Bundle B_σ					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
build quality	0,5870	pro	iso	0,2583	pro
mount	0,5840	pro	jpg	0,2260	pro
detail	0,5225	pro	display	0,2170	pro
autofocus	0,4877	pro	filter	0,2046	pro
price	0,4467	pro	picture	0,1992	pro
body	0,3600	pro	enthusiast	0,1835	pro
sensor	0,3586	pro	video	0,1773	pro
flash	0,3283	pro	function	0,1600	pro
white balance	0,3270	pro	grip	0,1600	pro
noise	0,3180	pro	dslr camera	0,1505	pro
performance	0,3082	pro	menu	0,1300	pro
focus	0,2907	pro	screen	0,1080	pro
memory	0,2805	pro	stereo	0,1015	pro
battery	0,2768	pro	raw	-0,2450	con
size	0,2760	pro	usb	-0,2590	con
lens	0,2713	pro	hdr	-0,4715	con

Table E.4: Agreement Argument Bundle B_σ of camera in Table E.1 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

Figure E.5: Cardinality Argument Bundle B_F of camera in Table E.1. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Cardinality Argument Bundle B_F					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
build quality	0,8571	pro	mount	0,3103	pro
price	0,5871	pro	flash	0,3000	pro
white balance	0,5593	pro	stereo	0,2857	pro
performance	0,5024	pro	sensor	0,2808	pro
lens	0,4829	pro	enthusiast	0,2778	pro
memory	0,4747	pro	filter	0,2766	pro
jpg	0,4737	pro	focus	0,2685	pro
noise	0,4682	pro	function	0,2551	pro
card	0,4590	pro	setting	0,2500	pro
size	0,4500	pro	viewfinder	0,2277	pro
display	0,4035	pro	picture	0,2188	pro
autofocus	0,4000	pro	detail	0,1786	pro
body	0,3484	pro	grip	0,1688	pro
photography	0,3431	pro	video	0,1674	pro
dslr camera	0,3208	pro	hdr	0,1000	pro
iso	0,3172	pro	usb	-0,4545	con

Table E.5: Cardinality Argument Bundle B_F of camera in Table E.1 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

E.2 Canon EOS Rebel T4i (DSLR)

Figure E.6: Canon EOS Rebel T4i DSLR camera.

Name:	Canon EOS Rebel T4i
Type:	Digital SLR
Date Online:	07/06/2012
Reviews:	732
Amazon Score :	4.6/5
Dpreview Score:	–/100

Table E.6: Information about Canon EOS Rebel T4i DSLR camera.

Figure E.7: Concept occurrences found in the reviews of camera presented in Table E.6 (larger means more frequent).

Concept	Occurrences	Concept	Occurrences
picture	216	stereo	13
lens	199	viewfinder	13
video	159	white balance	11
focus	147	grip	11
screen	118	filter	10
button	54	menu	9
iso	50	detail	9
noise	48	mount	7
autofocus	41	resolution	5
function	37	memory	4
sensor	27	display	3
photography	25	jpg	2
body	24	dslr camera	2
performance	23	build quality	1
raw	23	enthusiast	1
price	22	daylight	1
battery	21	usb	1
flash	19	wireless	1
setting	18		
card	15		
size	14		
hdr	14		

Table E.7: Quantity of concept occurrences in the reviews of camera in Table E.6.

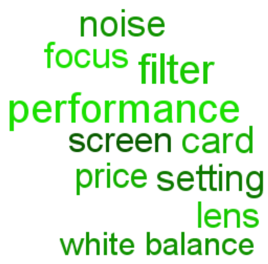


Figure E.8: Gini Argument Bundle B_G of camera in Table E.6. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Gini Argument Bundle B_G					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
performance	0,2129	pro	screen	0,1388	pro
filter	0,2105	pro	focus	0,1348	pro
card	0,1624	pro	lens	0,1312	pro
noise	0,1567	pro	price	0,1275	pro
setting	0,1548	pro	white balance	0,1071	pro

Table E.8: Gini Argument Bundle B_G of camera in Table E.6 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

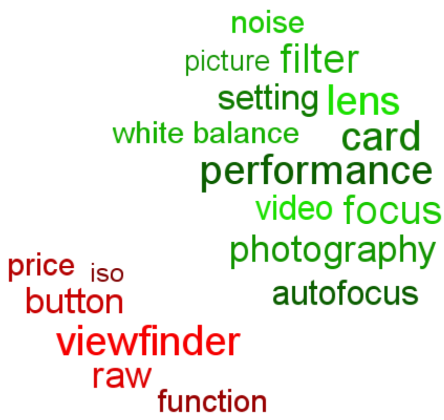


Figure E.9: Gini Argument Bundle B_G of camera in Table E.6. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Agreement Argument Bundle B_σ					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
card	0,3800	pro	noise	0,1880	pro
performance	0,3740	pro	white balance	0,1410	pro
lens	0,3715	pro	picture	0,1104	pro
filter	0,3655	pro	iso	-0,1087	con
focus	0,3604	pro	price	-0,1980	con
photography	0,2975	pro	function	-0,1989	con
setting	0,2300	pro	button	-0,2945	con
autofocus	0,2255	pro	raw	-0,3243	con
video	0,2155	pro	viewfinder	-0,4295	con

Table E.9: Agreement Argument Bundle B_σ of camera in Table E.6 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

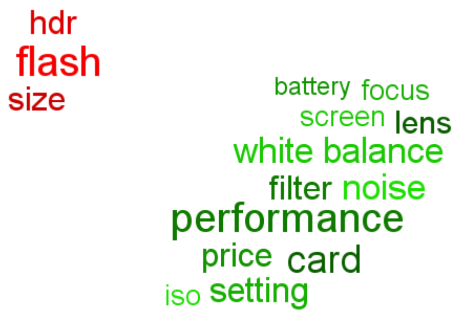


Figure E.10: Cardinality Argument Bundle B_F of camera in Table E.6. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Cardinality Argument Bundle B_F					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
performance	0,7391	pro	iso	0,2653	pro
card	0,6000	pro	screen	0,2414	pro
noise	0,5000	pro	focus	0,2245	pro
white balance	0,4545	pro	battery	0,1429	pro
setting	0,4118	pro	size	-0,1429	con
filter	0,4000	pro	hdr	-0,1429	con
price	0,4000	pro	flash	-0,2632	con
lens	0,3613	pro			

Table E.10: Cardinality Argument Bundle B_F of camera in Table E.6 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

E.3 Pentax K-5 (DSLR)

Figure E.11: Pentax K-5 DSLR camera.

Name:	Pentax K-5
Type:	Digital SLR
Date Online:	20/09/2010
Reviews:	147
Amazon Score :	4.7/5
Dpreview Score:	83/100

Table E.11: Information about Pentax K-5 DSLR camera.

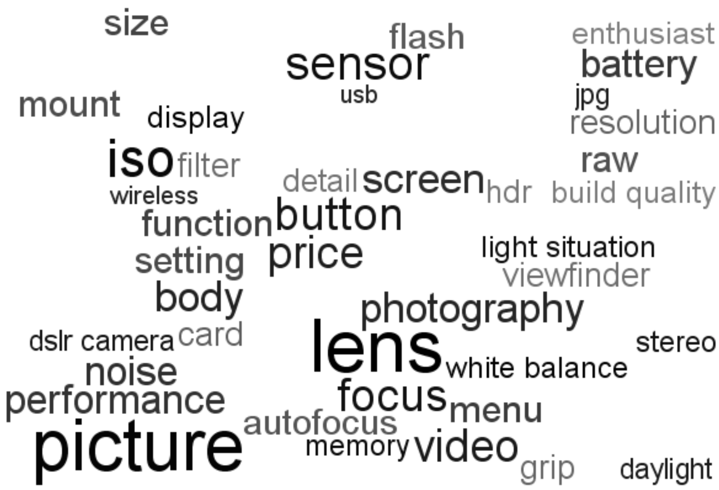


Figure E.12: Concept occurrences found in the reviews of camera presented in Table E.11 (larger means more frequent).

Concept	Occurrences	Concept	Occurrences
lens	465	card	20
picture	419	resolution	18
iso	153	grip	15
sensor	105	filter	15
button	97	viewfinder	13
focus	97	detail	12
video	88	hdr	10
price	85	build quality	7
body	69	display	6
photography	57	enthusiast	6
screen	54	white balance	5
noise	47	stereo	5
performance	46	memory	5
battery	42	light situation	5
menu	38	daylight	3
function	37	dslr camera	3
size	36	jpg	2
setting	35		
mount	30		
raw	28		
autofocus	24		
flash	22		

Table E.12: Quantity of concept occurrences in the reviews of camera in Table E.11.

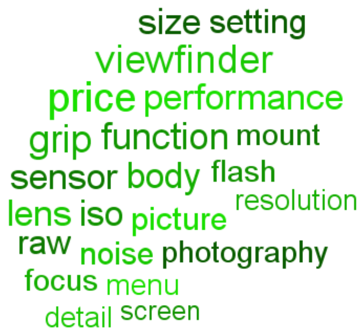


Figure E.13: Gini Argument Bundle B_G of camera in Table E.11. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Gini Argument Bundle B_G					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
price	0,3781	pro	setting	0,1950	pro
viewfinder	0,3372	pro	picture	0,1715	pro
grip	0,2933	pro	noise	0,1641	pro
function	0,2738	pro	focus	0,1533	pro
sensor	0,2687	pro	flash	0,1517	pro
size	0,2586	pro	mount	0,1486	pro
performance	0,2581	pro	photography	0,1481	pro
lens	0,2546	pro	menu	0,1335	pro
body	0,2453	pro	detail	0,1171	pro
iso	0,2315	pro	resolution	0,1110	pro
raw	0,2027	pro	screen	0,1078	pro

Table E.13: Gini Argument Bundle B_G of camera in Table E.11 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

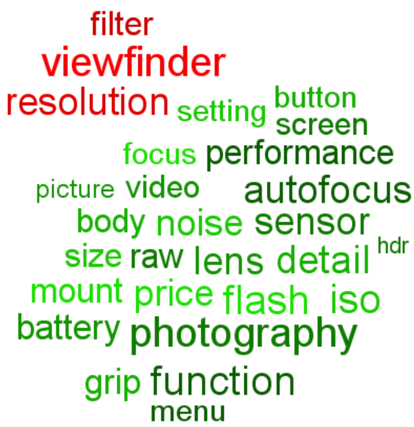


Figure E.14: Gini Argument Bundle B_G of camera in Table E.11. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Agreement Argument Bundle B_σ					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
photography	0,6312	pro	mount	0,3585	pro
function	0,6016	pro	size	0,3445	pro
flash	0,5995	pro	body	0,3380	pro
iso	0,5740	pro	video	0,2815	pro
detail	0,5470	pro	screen	0,2766	pro
lens	0,5328	pro	menu	0,2685	pro
sensor	0,5106	pro	focus	0,2554	pro
autofocus	0,5060	pro	button	0,2444	pro
price	0,4757	pro	setting	0,2420	pro
noise	0,4695	pro	picture	0,1442	pro
grip	0,4140	pro	hdr	0,1030	pro
battery	0,4090	pro	filter	-0,1192	con
raw	0,3770	pro	resolution	-0,1783	con
performance	0,3647	pro	viewfinder	-0,2270	con

