Supervised Identification of the User Intent of Web Search Queries

Cristina González-Caro

TESI DOCTORAL UPF / ANY 2011

DIRECTOR DE LA TESI
Ricardo Baeza-Yates, Departament of Information and Communication Technologies

UNIVERSITAT POMPEU FABRA

To God for being my rock and my fortress and to my mother who is my main motivation.

Acknowledgments

Completing a PhD is a truly challenge, and I would not have been able to complete this journey without the aid and support of countless people over the past five years.

I must first express my deep gratitude towards my advisor, Professor Ricardo Baeza-Yates. His leadership, support, attention to detail, hard work, and scholarship have set an example I hope to match some day.

I would also like to express all my appreciation to my friend, Liliana Calderón-Benavides, who was tolerant enough to live with me during these five years. She was not only my friend but my sister and we traveled together this long road pursuing the shared goal that is this Ph.D.

I want to thank also to Mari-Carmen Marcos, for her extremely valuable experiences, support, and insights. I learned many things from her and she has all my admiration.

I am particularly grateful to Aristides Gionis, who was so kind as to read the whole manuscript thoroughly and correct the English spelling and grammar and contributed also to the scientific discussion. I also thank him for share my happy moments and for being my support in the difficult times.

I also would like to express my gratitude to my good friends Francesca Guffoni (and all her family) and Cesar Priego. These individuals always helped me to keep my life in context. Graduate school isn't the most important thing in life, but good friends, good times and happiness are. This thesis also would not have been possible without the love of my family.

I also thank to Fundación Barcelona Media, Universitat Pompeu Fabra and Universidad Autónoma de Bucaramanga for funding the projects I was involved in during these last years.

Finally, although it is not a person, I thank to my beloved Barcelona. I never was a foreign here and I never will be.

Abstract

As the Web continues to increase both in size and complexity, Web search is a ubiquitous service that allows users to find all kind of information, resources, and activities. However, as the Web evolves so do the needs of the users. Nowadays, users have more complex interests that go beyond of the traditional *informational queries*. Thus, it is important for Web-search engines, not only to continue answering effectively informational and navigational queries, but also to be able to identify and provide accurate results for new types of queries.

This Ph.D. thesis aims to analyze the impact of the query intent in the search behavior of the users. In order to achieve this, we first study the behavior of users with different types of query intent on search engine result pages (SERP), using eye tracking techniques. Our study shows that the query intent of the user affects all the decision process in the SERP. Users with different query intent prefer different type of search results (organic, sponsored), they attend to different main areas of interest (title, snippet, URL, image) and focus on search results with different ranking position. To be able to accurately identify the intent of the user query is an important issue for search engines, as this will provide useful elements that allow them adapting their results to changing user behaviors and needs. Therefore, in this thesis we propose a method to identify automatically the intent behind user queries. Our hypothesis is that the performance of single-faceted classification of queries can be improved by introducing information of multi-faceted training samples into the learning process. Hence, we study a wide set of facets that can be considered for the characterization of the query intent of the user and we investigate whether combining multiple facets can improve the predictability of these facets. Our experimental results show that this idea can significantly improve the quality of the classification. Since most of previous works in query intent classification are oriented to the study of single facets, these results are a first step to an integrated query intent classification model.

Resumen

A medida que la Web sigue creciendo, tanto en tamaño como en complejidad, la búsqueda Web llega a ser un servicio ubicuo que permite a los usuarios encontrar todo tipo de información, recursos y actividades. Sin embargo, así como la Web evoluciona también lo hacen las necesidades de los usuarios. Hoy en día, los usuarios tienen intereses más complejos que van más allá de las tradicionales *consultas informacionales*. Por lo tanto, es importante para los motores de búsqueda Web, no solo continuar respondiendo efectivamente las consultas informacionales y navegacionales, sino también identificar y proveer resultados precisos para los nuevos tipos de consultas.

El objetivo de esta tesis es analizar el impacto de la intención de la consulta en el comportamiento de búsqueda de los usuarios. Para lograr esto, primero estudiamos el comportamiento de usuarios con diferentes intenciones en las páginas de resultados de motores de búsqueda (SERP). Nuestro estudio muestra que la intención de la consulta afecta todo el proceso de decisión en la SERP. Los usuarios con diferentes intenciones prefieren resultados de búsqueda diferentes (orgánicos, patrocinados), miran diferentes áreas de interés (título, snippet, URL, imagen) y se concentran en resultados con diferente posición en el ranking. Identificar automáticamente la intención de la consulta aportaría elementos valiosos que permitirán a los sistemas de búsqueda adaptar sus resultados a los comportamientos cambiantes del usuario. Por esto, esta tesis propone un método para identificar automáticamente la intención detrás de la consulta. Nuestra hipótesis es que el rendimiento de la clasificación de consultas basada en facetas simples puede ser mejorado con la introducción de ejemplos multi-faceta en el proceso de aprendizaje. Por lo tanto, estudiamos un grupo amplio de facetas e investigamos si la combinación de facetas puede mejorar su predictibilidad. Nuestros resultados muestran que esta idea puede mejorar significativamente la calidad de la clasificación. Dado que la mayoría de trabajos previos están orientados al estudio de facetas individuales, estos resultados son un primer paso hacia un modelo integrado de clasificación de la intención de la consulta.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Over the last two decades, the World Wide Web (Web) has been continuously growing, and it has become an indispensable tool that has permeated many aspects of the daily life of people. This rapid growth of the Web has resulted in an exponential growth of the data and information that can be found online. Thus, nowadays, the Web users have access to more information and more resources than they ever had. On the other hand, the capacity for producing information exceeds the human capacity for processing it. The need to examine large quantities of information in a limited period of time can cause the phenomenon known as *information overload*. This phenomenon can affect, among other aspects, the decision-making processes when people search for information [27, 47]. In their study, Farhoomand and Drury [30] identified that the major factors that lead to information overload are the following: excessive volume of information, difficulty or impossibility of processing it, irrelevance or non-importance of most of it, lack of time to understand it, and multiple sources containing the same information. To deal with this phenomenon of information overload information-retrieval techniques have gained a great deal of attention among scientists in academia and industry. Web information-retrieval systems appear as bridges between the users and the overwhelming amount of data and information contained in the Web.

According to Belkin and Croft [13], research in information retrieval can be classified in three main categories: text representation, design of retrieval techniques, and acquisition of information needs. Most of the initial and current research, has been concentrated in the first two categories. Many alternatives that explore diverse features and sources of information have been proposed in the area of representation of text and multimedia resources [62, 89, 82, 54, 8, 85, 41]. In the same way, we can find a plethora of research on information-retrieval techniques, ranging from the classic models (boolean models, vector-space model, probabilistic indexing, etc. [92, 28, 76, 102]) to more recent approaches that aim to take advantage of the underlying semantic relationships of the Web resource collection [90, 74, 73, 24]. Both text representation and design of retrieval techniques, are critical components of the information retrieval process and hence they have been well studied. However, despite all the advances in these two areas, information retrieval would be meaningless if the third main area of research, *acquisition of information needs*, is not well developed.

The main objective of any information-retrieval system (including Web-search systems), is to satisfy the needs of the users. The user is a crucial factor in the information retrieval process. For this reason, research on acquisition of user needs is necessary and must be taken into account in the planning and implementation of all processes in information-retrieval systems. But, what is meant by "user need" in an information retrieval system? If we think that the needs of the users are the target of every information retrieval system, then we should review the goal definitions of the information retrieval systems. Although there is not an unique concept, most of the definitions share some commonalities. Some of the most popular goals that are used to describe information-retrieval systems are the following:

"*The goal of an information retrieval system is to locate relevant documents in response to a user's query*" [60].

"*The goal of information retrieval systems is to extract documents that best match a query*" [70].

> "*The primary goal of an information retrieval system is to retrieve all the documents that are relevant to the user query*" [3].

The definitions above imply that the general goal of an information-retrieval system is to retrieve relevant documents in order to give answer to a query. We can say that in this context, the user need, in some way, is reduced to the words that the user choose to express her/his need. Traditional Web-retrieval systems and a big proportion of modern IR systems have worked with this conception of user need. However, as it was early stated by Belkin, Oddy and Brooks [12], the most general situation is that the user facing with a problem, recognizes her/his state knowledge is inadequate for resolving that problem (*Anomalous State of Knowledge*–ASK), and decides that obtaining information about the problem area is an appropriate means towards its resolution. Some times the user is able to specify precisely what information is required for the resolution of the problem, but the common situation is that the user does not know, and actually the user is using the Web-search system in an exploratory manner. In this situation, it appears obvious that the query is not the best descriptor of what the user needs. In order to really help the user, it would be more useful if the Web-search system attempts to determine the need underlying the user query than to ask the user to specify in some few words her/his need to the system. Thus, the real goal of a Web-search system is more concerned with the resolution of a problematic situation (derived from an ASK) than with the resolution of a simple query.

Hence, Web-search systems are evolving from a content-centric design (mainly keyword based) to a user-centric design. A new generation of Web-search engines, whose main challenge is to interpret properly the problems of the users (instead of answer single queries) and to be able to give right and comprehensive answers to these problems, is emerging. Several studies have begun to outline this new generation of Web-search systems [103, 93, 15]. To address the new challenges that the complex needs of the users pose to Web search systems, are needed more and better methods for inferring and characterizing the intent behind the user query, as well as mechanisms that directly address the appropriate content for that intent.

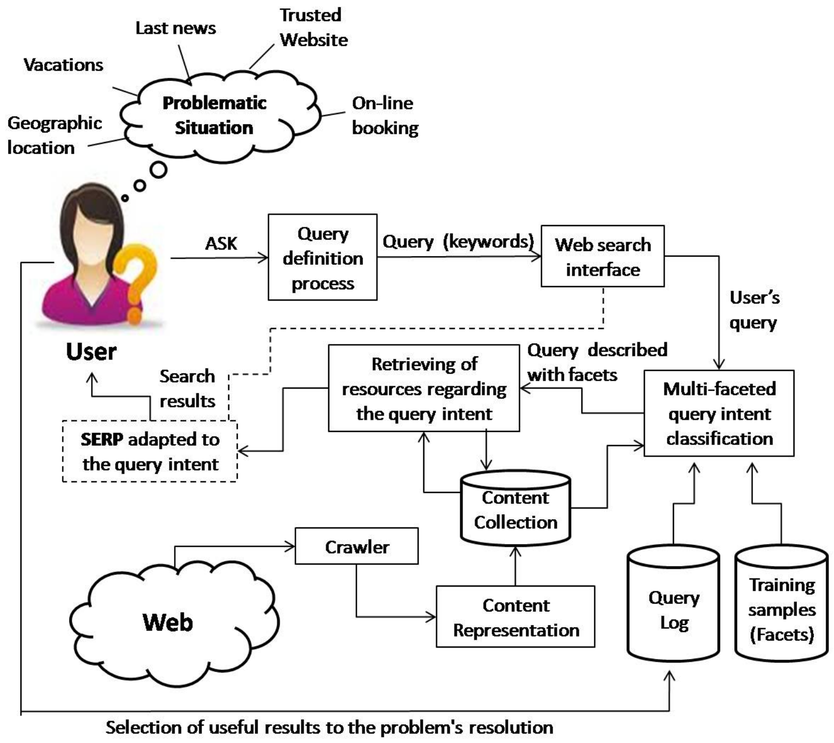


Figure 1.1: Model of Web search system using multi-faceted classification of query intent.

Attending to the crucial requirement of the modern Web-search engines to a deeper understanding of the intent of the user queries, this Ph.D. thesis aims to analyze the impact of the query intent in the search behavior of the users. Additionally, a method to identify automatically the intent behind user queries is proposed. The thesis addresses the problem in a comprehensive way. First, we present an analysis of the implications of identifying the intent of a user query. This analysis shows how the behavior of the user changes according to the intent of the query and it suggests that the search results should be adapted to this intent. Next, we explore a first model to the automatic classification of the query intent based on two perspectives, the type of resource and the topic category associated to the query. Finally, we study a wide set of facets that can be considered for the characterization of the query intent of the user. Then, we automatically classify queries into these facets, using a multi-faceted classification model. We discuss how this approach can be used to improve the Web-search performance and provide search results that are closer to real problems of the users.

As we mentioned before, one of the current problems with the acquisition of user needs in Web-search systems is that normally, the user is unable to specify in the few words of the query what is her/his real need. One option to deal with this problem is by learning from the past user search behavior using supervised learning methods and query logs. Query logs register the history of queries submitted to the search tool, and the pages selected, among other data [4]. The query may be the result of an anomalous state of knowledge, but a document (or a Web page) is a representation of a coherent state of knowledge. So, it is possible to use the information of the set of Web pages selected for a query, stored in query logs, to build a better representation of the user query. The enriched query representations together with labeled samples that connect queries with particular facets, would allow a better description of the intent of the query and hence a better selection and presentation of search results. Figure 1.1 shows a description of a Web search system based on this idea. In the figure we observe how the query goes first through a process of identifying the intent and then whether it goes to the retrieval of related Web resources. The query intent also affects the process of presentation of search results, as the search engine result page (SERP) is adapted to the intent of the query. In general, the suggestion is that the system begins with the identification of the intent of the query and that the rest of the processes take into account this intent.

A key difference of our work is to treat the problem of query intent identification as a multi-faceted problem. We provide an extensive experimental evaluation showing that the combination of facets proposed in this thesis can significantly improve the results of the query intent classification process. To the best of our knowledge, this is the first work that explores automatic multi-faceted classification of user's query intent. Previous work has considered the multi-faceted classification of the query intent an open research problem [49].

The main contributions of this thesis are:

- Chapter 3: Through a study based on eye tracking techniques, we analyze the user-browsing behavior in search engine results pages. We study whether this behavior is different for queries with different query intent. The results show that the query intent affects all the decision process in the SERP. These findings are useful to confirm that to identify the intent of the user query could improve the user search experience. The research findings are also useful for other areas related with Web search, like search-engine marketing. As result of this research two papers have been published:

  – User Behavior in Search Engine Result Pages: A study based on Eye-tracking (original in Spanish), *in El Profesional de la Información, 2010.* With M. C. Marcos [66].

  – Different Users and Intents: An Eye-tracking Analysis of Web Search, *in User modeling for web applications. A workshop at WSDM'2011.* With M. C. Marcos [34].

- Chapter 4: We introduce a model for automatic query intent classification based on two facets, the type of resource of the query and the topic category. Our results show the feasibility of the automatic classification of queries in more than one facet. We also show that describe the query intent with two facets provide a more comprehensive reference of the intent of the user. This work was published in the paper:

  – The Intention Behind Web Queries, *in 13th Edition of the Symposium on String Processing and Information Retrieval, SPIRE'2006.* With R. Baeza-Yates and L. Calderón-Benavides [5].

- Chapter 5: We study, analyze, and characterize a wide range of facets involved in the establishment of the intent the user query. We outline the relationships that exist among these facets as well as their contribution in the task of recognizing accurately the user intent in Web search. The results of this study were published in:

  - Towards a Deeper Understanding of the User's Query Intent, *in Query representation and understanding. A workshop at SIGIR'2010*. With R. Baeza-Yates and L. Calderón-Benavides [17].
  - Web Queries: The Tip of the Iceberg of User's Intent, *in User modeling for web applications. A workshop at WSDM'2011*. With R. Baeza-Yates and L. Calderón-Benavides [33].

- Chapter 6: We also investigate whether combining information of multiple facets can improve the predictability of the facets. We compare the results of multi-faceted classification with the conventional faceted classification to determine if the combination of related facets can improve the identification of the intent of the user queries. Our experimental results show that this idea can significantly improve the quality of the classification. Since most of previous works in query intent classification are oriented to the study of single facets, these results are a first step to an integrated query intent classification model. The results of this study were included in the paper:

  - A Multi-faceted Approach to Query Intent Classification. Submitted and accepted as full paper *in 18th edition of the Symposium on String Processing and Information Retrieval, SPIRE'2011*. With R. Baeza-Yates [32].

The processing of the data sets used in this thesis, as well as the designing of the facets considered for the description of the intent of the queries are part of a joint work that was done with Liliana Calderón-Benavides. In particular a comparison of our results and extensions have been recently submitted to a journal paper:

- A Comparative Study of Machine Learning Techniques Applied to Query Intent Prediction. *Submitted.* With R. Baeza-Yates, L. Calderón-Benavides, D. Dubhashi and L. Tansini.