Table E.14: Agreement Argument Bundle B_σ of camera in Table E.11 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

Figure E.15: Cardinality Argument Bundle B_F of camera in Table E.11. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Cardinality Argument Bundle B_F					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
price	0,6941	pro	hdr	0,4000	pro
viewfinder	0,6923	pro	picture	0,3525	pro
body	0,6765	pro	menu	0,3514	pro
raw	0,6429	pro	detail	0,3333	pro
setting	0,6000	pro	mount	0,3103	pro
grip	0,6000	pro	focus	0,3053	pro
sensor	0,5728	pro	card	0,3000	pro
performance	0,5652	pro	filter	0,2857	pro
function	0,5556	pro	flash	0,2727	pro
size	0,5429	pro	noise	0,2340	pro
lens	0,5216	pro	screen	0,2222	pro
photography	0,4386	pro	resolution	0,1111	pro
iso	0,4305	pro			

Table E.15: Cardinality Argument Bundle B_F of camera in Table E.11 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

E.4 Olympus OMD E-M5 (Compact)

Figure E.16: Olympus OMD E-M5 Compact camera.

Name:	Olympus OMD E-M5
Type:	Compact cameras
Date Online:	07/02/2012
Reviews:	256
Amazon Score :	4.6/5
Dpreview Score:	80/100

Table E.16: Information about Olympus OMD E-M5 Compact camera.

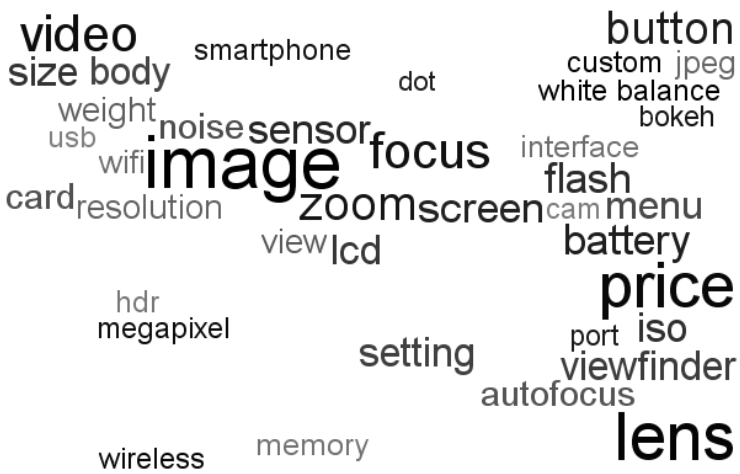


Figure E.17: Concept occurrences found in the reviews of camera presented in Table E.16 (larger means more frequent).

Concept	Occurrences	Concept	Occurrences
image	21649	weight	785
lens	15720	resolution	784
price	14086	jpeg	534
video	5806	wifi	499
focus	5700	interface	367
zoom	5038	memory	339
button	4586	hdr	287
battery	3115	cam	265
screen	3075	usb	245
sensor	2839	white balance	238
flash	2707	custom	183
iso	2437	wireless	181
size	2435	megapixel	173
lcd	2186	smartphone	167
viewfinder	2157	bokeh	113
setting	2123	port	106
body	2106	dot	78
menu	1907		
noise	1069		
card	1059		
autofocus	1054		
view	977		

Table E.17: Quantity of concept occurrences in the reviews of camera in Table E.16.

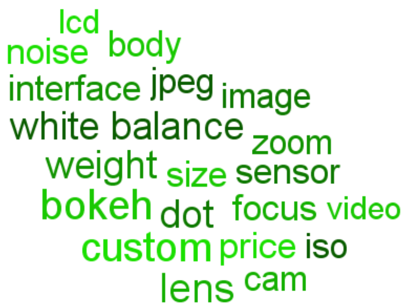


Figure E.18: Gini Argument Bundle B_G of camera in Table E.16. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Gini Argument Bundle B_G					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
custom	0,2340	pro	sensor	0,1461	pro
bokeh	0,2305	pro	jpeg	0,1451	pro
lens	0,2140	pro	noise	0,1439	pro
weight	0,1902	pro	body	0,1427	pro
dot	0,1843	pro	cam	0,1396	pro
white balance	0,1838	pro	zoom	0,1366	pro
price	0,1626	pro	image	0,1320	pro
size	0,1574	pro	iso	0,1315	pro
focus	0,1550	pro	lcd	0,1223	pro
interface	0,1544	pro	video	0,1103	pro

Table E.18: Gini Argument Bundle B_G of camera in Table E.16 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

Figure E.19: Gini Argument Bundle B_G of camera in Table E.16. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Agreement Argument Bundle B_σ					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
autofocus	0,3343	pro	price	0,1987	pro
bokeh	0,3000	pro	sensor	0,1968	pro
custom	0,2880	pro	interface	0,1955	pro
lens	0,2702	pro	body	0,1833	pro
iso	0,2610	pro	zoom	0,1816	pro
weight	0,2340	pro	cam	0,1660	pro
white balance	0,2310	pro	lcd	0,1613	pro
screen	0,2297	pro	menu	0,1510	pro
dot	0,2240	pro	setting	0,1190	pro
size	0,2130	pro	smartphone	0,1160	pro
image	0,2056	pro	jpeg	0,1145	pro
focus	0,2028	pro	memory	0,1070	pro
noise	0,2005	pro	megapixel	0,1060	pro

Table E.19: Agreement Argument Bundle B_σ of camera in Table E.16 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

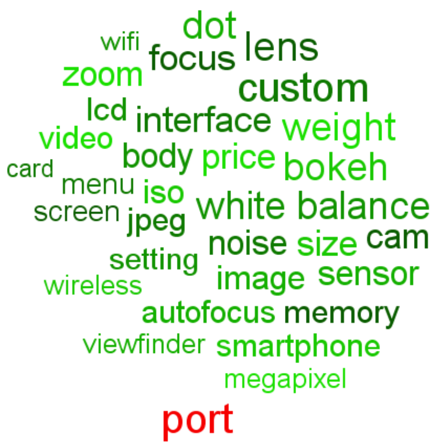


Figure E.20: Cardinality Argument Bundle B_F of camera in Table E.16. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Cardinality Argument Bundle B_F					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
custom	0,5519	pro	image	0,2889	pro
lens	0,5063	pro	lcd	0,2719	pro
weight	0,4802	pro	jpeg	0,2692	pro
bokeh	0,4643	pro	memory	0,2619	pro
dot	0,4545	pro	video	0,2445	pro
white balance	0,4359	pro	smartphone	0,2393	pro
interface	0,3881	pro	autofocus	0,2320	pro
focus	0,3866	pro	setting	0,2120	pro
price	0,3683	pro	menu	0,2013	pro
size	0,3612	pro	screen	0,1985	pro
sensor	0,3331	pro	megapixel	0,1481	pro
body	0,3285	pro	wireless	0,1461	pro
noise	0,3258	pro	viewfinder	0,1452	pro
cam	0,3154	pro	wifi	0,1399	pro
zoom	0,3108	pro	card	0,1094	pro
iso	0,3075	pro	port	-0,1346	con

Table E.20: Cardinality Argument Bundle B_F of camera in Table E.16 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

E.5 Samsung NX3000 (Compact)

Figure E.21: Samsung NX3000 Compact camera.

Name:	Samsung NX3000
Type:	Compact cameras
Date Online:	08/05/2014
Reviews:	31
Amazon Score :	4.4/5
Dpreview Score:	–/100

Table E.21: Information about Samsung NX3000 Compact camera.

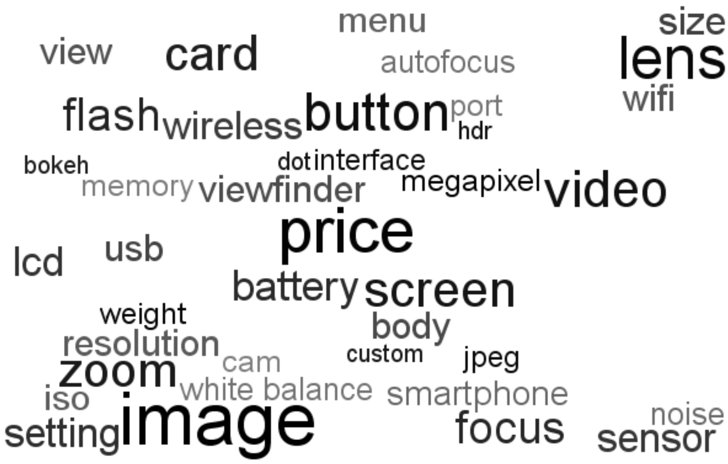


Figure E.22: Concept occurrences found in the reviews of camera presented in Table E.21 (larger means more frequent).

Concept	Occurrences	Concept	Occurrences
image	251	iso	14
price	158	menu	13
lens	131	smartphone	11
button	80	autofocus	6
video	75	noise	6
screen	67	white balance	5
zoom	63	memory	4
card	62	megapixel	4
flash	50	cam	4
focus	45	port	4
battery	36	weight	3
setting	27	interface	2
lcd	27	jpeg	2
sensor	25		
wireless	23		
size	22		
usb	20		
body	19		
viewfinder	18		
view	17		
resolution	15		
wifi	15		

Table E.22: Quantity of concept occurrences in the reviews of camera in Table E.21.