This Ph.D. thesis is organized into seven chapters, each of one presenting different aspects of the developing of the objective of the thesis. Each chapter is based on one or more published articles, which have been extended with the purpose to offer a more comprehensive view of the work. In Chapter 2 we make a review of the work related with this research. Chapter 3 presents an eye tracking study that analyzes the user browsing behavior in search engine results pages with the objective of detect factors that influence how users gaze different areas on a SERP during search tasks. This study shows that the intent of the query influences the search behavior of the user, hence, in Chapter 4, we explore the feasibility of automatic query intent prediction. The results of this first approach of automatic classification of the queries are promising, since they show it is possible to classify the queries according to their intent. Based on this reults, in Chapter 5 we introduce a deeper analysis of the intent of the user queries in order to obtain a more accurate description of the query intent. We study a set of nine facets that are useful to the identification of the intent of the query. We analyze the relationships among this group of facets, as well as the contribution of each facet into the task of identify the intent of the user query. In Chapter 6 we explore the feasibility of automatic faceted-classification. We perform single-facet and multi-facet classification of queries. Then, we compare the performance of the two classification methods. The results of these experiments as well as the discussion of the results are described in this chapter. Finally, Chapter 7 presents the conclusions, an outlook of this work and its potential improvements.

Chapter 2

RELATED WORK

In recent years, there have been a growing interest in understanding the underlying intent of the Web search queries. However, due to the natural diversity of the needs and behavior of the users, capture all the characteristics involved in the query intent it has not been an easy task. According to Cool and Belkin [19], users engage in multiple information seeking behavior within the context of accomplishing a single task. Therefore, it is important to have Web-search systems able to support multiple information seeking behaviors, and multiple interactions with the information. In this direction, many efforts have been devoted in trying to understand the intent of user queries. The understanding of user queries has been conducted from different perspectives, in this chapter we discuss an excerpt of the most important ones.

A relevant line of research has focused on the classification of queries regarding the type of resource associated with the queries. In this case, most of the works follow the taxonomy introduced by Broder in 2002 [14]. In this work, three types of search intents are identified: navigational, informational, and transactional. Navigational queries were defined as those intended to reach a particular site that the user has in mind. Informational queries are intended to find information on a specific topic and transactional queries are intended to find sites where some kind of interaction will be carry out. Based on the manual classification of 400 queries from a commercial query log, Broder reported 48% of informational queries, 20% of navigational queries and 30% of transactional queries. An interesting observation of the study is that although the proposed taxonomy is concerned with only one facet of the query intent (type of resource), some other facets are mentioned indirectly. For instance, the research recognizes that transactional queries are associated with particular topic-categories, like shopping or entertainment.

Rose and Levinson [81] presented a fine-grained taxonomy for search queries, based on Broder's taxonomy. In this taxonomy, the authors considered that the entries within transactional category could be classified into more subclasses. So they replaced 'transactional category' with 'resource category' and they also added hierarchical sub-categories for informational and resource categories. A set of approximately 1,500 queries from a commercial search engine were manually labeled using a query classification tool that provided information about the query, the results returned by the search engine for that query, the results clicked by the users and some additional information about the user's actions. As the taxonomy of Rose and Levinson has more categories, the manual classification results show that the query intent goes beyond just getting information resources. Nearly 40% of the queries were non informational, with categories like 'download', 'obtain' or 'interact'.

Based on the early taxonomies of Broder and Rose and Levinson, many works have attempted automatic classification of queries. A group of studies worked on the classification of informational and navigational queries, without to consider transactional queries. Although these works are relevant because they present different approaches to automatic classification of queries, they were not classifying transactional queries, which according with Broder, must be one of the most relevant and interesting type of queries. Kang and Kim [58] presented one of the first approaches that implemented automatic classification of queries. They classified queries into informational (topic relevance task) and navigational (home page finding task) categories. The automatic classification was made using a combination of features: part-of-speech tags information and usage rate of query terms in anchor texts. Although they provide classification results based on the individual features, the best performance classification is reached with combination of features where they got 91.7% precision and a 61.5% recall. Lee, Liu and Cho [61] also automatically classified queries into informational and navigational categories. The automatic classification of this work is based on two features: anchor link distribution and user click behavior. They used a data set of 50 queries collected from computer science students at a US university. Best results are reported with a combination of the proposed features: 90% precision. Although this work identified occurrence of ambiguous queries, and even suggested that these type of queries could be recognized using topic categories, automatic classification for ambiguous queries is not implemented. Using a test set of 233 queries extracted from a Chinese search engine log, Liu *et al.* [64] automatically classified navigational and transactional queries. For the automatic classification of queries they used a decision tree classification algorithm and two features derived from click through data, with precision and recall values of 80%.

Another group of works classified queries using the original Broder's taxonomy, informational, navigational and transactional. Jansen, Booth and Spink [48] present an automatic classification method based on attributes of each type of query, using a decision tree approach. Although they processed a large query log of a million and half queries, they based the effectiveness measures on the evaluation of only 400 queries. The distribution of queries into categories of their results, differs from the earlier works. While Broder estimated informational queries between 39% and 48%, Jansen *et al.*, stated that the percentage of these type of queries is above 80%. It is a big difference if we consider that with the diversity of the current Web resources, people are not only interested in information, instead they are moving to other kinds of resources like multi-media content. Mendoza and Zamora [68] used three different vector representations of queries based on click through data. They then use these vector representations to automatically classify informational, navigational and transactional queries through support vector machines. For this work Mendoza *et al.*, processed a query log of 594,564 queries. From this dataset, the top-2000 most frequent queries were manually classified into the three types of query and posteriorly used to test the effectiveness of the automatic classification of queries. The best results were obtained with two vector representations, representation based on descriptive text and clicks and representation based on descriptive text, reading time and clicks, in both cases the average precision for the three query intent categories was of 0.89. More recently, Herrera *et al.* [42], also explored a variety of features for the representation of queries and they tested the affectivity of these features in the task of automatically classify informational, navigational and transactional queries. The results of this study suggested that query popularity is a good feature to incorporate in the representation of queries for automatic classification. The best classification results were obtained for transactional queries.

Overall, the research works that classify queries according to the type of resource, are good approaches to automatic classification of queries. They are mainly concentrated in the study of best features to represent the queries in order to obtain better classification results. These works present methods to automatically classify queries into the individual categories of Broder's taxonomy. However, despite that this taxonomy has around ten years of proposed, there are not many studies attempting to perform deeper analysis of it. Studies analyzing the relationship between the categories or using fine-grained categories like those proposed by Rose and Levinson for informational and resource queries, are not common.

Other approaches to identify query intent include classifying queries based on the topic. Topical associations of queries are important because they allow to place the queries in a particular context. There are several studies in the literature that deal with the automatic classification of queries into topic categories. For instance, in 2005 the ACM Conference on Knowledge and Data Discovery (KDD) devoted its annual competition known as the KDD Cup to the task of topical query categorization [63]. The dataset provided for the competition consisted of 800,000 Web queries, and 67 possible categories, with each category pertaining to a specific topic. Each participant was to classify all queries into as many as five categories. An evaluation set was created by having three human assessors independently judge 800 queries that were randomly selected from the sample of 800,000. On average, the assessors assigned each query to 3.2 categories for their judgments. In all, there were 37 classification runs submitted by 32 individual teams. The winner solution was the presented by Shen *et al.* [83]. For the automatic topic classification they used an ensemble-search based on the combination of two type of classifiers: synonym-based classifiers which uses the category information associated with Web pages collected for each query and statistical classifiers in which support vector machine (SVM) is employed. The average precision reached by the participants of the competition was 0.24 (0.75 max). Shen *et al.*presented an improved version of their winning solution in [84]. Another significant works in this line are those presented by Beitzel *et al.* [10, 11]. These works explored the automatic classification of queries into topic categories combining multiple classifiers, including exact lookup and partial matching in databases of manually classified frequent queries, linear models trained by supervised learning and an approach based on mining selectional preferences from a large unlabeled query log.

Another approaches that have explored alternative facets for query intent classification include: geographic locality [57], time sensitivity [69], and ambiguity [94]. However, each of these works has been restricted to the analysis of only one facet. Meanwhile, there have been few works attempting to classify query intent in more than one facet. Inguen and Kat [72] analyzed a set of four facets (ambiguity, authority sensitivity, temporal sensitivity and spatial sensitivity). However, despite that they presented a query log analysis of the four facets, they only provided automatic classification results for one of the facets: authority sensitivity. They based their automatic classification results on a reduced set of 50 queries and they excluded from the analysis the ambiguous queries. The best precision result obtained with the automatic classification for the facet authority sensitivity was 0.79.

The research presented in this thesis differs from the related work in several aspects. First of all, we analyze more than one facet (until nine facets) and we study the relationships between these facets and how these facets can be combined in order to obtain a better description of the query intent. We also characterize and evaluate the predictability of the ambiguous queries. Although other works have considered ambiguous queries, they have made this in an isolated way, that is, they have not contrasted ambiguous queries with other type of queries as we do in this work. Additionally, although most of the approaches presented until now have used supervised learning methods in their automatic query classifications, none of this works have tried to predict simultaneously several facet-values for a same query. They only implemented automatic classification of query intent based on the information of individual facets. In this thesis we present a multi-faceted approach for query intent classification that shows that outperforms significantly the results of the traditional single-label classification.

We introduce details about the state of the art of more specific topics along the rest of chapters of the thesis.

Chapter 3

THE RELEVANCE OF THE UNDERLYING QUERY INTENT: AN EYE-TRACKING APPROACH

Searching for information is not as a straightforward process as people would like it to be. Web search engines offer multiple types of searches and produce long lists of results. However, despite the diversity and complexity of the information-seeking process [88, 101], the web-search interface and web-search-results presentation are usually the same for all the queries. The particularities associated with each kind of information need and the intents behind them are not considered. A better understanding of the user behavior during the search process is needed in order to determine which is the impact of the intent behind the query into the user's decision-making process.

This chapter presents an eye tracking study that analyzes both of the following aspects: the user browsing behavior in search engine results pages (SERBS) as well as whether this behavior is different for queries with an informational, navigational and transactional intent. The study considers a diverse set of variables that influence the scanning behavior of the users. The main goal is to detect factors that influence how users gaze different areas on a SERP during search tasks.

The results show that the query intent affects all the decision process in the SERP. Users with different intent preferred different type of search results, they attended to different main areas of interest and focused on search results with different ranking position.

After presenting an overview of the related work on web search user behavior (Section 3.1), we give a description of the methodology of the study (Section 3.2). We then provide an analysis and discussion of factors influencing the ocular behavior of the users on SERBS (Section 3.3). Finally a summary of the implications of the results is presented in Section 3.4.

3.1 Related Work

Most of the previous research on user search behavior analysis have been focused on log files, mouse tracking and clickthrough data. Log files can give useful information about the user behavior in SERBS [7, 67, 52, 100, 56] but have a main limitation: they can reveal what users click but not what they look. Mouse tracking and clickthrough tools capture, partially, the search actions of the users [55, 37, 39, 38] but they entail lack of information if the user is just browsing and not moving the mouse. Searching implies both mental and physical actions. There are many activities of the search process that do not involve interacting directly with the search interface. If the users are experiencing problems to find what they are looking, they might not click or type, but instead they might scan the screen or read carefully a piece of information.

Eye tracking is a promising technique that appears as one of the best current alternatives to study user behavior in SERBS. The reason is that eye-tracking mechanisms register the gaze activity of the users. Gaze-related activities represent the instances in which most of the information acquisition and processing occurs [80]. Because of these reasons, there are several works that have analyzed the user search behavior through the eye tracking technique. The first eye tracking studies applied to web search were focused exclusively on organic results: Pan *et al.* [75] analyzed different factors that might affect the viewing behavior of the users such as gender and site type, which were factors that shown to have influence in the behavior of the users. Joachims *et al.* [55] studied how users browse the abstracts of a SERP and how users select the associated links; the clicks on search results were assumed as implicit feedback and were compared with the explicit relevance judgments of the users. Cutrell and Guan [21] experimented with changes in the presentation of search results, specifically, they modified the size of the snippets and reported the effects of these changes for queries with two types of intent. Only few studies have evaluated both types of results offered by search engines, organic and sponsored results; and most of these studies are based on log files and clickthrough data [50, 48, 22]. Although, Busher *et al.* [16] analyze organic and sponsored results with eye tracking technique, the research is specially oriented to sponsored results.

Some of the studies that analyze user behavior on SERBS have included the intent of user query in their analysis. These works compare only two possible intents; the most popular combinations are: informational vs. navigational [35, 55, 21, 37, 16, 2], followed by informational vs transactional [97, 95, 39] and in some particular cases, navigational vs non navigational [22] or commercial vs non commercial [44].

On the other hand, the study of more fine-grained analysis of user-search behavior has received considerably less attention. Studies with specific AOIs labeled by each single search result, further than organic and sponsored areas, are not common. Busher *et al.* [16] labeled three AOIs (title, snippet and url) but did not present results for these areas; Hotchpot [44] also labeled AOIs, but did not present the fixation data; and finally Cutrell *et al.*[21] mainly focused on snippets and do not analyze the other areas.

To the best of our knowledge, no study about user behavior in SERBS using eye tracking and considering the intent behind the query has combined all the elements that are included in this research: both types of search results (organic and sponsored results), four special areas of interests for each single search result (title, snippet, url and image) and the position in the ranked list of the search results.

3.2 Experimental Design and Procedure

The study analyzes the ocular behavior on SERBS using eye tracking metrics in order to obtain evidence about how eye movement is determined by the intention behind queries. We have run an experiment in which 58 participants had to complete six tasks using Googly or Yahoo! search engines.

3.2.1 Tasks

We designed eighteen questions, eleven questions (61%) with informational intent: the users had to find some specific information about a concrete item; three questions (17%) with navigational intent: the users had to find a specific Web page; and four questions (21%) with transactional intent: the users had to find a Web page where performing an action. The proportion of queries for each intention is based on the results of previous works [14, 81, 49] which reveal that more than half of the total number of queries in search engines are informational and around 25% are navigational and transactional.

In order to avoid specialization bias, the questions were designed about topics of common interests of the people. Table 3.1 shows a sample of three search tasks and their respective queries (original in Spanish). In total, the users performed 348 tasks: 216 with informational intent, 58 with navigational intent and 74 with transactional intent. Half of the queries were run in Googly and the other half in Yahoo!.

3.2.2 Web Search Sessions

At the beginning of the experiment the users were informed about the purpose and the procedure of the study. Next, they filled a survey about

Query Intent	Search Task	Query example
Informational	Which is the timetable of the Louvre Museum in Paris?	horrid Louvre
Navigational	Find the official web page of the Grandma Town Hall	ayuntamiento de Granada
Transactional	Find a web page where you can book a room in a hotel in Grandma	hotel Granada

Table 3.1: Sample of search tasks used in our study.

demographic information, level of studies, use of Internet and skills in using search engines. Afterwards, the eye tracker was calibrated. Then, the participants started with the search tasks using the search engine as they normally would. For each task, we provided the participants with a written task description in a previous screen. After reading the description, the participants pressed a "continue" button to begin the task in a search engine that the researchers had selected. In order to obtain comparable data, all the users worked on the same first search result page. The first SERP of each query was cached locally, and then this SERP was showed to the users to start the search process. From here on, participants were free to interact with search results.

The tasks were considered finished when the user navigated to a web page clicking on the SERP or after five minutes browsing the SERP. The experiment took about half an hour per participant. During the time of task's performance, eye tracking equipment was used to follow eye movement of the users throughout the SERP. Data about individual eye movement and overall behavior were collected.

3.2.3 Apparatus

The device used for measuring eye movements was a Torii T120 Eye Tracker. The eye tracker is integrated into a 17" TAFT monitor (Figure 3.1). The tracker illuminates the user with two near infrared projections to generate reflection patterns on the corneas of the user. A video camera

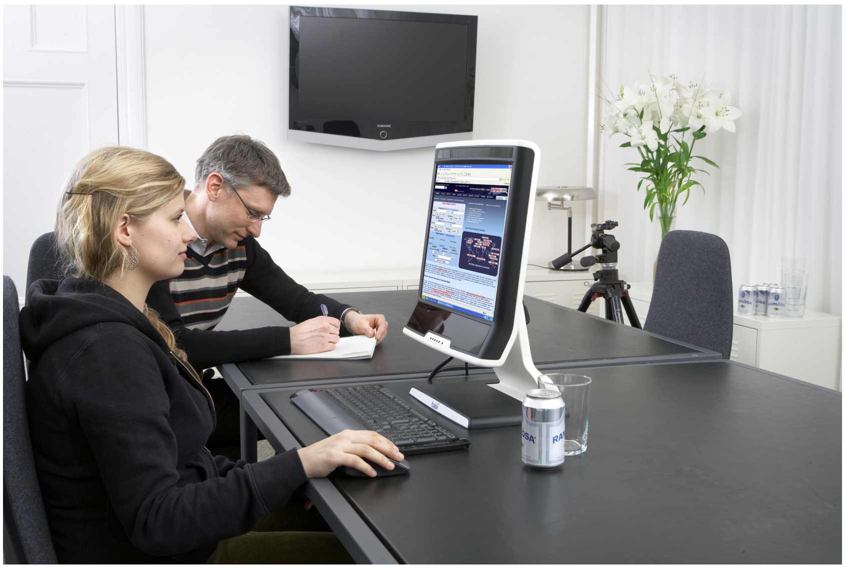
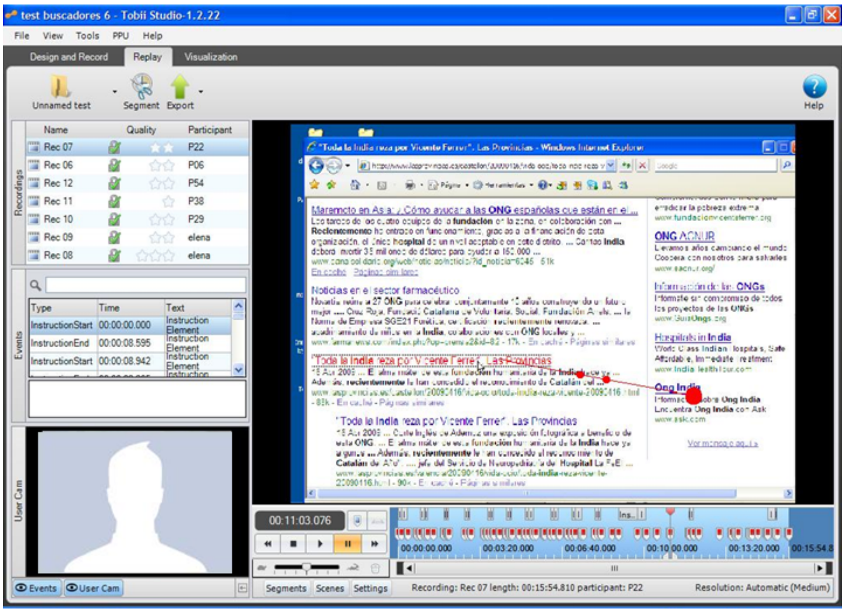


Figure 3.1: Subject with eye tracking apparatus.

then gathers these reflection patterns as well as the stance of the user. Digital image processing is then carried out for extracting the pupils from the video signal. The system tracks pupil location and pupil width at the rate of 50 Hz. The resolution of the tracker is 1280x1024. The system allows free head motion in a cube of 30x15x20 cm at 60 cm from tracker. Logging of gaze data and maps were done by the software Torii Studio Enterprise Edition, version 1.2. Figure 3.2 shows a screen of the interface of Torii software showing a participant's scanpath map.

3.2.4 Participants

Participants were people with a diverse range of professions, background and education levels (undergraduate student was the minimum education level). In total, fifty eight (58) people participated in the study, 25 men and 33 women. The average age of the participants was 28 years in a range of 18 to 61 years. All participants were moderately experienced at Web



Name	Quality	Participant
Rec 07	★★	P22
Rec 06	☆☆	P06
Rec 12	☆☆	P54
Rec 11	☆	P38
Rec 10	☆☆	P29
Rec 09	☆☆	elena
Rec 08	☆☆☆	elena

Type	Time	Text
InstructionStart	00:00:00.000	Instruction Element
InstructionEnd	00:00:08.595	Instruction Element
InstructionStart	00:00:08.942	Instruction Element

Recording: Rec 07 length: 00:15:54.810 participant: P22

Resolution: Automatic (Medium)

Figure 3.2: Screen of Eye-Tracker Software.

search, 45% of the people use the computer between 6 to 10 hours per day and all of them reported that they searched the Web for information at least once at week. None of them had experience using an eye-tracker.

3.2.5 Measures

Since all SERBS have similar elements, we created common areas of interest (AOI). AOIs are specific rectangular areas that contain lookzones of potential interest to the users. Figure 3.3 shows the distribution of AOIs that we made in the SERP. First of all, we distinguished two main areas: organic results and sponsored links. As sponsored links are in two zones of the interface, we have done another division: ads on the top and on the side. Eventually, for each single search result, we have labeled four AOIs: title, snippet, URL and image.

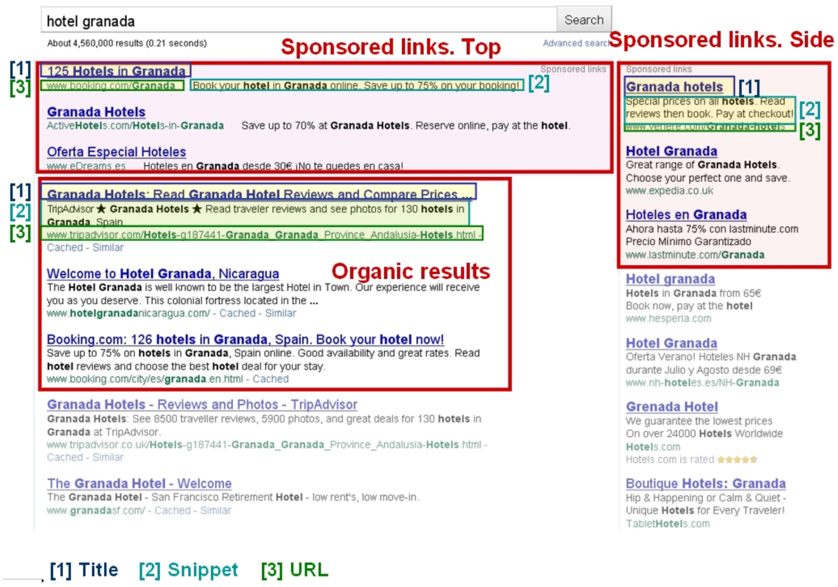
In order to explore the determinants of ocular behavior on web pages, the study incorporates an analysis of two dependent variables: fixation frequency and fixation duration. Fixations represent the moment when the eyes are relatively stationary, taking in or encoding information, 200-300 milliseconds on average [77]. Fixation frequency refers to the times that user fixes his attention on a specific point of the screen. We measured the number of fixations in each special area of interest and the duration (ms) of each fixation.

Considering the intent of the query: informational, navigational and transactional, both measures, fixation frequency and fixation duration, were explored as function of independent variables:

- Type of search result: organic and sponsored.
- Areas of Interest in each single search result: title, snippet, url and image.
- Ranking position of the three first results in each zone (organic and sponsored results).

We decided to classify the tasks in the three types of intent indicated by [14]. As we have seen in Section 3.1, previous studies have considered only two of these types of intent: informational and navigational, informational and transactional, and navigational and non-navigational, but not the three types of intent as we did in this study.

We chosen only the three first results to label with AOIs because of ”the golden triangle” [44], that is, the area of the first results receives most of the gazes in a SERP. As we wanted to analyze this triangle zone, we chose the three first ads and the three first organic results, and to have a similar number of results in each zone we labeled the three first side ads too. We also considered in the analysis, the rank position of the results, as previous research has shown its importance in the scanning behavior of the users [35, 44].



hotel granada

About 4,560,000 results (0.21 seconds)

Sponsored links

125 Hotels in Granada
www.booking.com/Granada Book your **hotel** in **Granada** online. Save up to 75% on your booking!

Granada Hotels
ActiveHotels.com/Hotels-in-Granada Save up to 70% at **Granada Hotels**. Reserve online, pay at the **hotel**.

Oferta Especial Hoteles
www.eDreams.es Hoteles en **Granada** desde 30€ ¡No te quedes en casa!

Granada Hotels: Read Granada Hotel Reviews and Compare Prices ...
TripAdvisor ★ **Granada Hotels** ★ Read traveler reviews and see photos for 130 **hotels** in **Granada**, Spain.
www.tripadvisor.com/Hotels-g187441-Granada_Granada_Province_Andalusia-Hotels.html - Cached - Similar

Welcome to Hotel Granada, Nicaragua
The **Hotel Granada** is well known to be the largest Hotel in Town. Our experience will receive you as you deserve. This colonial fortress located in the **...**
www.hotelgranadanicaragua.com/ - Cached - Similar

Booking.com: 126 hotels in Granada, Spain. Book your hotel now!
Save up to 75% on **hotels** in **Granada**, Spain online. Good availability and great rates. Read **hotel** reviews and choose the best **hotel** deal for your stay.
www.booking.com/city/es/granada.en.html - Cached

Granada Hotels - Reviews and Photos - TripAdvisor
Granada Hotels: See 8500 traveller reviews, 5900 photos, and great deals for 130 **hotels** in **Granada** at TripAdvisor.
www.tripadvisor.co.uk/Hotels-g187441-Granada_Granada_Province_Andalusia-Hotels.html - Cached - Similar

The Granada Hotel - Welcome
The **Granada Hotel** - San Francisco Retirement **Hotel** - low rent's, low move-in.
www.granadasf.com/ - Cached - Similar

Sponsored links

Granada hotels
Special prices on all **hotels**. Read reviews then book. Pay at checkout!
www.venere.com/Granada-hotels

Hotel Granada
Great range of **Granada Hotels**.
Choose your perfect one and save.
www.expedia.co.uk

Hoteles en Granada
Ahora hasta 75% con lastminute.com
Precio Mínimo Garantizado
www.lastminute.com/Granada

Hotel granada
Hotels in **Granada** from 65€
Book now, pay at the **hotel**
www.hesperia.com

Hotel Granada
Oferta Verano! Hoteles NH **Granada**
durante Julio y Agosto desde 69€
www.nh-hoteles.es/NH-Granada

Grenada Hotel
We guarantee the lowest prices
On over 24000 **Hotels** Worldwide
Hotels.com
Hotels.com is rated ★★★★★

Boutique Hotels: Granada
Hip & Happening or Calm & Quiet -
Unique **Hotels** for Every Traveler!
TabletHotels.com

Figure 3.3: Sample of Search Engine Result Page with Areas of Interest.

3.3 Results

In this section, we analyze how the user behavior on a SERP varies for different query intents. From the 58 participants recruited for the study we recorded usable eye tracking data for 301 of the 348 search tasks that they performed. We discarded some search tasks because of stability problems with the eye tracking tool and/or incomplete data.

Table 3.2 shows the distribution of user's fixation in the SERP's, as well as the mean time that users spent in these pages. The duration of the fixations on SERBS is longer for informational and transactional queries, while for navigational queries users show less time fixating information. Since mean fixation duration and fixation frequency are taken as indicators of task difficulty and information complexity [31], the results of Table 3.2 suggest that informational and transactional queries demand from the users a harder examination of the SERP. The results of the ANOVA test

	Fixation Freq.	Fixation Dur.
All	8.37	5991.13
Informational	**9.50**	**6484.31**
Navigational	6.80	5355.11
Transactional	8.81	6133.98

Table 3.2: Mean fixation frequency and fixation duration for task on a SERP.

indicated that there is a significant difference between the intent of the query and the total processing time (F= 619.85, $p < .005$) and the total number of fixations (F= 253.04, $p < .005$) of the users on SERBS.

From these initial results we establish some facts: first, the number of fixations and the time that a user invests examining a Web search result page is not high and second, the underlying intent of the queries influences the little attention that users devote to the SERP. It is important for Search Engines and related services to understand how the users are examining SERP's and extracting the maximum knowledge from the user's interactions with Web search results. In the following sections we present an analysis of several elements of Web search results that might affect the user behavior, focusing in how the effect of these elements varies regarding the intent of the queries.

3.3.1 Query Intent and Type of Search Result

Since one of the most important questions in the Web search process is the effectiveness of the search results, in this section we analyze the behavior of the users with respect to the two main types of search results: organic and sponsored links. Despite the general idea that organic results are more relevant to queries with informational intent and sponsored results are more relevant to queries with transactional intent (mainly for commercial oriented queries) [48]; it is important to establish which is the real behavior of the users on SERBS when they place queries with different intent on the search engine. Are users of transactional queries

	Organic (ms)	Sponsored (ms)	F (*p*< *.005)*
Informational	**5814**	447	1676.9
Navigational	4757	530	305.28
Transactional	2840	**1505**	282.54

Table 3.3: Average gaze duration for type of intent broken down by type of search result.

more interested on sponsored links? What about users with navigational queries, what kind of results are they interested in?

Organic vs Sponsored Links

Among the total of evaluated SERBS, only 36% of the pages included sponsored links, it is reflected in the distribution of the attention of the users on these areas of the SERP: informational (organic=97.4%, sponsored=2.6%), navigational (organic=92.4%, sponsored=7.6%) and transactional (organic=71%, sponsored=29%). In Table 3.3 we can see the average gaze duration for each type of query intent broken down by type of search result. As we can observe, organic results capture most of the attention of the users when examining SERBS. Among the three types of query intent, organic results have at least 70% of the accumulated number of fixations. These results agree with previous work concerning organic and sponsored links [22, 16]. Informational and navigational queries are the two types of query intent with highest proportion of fixation and gaze duration on organic results. Although informational and navigational intent show almost the same distribution of fixations on organic results, the average gaze duration for each type of intent is not the same. Users spend more time in SERBS when they are dealing with informational tasks, indicating that these tasks require a deeper analysis of the SERP. For navigational tasks, users have also long number of fixations on organic results but a lower time of gaze duration; usually users tend to go directly to the site they have in mind and hence their fixation time on search results is shorter.

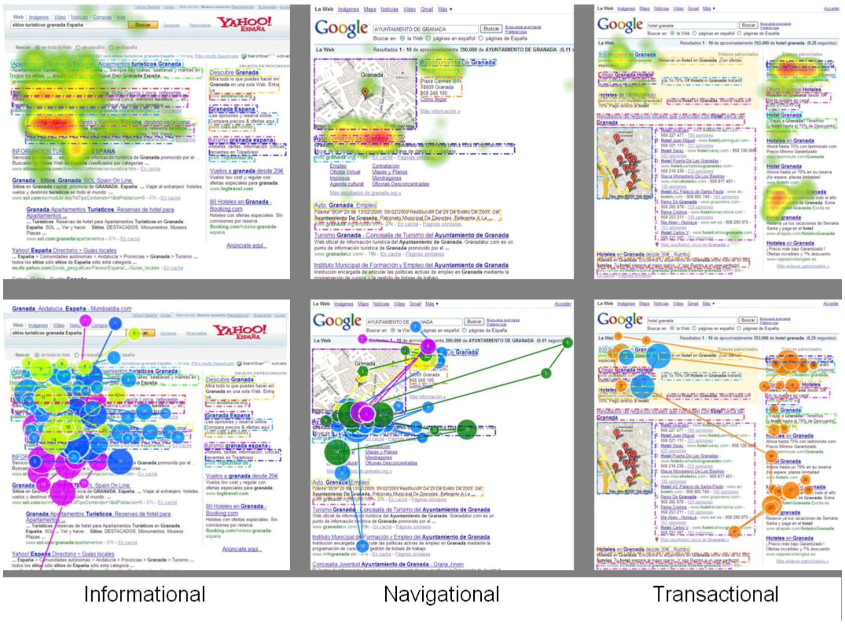


Figure 3.4: Heat maps (top) and gaze plots (bottom) for three queries with different intent.

Although the accumulated attention for sponsored links is substantially lower than for organic links, there is an interesting behavior of transactional queries with this kind of results. Transactional queries are the only queries that show a considerable percentage of fixation attention on sponsored links (almost 30%). The percentage of attention of navigational queries on sponsored links is very low (less than 10%) and for informational queries that percentage is not representative (less than 3%). If we observe the average gaze duration of all the queries on sponsored links (Table 3.3), the trend that we found in the number of fixations is confirmed, the users invest an important time on sponsored results only for transactional queries. While for organic links informational queries are in the first place of number of fixations, for sponsored links is the opposite, informational queries are in the last place for time of visual attention. Figure 3.4 shows heat maps and gaze plots for queries with different intents.

Overall, our findings show that organic results are the main focus of attention for informational, navigational and transactional queries. Although transactional queries exhibit an important proportion of attention on sponsored links, it does not mean that sponsored links are more important than organic links for queries with this type of intent. With exception of transactional queries, the users do not spend too much time exploring sponsored links.