Figure E.23: Gini Argument Bundle B_G of camera in Table E.21. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Gini Argument Bundle B_G					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
wifi	0,2972	pro	focus	0,2108	pro
lens	0,2912	pro	usb	0,1952	pro
menu	0,2705	pro	zoom	0,1697	pro
smartphone	0,2619	pro	setting	0,1082	pro
sensor	0,2206	pro	lcd	0,1022	pro
price	0,2119	pro			

Table E.23: Gini Argument Bundle B_G of camera in Table E.21 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

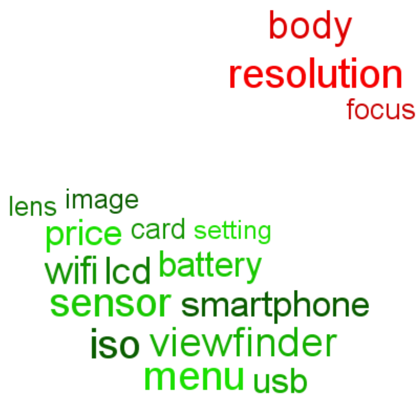


Figure E.24: Gini Argument Bundle B_G of camera in Table E.21. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Agreement Argument Bundle B_σ					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
menu	0,5650	pro	usb	0,3335	pro
iso	0,5630	pro	card	0,1767	pro
sensor	0,5565	pro	lens	0,1608	pro
viewfinder	0,5370	pro	image	0,1507	pro
wifi	0,4500	pro	setting	0,1410	pro
lcd	0,4107	pro	focus	-0,1685	con
price	0,4086	pro	body	-0,4175	con
smartphone	0,4060	pro	resolution	-0,4855	con
battery	0,3637	pro			

Table E.24: Agreement Argument Bundle B_σ of camera in Table E.21 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

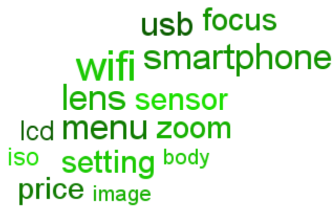


Figure E.25: Cardinality Argument Bundle B_F of camera in Table E.21. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Cardinality Argument Bundle B_F					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
wifi	0,8571	pro	zoom	0,3968	pro
lens	0,6031	pro	focus	0,3778	pro
smartphone	0,6000	pro	price	0,3503	pro
menu	0,5385	pro	lcd	0,2308	pro
usb	0,4444	pro	iso	0,1429	pro
sensor	0,4167	pro	body	0,1111	pro
setting	0,4074	pro	image	0,1093	pro

Table E.25: Cardinality Argument Bundle B_F of camera in Table E.21 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

E.6 Sony a7 (Compact)

Figure E.26: Sony a7 Compact camera.

Name:	Sony a7
Type:	Compact cameras
Date Online:	15/10/2013
Reviews:	83
Amazon Score :	4.3/5
Dpreview Score:	–/100

Table E.26: Information about Sony a7 Compact camera.

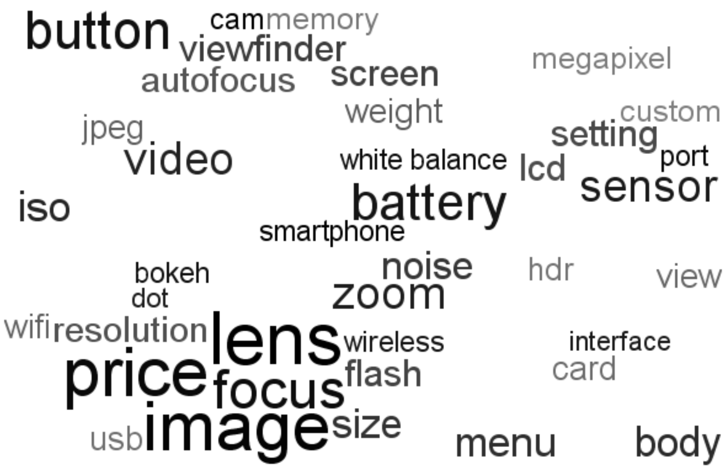


Figure E.27: Concept occurrences found in the reviews of camera presented in Table E.26 (larger means more frequent).

Concept	Occurrences	Concept	Occurrences
lens	552	jpeg	19
image	458	wifi	19
price	426	card	17
focus	221	view	17
button	177	usb	15
battery	172	hdr	12
sensor	113	memory	9
zoom	112	custom	8
video	112	megapixel	8
iso	82	cam	5
body	67	smartphone	4
menu	63	port	4
size	56	bokeh	3
noise	47	white balance	3
viewfinder	43	wireless	3
screen	38	interface	2
flash	37	dot	2
setting	36		
lcd	32		
resolution	29		
autofocus	25		
weight	22		

Table E.27: Quantity of concept occurrences in the reviews of camera in Table E.26.

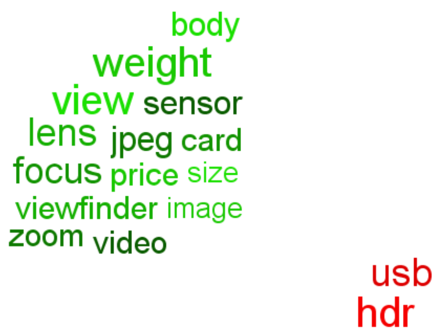


Figure E.28: Gini Argument Bundle B_G of camera in Table E.26. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Gini Argument Bundle B_G					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
view	0,3291	pro	viewfinder	0,1543	pro
weight	0,3159	pro	card	0,1525	pro
lens	0,2569	pro	zoom	0,1499	pro
focus	0,2390	pro	video	0,1395	pro
jpeg	0,1947	pro	size	0,1121	pro
sensor	0,1649	pro	image	0,1028	pro
body	0,1627	pro	usb	-0,1061	con
price	0,1573	pro	hdr	-0,1357	con

Table E.28: Gini Argument Bundle B_G of camera in Table E.26 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

Figure E.29: Gini Argument Bundle B_G of camera in Table E.26. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Agreement Argument Bundle B_σ					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
view	0,5610	pro	image	0,1877	pro
lens	0,5313	pro	focus	0,1714	pro
sensor	0,5278	pro	card	0,1505	pro
weight	0,4700	pro	battery	0,1220	pro
viewfinder	0,3985	pro	body	-0,1053	con
zoom	0,3584	pro	usb	-0,1455	con
iso	0,2940	pro	hdr	-0,2230	con
price	0,2762	pro	size	-0,3177	con

Table E.29: Agreement Argument Bundle B_σ of camera in Table E.26 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

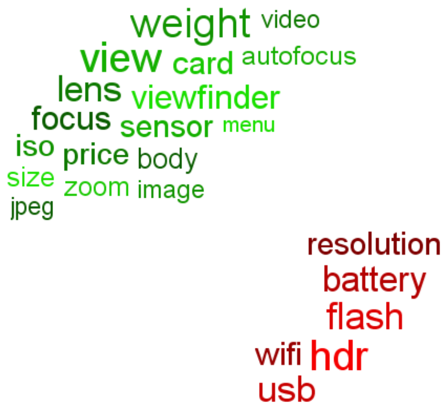


Figure E.30: Cardinality Argument Bundle B_F of camera in Table E.26. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Cardinality Argument Bundle B_F					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
view	0,8824	pro	autofocus	0,2000	pro
weight	0,8182	pro	image	0,1938	pro
lens	0,6015	pro	video	0,1892	pro
focus	0,4909	pro	jpeg	0,1579	pro
viewfinder	0,4884	pro	menu	0,1111	pro
card	0,4118	pro	resolution	-0,1429	con
sensor	0,3982	pro	wifi	-0,1579	con
iso	0,3659	pro	battery	-0,2096	con
price	0,3398	pro	usb	-0,2308	con
body	0,2836	pro	flash	-0,2571	con
size	0,2500	pro	hdr	-0,3333	con
zoom	0,2364	pro			

Table E.30: Cardinality Argument Bundle B_F of camera in Table E.26 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

E.7 Canon PowerShot ELPH 300 HS (Point & Shoot)

Figure E.31: Canon PowerShot ELPH 300 HS Point & Shoot camera.

Name:	Canon PowerShot ELPH 300 HS
Type:	Point & Shoot Digital Cameras
Date Online:	06/02/2011
Reviews:	1112
Amazon Score :	4.2/5
Dpreview Score:	–/100

Table E.31: Information about Canon PowerShot ELPH 300 HS Point & Shoot camera.

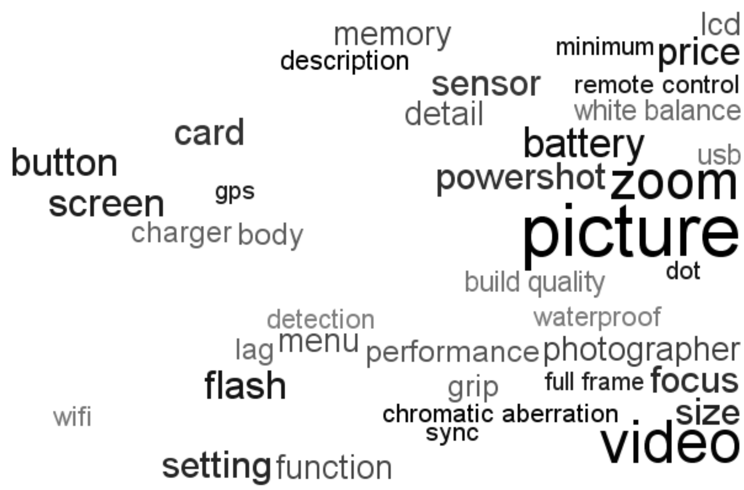


Figure E.32: Concept occurrences found in the reviews of camera presented in Table E.31 (larger means more frequent).

Concept	Occurrences	Concept	Occurrences
picture	2922	body	42
video	1567	charger	36
zoom	1003	lag	23
battery	341	white balance	20
price	265	usb	16
flash	257	build quality	13
button	252	wifi	9
screen	248	waterproof	8
size	237	detection	6
setting	231	description	5
focus	217	chromatic aberration	4
powershot	210	sync	3
card	200	minimum	2
sensor	168	dot	1
photographer	101	gps	1
memory	97	remote control	1
function	89		
menu	88		
detail	79		
lcd	66		
performance	43		
grip	42		

Table E.32: Quantity of concept occurrences in the reviews of camera in Table E.31.