Location of Sponsored Links on SERBS

Sponsored links are not on the top preferences of the users when they examine SERBS, as we saw in the previous section. However, sponsored links are not originally designed to fully satisfy all the needs of the users, sponsored links are created to cover a target group of user needs: those that are related with commercial transactions. From this point of view and considering our previous findings, we could say that sponsored links serve to the purpose for which they were created, that is, ads are drawing the attention of users who are involved in transactional tasks. One of the major challenges for search advertising is extracting the maximum utility of the attention that they are drawing from search engine users. In this direction, a relevant factor is knowing what are the preferences of the users with respect to the sponsored links: what type of sponsored links would users prefer: top-listed or side-listed sponsored links?

Figure 3.5 shows the distribution of visual fixation of the users on each type of sponsored search result: top-listed and side-listed sponsored links and Figure 3.6 shows the mean gaze duration for each task on sponsored links. The results suggest that top-listed sponsored links are more relevant than side-listed sponsored links for all types of queries. As we already established, transactional queries appear as the group of queries with a highest proportion of attention on both types of sponsored links. Despite that participants even have fixed their attention on side-sponsored links (Figure 3.5), their fixations are very short.

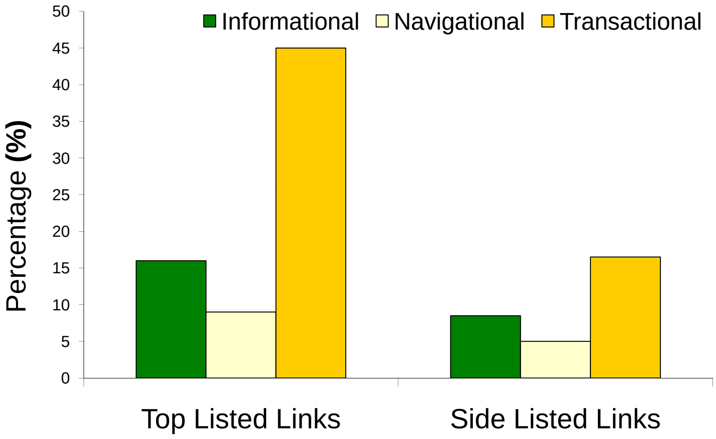


Figure 3.5: Percentage of number of fixations for each type of sponsored link broken down by query intent.

One interpretation of why users exhibit more visual attention on sponsored links placed on the top of the SERP than in those placed in the lateral part, it could be that the users confuse top-listed sponsored links with organic search results. However, when we analyze the Figure 3.6, we observe that the average time invested in top sponsored links is not the same for tasks of all types of intents. We see that for informational and navigational tasks, the accumulated time is around 200 ms and for transactional tasks is over 1000 ms. Fixations are commonly interpreted as signs of processing new information. When processing information, eye fixations are estimated to last between 200 and 250 ms (on average) [80, 65]. Since users with informational and navigational queries, usually are not interested on sponsored results, they are more likely to confuse a top sponsored link with a organic result and abandon it when realize it is sponsored information. The average gaze time of 200 ms is just the enough time to recognize the ‘fake’ organic result. In the case of transactional queries, an average gaze time of 1000 ms indicates that the user really evaluate

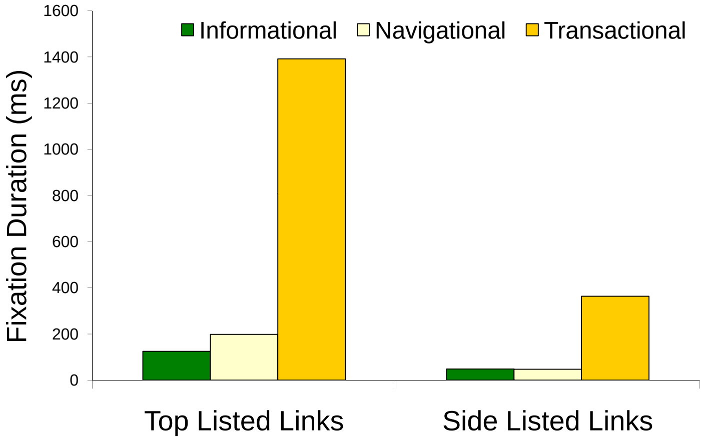


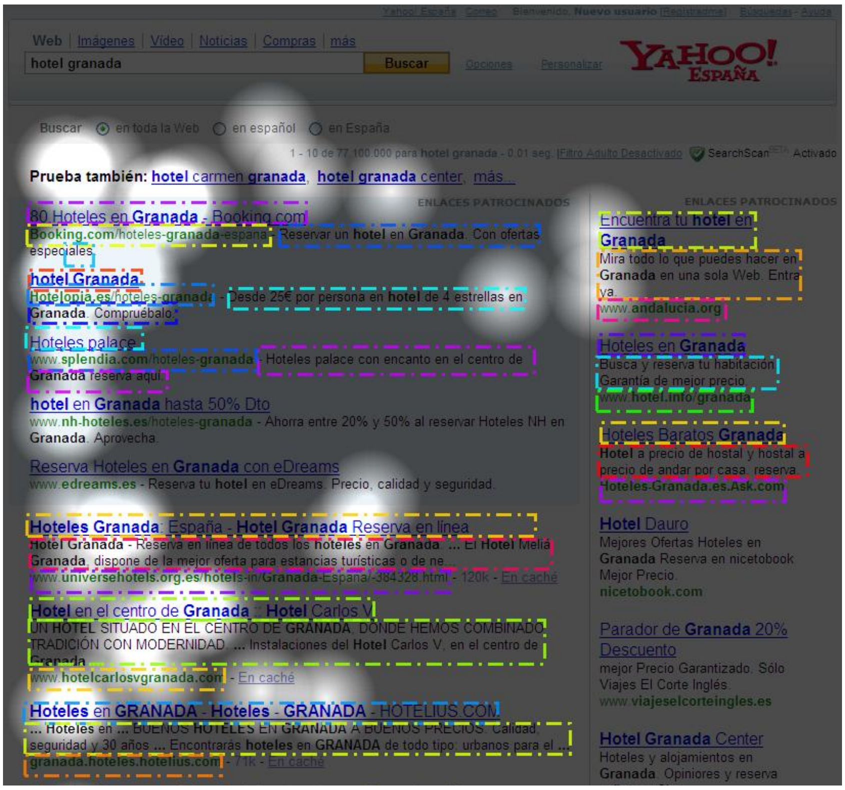
Figure 3.6: Average gaze duration (ms) for each type of sponsored link broken down by query intent.

the sponsored result and suggests this evaluation is not due to a confusion with organic results.

Overall, 70% of the visual fixation on sponsored search results were on top-sponsored links and 30% on side-sponsored links; however, with exception of transactional queries, gaze durations on side-sponsored links were minimum. The mean gaze duration of the three types of query intent with the two types of sponsored links are statistically significant: informational (F=62.59, $p < 0.001$), navigational (F=37.79, $p < 0.001$), transactional (F=81.92, $p < 0.001$). Figure 3.7 shows the distribution of visual attention on a SERP for a transactional task, where white areas indicates accumulation of user's fixations.

3.3.2 Query Intent and Areas of Interest

In previous sections we have analyzed and delineated the general user behavior on different types of search results. In this section we show the



Yahoo! España Correo Bienvenido, **Nuevo usuario** [Registrarme] Búsquedas - Ayuda

Web | Imágenes | Vídeo | Noticias | Compras | más

hotel granada Buscar Opciones Personalizar

Buscar ● en toda la Web ○ en español ○ en España

1 - 10 de 77.100.000 para **hotel granada** - 0.01 seg. |Filtro Adulto Desactivado SearchScan BETA Activado

Prueba también: **hotel** carmen **granada**, **hotel granada** center, más...

ENLACES PATROCINADOS

80 Hoteles en **Granada** - Booking.com
Booking.com/hoteles-**granada**-espana - Reservar un **hotel** en **Granada**. Con ofertas especiales.

hotel Granada
Hotelopia.es/hoteles-**granada** - Desde 25€ por persona en **hotel** de 4 estrellas en **Granada**. Compruébalo.

Hoteles palace
www.**splendia.com**/hoteles-**granada** - Hoteles palace con encanto en el centro de **Granada** reserva aquí.

hotel en **Granada** hasta 50% Dto
www.**nh-hoteles.es**/hoteles-**granada** - Ahorra entre 20% y 50% al reservar Hoteles NH en **Granada**. Aprovecha.

Reserva Hoteles en **Granada** con eDreams
www.**edreams.es** - Reserva tu **hotel** en eDreams. Precio, calidad y seguridad.

Hoteles Granada: España - **Hotel Granada** Reserva en línea
Hotel Granada - Reserva en línea de todos los **hoteles** en **Granada**. ... El **Hotel** Meliá **Granada**, dispone de la mejor oferta para estancias turísticas o de ne...
www.**universehotels.org.es/hotels-in/Granada**-Espana/-384328.html - 120k - En caché

Hotel en el centro de **Granada** :: **Hotel** Carlos V
UN **HOTEL** SITUADO EN EL CENTRO DE **GRANADA**. DONDE HEMOS COMBINADO TRADICIÓN CON MODERNIDAD. ... Instalaciones del **Hotel** Carlos V, en el centro de **Granada**
www.**hotelcarlosvgranada.com** - En caché

Hoteles en **GRANADA** - **Hoteles** - **GRANADA** - HOTELIUS.COM
... **Hoteles** en ... BUENOS **HOTELES** EN **GRANADA** A BUENOS PRECIOS. Calidad, seguridad y 30 años ... Encontrarás **hoteles** en **GRANADA** de todo tipo, urbanos para el ...
granada.hoteles.hotelius.com - 71k - En caché

ENLACES PATROCINADOS

Encuentra tu **hotel** en **Granada**
Mira todo lo que puedes hacer en **Granada** en una sola Web. Entra ya.
www.**andalucia.org**

Hoteles en **Granada**
Busca y reserva tu habitación Garantía de mejor precio
www.**hotel.info/granada**

Hoteles Baratos **Granada**
Hotel a precio de hostal y hostal a precio de andar por casa, reserva.
Hoteles-Granada.es.Ask.com

Hotel Dauro
Mejores Ofertas Hoteles en **Granada** Reserva en nicetobook Mejor Precio.
nicetobook.com

Parador de **Granada** 20% Descuento
mejor Precio Garantizado. Sólo Viajes El Corte Inglés.
www.**viajeselcorteingles.es**

Hotel Granada Center
Hoteles y alojamientos en **Granada**. Opiniores y reserva

Figure 3.7: Gaze opacity map for a transactional query.

distribution of visual attention of users on different areas of interest of a SERP, regarding the underlying intent of the query. Do users attend to different AOIs of SERP for queries with different intent? Figure 3.8 shows the distribution of number of fixations and gaze durations for each type of task on AOIs: title, snippet, URL and image. Since most of the attention of the three types of query intent is focused on organic search results, the Figure 3.8 is based on this type of results.

There are several differences for the distribution of visual attention of the users (fixations and gazes) on each AOI with respect to the intent. As we expected, because its size, the area that attract highest fixations and gaze durations for all the queries is the snippet. Specifically the distri-

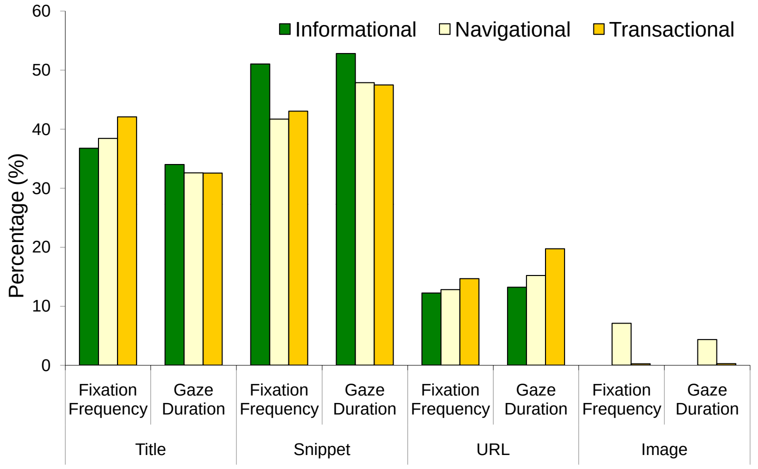


Figure 3.8: Distribution of visual attention on AOIs broken down by query intent.

bution on the snippet is: informational (51%), transactional (43%) and navigational (42%). Users spend more time on the snippet for informational queries, confirming that the snippet is an important element to help users to determine if a given result is likely to have the information they are looking for. These findings are consistent with the results reported on [21]. The second area with high number of fixations and gaze durations for the queries is the title, the distribution of visual attention for this AOI is different with respect to the snippet. Titles are more relevant for transactional and navigational queries and in lower proportion for informational queries. The URLs show the same distribution of visual attention than titles, that is, transactional and navigational queries in the first place of the user preferences followed by the informational queries. The attention bias on the tittle and the URL for transactional and navigational queries suggest that users with this kind of query intent are specially interested on the identification of the web sites to which search results belong. The only type of intent that devoted a portion of time examining images was the navigational intent. Despite that the proportion of total time and fixations invested for navigational queries on images is not high, it is significant, if we consider that informational and transactional queries did not register gazes on images.

Finally, considering that only the transactional queries showed significant visual attention on sponsored search results, we show the distribution of the attention for these queries on sponsored links regarding the AOIs (Figure 3.9). The most important AOIs when users examine sponsored links are the title, followed by the snippet and the URL. Although the snippet exhibit a higher number of fixations than the URL, it is important to notice that the gaze duration is almost the same for the two AOIs. Given that the snippets have a longer size, this result suggest users spend more time examining each individual URL than with each snippet. If we normalize the snippet by the number of its lines, the correlation between the number of fixations and the duration of fixations becomes similar to the corresponding correlation for the AOI's title and URL (that normally have one line). In other words, in cases where the snippet is the area with the greatest concentration of attention, its popularity would drop and it would be become closer to the popularity of other AOI's, while in cases where the popularity of the snippet is below the popularity of other AOI's, its importance would decline further.

3.3.3 Query Intent and SERP Ranking Position

Overall, we can say that the ranking position of the result is relevant for the users (see Table 3.4). For all types of search results, the first ranking position collected the highest proportion of attention. For organic results this proportion was more than 40% for informational and navigational queries and 30% for transactional queries.

As we mentioned in Section 3.3.1, organic results are the most valuable search results for the users. In Table 3.4 we can see how the attention of users is distributed along the first three organic results. Users devoted a big part of their fixations on rank-1 and rank-2 results. For the three query intents the accumulated fixations on the first two results are: navigational

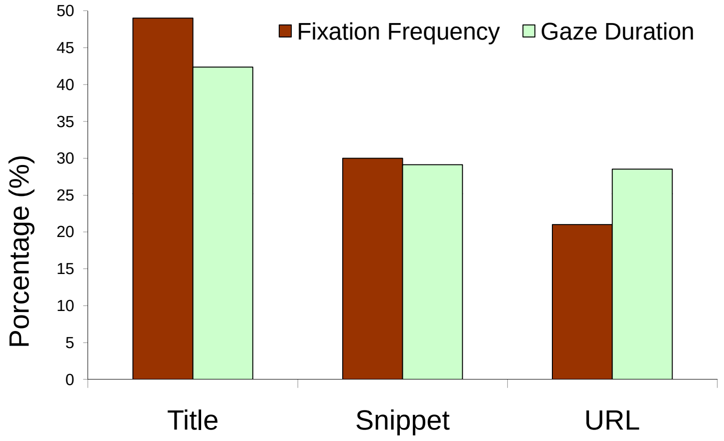


Figure 3.9: Distribution of visual attention of transactional queries on AOIs for sponsored links.

(76.84%), informational (75.46%) and transactional (56.74%). Although navigational queries were not the type of queries that accumulated the major number of fixations in organic results, they were the tasks with the biggest number of fixations on organic results 1 and 2. This suggests that these two search results are particularly relevant for navigational tasks. With exception of informational tasks, the interest of the users on organic rank-3 result with respect to the other two results, decreases considerably. However, overall, the rank-3 result remains an important part of the user fixations.

With respect to the sponsored results, as we also mentioned in Section 3.3.1, top-listed sponsored results attract more visual attention from users than side-listed sponsored results. Concerning with the ranking position of the search results, the distribution of fixations is different to the fixations on organic results. In sponsored results, the fixations are almost totally in the first result. Informational and navigational queries do not present significative fixations on side-listed sponsored results and for top-

	User's Intent		
Rank	Informational	Navigational	Transactional
Organic1	47.22	47.96	30.02
Organic2	28.24	28.88	26.71
Organic3	21.95	15.53	14.18
Top-Ad1	0.85	3.00	10.87
Top-Ad2	0.85	0.82	5.67
Top-Ad3	0.00	1.09	4.73
Side-Ad1	0.69	1.36	5.67
Side-Ad2	0.11	0.54	1.65
Side-Ad3	0.11	0.82	0.47
Total	100	100	100

Table 3.4: Percentage of number of fixations on search results broken down by rank and query intent.

listed sponsored results the percentage of fixations is bigger than 1% only for the first ranking position. On the other hand, 29.08% of the fixations of transactional queries were for sponsored results, 21.28% of these fixations were on top-listed results and 7.80% on side-listed results. Again, the most relevant ranking position is number 1. Transactional queries distributed the visual attention in all types of results; the fixations are not concentrated in only one type of results like in informational queries. Unlike to [50], where they found that the ranking position did not affect sponsored result viewing, our findings suggest that ranking position and the location on the SERP of the sponsored search results influence sponsored result viewing. The lower the rank (closer to the top of the page) the more likely a user will check the search result.

Figure 3.10 shows the distribution of fixations on organic results broken down by AOIs. The specific areas of interest are influenced by the query intent. The viewing pattern for informational queries is homogenous along the three organic results: the snippet with the highest number of fixations, followed by the title and finally the URL.

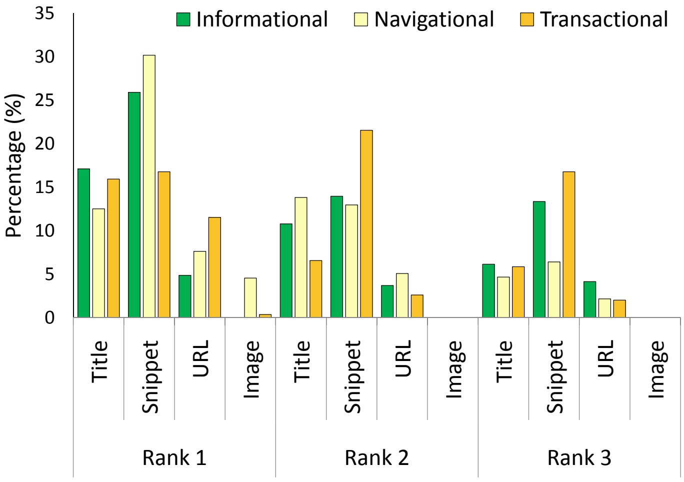


Figure 3.10: Percentage of number of fixations on AOIs for organic results broken down by rank and query intent.

Snippet and title are also the most important AOIs for navigational queries, but in this case the distribution of fixations varies from the first to the second organic result. In the first organic result the snippet is remarkably more important than the title for navigational queries and in the second organic result the title is slightly more relevant than the snippet. The URL is also an important element for navigational queries, especially in the first result. Transactional queries have a balanced distribution of fixations on the snippet, tittle and URL for the first organic result. Figure 3.11 shows the distribution of fixations on AOIs for sponsored results. Compared with the organic results, there are some changes in the order of relevance of the AOIs for each query intent. In the first top-listed sponsored result for example, informational queries devote a considerable percentage of visual fixations to the URL and the attention on the title decreases. The snippet is still the area with the largest number of fixations for informational queries in this result.

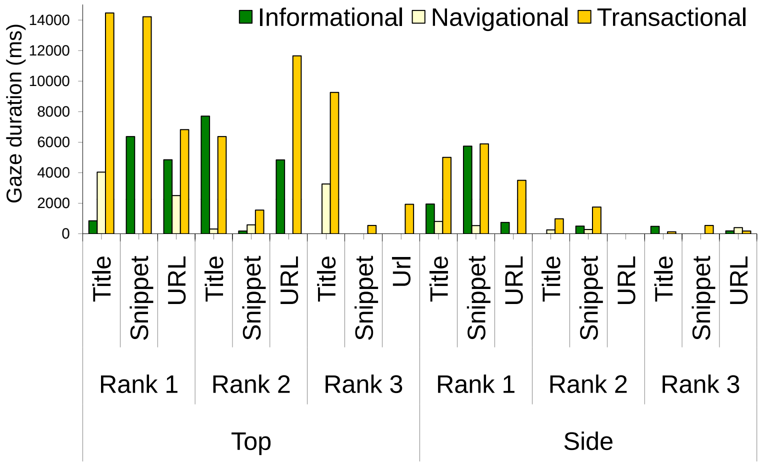


Figure 3.11: Accumulated gaze time on AOIs for sponsored results broken down by ranking position and query intent.

Navigational queries focused all their fixations in the title and the URL, there are not fixations on the snippet in the first result for these queries. In the case of transactional queries, snippet and title with almost the same percentage of number of fixations, accumulated most of the visual attention in the first top-listed result. In the second and third results, transactional queries focused the attention on the URL and the title, while the snippet is almost ignored. Overall, for top-listed sponsored results the title and the URL are the most important areas of interest for the users.

Although the number of fixations on side-listed sponsored results is very low, we can analyze the visual attention that users devoted to the first result. In general, the viewing behavior of users for the first side-listed sponsored result is similar to the viewing behavior of users for the organic results (keeping the proportion on the number of fixations): for informational queries the snippet captures most of the visual fixations of the users, followed by title and at the last place the URL. For transactional queries, we observe again, a balanced distribution of fixations on the snippet, tittle and URL for the first side-listed result and in the second and third side-listed results the most important AOI was the snippet. Interestingly, organic and side-listed sponsored results have similar trends for visual attention on AOIs, while top-listed sponsored results exhibit a different distribution of the user's visual attention.

3.4 Summary and Implications

This study presents an analysis of how the behavior of the user who examines a SERP varies for different query intents. Users with different query intent preferred different type of search results, they attended to different main areas of interest and focused on search results with different ranking position.

Overall, our findings show that organic results are the main focus of attention for informational, navigational and transactional queries. Sponsored links are important for transactional queries but not more important than organic links and with exception of transactional queries, the users do not spend too much time exploring sponsored links. The location of the search results is also an important factor that influences the users viewing behavior on SERBS, this factor specially affects sponsored results. Top-listed sponsored results are more relevant than side-listed sponsored results for all types of queries. Informational and navigational queries do not present significative fixations on side-listed sponsored results and for top-listed sponsored results the percentage of fixations is low. Transactional queries distributed the visual attention in all types of results; the fixations are not concentrated in only one type of results like informational queries.

We confirmed that the ranking position of the result is relevant for the users. For all types of search results, the first ranking position collected the highest proportion of visual attention. Navigational queries pay special attention to the first two organic results and with exception of informational queries, the interest of the users on the third organic result, decreases considerably. The specific areas of interest and the relevance

Query intent	Type of result	Areas of Interest	Ranking position
Informational	Most of the attention on organic results: 97.4%	**Snippet,** title	First organic result: 47.22% of attention
Navigational	Important attention on organic results: 92.4% and weak attention on sponsored results: 7.6%	**Snippet,title,** URL, image. *The only one with fixations on images.*	First organic result: 47.96% of attention and first sponsored result: 3%
Transactional	Balanced distribution of attention: organic results: 71% and sponsored results: 29%%	**Snippet, title,** URL For sponsored results *title* is preferred	First organic result: 30%% of attention and first sponsored result: 10%

Table 3.5: Main observed differences between each type of query intent.

order of the AOIs change depending of the query intent. We show a summary of the main differences between each type of intent in the Table 3.5.

In summary, the results of the study show that the query intent of the user affects all the decision process in the SERP; different types of query intent have different implications on users' search behavior. As a consequence, knowing the intent of the user's Web queries could help Web search engines in the task of associating resources available on the Web with the user's goals. To be able to accurately identify the intent of the user query, will provides useful elements that allow search engines adapt their results to changing user behaviors and needs.

This chapter shown how the search behavior of the user changes according to his query intent. Hence, the interfaces of Web search engines as well as the presentation of the search results could be adapted to different goals of the users. However, in order to do that, it is necessary to being able to identify the user query intent. Addressing this important requirement of identify the intent behind the Web queries, the next chapter presents a first approach to the automatic prediction of the user's query intent.

Chapter 4

FEASIBILITY OF AUTOMATIC QUERY INTENT PREDICTION

Users utilize the Web in different ways to achieve their goals and one of the most important tools they use to find resources that match these goals are search engines. Unfortunately, the Web in its basic form is a non-intentional repository. The users are obligated to express their goals in few words (queries) that are understandable for search engines, in order to obtain resources close to their needs. A natural way to improve the user's search experience is infer the intent of the user's queries and adapt the search results to the specific goals of the users.

This chapter explores the feasibility of automatic query intent prediction. We take a supervised learning approach to automatically identify the intent of the queries from two perspectives: the type of intent of the user's queries and the topic categories in which this type of intent could be placed. In order to measure the effectiveness of the automatic classification and establish the feasibility of the query intent prediction, a manual classification of a set of queries was carried out and the results were compared with the results of the automatic classifier. This provided a measure of the accuracy of the prediction process of the query intents. The organization of the chapter is as follows. In Section 4.1 we describe the query intent taxonomy and topic categories used in this study. In Section 4.2 we introduce the methodology followed in the experiments. Section 4.3 presents results of the manual classification of queries. In Section 4.4 we present and analyze the results of the automatic query intent prediction. Finally, in Section 4.6 we introduce the discussion and conclusions of the study.

4.1 Query Intent and Topic Categories

From the content of the queries we established three categories for the reasons or intents that motivate a user to make a search:

- *Informational*: An informational query is one in which the user exhibits an interest to obtain information about a specific topic and this information can be located in one or more Web sites, e.g., “foreign currencies” or “fermat theorem proof”.

- *Not informational*: As not informational queries, we categorize queries that involve finding some kind of resource or target a specific transaction (e.g. buy, download, reserve, etc.) or a specific Web site: “ringtones mobile phone”, “motorcycles sales” or “hotmail”.

- *Ambiguous*: ambiguous queries are those that their intent can not be inferred directly from the query, that is, query words are not informative enough to associate the query to a single type of intent, e.g., “ice” or “snorkel”.

In contrast with the taxonomies of queries proposed by [14, 81], we consider the *Not Informational* category, which groups *Transactional* and *Navigational* queries. In this case, *Navigational* queries are considered a kind of *Transactional* query, the target Web site is the resource that the user is looking for. We also consider the *Ambiguous* category, as there are many queries submitted to search engines that are ambiguous by nature.

Arts	Business	Computers
Education	Games	Health
Home	News	Recreation
Reference	Science	Shopping
Society	Sports	World

Table 4.1: ODP Topic Categories used in the classification process.

Most of previous work have realized the existence of *Ambiguous* queries only after performing the classification of the queries, when they find that there is an overlap between different types of queries. For example, in Lee *et al.* [61], they found that some queries can not be clearly associated with a specific type of query, they calling these queries 'unpredictable queries'. Our idea is to have a particular category for the *Ambiguous* queries. Identifying this type of queries since the beginning of the classification process and evaluating the predictability of *Ambiguous* queries, in the same way that we evaluate the predictability of the other two types of queries.

Additionally to the query intent, we also propose to associate the queries with corresponding topic categories. One of the conclusions in [61] after the analysis of the unpredictable queries (*Ambiguous*), was that these queries belonged to a couple of topic categories. These initial findings suggest that *Ambiguous* queries could be associated to some topic categories and knowing the topic of a query could bring useful evidence to identify if the query is ambiguous or not. In this work, we use the query intent in combination with topic categories in order to establish how the relation between these features (query intent and topic category) could improve the identification of the user' interests. The topic categories used to classify the queries are based on the most general categories of the Open Directory Project (see Table 4.1).

In addition to the topic categories of ODP , we considered three additional categories:

- *Various*: For queries that could belong to several topic categories.
- *Other*: The query does not belong to any of the considered topic categories.
- *Adult*: We included this topic category considering the amount of queries in the data set that are related with adult content.

In summary, we have eighteen topic categories, fifteen categories extracted from ODP (Table 4.1) and three additional categories proposed from the observations of the data set.

Based on the query intents and the topic categories introduced before, we designed an automatic identification process to determine the feasibility of the query intent prediction. In the next sections, we will show the procedure and the results obtained with the automatic identification experiments.

4.2 Methodology

This section describes the methodology we employed in the automatic identification of the user query intent. It begins with the description of the data set of queries we used in the experiments. It then presents a description of the preprocessing of data that we applied, in order to obtain the necessary input to our methods. Finally, we describe the manual classification process that we carried out to obtain a labeled data set against which we can evaluate our automatic identification process of the query intent.

4.2.1 Data Set and Data Preprocessing

For this work a log sample from the Chilean Web search engine TodoCL[1] were processed. The sample contains 6,042 popular queries having clicks in their answers. There are 22,190 clicks registered in the log, and these

[1] TodoCL. http://www.todocl.com

clicks are over 18,527 different URLs. Thus, in average users clicked 3.67 URLs per query.

In a similar way than [6, 36], each query was represented as a vector of terms that appeared in the documents giving an answer to the query (stop words were removed):

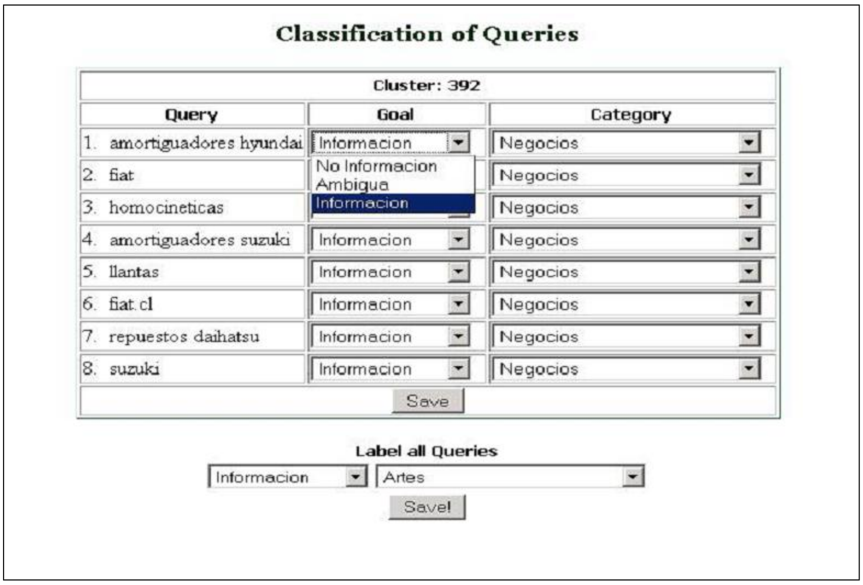
$$Q_i = w(t_1), w(t_2)...w(t_n)$$

Where $w(t_j)$ is the associated weight of term j inside query Q_i. The classical TF-IDF weighting scheme was used to assign the weight to each query term and clicked page, replacing IDF by the number of clicks on each page.

4.2.2 Semi-Manual Classification of Queries

In this section we describe the semi-manual classification of the queries. After the representation process, a clustering process was applied over the data. We obtained query groups with similar characteristics, i.e. they belong to the same subject, are related with specific topics or describe the same situation using different terms. To do this, we used a simple K-means clustering method [59]. The groups of queries obtained through the clustering process, were used to facilitate the manual classification of the queries. We use the term *semi-manual classification*, given that, through the clustering process, implicit information from a group of queries is automatically added to a query. This information can be seen as the context that a set of similar queries provides to a single one. Once the queries were manually classified, the clusters were dissolved, and the queries were considered again as individual items belonging to the general set. In this way we could speedup a direct manual classification by at least one order of magnitude.

As a way to facilitate the manual classification process, we created a software tool which offered to users the possibility to select the intents and the topic categories from a list and save them in an organized and fast way (See Figure 4.1). The query groups obtained with the clustering process that we mentioned before, were used to present the queries to



Classification of Queries

Cluster: 392		
Query	Goal	Category
1. amortiguadores hyundai	Informacion	Negocios
2. fiat		Negocios
3. homocineticas		Negocios
4. amortiguadores suzuki	Informacion	Negocios
5. llantas	Informacion	Negocios
6. fiat.cl	Informacion	Negocios
7. repuestos daihatsu	Informacion	Negocios
8. suzuki	Informacion	Negocios

Label all Queries: Informacion, Artes

Figure 4.1: Manual Classification Tool.

the annotators in a structured way. This cluster based structure provided our team with information about the context of the query and at the same time, the organization of the queries in clusters gave to the annotators a global idea about the class to which each query belongs. This information is used to facilitate the human classifiers in the case that a query did not suggest a complete idea by itself. The observation of the rest of queries in the cluster could bring a more comprehensive idea of the intent of that query. In any case, the task of a human classifier is to select the type of intent for a query and the topic category in which this intent can be situated.