Figure E.33: Gini Argument Bundle B_G of camera in Table E.31. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Gini Argument Bundle B_G					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
price	0,3369	pro	photographer	0,1215	pro
build quality	0,2149	pro	white balance	0,1119	pro
grip	0,1565	pro	card	0,1020	pro
size	0,1351	pro			

Table E.33: Gini Argument Bundle B_G of camera in Table E.31 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

Figure E.34: Gini Argument Bundle B_G of camera in Table E.31. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Agreement Argument Bundle B_σ					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
memory	0,5707	pro	white balance	0,1610	pro
build quality	0,4230	pro	photographer	0,1605	pro
card	0,4140	pro	sensor	0,1462	pro
body	0,3585	pro	video	0,1342	pro
price	0,3000	pro	performance	0,1240	pro
size	0,2773	pro	powershot	0,1185	pro
battery	0,2390	pro	charger	0,1185	pro
grip	0,2060	pro	focus	0,1042	pro
picture	0,1924	pro	button	-0,1350	con
setting	0,1710	pro			

Table E.34: Agreement Argument Bundle B_σ of camera in Table E.31 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

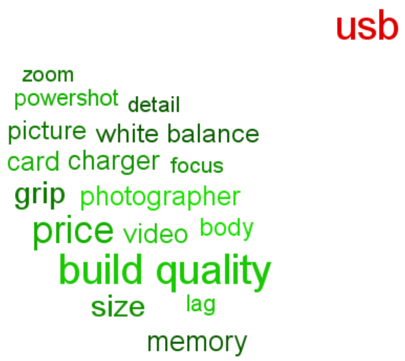


Figure E.35: Cardinality Argument Bundle B_F of camera in Table E.31. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Cardinality Argument Bundle B_F					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
build quality	0,8333	pro	picture	0,2039	pro
price	0,6604	pro	white balance	0,2000	pro
size	0,3391	pro	body	0,1707	pro
grip	0,3171	pro	lag	0,1304	pro
memory	0,2708	pro	powershot	0,1275	pro
photographer	0,2292	pro	focus	0,1111	pro
video	0,2291	pro	zoom	0,1043	pro
charger	0,2222	pro	detail	0,1026	pro
card	0,2222	pro	usb	-0,2000	con

Table E.35: Cardinality Argument Bundle B_F of camera in Table E.31 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

E.8 Panasonic Lumix DMC-GH3K (Point & Shoot)

Figure E.36: Panasonic Lumix DMC-GH3K Point & Shoot camera.

Name:	Panasonic Lumix DMC-GH3K
Type:	Point & Shoot Digital Cameras
Date Online:	17/09/2012
Reviews:	88
Amazon Score :	4.5/5
Dpreview Score:	–/100

Table E.36: Information about Panasonic Lumix DMC-GH3K Point & Shoot camera.

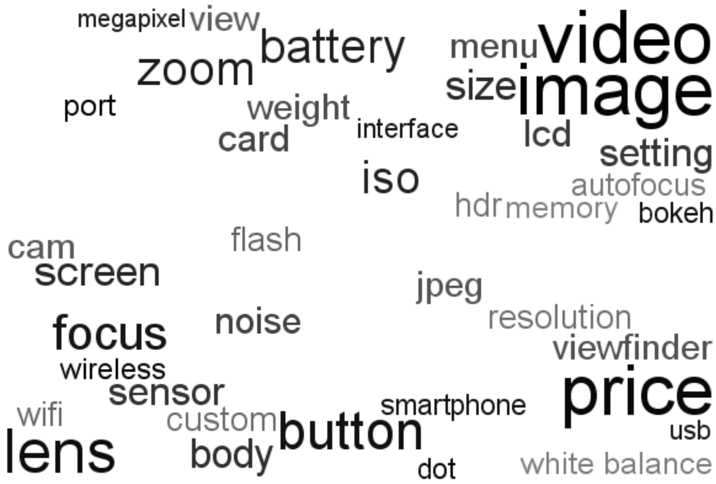


Figure E.37: Concept occurrences found in the reviews of camera presented in Table E.36 (larger means more frequent).

Concept	Occurrences	Concept	Occurrences
picture	384	flash	13
video	305	full frame	7
zoom	229	memory	6
button	96	build quality	6
focus	75	white balance	5
screen	62	usb	5
body	56	charger	3
battery	41	remote control	3
price	40	description	2
detail	39	dot	2
photographer	36	minimum	1
size	34	lag	1
lcd	33		
sensor	28		
setting	26		
card	25		
function	25		
wifi	24		
performance	20		
menu	16		
powershot	16		
grip	15		

Table E.37: Quantity of concept occurrences in the reviews of camera in Table E.36.

Figure E.38: Gini Argument Bundle B_G of camera in Table E.36. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Gini Argument Bundle B_G					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
price	0,3093	pro	screen	0,1612	pro
performance	0,2356	pro	zoom	0,1510	pro
setting	0,2279	pro	video	0,1380	pro
powershot	0,1963	pro	body	0,1192	pro
size	0,1877	pro	picture	0,1161	pro
focus	0,1760	pro			

Table E.38: Gini Argument Bundle B_G of camera in Table E.36 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

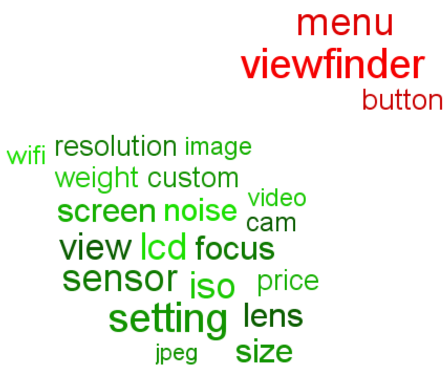


Figure E.39: Gini Argument Bundle B_G of camera in Table E.36. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Agreement Argument Bundle B_σ					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
setting	0,6520	pro	powershot	0,2470	pro
focus	0,5062	pro	body	0,2357	pro
screen	0,4864	pro	menu	0,1340	pro
price	0,3405	pro	photographer	0,1315	pro
performance	0,3020	pro	picture	0,1240	pro
sensor	0,2930	pro	grip	-0,1140	con
size	0,2520	pro	lcd	-0,2820	con

Table E.39: Agreement Argument Bundle B_σ of camera in Table E.36 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

Figure E.40: Cardinality Argument Bundle B_F of camera in Table E.36. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Cardinality Argument Bundle B_F					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
setting	0,6154	pro	detail	0,3333	pro
price	0,5500	pro	photographer	0,3143	pro
size	0,5152	pro	sensor	0,3077	pro
performance	0,5000	pro	picture	0,2368	pro
powershot	0,5000	pro	flash	0,2308	pro
zoom	0,4779	pro	screen	0,2258	pro
focus	0,3514	pro	lcd	0,2121	pro
body	0,3455	pro	card	0,1200	pro
video	0,3333	pro	battery	-0,1000	con

Table E.40: Cardinality Argument Bundle B_F of camera in Table E.36 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

E.9 Fujifilm FinePix Real 3D W3 (Point & Shoot)

Figure E.41: Fujifilm FinePix Real 3D W3 Point & Shoot camera.

Name:	Fujifilm FinePix Real 3D W3
Type:	Point & Shoot Digital Cameras
Date Online:	27/08/2010
Reviews:	285
Amazon Score :	4.5/5
Dpreview Score:	–/100

Table E.41: Information about Fujifilm FinePix Real 3D W3 Point & Shoot camera.

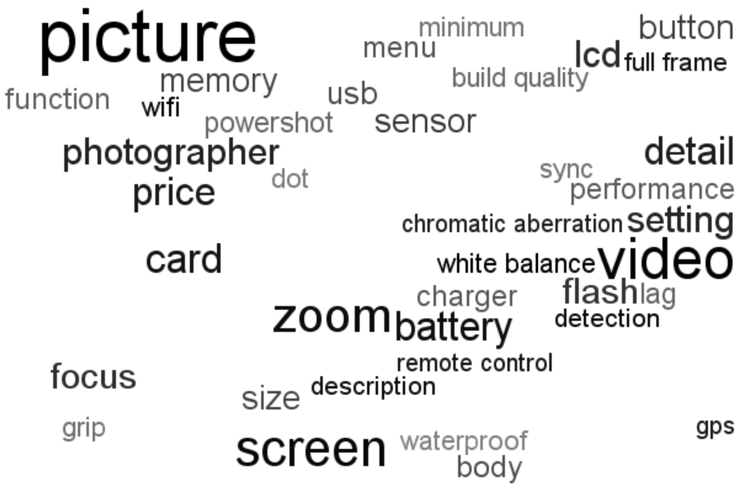


Figure E.42: Concept occurrences found in the reviews of camera presented in Table E.41 (larger means more frequent).

Concept	Occurrences	Concept	Occurrences
picture	1414	performance	11
video	622	lag	9
screen	444	powershot	8
zoom	344	grip	4
battery	155	build quality	4
card	153	sync	3
price	124	minimum	3
detail	123	dot	3
photographer	95	waterproof	3
setting	94	white balance	2
lcd	90	wifi	2
focus	73	description	1
flash	65	detection	1
sensor	60		
button	56		
memory	47		
size	38		
usb	32		
body	19		
menu	17		
charger	14		
function	11		

Table E.42: Quantity of concept occurrences in the reviews of camera in Table E.41.

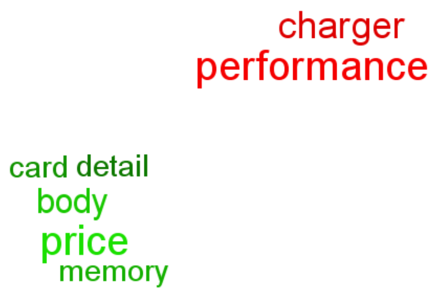


Figure E.43: Gini Argument Bundle B_G of camera in Table E.41. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Gini Argument Bundle B_G					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
price	0,3024	pro	detail	0,1153	pro
body	0,1748	pro	charger	-0,1059	con
memory	0,1269	pro	performance	-0,1434	con
card	0,1188	pro			

Table E.43: Gini Argument Bundle B_G of camera in Table E.41 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

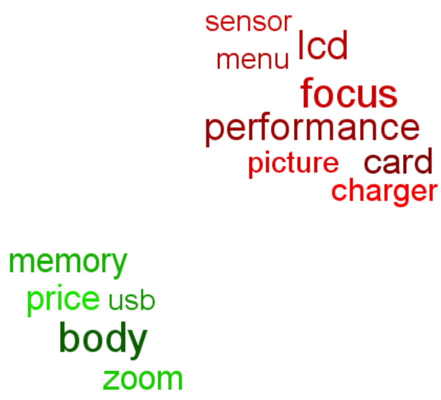


Figure E.44: Gini Argument Bundle B_G of camera in Table E.41. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Agreement Argument Bundle B_σ					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
body	0,3283	pro	picture	-0,1278	con
price	0,2125	pro	charger	-0,1550	con
zoom	0,1928	pro	card	-0,2283	con
memory	0,1903	pro	performance	-0,2520	con
usb	0,1155	pro	lcd	-0,2818	con
sensor	-0,1030	con	focus	-0,3173	con
menu	-0,1060	con			

Table E.44: Agreement Argument Bundle B_σ of camera in Table E.41 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

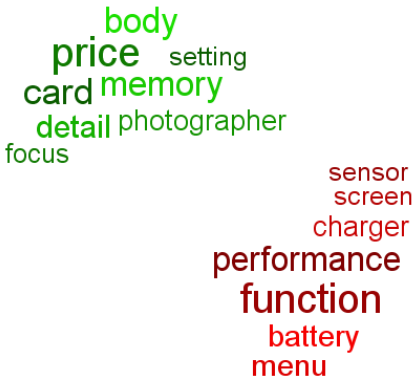


Figure E.45: Cardinality Argument Bundle B_F of camera in Table E.41. Pro arguments are presented in green and con arguments in red. Size corresponds with the strength of the argument sentiments, the larger the stronger.

Cardinality Argument Bundle B_F					
Concept	Sentiment	Argument	Concept	Sentiment	Argument
price	0,5806	pro	sensor	-0,1034	con
card	0,4172	pro	screen	-0,1116	con
body	0,3684	pro	charger	-0,1429	con
memory	0,3617	pro	menu	-0,1765	con
detail	0,2881	pro	battery	-0,1812	con
photographer	0,1915	pro	performance	-0,2727	con
focus	0,1507	pro	function	-0,4545	con
setting	0,1494	pro			

Table E.45: Cardinality Argument Bundle B_F of camera in Table E.41 showing an argument at each row composed of concept name, sentiment value and argument polarity (pro or con).

Appendix F

Bundle Comparison for DSLR, Compact, and Point & Shoot cameras

In this appendix we show the quantity of pros shared by two or three bundle types (B_G, B_σ and B_F) for each of the top 50 DSLR, Compact and Point & Shoot products with more reviews. Here "shared" means the intersection, i.e. those pro concepts that appear in the three bundle types of a product p. For instance, concept "sensor" is considered a pro argument in the Gini bundle B_G, the agreement bundle B_σ, and the cardinality bundles of arguments B_F of the Nikon D5300. As such, we say that pro argument is "shared" between the three bundles of arguments of the Nikon D5300.

Figure F.1 focuses on the top 50 products with more reviews from the DSLR corpus K_D. The x axis corresponds with the product names, and the y axis shows the quantity of pros shared by two or three bundle types for each product. The results show that almost 8 out of 10 pro arguments are shared between the three types of bundles of the top 50 DSLR products with more reviews, indicating a strong consistency between the three types of bundles of a product. This means that a concept of a pro argument in a B_G is also likely present in the B_σ pros and the B_F pros. Furthermore, we show that the quantity of pro arguments (and also con arguments, not included in this figure because results are similar) of a product p is directly related with the quantity of concept occurrences $Occ(C, p)$ in the reviews: the more reviews the richer in pro and cons the bundles are. With more user judgments related to the various concepts of the concept vocabulary, we have more knowledge to make an informed decision to consider an argument as a pro or con instead of a moot. Notice that we are only analyzing whether a concept is part of a pro argument in more than one of bundle types of a product, but we are not comparing the concrete positive sentiment values of the arguments.

Figure F.2 studies the relation between the quantity of user judgments in the reviews of DSLR products and the quantity of moot arguments in the bundles of that product. The x axis corresponds with the product names, the left y axis shows the total added quantity of moot arguments of the three types of bundles of arguments of each product, and the right y axis shows the number of concept occurrences $Occ(C,p)$ in the reviews of each product. The total added quantity of moot arguments corresponds to the union of all moot arguments of the three bundles of arguments B_G, B_σ, and B_F of a product. Since the DSLR concept vocabulary $\mathcal{C}_D$ has 41 different concepts, and in each of the three bundles of arguments of a DSLR product we define an argument per each concept in $\mathcal{C}_D$, the maximum quantity of moot arguments, considering that all arguments are moots, is 123 (41 arguments per bundle type).

The results show that there is an inverse correlation between the quantity of user judgments and moot arguments in the bundles of arguments of a product. For instance, the reviews of the Canon EOSM contain 275 judgments related to the concepts of the DSLR concept vocabulary $\mathcal{C}_D$. This quantity of user judgments is low if we compare them with the quantity of judgments of other cameras in the figure. As a consequence, most of the arguments of the three bundles of arguments of the Canon EOSM (121 arguments) are moot arguments, because we do not have enough evidence to define those arguments as pro or cons. On the other hand, the reviews of the Nikon D7100 contain 11.652 judgments related to the concepts of the concept vocabulary $\mathcal{C}_D$. Considering the three bundles of arguments of that product, only 16 arguments out of 123 are moots. The rest of arguments are either pro or con arguments.

Figure F.3 focuses on the top 50 products with more reviews from the Compact corpus K_C. We observe similarities between the results for Compact cameras presented in Figure F.3 and the results for the DSLR cameras presented in Figure F.1. Almost 70% of pro arguments are shared between two or three different bundles of arguments of products, and the quantity of pro arguments is also directly correlated with the quantity of concept occurrences in the reviews of the products. For Compact cameras, the product with more concept occurrences is the Olympus OMD EM5 (6.195 concept occurrences), with 26 pro arguments (out of 39 concepts in the Compact concept vocabulary $\mathcal{C}_C$) shared between two or three bundles B_G, B_σ, and B_F. Notice that the reviews of this product contain almost half of the concept occurrences of the Nikon D7100, the DSLR product with most concept occurrences (11.652). However, the difference of pro arguments shared by two or more bundles between those cameras is only 7 (26 for the Olympus OMD EM5, and 33 for the Nikon D7100).

Finally, Figure F.4 shows the top 50 products with more reviews from the Point & Shoot cameras corpus K_P. As expected, we also have a strong consistency between the bundles of arguments of Point & Shoot cameras, being more than 8 out of 10 pro concepts shared between bundles. Notice that, since we have more reviews for Point & Shoot cameras than for Compact, the product with less concept occurrences of the Point & Shoot top 50, the Fujifilm X20 (with 3.262 concept occurrences), has more than 6 times the concept occurrences of the Nikon 1 AW1 (495), the Compact camera with less concept occurrences of the ranking presented in Figure F.3. As such, the bundles of arguments of the top 50 Point & Shoot cameras have, in average, more pro and con arguments than the bundles of arguments of the top 50 Compact cameras, and a similar average quantity of pro concepts than the bundles of arguments of the DSLR cameras.

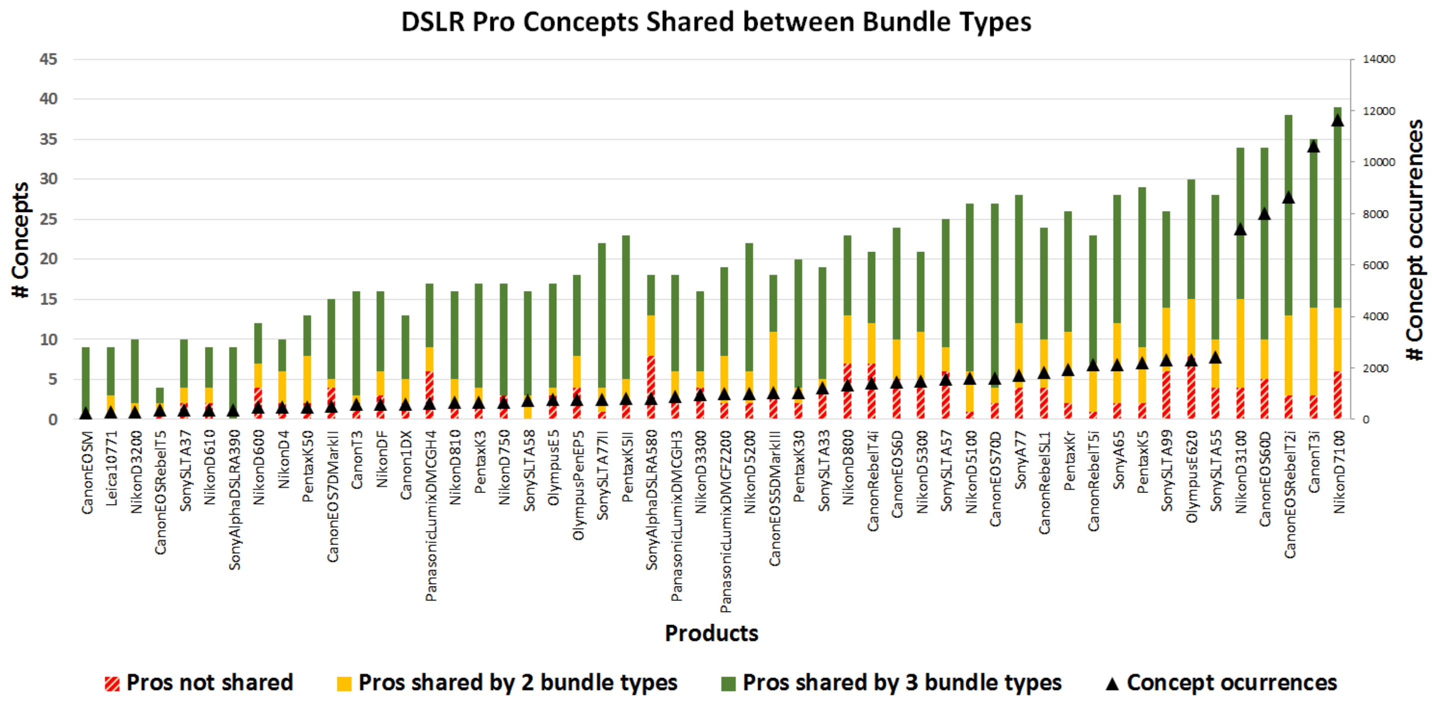


Figure F.1: Quantity of pros shared between the three bundles of arguments B_G, B_σ and B_F for DSLR cameras, together with the number of occurrences of the pro concepts in the reviews of the product.