4.3 Manual Classification Results

By looking at the manual classification results, we can make a first assessment of the predictability of the query intent and topic categories. The

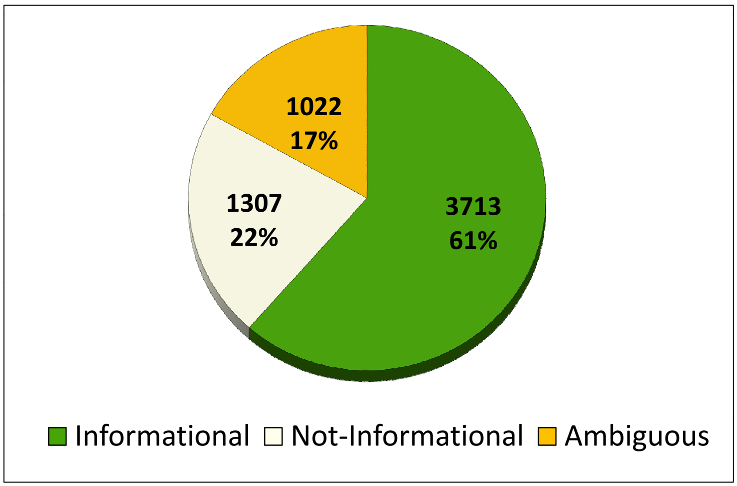


Figure 4.2: Distribution of Queries into Intents.

number of queries in each type of intent and its relation with the topic categories allow us to establish an initial idea of the distribution of the interest of the users.

First we introduce separately, the results of the manual classification for query intents and for topic categories. Then, we present the results of the query intents combined with the results of the topic categories, and we analyze the relations between these two features.

4.3.1 Queries and Search Intents

As it was previously described, we established three types of intent to which a query can belong: *Informational, Not Informational* and *Ambiguous*. The human annotators classified all the queries of our data set (6042) into these three types of intent. The results are shown in Figure 4.2.

Most of the queries were classified as *Informational*, showing that users concentrate most of their activities around obtaining information

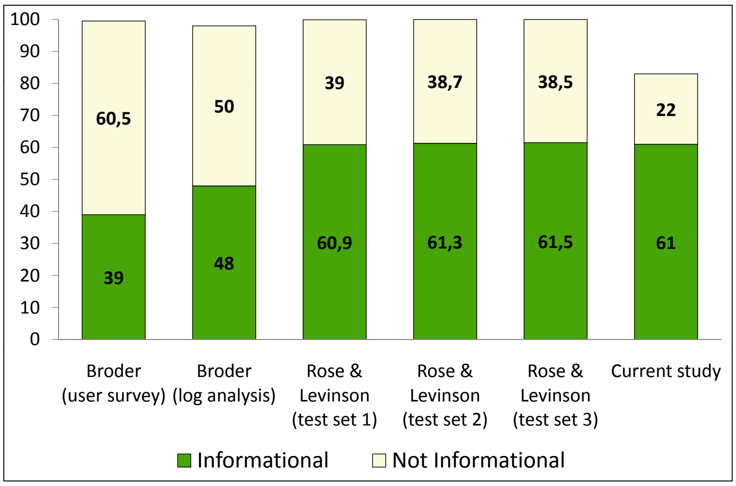


Figure 4.3: Comparison of search taxonomies of Broder and Rose and Levinson to the query intent taxonomy of the current work.

and hence they use search engines as information retrieval tools. In the second place of the preferences of the users we have *Not Informational* queries. After *Informational* queries, the *Not Informational* queries grouped 22 % of the total of queries, which is an important quantity and reveals the trend of users to diversify the use of search engines. Users are not only looking for information, they are also interested in obtaining resources and are involved in many kind of activities where search engines help to locate what they want. In the last position we place *Ambiguous* queries. This group of queries has special characteristics, as we mentioned before, because could be both *Informational* or *Not Informational*. Although, *Ambiguous* queries collected the lowest number of queries, its proportion is considerable and it is important to evaluate the characteristics that distinguish these type of queries. In the next sections we will analyze with more detail this case.

However the types of intent considered to label a query are different between this work and the work done by Broder [14] and Rose and

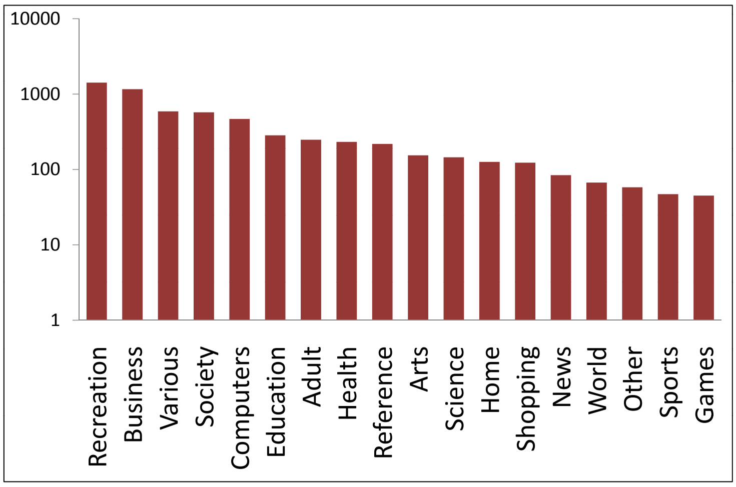


Figure 4.4: Distribution of queries into topic categories.

Levinson [81]. We can compare our results with them if we consider some adjustments on the categories of their taxonomies. We can consider the Broder's taxonomy (*Informational, Navigational* and *Transactional*) as an *Informational - Not Informational* (*Navigational* and *Transactional*) taxonomy. Similarly, we can consider the top-level goal taxonomy of Rose and Levinson (*Informational, Navigational* and *Resource*) as an *Informational - Not Informational* (*Navigational* and *Resource*) taxonomy.

Figure 4.3 compares our manual classification results with results reported by Broder and Rose and Levinson. Since *Ambiguous* is a different type of query intent, which can contain queries with mixed characteristics, we did not include these queries in the comparison. For this reason, our results do not total 100% in the Figure 4.3. We agree with Rose and Levinson in the proportion of *Informational* queries, the percentage of these queries is the same in both studies: (61%). The percentage of *Informational* queries reported by Broder is between 40% and 50%, which is also a high proportion of queries. From the results the most heterogenous group is the one formed by the *Not Informational* queries. Given that *Not*

Informational queries could imply actions with different characteristics, the studies differ in the proportion of these queries, depending of the definition that they use for this group of queries. In our case, we reported 22% of *Not Informational* queries, while Broder and Rose and Levinson reported an average of 45% of *Not Informational* queries between them. Despite that there is an important difference in the proportion of *Not Informational* queries of the current study with respect to earlier studies, it is important to consider the *Ambiguous* queries. Since earlier studies did not consider *Ambiguous* queries, it is possible that some queries from the *Not Informational* group, actually belonged to the *Ambiguous* group.

Overall, we can say that there is a reasonable agreement in the characteristics of *Informational* queries and a deeper analysis is required to understand the nature of *Not Informational* queries.

4.3.2 Queries and Topic Categories

Most of the current Web search systems include some type of topical classification of user queries. Knowing topical associations of the queries can bring improvements in the search experience of the users. For instance, the content and presentation of search results may be adapted according to the topic categories of the queries. Figure 4.4 and Table 4.2 show the results of the manual classification of queries into topic categories. The topic categories with higher amount of queries are *Recreation* (*Entertainment*) and *Business*, which confirm the search behavior of people that have been well described by Spink and Jansen in their works [91, 51].

After *Recreation* and *Business*, the topic categories that concentrated most quantity of queries were: *Society*, *Computers* and *Education*. The topic *Various* also concentrated an important number of queries. This topic category has an ambiguous nature and we introduce it because we wanted to test how many queries were related with more than one topic at the same time. The manual classification results for *Various* suggest that there is a considerable group of queries with complex topical associations, beyond a single topic.

Topic	**Query Count**
Adult	4.1%
Arts	2.5%
Business	**19.2%**
Computers	7.7%
Education	4.7%
Games	0.7%
Health	3.8%
Home	2.1%
News	1.4%
Other	1.0%
Recreation	**23.5%**
Reference	3.6%
Science	2.4%
Shopping	2.0%
Society	9.5%
Sports	0.8%
Various	9.8%
World	1.1%

Table 4.2: Percentage distribution of manual classification of queries into topic categories.

The rest of the queries are distributed along the rest of topic categories such as *News, Arts, Health* and so on. Although these topics did not concentrate the same proportion of queries than *Business* or *Recreation*, the queries were clearly recognized by the annotators to belong to this group of topics.

4.3.3 Relations between Intents and Topic Categories

In this section we examine the manual classification results, considering the association between the three levels of user query intent and the eighteen topic categories. Table 4.3 presents the percentage distribution of the type of query intent into topic categories. Notable percentages, over 30%, were bolded. In Figure 4.5, we see the distribution of number of queries along the intents and topic categories. Results show that the associations of topics vary for each query intent.

As we can see from Table 4.3, there are topic categories which main component is *Informational*. In this group we have topics such as *News* (93%), *Science* (89%), *Society* (87%), *Education* (82%) and *Health* (74%). All these topics suggest informational purposes, as people are searching for resources answering in many of the cases to a specific information need. Another topic that was clearly associated with the *Informational* intent was *Business* (83%). *Business* was also one of the topics that grouped the highest numbers of queries (see Figure 4.5).

Queries grouped as *Not Informational* belong to topics such as *Adult* (72%), *Computers* (44%), *Games* (58%) or *Recreation* (34%). The intent of these queries is, in most of the cases, to visit a Website to download resources (e.g, software) or perform activities related with the mentioned topics.

Finally, as it was expected, the queries grouped as *Ambiguous*, were mostly related with the topics *Various* (57%) and *Other* (57%). This relation of *Ambiguous* queries with the topics *Various* and *Other*, suggest that when the nature of a query is ambiguous, this ambiguity is extended to the rest of characteristics of the query, including its topical associations. Another topics with important proportions of *Ambiguous* queries were *Home* (33%), *Shopping* (32%) and *World* (22%).

4.4 Automatic Query Intent Prediction

In this section, we used the manually classified data set to train a supervised machine learning model that can predicts the type of intent and the topic category of a query. The idea is that the machine learning model can abstract from the particular combinations of features seen in the man-

Topic	**Info.**	**Not Info.**	**Amb.**
Adult	15%	**72%**	13%
Arts	**66%**	15%	19%
Business	**83%**	9%	8%
Computers	**37%**	**44%**	18%
Education	**82%**	10%	8%
Games	24%	**58%**	18%
Health	**74%**	9%	17%
Home	**40%**	28%	**33%**
News	**93%**	6%	1%
Other	28%	16%	**57%**
Recreation	**56%**	**34%**	10%
Reference	**49%**	**39%**	12%
Science	**89%**	5%	6%
Shopping	**45%**	24%	**32%**
Society	**87%**	2%	10%
Sports	**66%**	23%	11%
Various	**38%**	5%	**57%**
World	**69%**	9%	22%
Average	58%	23%	20%
Sd Dev	24%	20%	16%

Table 4.3: Percentage distribution of queries into intents and topic categories.

ual classified data and uses this abstraction to identify query intent and topic categories for new queries, that have not seen before in the training process.

4.4.1 Supervised Machine Learning

In this study, Support Vector Machines (SVM) have been used to build classification models for queries. SVM is a classifier, originally proposed by Vapnik [98]. This classification algorithm is based on the margin of a training sample. The margin of a training example is a number that is

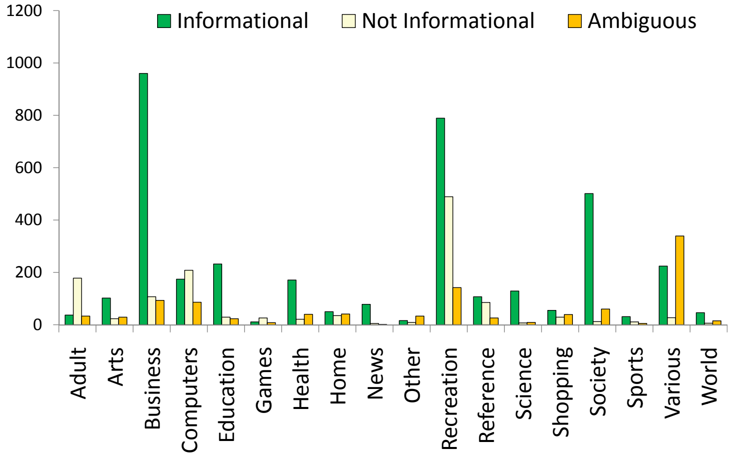


Figure 4.5: Distribution of number of queries into intents and topic categories.

positive if and only if the example is correctly classified by a given classifier and whose magnitude is a measure of confidence in the prediction. The SVM algorithm attempts to maximize the minimum margin of any training example; that is, it finds a maximal margin separating hyperplane between two classes of data. A detailed description of this model can be found in [20].

The SVM classifier was selected due to its proven effectiveness in different scenarios with a high feature dimensionality, including text classification [9]. Considering that the queries were represented by the words of the pages selected by the users, this characteristic is quite useful.

To handle the multiclass problem, we combine SVM with Error-Correcting Output Coding (ECOC). This is an approach for solving multiclass categorization problem originally introduced by Dietterich and Bakiri [25]. It reduces the multiclass problem to a group of binary classification tasks and combine the binary classification results to predict multiclass labels. The RBF (Radial Basis Function) kernel was used to the

SVM's setup, and we choose the kernel's parameters through a standard cross-validation process.

We performed the classification experiments using the Weka toolkit [40]. To build the models and make the predictions for topic categories, SVM spent about thirty minutes. In the case of query intents, considering the low amount of labels involved, the classifier took about ten minutes. The experiments were run on a Pentium III computer with 1.28 GBs of RAM, under a Linux OS.

4.4.2 Automatic Prediction Results

After the manual classification of the queries was made, part of these labeled data was used as input to train automatic classifiers. We trained on 70% of labeled queries, and tested on the remaining 30%.

In order to measure the accuracy of the automatic prediction results with respect to the original labels provided by our team, we report the recall and precision measures in the same way as were reported in [1]. Precision is then, the fraction of predicted query labels that agree with the labels obtained from the explicit human judgments. Recall is the fraction of query labels obtained from the explicit human judgments that were correctly predicted.

With respect to the types of query intent, the precision is over 50% for two of the intents, *Informational* and *Not Informational* (see Figure 4.6). The best results were obtained with *Informational* queries for which the recall and precision values are high. As it was observed in Section 4.3.1, there is a general agreement in the concept of *Informational* queries, and this is reflected in the automatic classification process. Pages selected by the users of *Informational* queries have more homogeneous purposes, unlike those that belong to *Not Informational* or *Ambiguous*, where the diversity of objectives of the queries is higher, and hence the nature of the pages selected by the users is more heterogeneous.

Although not in all the cases the query intent predictions were agree with the human judgments, the prediction in most cases is related with the subject of the query. Despite that the human annotators tried to identify

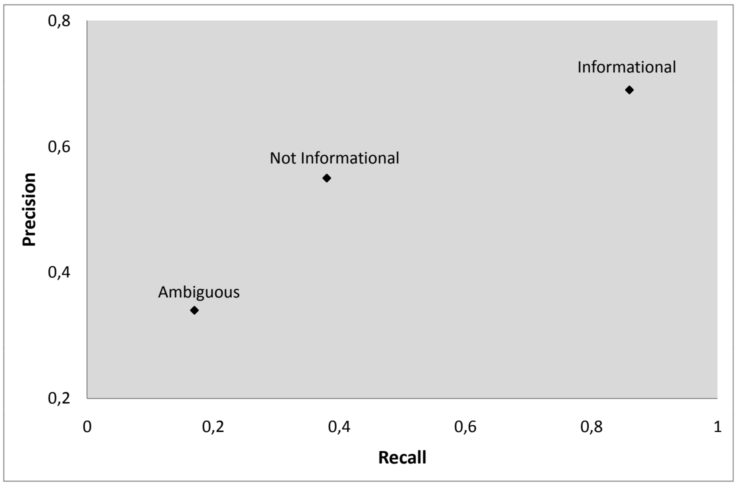


Figure 4.6: Precision vs Recall of automatic query intent prediction

all the *Ambiguous* queries, some of them that were labeled as *Not Ambiguous* (*Informational* or *Not Informational*), were automatic classified in a different type of intent. This fact indicates that the representation of these queries suggests a different intent than the indicated by the manual label.