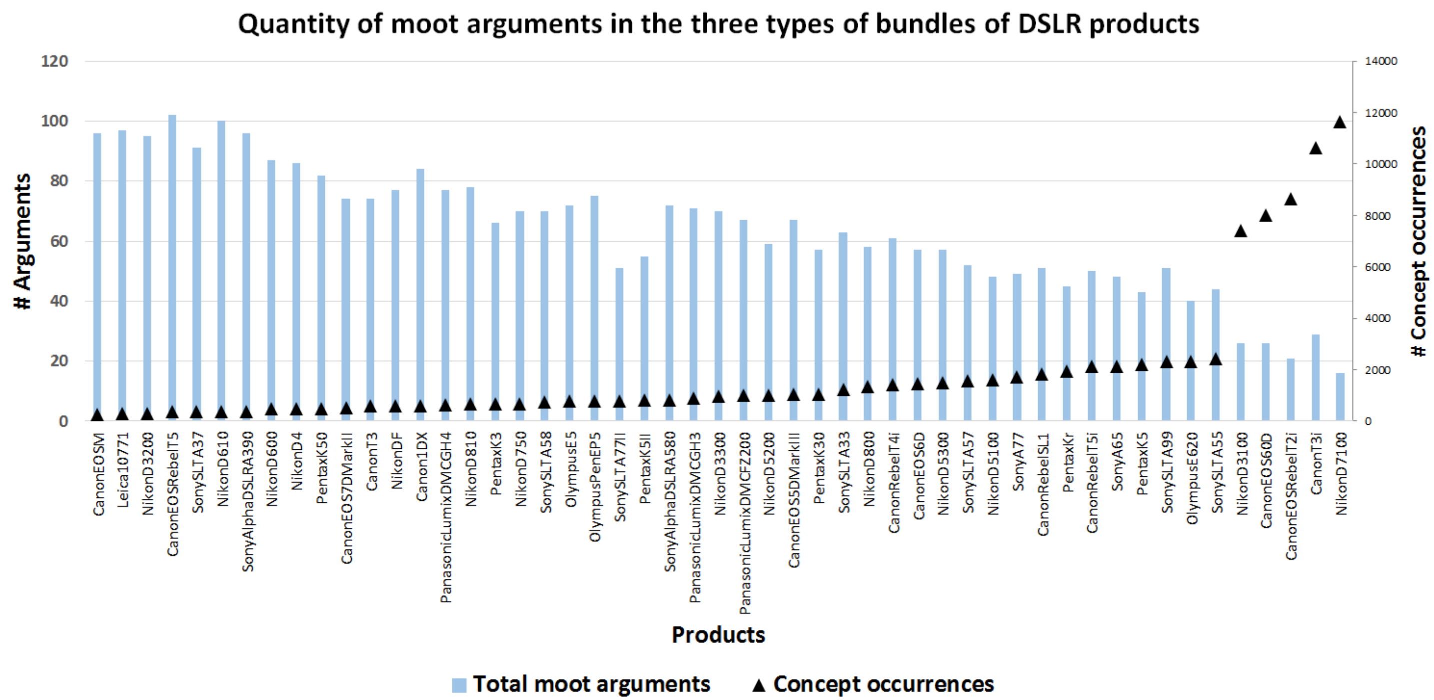


Figure F.2: Total quantity of moot arguments in the three bundles of arguments B_G, B_σ and B_F for the top 50 DSLR cameras

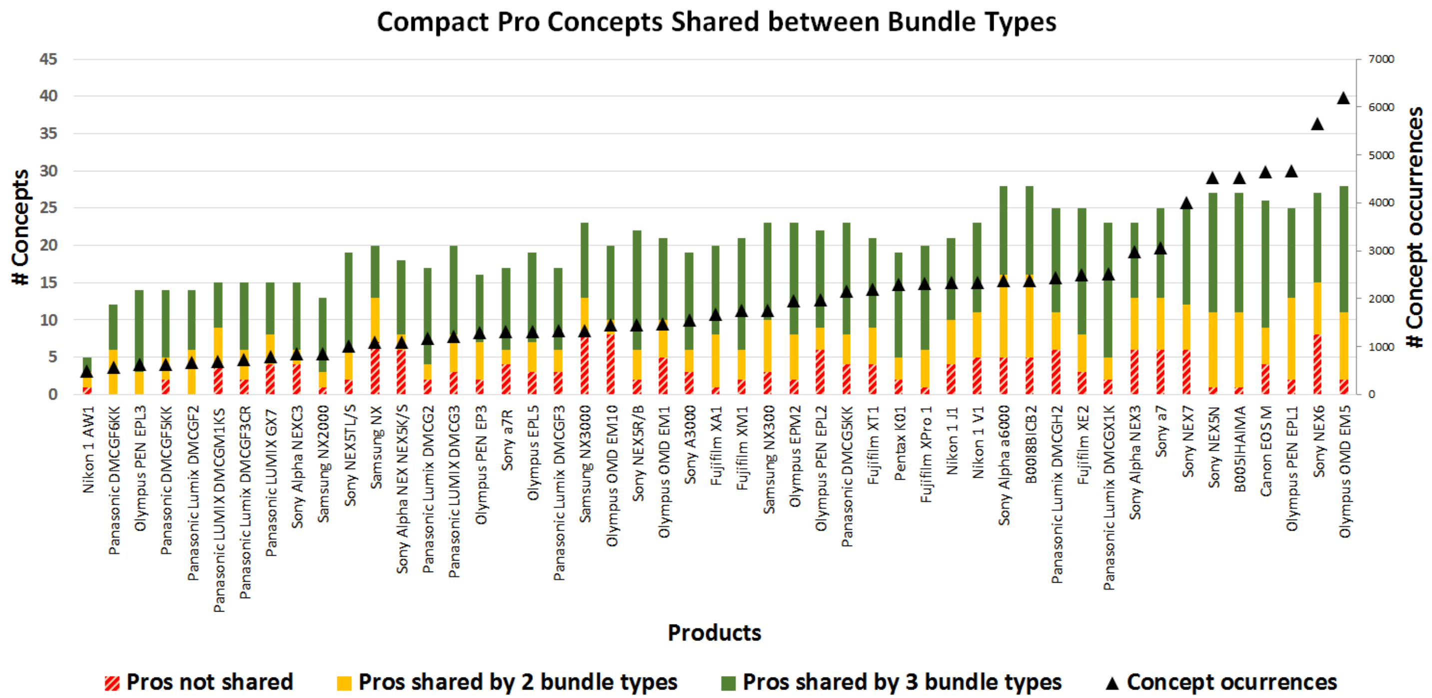


Figure F.3: Quantity of pros shared between the three bundles of arguments B_G, B_σ and B_F for Compact cameras, together with the number of occurrences of the pro concepts in the reviews of the product.

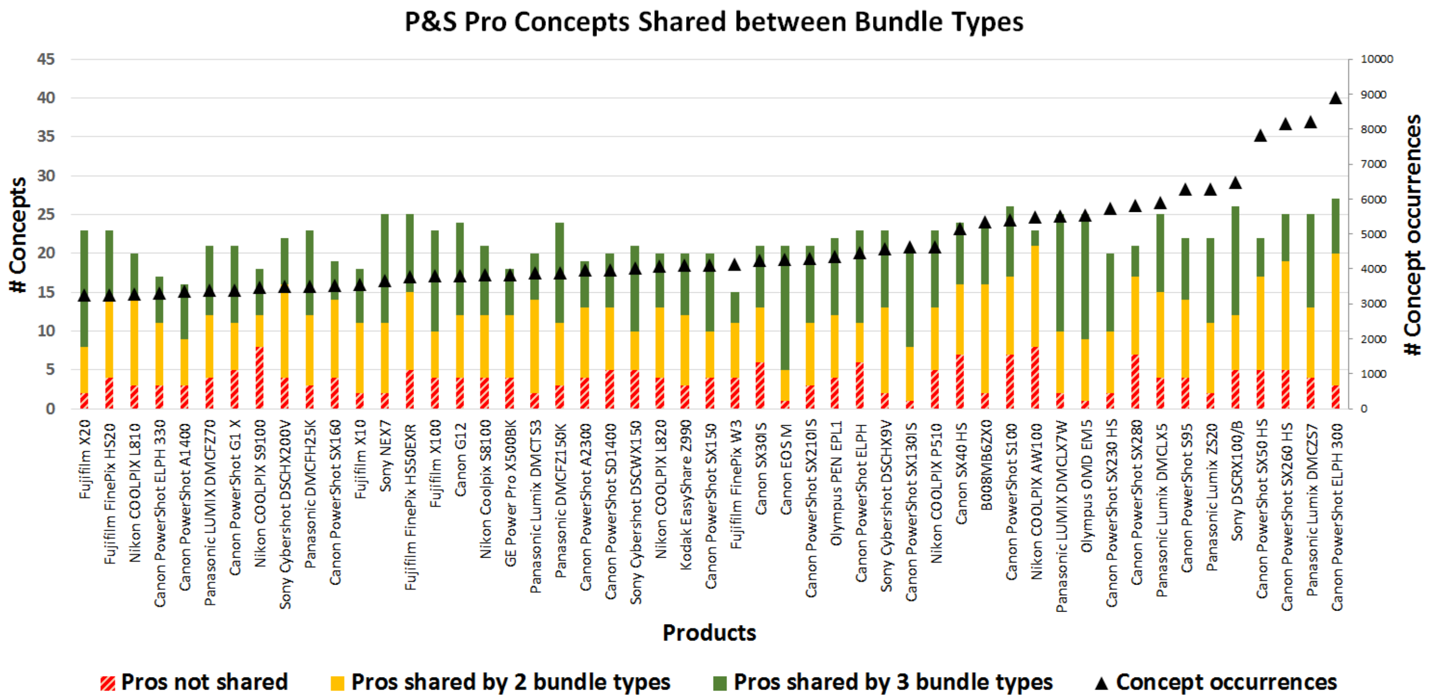


Figure F.4: Quantity of pros shared between the three bundles of arguments B_G, B_σ and B_F for Point & Shoot cameras, together with the number of occurrences of the pro concepts in the reviews of the product.

Bibliography

Abdi, H. (2007). The Kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA*, pages 508–510.

Amatriain, X., Pujol, J., and Oliver, N. (2009). I like it... I like it not: Evaluating user ratings noise in recommender systems. In *Proceedings of the International Conference on User Modeling, Adaptation, and Personalization (UMAP'09)*, pages 247–258. Springer.

Andersen, P. (2007). *What is Web 2.0?: Ideas, technologies and implications for education*. JISC Bristol.

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation Conference (LREC'10)*, volume 10, pages 2200–2204.

Baccigalupo, C. and Plaza, E. (2006). Case-based sequential ordering of songs for playlist recommendation. In *Proceedings of the 8th European Conference on Case-Based Reasoning (ECCBR'06)*, pages 286–300. Springer.

Baccigalupo, C. and Plaza, E. (2007). A case-based song scheduler for group customised radio. In *Proceedings of the 7th International Conference on Case-Based Reasoning (ICCBR'07)*, pages 433–448. Springer.

Baddeley, A. (1992). Working memory and conscious awareness. *Theories of Memory*, pages 11–20.

Basler, H. (1988). Equivalence between tie-corrected Spearman test and a chi-square test in a fourfold contingency table. *Metrika*, 35(1):203–209.

Bench-Capon, T. and Dunne, P. (2007). Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641.

Berger, A. L., Pietra, V., and Pietra, S. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.

Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G. A., and Reynar, J. (2008). Building a sentiment summarizer for local service reviews.

In *Proceedings of the 10th International Conference on World Wide Web, Workshop on Natural Language Processing in the Information Explosion Era (WWW'08)*, volume 14, pages 339–348.

Blei, D. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

Blei, D., Ng, A., and Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022.

Bondarenko, A., Dung, P., Kowalski, R., and Toni, F. (1997). An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1-2):63–101.

Bosco, C., Patti, V., and Bolioli, A. (2013). Developing corpora for sentiment analysis: The case of irony and senti-TUT. *IEEE Intelligent Systems*, 28(2):55–63.

Bridge, D. and Healy, P. (2012). The Ghostwriter-2.0 case-based reasoning system for making content suggestions to the authors of product reviews. *Knowledge-Based Systems*, 29:93–103.

Bridge, D. and Waugh, A. (2009). Using experience on the read/write web: The Ghostwriter system. In *Proceedings of the International Conference on Case-Based Reasoning, Workshop on WebCBR: Reasoning from Experiences on the Web (ICCBR'09)*, pages 15–24.

Carroll, J., Minnen, G., and Briscoe, T. (1999). Corpus annotation for parser evaluation. *ArXiv Preprint CS/9907013*.

Carvalho, P., Sarmento, L., Silva, M., and De Oliveira, E. (2009). Clues for detecting irony in user-generated contents: Oh...!! It's so easy;-). In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management, Workshop on Topic-sentiment Analysis for Mass Opinion (CIKM'09)*, pages 53–56. ACM.

Ceci, F., Weber, R., Gonçalves, A., and Pacheco, R. (2015). Adapting sentiments with context. In *Proceedings of the 23rd International Conference on Case-Based Reasoning (ICCBR'15)*, pages 44–59. Springer.

Chen, L., Chen, G., and Wang, F. (2015a). Recommender systems based on user reviews: the state of the art. *User Modeling and User-Adapted Interaction*, 25(2):99–154.

Chen, Y., Ferrer, X., Wiratunga, N., and Plaza, E. (2014). Sentiment and preference guided social recommendation. In *Proceedings of the 22nd International Conference on Case-Based Reasoning (ICCBR'14)*, pages 79–94. Springer.

Chen, Y., Ferrer, X., Wiratunga, N., and Plaza, E. (2015b). Aspect selection for social recommender systems. In *Proceedings of the 23rd International Conference on Case-Based Reasoning (ICCBR'15)*, pages 60–72. Springer.

Chesñevar, C., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., and Willmott, S. (2006). Towards an argument interchange format. *The Knowledge Engineering Review*, 21(04):293–316.

Cho, Y., Cho, Y., and Kim, S. (2005). Mining changes in customer buying behavior for collaborative recommendations. *Expert Systems Applications*, 28:359–369.

De Marneffe, M., MacCartney, B., and Manning, C. (2006). Generating typed dependency parses from phrase structure parses. In *Proceedings of the 5th International Conference on Language Resources and Evaluation Conference (LREC'06)*, volume 6, pages 449–454. Genoa.

Denecke, K. (2008). Using sentiwordnet for multilingual sentiment analysis. In *Proceedings of the Institute of Electrical and Electronics Engineers, Workshop on Data Engineering (IEEE'08)*, pages 507–512. IEEE.

Ding, X., Liu, B., and Yu, P. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM'08)*, pages 231–240. ACM.

Ding, Y. and Li, X. (2005). Time weight collaborative filtering. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM'05)*, CIKM '05, pages 485–492.

Dong, R., O'Mahony, M., and Smyth, B. (2014). Further experiments in opinionated product recommendation. In *Proceedings of the 22nd International Conference on Case-Based Reasoning (ICCBR'14)*, pages 110–124. Springer.

Dong, R., Schaal, M., O'Mahony, M., McCarthy, K., and Smyth, B. (2013a). Mining features and sentiment from review experiences. In *Proceedings of the 21st International Conference on Case-Based Reasoning (ICCBR'13)*, pages 59–73.

Dong, R., Schaal, M., O'Mahony, M., McCarthy, K., and Smyth, B. (2013b). Opinionated product recommendation. In *Proceedings of the 21st International Conference on Case-Based Reasoning (ICCBR'13)*, volume 7969 of *Lecture Notes in Computer Science*, pages 44–58. Springer.

Dong, R. and Smyth, B. (2016). Personalized opinion-based recommendation. In *Proceedings of the 24th International Conference on Case-Based Reasoning (ICCBR'16)*, pages 93–107. Springer.

Dumais, S. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, 38(1):188–230.

Dumais, S., Platt, J., Heckerman, D., and Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. In *Proceedings of the 7th ACM International Conference on Information and Knowledge Management (CIKM'98)*, pages 148–155. ACM.

Dung, P. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357.

Esuli, A. and Sebastiani, F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation Conference (LREC'06)*, pages 417–422.

Ferrer, X., Chen, Y., Wiratunga, N., and Plaza, E. (2014). Preference and sentiment guided social recommendations with temporal dynamics. In *Research and Development in Intelligent Systems XXXI*, pages 101–116. Springer.

Ferrer, X. and Plaza, E. (2016). Concept discovery and argument bundles in the Experience Web. In *Proceedings of the 24th International Conference on Case-Based Reasoning (ICCBR'16)*, pages 108–123. Springer.

Freyne, J. and Smyth, B. (2009). Many cases make light work for visualization in many eyes. In *Proceedings of the International Conference on Case-Based Reasoning, Workshop on WebCBR: Reasoning from Experiences on the Web (ICCBR'09)*.

Fu, G. and Luke, K. (2005). Chinese named entity recognition using lexicalized HMMs. *ACM SIGKDD Explorations Newsletter*, 7(1):19–25.

Ganu, G., Elhadad, N., and Marian, A. (2009). Beyond the stars: Improving rating predictions using review text content. In *Proceedings of the 12th International Workshop on the Web and Databases (WebDB'09)*, volume 9, pages 1–6. Citeseer.

Goel, V., Gold, B., Kapur, S., and Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *NeuroReport*, 8(5):1305–1310.

Grishman, R. (1997). Information extraction: Techniques and challenges. In *Information extraction a multidisciplinary approach to an emerging information technology*, pages 10–27. Springer.

Gupta, S. and Manning, C. (2011). Analyzing the dynamics of research by extracting key aspects of scientific papers. In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP'11)*, pages 1–9.

Hagoort, P., Baggio, G., and Willems, R. (2009). Semantic unification. In *The cognitive neurosciences, 4th ed.*, pages 819–836. MIT press.

Hai, Z., Chang, K., and Kim, J. (2011). Implicit feature identification via co-occurrence association rule mining. In *International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'11)*, pages 393–404. Springer.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques.* Elsevier.

Healy, P. and Bridge, D. (2010). The Ghostwriter-2.0 system: Creating a virtuous circle in web 2.0 product reviewing. In *Proceedings of the International Conference on Case-Based Reasoning, Workshop on WebCBR: Reasoning from Experiences on the Web (ICCBR'10)*, pages 121–130.

Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence (UAI'99)*, pages 289–296. Morgan Kaufmann Publishers Inc.

Hong, W., Li, L., and Li, T. (2012). Product recommendation with temporal dynamics. *Expert Systems Applications*, 39:12398–12406.

Hu, M. and Liu, B. (2004a). Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*, pages 168–177. ACM.

Hu, M. and Liu, B. (2004b). Mining opinion features in customer reviews. In *Proceedings of the 19th National Conference on Artificial Intelligence (AAAI'04)*, pages 755–760.

Huang, J., Etzioni, O., Zettlemoyer, L., Clark, K., and Lee, C. (2012). Revminer: An extractive interface for navigating reviews on a smartphone. In *Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST'12)*, pages 3–12. ACM.

Ihle, N., Hant, A., and Althoff, K. (2009). Extraction of adaptation knowledge from internet communities. In *Proceedings of the International Conference on Case-Based Reasoning, Workshop on WebCBR: Reasoning from Experiences on the Web (ICCBR'09)*, pages 35–44.

Izquierdo, R., Suárez, A., and Rigau, G. (2007). Exploring the automatic selection of basic level concepts. In *Proceedings of Recent Advances in Natural Language Processing (RANLP'07)*, volume 7. Citeseer.

Izquierdo, R., Suárez, A., and Rigau, G. (2015). Word vs. class-based word sense disambiguation. *Journal of Artificial Intelligence Research*, 54:83–122.

Jakob, N. and Gurevych, I. (2010). Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP'10)*, pages 1035–1045. Association for Computational Linguistics.

Janssen, D. and Kies, R. (2005). Online forums and deliberative democracy. *Acta Política*, 40(3):317–335.

Jiang, L., Yu, M., Zhou, M., Liu, X., and Zhao, T. (2011). Target-dependent Twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-11)*, pages 151–160. Association for Computational Linguistics.

Jin, O., Liu, N., Zhao, K., Yu, Y., and Yang, Q. (2011). Transferring topical knowledge from auxiliary long texts for short text clustering. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM'11)*, pages 775–784. ACM.

Jin, W., Ho, H., and Srihari, R. (2009). Opinionminer: A novel machine learning system for web opinion mining and extraction. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, pages 1195–1204. ACM.

Jo, Y. and Oh, A. (2011). Aspect and sentiment unification model for online review analysis. In *Proceedings of the 2011 International Conference on Web Search and Data Mining (WSDM'11)*, pages 815–824. ACM.

Johnson, R. (2012). *Manifest rationality: A pragmatic theory of argument*. Routledge.

Kao, A. and Poteet, S. R. (2007). *Natural language processing and text mining*. Springer Science & Business Media.

Kessler, J. and Nicolov, N. (2009). Targeting sentiment expressions through supervised ranking of linguistic configurations. In *Proceedings of the 3rd International Conference on Weblogs and Social Media (ICWSM'09)*.

Kim, S. and Hovy, E. (2006). Automatic identification of pro and con reasons in online reviews. In *Proceedings of the 21st International Conference on Computational Linguistics (COLING'06)*, pages 483–490. Association for Computational Linguistics.

King, T., Crouch, R., Riezler, S., Dalrymple, M., and Kaplan, R. (2003). The parc 700 dependency bank. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL'03), Workshop on Linguistically Interpreted Corpora (LINC'03)*, pages 1–8.

Klein, D. and Manning, C. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL-03)*, pages 423–430. Association for Computational Linguistics.

Kobayashi, N., Iida, R., Inui, K., and Matsumoto, Y. (2006). Opinion mining on the web by extracting subject-aspect-evaluation relations. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI'06)*, pages 86–91.

Kolter, J. and Maloof, M. (2003). Dynamic weighted majority: A new ensemble method for tracking concept drift. In *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM'03)*, pages 123–130.

Koren, Y. (2010). Collaborative filtering with temporal dynamics. *Communications of Association for Computing Machinery*, 53:89–97.

Lafferty, J., McCallum, A., and Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning (ICML'01)*, volume 1, pages 282–289.

Lakkaraju, H., Bhattacharyya, C., Bhattacharya, I., and Merugu, S. (2011). Exploiting coherence for the simultaneous discovery of latent facets and associated sentiments. In *Proceedings of the 2011 SIAM international conference on data mining (SIAM'11)*, pages 498–509. SIAM.

Lancichinetti, A., Sirer, M., Wang, J., Acuna, D., Körding, K., and Amaral, L. (2015). High-reproducibility and high-accuracy method for automated topic classification. *Physical Review X*, 5(1):011007.

Landauer, T. (2006). *Latent semantic analysis*. Wiley Online Library.