For the topic categories, overall results are good; nevertheless, for some categories the results are better than for others. In the Figure 4.7 we can see a sample of the most representative categories: The categories that show better precision are those that have greater popularity, that is to say, those related to subjects that the people consult most frequently, like for example: *Recreation*. Since most of pages of this type of subjects handle a moderately similar vocabulary, the queries are more identifiable. Topic categories as *Computers* have, relatively, a specialized vocabulary, which allows to identify more accurately the related queries. Another particular case is the category *Adult*, where the users do not change the words used to make their queries. Most of these queries are built using similar words, without counting that, even though the users use different

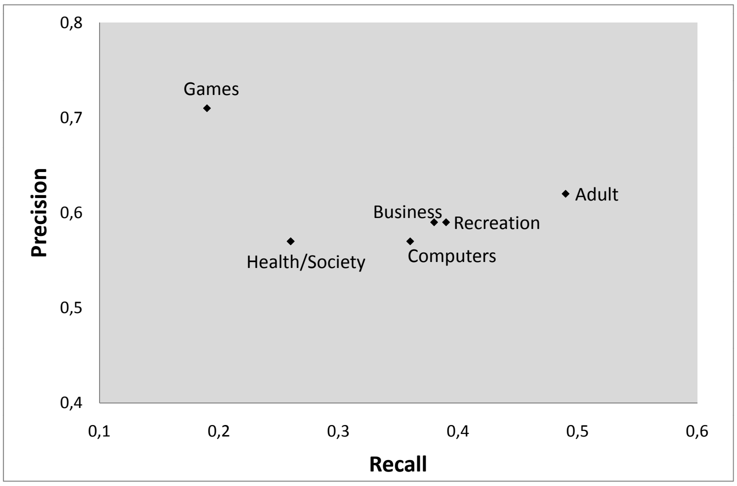


Figure 4.7: Precision vs recall of automatic topic category prediction

words to describe their queries, the pages which they choose as answer do not change the vocabulary, so the queries are represented from the same words again.

When we analyze the relationship between query intents and topic categories (see Table 4.4), we can observe the coherence that exists between the informational objectives of the users and the topics in which their queries are located. For the *Informational* intent, the greater distribution of queries are categories like *Business, Education, Society* and *News*. Whereas the categories with smaller concentration of queries in this intent are: *Adult, Other, Shopping* and *Games*. These last categories, suggest different motivations for the user that are not related to obtain information, for example *Games*, where the users are more interested in downloading software and resources. A particular case is the category *News*, that concentrates the 100% of its queries in the *Informational* intent.

Similarly, the *Not Informational* intent shows enough coherence with its related topic categories. In this case, the categories with greater con-

Topic	**Info.**	**Not Info.**	**Amb.**
Adult	9%	**84%**	7%
Arts	**86%**	9%	5%
Business	**85%**	8%	7%
Computers	**48%**	**38%**	14%
Education	**93%**	6%	1%
Games	29%	**57%**	14%
Health	**75%**	7%	18%
Home	**36%**	24%	**40%**
News	**100%**	0%	0%
Other	0%	**33%**	**67%**
Recreation	**72%**	20%	7%
Reference	**44%**	**47%**	9%
Science	**78%**	8%	15%
Shopping	21%	**37%**	**42%**
Society	**89%**	2%	9%
Sports	**86%**	7%	7%
Various	**41%**	10%	**49%**
World	**88%**	8%	4%
Average	60%	23%	17%
Sd Dev	32%	23%	19%

Table 4.4: Distribution of automatic prediction results into intents and topic categories.

centration of queries are: *Adult, Games, Reference* and *Computers*. Those with smaller concentration of queries are: *Society, Education, Health, Science* and *News* (as there are not queries of this topic in this type of intent).

Finally, for the *Ambiguous* query intent, the greater concentration of queries is in the topics *Other* and *Various*, which is logical, given that the nature of these topics is also ambiguous. The categories with smaller concentration of queries in this intent are: *News* (0%), *Education* (1%), *World* (4%), *Arts* (5%) and *Business*, *Recreation* and *Adult* (each one with 7%). These topic categories, that exhibit a very low proportion of queries with *Ambiguous* intent and a clear association with one of the other two

	Supervised			Semi–supervised			Unsupervised		
Class	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.
Informational	0.646	0.985	0.780	**0.789**	0.771	0.785	0.626	0.723	0.671
Not Inf.	**0.843**	0.219	0.348	0.556	0.605	0.579	0.346	0.421	0.380
Ambiguous	**0.478**	0.021	0.040	0.438	0.423	0.461	0.476	0.277	0.350

Table 4.5: Comparison of performance of automatic prediction for the query intent.

query intents (*Informational*, *Not Informational*) are an important finding. It suggest that queries associated with these topics are not ambiguous and could be managed in a different and more precise way than queries from those topics that show high levels of association with the *Ambiguous* intent, like *Home* (40%).

4.5 Comparison with other Approaches

In order to contrast the results of the automatic classification of queries that we obtained with the supervised classifier with the results of other methods, we compare these results against the results obtained with two machine learning techniques: semi-supervised and unsupervised learning.

Semi–supervised learning attempts to combine features of both supervised and unsupervised learning. The classification depends on both labeled training examples and a similarity based clustering involving both labeled and unlabeled data points. The algorithms used for this comparison are improved semi–supervised linear support vector classifiers [86]. We used Probabilistic Latent Semantic Analysis (PLSA) [43] as the unsupervised learning technique in the comparison.

In the Table 4.5 we can see the automatic classification results for the query intent. As we can observe, the SVM classifier obtained the highest values of precision for *Not Informational* and *Ambiguous* queries, while the unsupervised learning technique obtained the lowest values of precision for the three types of intent. Semi-supervised learning obtained the best recall values. Overall, both supervised and semi-supervised learning

	Supervised			Semi–supervised			Unsupervised		
Category	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.
News	**1**	0.07	0.13	0.6	0.32	0.41	0	0	0
Sports	**1**	0.09	0.17	0.64	0.19	0.29	0	0	0
Computers	**0.88**	0.03	0.06	0.52	0.63	0.57	0.13	0.41	0.2
Reference	**0.8**	0.03	0.07	0.44	0.37	0.4	0	0	0
Education	**0.8**	0.13	0.23	0.54	0.58	0.56	0.24	0.21	0.22
Society	**0.76**	0.12	0.2	0.6	0.58	0.59	0.25	0.48	0.33
Sex	**0.67**	0.34	0.45	0.54	0.75	0.63	0.45	0.44	0.44
Health	0.62	0.04	0.07	**0.82**	0.64	0.72	0.22	0.17	0.19
Business	0.46	0.54	0.5	0.55	0.75	0.64	**0.65**	0.26	0.37
Recreation	0.28	0.90	0.43	**0.58**	0.72	0.64	0.45	0.35	0.39

Table 4.6: Comparison of performance of automatic prediction for the topic categories.

obtain better results than unsupervised learning. This last one technique is more useful to discover new relations between the data but is not as good as supervised and semi-supervised learning techniques in the task of classifying queries into intents.

In the Table 4.6 we show the automatic classification results obtained by the three techniques for the most representative topics. As we can observe, the SVM classifier obtained the best precision values, while semi-supervised learning has a good balance between the precision and recall values. The precision results of unsupervised learning are not high, however, the values of the recall are not bad.

In general, supervised learning obtains good precision values contrasted with the other two techniques. However, it is important to improve the recall values.

4.6 Summary

In this chapter we have presented a study of the feasibility of the query intent prediction. An analysis was made from two perspectives: the user's query intent and the topic categories in which the queries within these intents can be located. We used a taxonomy of three query intents, *Informational, Not Informational* and *Ambiguous* and a group of eighteen topic categories. A query log from a commercial search engine was labeled into query intents and topic categories by a group of experts. We used this labeled data set to train a supervised learning model and evaluate the effectiveness of the automatic query intent prediction. Support Vector Machine classifier was used to accomplish the prediction task.

The results of the automatic query intent prediction show that about 60% of the queries we evaluated have predictable query intents. The user interests were identifiable through a particular representation of queries, together with the automatic classifier. This was a good combination, since representing the queries by the terms of the documents that gave good answers to them, reduces the problem of the low number of words that the users use to make their queries (and hence the sparsity of the query space), and additionally because the pages that belong to a same category share a similar vocabulary that allowed us to make a better classification.

From the automatic prediction results we observe that a large proportion of queries can be associated with a particular query intent, most of them to the *Informational* query intent. Although it is easier to predict the *Informational* intent, it is more difficult to predict the information topic. We can observe associations between some topic categories and query intents. We identify groups of topic categories with more than 80% of their queries in a specific query intent. This findings are important, because identify the intent behind the query is not a trivial task, but if we know that some topics are directly related with a specific query intent, the identification of this query intent could be improved substantially.

Although we were able to accurately predict the intent of most of the queries, which was the objective of this initial study, there is still an important work to do with the *Ambiguous* queries. In this study we gave a first step, most of the previous work have recognized the existence of *Ambiguous* queries, but few of them have reported concrete information about these queries, like which percentage of the total of queries are *Ambiguous* or which is the predictability of these queries. Our automatic prediction results for *Ambiguous* queries, show that, despite we can even determine associations with some particular topics —*Other* and *Various*—, these topics have an ambiguous nature too.

Automatically learning to predict user interests requires a deeper understanding of the user query intent. Terms like *Ambiguous*, *Other* or *Various*, suggest more complex interactions with the information, that may require a richest modeling of the intent of the user query. Hence, in the next chapter, we introduce a study of several facets that could improve the understanding of the intent of user queries. At the time when the results of this study were published [5], they were the state of the art on this problem.

Chapter 5

FACETED QUERY INTENT MODEL

As we saw in the past chapter, users who search for information on the Web may have very different intents: shopping, researching on a topic, planning a trip, looking for a quick reference, and so on. Characterizing and identifying the user's query intent is one of the most important challenges of modern information-retrieval systems. Most of the studies characterizing the user's intent have conceived this intent as a unique dimension. However, such a simple model is often inadequate for capturing the complexity of information seeking in the real world (e.g. ambiguous queries), and search engines need a better understanding of user intent and its multi-dimensional nature.

In this context, a query can be considered as the *tip of the iceberg* (Figure 5.1), which only shows –in a direct way– a small part of its content, while hiding an important proportion of its potential information.

In order to improve the quality of results, while increasing the users satisfaction, Web IR systems current efforts are focused on representing the information needs of its users in a way that not only involves the query but to get as much information related to this need, in an automatic way, transparent to the user and as accurately as possible. From this arises the need to determine the intents that lead a user to perform a search on

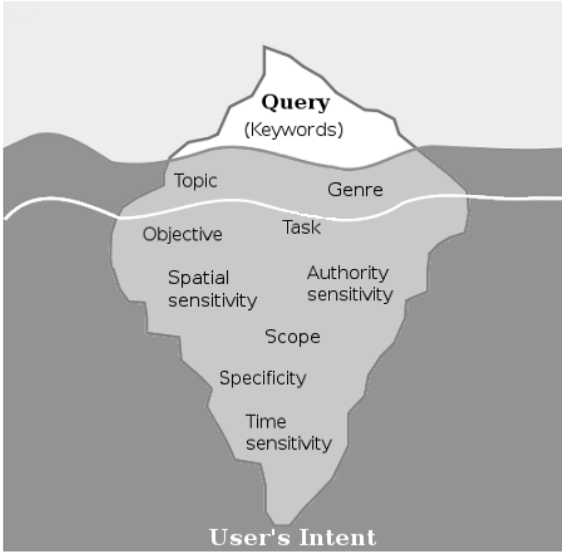


Figure 5.1: Facets of user's intent

the Web. Search Engines require the use of different strategies to meet the user's intents. For this, it is necessary to combine several factors in order to gain insight into the real goal of users' search. For example, is the user searching for a document to read it? Or does he/she want to perform a transaction such as buying a product? Is it important that the retrieved information comes from trusted Web sites? –Or the source of the information is not a relevant feature, as long as the content is good?– How important is the information recency? What kind of resource is the user searching for? What the user wants to do with such resource? – download it, read it online, or just find a reference. These questions help us to build a picture of the multi–dimensional nature of the user's intent, and show why is it necessary to place the user's requirements in a wider context. Hence, in this chapter we study, analyze, and characterize a wide range of facets involved in user's intentions. We outline the relationships that exist among these facets as well as their contribution in the task of recognizing accurately the user intention in Web search.

The selection of the facets was done by means of the observation of a representative set of queries, the characteristics of such queries and the main features that are noticeable in the whole set, or in subsets, of queries. Hence, although this is a high number of facets, in this work we wish to ascertain which of them can be considered to be good –or effective– descriptors of the user's information needs, and using this information to filter out the dimensions leaving the most representative ones. The studied facets are: genre, objective, specificity, scope, topic, task, authority sensitivity, spatial sensitivity and time sensitivity. This research introduces the first four facets, and analyze them together with the others that were proposed by different authors [5, 14, 72].

The reminder of this chapter is organized as follows: in Section 5.1 we describe the facets we are analyzing and describe some possible applications of them from a search engine's perspective. Section 5.2 describes the data we used and the semi-manual classification of queries. In Section 5.3 we delineate the correlations among the facets and present an analysis of the combination of the most correlated facets. Finally, in Section 5.4 we present a discussion of our findings and the conclusions of the study.

5.1 Facets for the User Intent

In this study we analyze a set of facets that can be extracted from a user's query and that may be considered in order to determine the user's intents. Each of the facets covers a specific part of the user need, being seen as a compound of a wide spectrum of facets. Below, we define each facet and the possible values they can take.

Genre {*News, Business, Reference, Community*} This is the broadest facet from the considered set. Genre provides a generic context to the user's query intent, and can be thought as a meta-facet. We say a query is considered as *News* when the user interest is related to current events or people (e.g., "Palestinian conflict", "catholic university hunger's strike"). Queries marked as *Business* are those that reflect a need of making a transaction such as a purchase, sale, or lease of a good or service (e.g., “sale of furniture”, “Internet housing”). *Reference* queries are those related to seeking information about a data item, a name, a location, or a resource, with the purpose to get the picture of it (e.g., “history of the Chilean air force”, “contraceptive method”). *Community* queries are those related to society, lifestyle, celebrities, Web social groups and interaction tools (e.g., “video chat”, “Ricky Martin”).

Topic {*Adult, Arts & Culture, Beauty & Style, Cars & Transportation, Charity, Computers & Internet, Education, Entertainment, Music & Games, Finance, Food & Drink, Health, Home & Garden, Industrial Goods & Services, Politics & Government, Religion & belief systems, Science & Mathematics, Social Science, Sports, Technology & Electronic, Travel, Undefined, Work*}. This facet was taken from our prior work (see Chapter 4). Based on those results, we introduced some changes to this facet. First we removed some topics that were not representative or that seemed to be part of other more general topics (i.e. *Shopping, Games, World*). Then, in order to have a more comprehensive representation of the topical associations of the queries, we compound a list of topics based on some well known taxonomies: ODP[1], Yahoo![2] and Wikipedia[3]. We also removed the ambiguous topics *Other* and *Various*. Finally, as a result of the analysis of our initial findings, we took out from the list of topic categories the next items: *Business, News, Reference and Society*. We consider these categories are not simple topics, instead of that, we consider these categories are meta-topics that can be related with all the primary topics. With these four topics we created the meta-facet genre.

Task {*Informational, Not Informational, Ambiguous*} This facet was also taken from our prior work. We consider task as the primary

[1] www.dmoz.org/

[2] www.yahoo.com

[3] en.wikipedia.org/

need that is reflected by the query. Queries marked as *Informational* reflect an interest to obtain information from the Web, irrespectively from the knowledge area of the resource to be retrieved (e.g., “what is thermodynamic”, “article 150 of the civil code”). Queries marked as *Not Informational* reflect an interest to obtaining resources, or target a specific transaction, such as buy, download, or booking (e.g., “web hosting free”, “notebook sales”, “online TV”). Queries marked as *Ambiguous* reflect queries which their task can not be determined and may be either obtaining information or performing an activity (e.g., “airbrush”, “peeling”). To determine the correct intent for this kind of queries we need to know the context of it.

Objective {*Resource, Action*}. Represents if the query is aimed to do some action or to obtain a Resource. We understand as *Resource* any information represented in formats such as PDF documents, videos or slides (e.g., “street map of Santiago”, “Pablo Neruda poems”), otherwise it may be to perform an *Action* such as filling a form or subscribing to a mailing list (e.g., “rent of apartments”, “buy in the supermarket by internet”).