Lee, S., Tsujii, J., and Rim, H. (2000). Lexicalized hidden Markov models for part-of-speech tagging. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING'00)*, volume 1, pages 481–487. Association for Computational Linguistics.

Li, Z., Zhang, M., Ma, S., Zhou, B., and Sun, Y. (2009). Automatic extraction for product feature words from comments on the web. In *Asia Information Retrieval Symposium (AIRS'09)*, pages 112–123. Springer.

Liu, B. (2007). *Web data mining: Exploring hyperlinks, contents, and usage data*. Springer Science & Business Media.

Liu, B., Hu, M., and Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of the 14th International Conference on World Wide Web (WWW'05)*, pages 342–351. ACM.

Loui, R. (2005). A citation-based reflection on toulmin and argument. *Argumentation*, 19(3):259–266.

Lu, B., Ott, M., Cardie, C., and Tsou, B. (2011). Multi-aspect sentiment analysis with topic models. In *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM'11)*, pages 81–88. IEEE.

Malouf, R. (2002). A comparison of algorithms for maximum entropy parameter estimation. In *Proceedings of the 6th Conference on Computational Natural Language Learning (CoNLL'02)*, volume 20, pages 1–7. Association for Computational Linguistics.

Maynard, D. and Greenwood, M. (2014). Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis. In *Proceedings of the 9th International Conference on Language Resources and Evaluation Conference (LREC'14)*, pages 4238–4243.

McCarthy, K., Salem, Y., and Smyth, B. (2010). Experience-based critiquing: Reusing critiquing experiences to improve conversational recommendation. In *Proceedings of the 18th International Conference on Case-Based Reasoning (ICCBR'10)*, pages 480–494.

Meng, L., Huang, R., and Gu, J. (2013). A review of semantic similarity measures in Wordnet. *International Journal of Hybrid Information Technology*, 6(1):1–12.

Moghaddam, S. and Ester, M. (2010). Opinion digger: An unsupervised opinion miner from unstructured product reviews. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM'10)*.

Moghaddam, S. and Ester, M. (2011). ILDA: Interdependent LDA model for learning latent aspects and their ratings from online product reviews. In *Proceedings of the 34th ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR'11)*, pages 665–674. ACM.

Moghaddam, S. and Ester, M. (2012). On the design of LDA models for aspect-based opinion mining. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM'12)*, pages 803–812. ACM.

Muhammad, A., Wiratunga, N., Lothian, R., and Glassey, R. (2013). Contextual sentiment analysis in social media using high-coverage lexicon. In *Proceedings of the 33rd International Conference of the British Computer Society's Specialist Group on Artificial Intelligence (AI'13)*, pages 79–93.

Muhammad, K., Lawlor, A., Rafter, R., and Smyth, B. (2015). Great explanations: Opinionated explanations for recommendations. In *Proceedings of the 23rd International Conference on Case-Based Reasoning (ICCBR'15)*, pages 244–258. Springer.

O'reilly, T. (2005). What is web 2.0.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP'02)*, pages 79–86. Association for Computational Linguistics.

Peng, F. and McCallum, A. (2006). Information extraction from research papers using conditional random fields. *Information Processing & Management*, 42(4):963–979.

Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Střítesky̌, V., and Holzinger, A. (2013). Opinion mining on the web 2.0–characteristics of user generated content and their impacts. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*, pages 35–46. Springer.

Pinkwart, N., Aleven, V., Ashley, K., and Lynch, C. (2006). Toward legal argument instruction with graph grammars and collaborative filtering techniques. In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems (ITS'06)*, pages 227–236. Springer.

Plaza, E. (2008). Semantics and experience in the future web. In *Proceedings of the 9th European Conference on Case-Based Reasoning (ECCBR'08)*, volume 5239 of *Lecture Notes in Artificial Intelligence*, pages 44–58. Springer Verlag.

Plaza, E. (2009). On reusing other people's experiences. In *Künstliche Intelligenz*, volume 21, pages 18–23.

Plaza, E. and Baccigalupo, C. (2009). Principle and praxis in the experience web: A case study in social music. In *Proceedings of the International Conference in Case-Based Reasoning, Workshop on WebCBR: Reasoning from Experiences on the Web (ICCBR'09)*, pages 55–63.

Popescu, A. and Etzioni, O. (2007). Extracting product features and opinions from reviews. In *Natural Language Processing and Text Mining*, pages 9–28. Springer.

Rahwan, I., Zablith, F., and Reed, C. (2007). Laying the foundations for a world wide argument web. *Artificial Intelligence*, 171(10-15):897–921.

Riloff, E., Wiebe, J., and Phillips, W. (2005). Exploiting subjectivity classification to improve information extraction. In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI'05)*, volume 20, page 1106. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Ronen, R., Koenigstein, N., Ziklik, E., and Nice, N. (2013). Selecting content-based features for collaborative filtering recommenders. In *Proceedings of the 7th ACM International Conference on Recommender Systems (RecSys'13)*, pages 407–410. ACM.

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4(3):328–350.

Rosch, E. (1977). Human categorization. *Studies in Cross-Cultural Psychology*, 1:1–49.

Rosch, E. and Lloyd, B., editors (1978). *Cognition and Categorization*. Lawrence Erlbaum.

Rosch, E., Mervis, C., Gray, W., Johnson, D., and Boyes-Braem, P. (2004). Basic objects in natural categories. *Cognitive Psychology: Key Readings*, 448.

Saif, H., He, Y., Fernandez, M., and Alani, H. (2014). Adapting sentiment lexicons using contextual semantics for sentiment analysis of Twitter. In *Proceedings of the 11th European Semantic Web Conference (ESWC'14)*, pages 54–63. Springer.

Sani, S., Wiratunga, N., Massie, S., and Lothian, R. (2011). Term similarity and weighting framework for text representation. In *Proceedings of the 19th International Conference on Case-Based Reasoning (ICCBR'11)*, pages 304–318. Springer.

Sauper, C. and Barzilay, R. (2013). Automatic aggregation by joint modeling of aspects and values. *Journal of Artificial Intelligence Research*, 46:89–127.

Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H., and Jin, C. (2007). Red opal: Product-feature scoring from reviews. In *Proceedings of the 8th ACM Conference on Electronic Commerce (EC'07)*, pages 182–191. ACM.

Scheuer, O., Loll, F., Pinkwart, N., and McLaren, B. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5(1):43–102.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys (CSUR)*, 34(1):1–47.

Sierra, C., Jennings, N., Noriega, P., and Parsons, S. (1997). A framework for argumentation-based negotiation. In *Proceedings of the International Workshop on Intelligent Agents, Workshop on Agent Theories, Architectures, and Languages (ATAL'97)*, pages 177–192. Springer.

Spina, D., Meij, E., De Rijke, M., Oghina, A., Bui, M. T., and Breuss, M. (2012). Identifying entity aspects in microblog posts. In *Proceedings of the 35th ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR'12)*, pages 1089–1090. ACM.

Spink, A., Wolfram, D., Jansen, M., and Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3):226–234.

Suthers, D. (2001). Architectures for computer supported collaborative learning. In *Proceedings of the International Conference on Advanced Learning Technologies (ICALT'01)*, pages 25–28. IEEE.

Thagard, P. (2006). Evaluating explanations in law, science, and everyday life. *Current Directions in Psychological Science*, 15(3):141–145.

Titov, I. and McDonald, R. (2008a). A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL-08)*, volume 8, pages 308–316. Citeseer.

Titov, I. and McDonald, R. (2008b). Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th International Conference on World Wide Web (WWW'08)*, pages 111–120. ACM.

Toulmin, S. (1972). *Human Understanding*, volume Volume I. Princeton University Press.

Toulmin, S. (2003). *The uses of argument*. Cambridge university press.

Tulving, E. (1985). *Elements of episodic memory*. Oxford University Press.

Tulving, E. (1986). Episodic and semantic memory: Where should we go from here? *Behavioral and Brain Sciences*, 9(03):573–577.

Turney, P. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02)*, pages 417–424.

Vasudevan, S. and Chakraborti, S. (2014). Mining user trails in critiquing based recommenders. In *Proceedings of the 23rd International Conference on World Wide Web (WWW'14)*, pages 777–780.

Wallach, H., Mimno, D., and McCallum, A. (2009). Rethinking LDA: Why priors matter. In *Proceedings of the 23rd Conference on Neural Information Processing Systems (NIPS'09)*, pages 1973–1981.

Walton, D. (2008). *Informal logic: A pragmatic approach*. Cambridge University Press.

Wang, H., Lu, Y., and Zhai, C. (2011). Latent aspect rating analysis without aspect keyword supervision. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*, pages 618–626. ACM.

Wiebe, J. and Rapaport, W. (1988). A computational theory of perspective and reference in narrative. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL-07)*, pages 131–138. Association for Computational Linguistics.

Wilks, Y. and Bien, J. (1983). Beliefs, points of view, and multiple environments. *Cognitive Science*, 7(2):95–119.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing (EMNLP'05)*, pages 347–354. Association for Computational Linguistics.

Winkler, W. (1999). The state of record linkage and current research problems. In *Statistical Research Division, US Census Bureau*. Citeseer.

Wiratunga, N., Lothian, R., and Massie, S. (2006). Unsupervised feature selection for text data. In *Proceedings of the 8th European Conference on Case-Based Reasoning (ECCBR'06)*, pages 340–354.

Wu, Y., Zhang, Q., Huang, X., and Wu, L. (2009). Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP'09)*, pages 1533–1541. Association for Computational Linguistics.

Xiang, L., Yuan, Q., Zhao, S., Chen, L., Zhang, X., Yang, Q., and Sun, J. (2010). Temporal recommendation on graphs via long- and short-term preference fusion. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10)*, pages 723–732.

Yi, J., Nasukawa, T., Bunescu, R., and Niblack, W. (2003). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM'03)*, pages 427–434. IEEE.

Yitzhaki, S. (1979). Relative deprivation and the Gini coefficient. *The Quarterly Journal of Economics*, pages 321–324.

Zha, Z., Yu, J., Tang, J., Wang, M., and Chua, T. (2014). Product aspect ranking and its applications. *IEEE Transactions on Knowledge and Data Engineering*, 26(5):1211–1224.

Zhan, T. and Li, C. (2011). Semantic dependent word pairs generative model for fine-grained product feature mining. In *Proceedings of the 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'11)*, pages 460–475. Springer.

Zhang, L., Liu, B., Lim, S., and O'Brien-Strain, E. (2010). Extracting and ranking product features in opinion documents. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10)*, pages 1462–1470. Association for Computational Linguistics.

Zhuang, L., Jing, F., and Zhu, X. (2006). Movie review mining and summarization. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM'06)*, pages 43–50. ACM.