Specificity {*Specific, Medium, Broad*}. This facet describes how specialized is a query. If the query contains a name, date, place, an acronym, or even an URL, such query is regarded as *Specific* (e.g., “university of Santiago de Chile”, “windows media player 6.4”); if the terms of the query are more general, then it is evaluated as *Medium* (e.g., “private universities”, “agricultural machinery”). Finally, if the query is represented as a very general term, it is regarded as *Broad* (e.g., “universities”, “maps”).

Scope {*Yes, No*}. The scope aims at capturing whether the query contains polysemic words or not. A positive scope will hint to the possibility that the user is looking for a certain information issuing a query whose words have different meanings (e.g., “wood”, “dove”).

Authority Sensitivity $\{$*Yes, No*$\}$. Through this facet it is possible to determining whether the query is designed to retrieve authoritative and trusted answers [72] (e.g., “clinical hospital of the university of Chile”, “hotmail.com”). According to Broder’s taxonomy, *navigational* queries are sensitive to authority.

Spatial Sensitivity $\{$*Yes, No*$\}$. This facet reflects the interest of the user to get a resource related to an explicit spatial location, such a city or a place (e.g., “hotels in Valparaiso”), or to find a location that is not mentioned, but that might be inferred from the query terms [72](e.g., “nearest hotel”, “restaurant”).

Time Sensitivity $\{$*Yes, No*$\}$. This facet captures the fact that some queries require different results when posed at different times [72], such as “US president” vs. “US president 2009”.

The reader may wonder why these facets are useful. Although would be impossible to claim facet completeness, how could a search engine take advantage of the knowledge of these facets? Below, we delineate some useful applications of these facets, and the way a search engine can benefit from them.

Genre. This facet clearly delimits the search engines area of search, and is expected to make the search faster and more accurate. For example if the genre is *Community*, the pages that have to be looked at are blogs, forums, chats, and social networks, such as Facebook.

Topic. Although this is one of the first features that are considered by the search engines, they should ideally present the answers to the user organized by topic.

Task. If the query is *Informational*, then Web resources to be recommended should avoid those with transactional content, or the ones where the user has to assume an active role of interaction, such as filling a form.

Objective. If the user wants to perform an *Action*, then he/she may also be interested in commercial sites that offer a product or service, hence appropriate ads may be presented to the user.

Specificity. The search engines can help the user by presenting many similar pages, for example on the same topic, or a broad variety of pages.

Scope. The answers presented to the user by the search engine should ideally be presented by topic, and trying to cover most of the topics related to the polysemic words, allowing the user to select the right one.

Authority Sensitivity. For this dimension the search engines have the opportunity to be very selective regarding the authority of the web pages presented to the user. For example, it is critical to trust the pages from where to download antivirus software, and it is essential to reach the right page for all the navigational queries.

Spatial Sensitivity. Although a query does not mention a place, search engines should be able to identify spatial sensitivity to offer information related to the local area from which the query was submitted, emphasizing in queries that reflect a commercial need. For example: “school”, “tai chi classes”, “cars selling”.

Time Sensitivity. Search engines can take advantage of the time sensitivity of queries to suggest related information that occurred in the same period as what it is being sought for. Further more, this information can be used to recommend resources following a temporal order, as it is done with news.

5.2 Semi-manual Classification of Queries

To classify queries into facets, we follow the same semi-manual classification method that we introduced in the Chapter 4 (see Section 4.2.2). We describe below, the new data-set that used for this study and an improved version of the classification tool that we created to facilitate the task of the annotators.

5.2.1 Query Sampling

We processed and then manually classified a sample of 5,249 queries from a TodoCL search engine query log. For this study we used a different sample of queries to the one used in Chapter 4. The queries were randomly selected in a way that allowed us to work with queries with different level of popularity according to the Zipf's law distribution. The selection process was the following: 15% were taken from the set of the most popular queries, 15% from the (long) tail set, and the remaining 70% was sampled from the middle set (i.e., queries with average popularity).

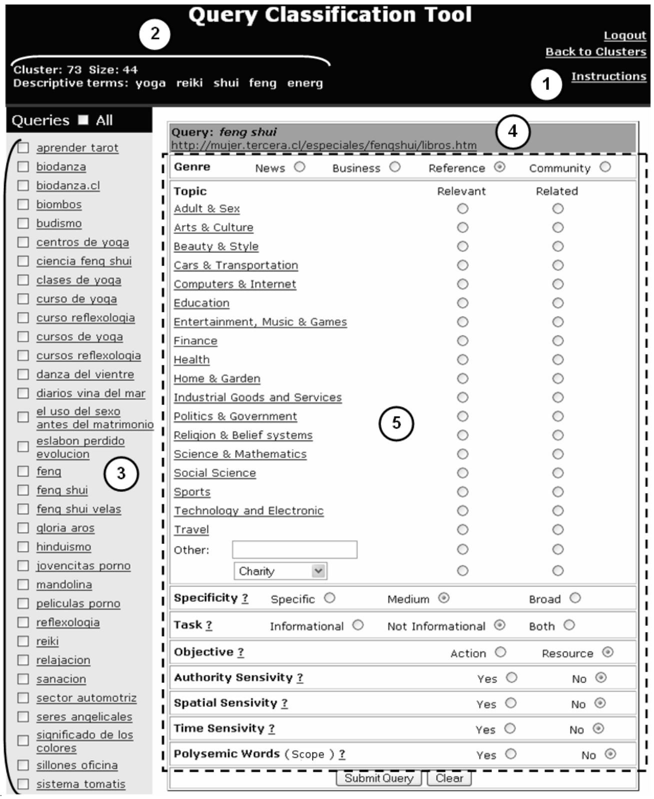
In the same way as in Chapter 4, each query was represented as a term-weight vector compounded by the terms appearing in the selected Web pages for such query.

5.2.2 Query Classification Tool

To facilitate the task of manual classification, we created a classification tool that provided to the human annotators with an interface that contains all the information about the facets, necessary to classify the queries. Figure 5.2 is a screen shot of the software tool that was developed for the classification task. The main screen of the tool was divided in five areas (enumerated in a circle), where each one contains valuable information to the classifiers. The five areas are:

1. Instructions. Contains information about the classification process as well as some illustrative examples.

2. Information about the cluster. Shows the number of queries grouped in a cluster, as well as the descriptive terms of such a cluster. With this information the judge can make a picture of the subject of the queries contained in this group. In Figure 5.2 the cluster 73, contains



Cluster: 73 Size: 44
Descriptive terms: yoga reiki shui feng energ

Queries ■ All

- □ aprender tarot
- □ biodanza
- □ biodanza.cl
- □ biombos
- □ budismo
- □ centros de yoga
- □ ciencia feng shui
- □ clases de yoga
- □ curso de yoga
- □ curso reflexologia
- □ cursos de yoga
- □ cursos reflexologia
- □ danza del vientre
- □ diarios vina del mar
- □ el uso del sexo antes del matrimonio
- □ eslabon perdido evolucion
- □ feng
- □ feng shui
- □ feng shui velas
- □ gloria aros
- □ hinduismo
- □ jovencitas porno
- □ mandolina
- □ peliculas porno
- □ reflexologia
- □ reiki
- □ relajacion
- □ sanacion
- □ sector automotriz
- □ seres angelicales
- □ significado de los colores
- □ sillones oficina
- □ sistema tomatis

Query: *feng shui*
http://mujer.tercera.cl/especiales/fengshui/libros.htm

Genre News ○ Business ○ Reference ● Community ○

Topic	Relevant	Related
Adult & Sex	○	○
Arts & Culture	○	○
Beauty & Style	○	○
Cars & Transportation	○	○
Computers & Internet	○	○
Education	○	○
Entertainment, Music & Games	○	○
Finance	○	○
Health	○	○
Home & Garden	○	○
Industrial Goods and Services	○	○
Politics & Government	○	○
Religion & Belief systems	○	○
Science & Mathematics	○	○
Social Science	○	○
Sports	○	○
Technology and Electronic	○	○
Travel	○	○
Other:	○	○
Charity	○	○

Specificity ? Specific ○ Medium ● Broad ○

Task ? Informational ○ Not Informational ● Both ○

Objective ? Action ○ Resource ●

Authority Sensivity ? Yes ○ No ●

Spatial Sensivity ? Yes ○ No ●

Time Sensivity ? Yes ○ No ●

Polysemic Words (Scope) **?** Yes ○ No ●

Figure 5.2: Query classification tool

44 queries, and its descriptive terms are: “yoga”, “reiki”, “shui”, “feng”, “energ”.

3. Queries in the cluster. Contains the set of queries that belong to the cluster that is being classified. This scheme allowed the annotators to contextualize a query through finding sets of related queries, as well as to overcome the problem of having to infer the meaning of queries only from their terms.

4. Query to classify. In this area appears the query that is being classified together with the URL of the most popular Web page that was clicked as answer for such query. In Figure 5.2 the current query

Genre	Query Count	%
Business	1381	29.2%
Community	2351	49.7%
News	122	2.6%
Reference	872	18.5
Total	4726	100%

Table 5.1: Distribution of queries into the facet genre.

is “feng shui”, and its related URL is `http://mujer.tercera.cl/especiales/fengshui/libros.htm`.

5. Facets. The facets and their values are contained in this area. If the judge has a doubt about the meaning or content of a facet, he/she can find related information by clicking on the interrogation mark that appears next to the name of the dimension. For example, by clicking on the topic=*Education* a list with the subtopics related to *Education* appears (i.e., Higher Education, Teaching, Homework Help, Preschool, among others).

In spite that using the classification tool is very intuitive, and that it contains basic information about its usage, the annotators were trained about how to perform the classification of the queries, as well as the meaning and content of each facet.

5.2.3 Distribution of Queries into Facets

From the total of 5,249 processed queries, a set of 4,726 queries was manually classified using the classification tool described in Section 5.2.2. In Tables 5.1 to 5.6 we report the faceted classification results. From the manual classification results we have a first picture of the influence of the facets in the intents of the users.

For the facet genre (Table 5.1), we observe that most of the queries are related with *Community* and *Business* and to a lesser extent with *Reference* and *News*. On the other hand, the facet objective (Table 5.2) shows that most of users are involved with queries oriented to obtain/find some

Objective	**Query Count**	**%**
Action	1147	24.3%
Resource	3579	75.7%
Total	4726	100%

Table 5.2: Distribution of queries into the facet objective.

Specificity	**Query Count**	**%**
Broad	218	4.6%
Medium	3716	78.6%
Specific	792	16.8
Total	4726	100%

Table 5.3: Distribution of queries into the facet specificity.

Facet	**Yes**		**Not**	
	Query Count	%	Query Count	%
Authority Sen.	886	18.7%	3840	81.3%
Scope	113	2.4%	4613	97.6%
Spatial Sen.	1998	42.3%	2728	57.7%
Time Sen.	71	1.5%	4655	98.5%

Table 5.4: Distribution of queries into facets: authority sen., scope, spatial sen. and time sen.

kind of *Resource* more than with performing some *Action*. In the case of specificity (Table 5.3), global results indicates that queries have a moderate degree of specificity, being *Medium* the specificity degree that concentrates the highest number of queries. *Broad* queries are only 4.6% of the total of queries, indicating that in most of the cases, the terms of the queries describe moderately specific purposes.

Among the facets measuring some kind of sensitivity (see Table 5.4), spatial sensitivity reports the highest number of queries (*Yes*=42.3%). That is, annotators were able to identify clearly when a query is spatially sensitive. This has potentially huge implications for searchers. Identi-

Task	Query Count	%
Informational	2822	59.7%
Not Informational	1515	32.1%
Ambiguous	389	8.2%
Total	4726	100%

Table 5.5: Distribution of queries into the facet task.

fying spatially sensitive queries would allow search engines return better results considering user location or geographic features and it is also of great commercial value as it enables location specific advertising and improved search for local businesses. After spatial sensitivity, authority sensitivity is the facet that collects more sensitive queries (*Yes*=18.3%). Although the percentage of authority-sensitive queries is not very high, it is important to be able to recognize this kind of queries because they require trusted information. The intent of users with authority-sensitive queries can not be satisfied with not authoritative answers. In the last place of the three sensitivity facets, we find time sensitivity. The distribution of queries for this facet indicates that there is not a high percentage of time-sensitive queries. This could be related to the sample of queries we are analyzing, but it is also possible that this facet is not a representative feature for queries. We will see more information about this facet in next sections.

For the facet scope we observe a big concentration of queries in one of its two possible values (see Table 5.4). Most of the queries were assigned to the group without polysemic words, that is, the terms of the queries tend to have defined meanings. This is consistent with the results reported for specificity where the proportion of broad queries is low.

The distribution of queries for topic varies with respect to the prior work, presented in the Chapter 4. The distribution of queries along all the topics is more balanced (see Table 5.6). The maximum percentage of queries that was concentrated by a single topic was 9.9% (as a reference, in the previous work just for Business, we had 19.2% of queries). We attribute this more balanced distribution of queries to two reasons. First,

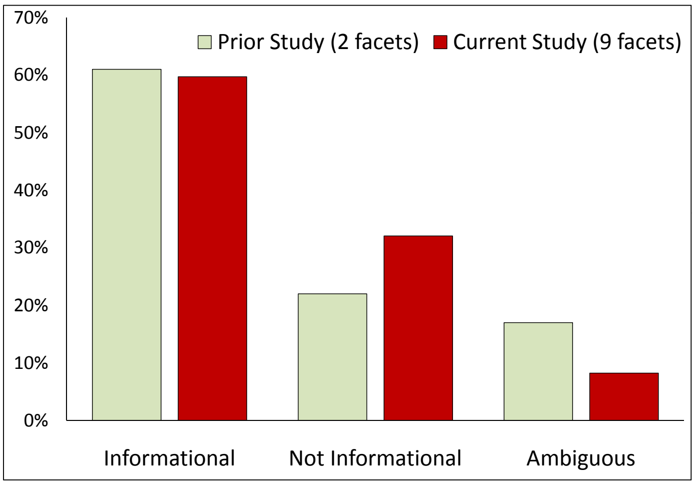


Figure 5.3: Comparison of manual classification results for query intent/task from our prior study and the current study.

we considered a more comprehensive list of topics categories, with not too general topics, that tend to accumulate a major quantity of queries with very different characteristics (high diversity). For example, we replaced the topic *Recreation* (23.5% of queries in previous study) by categories like *Entertainment, Music & Games* and *Travel*, which are more specific topics. In this way we have a more fine grained classification of the queries. Another reason for the balanced distribution of the queries is that we took out the meta-topics *Business, News, Reference and Society* and leave only primary topics. With this we reduced the ambiguity of the classification of queries into topics. A meta-topic like *Business* generates a sense of duality in the queries, for example the query '*sales classic cars*', people tend to classify it as Business, but with our new topic scheme this query was classified as follows: genre=*Business* and topic=*Cars & Transportation*.

We also got interesting results for the facet task (see Table 5.5). From the distribution of queries we observe a low proportion of *Ambiguous*

Topic	**Query Count**	**%**
Adult	177	3.7%
Arts & Culture	264	5.6%
Beauty & Style	45	1.0%
Cars & Transportation	321	6.8%
Computers & Internet	292	6.2%
Education	279	5.9%
Entertainment, Music & Games	392	8.3%
Finance	461	9.8%
Food & Drink	227	4.8%
Health	373	7.9%
Home & Garden	295	6.2%
Industrial Goods & Services	37	0.8%
Politics & Government	442	9.4%
Religion & Belief Systems	82	1.7%
Science & Mathematics	466	9.9%
Social Science	176	3.7%
Sports	25	0.5%
Technology & Electronic	32	0.7%
Travel	314	6.6%
Work	26	0.6%
Total	4726	100%

Table 5.6: Distribution of queries into the facet topic.

queries (only 8.2%). If we compare the manual classification results of the prior study with the results of the current study (see Figure 5.3), we can observe that we have reduced the proportion of *Ambiguous* queries in more than 50%. The proportion of *Informational* queries remains stable while the proportion of *Not Informational* queries has increased. As we analyzed in the past chapter (see Section 4.3.3), there is agreement in the definition of Informational queries, but not with Not Informational queries. As it was shown in Figure 5.2, the task of the annotators was to assign a value for each of the nine facets to each query at the same time, that is, the annotators had a complete picture of the query, with values for

Facet	**Overall Agreement**	**Kappa Coefficient**
Time Sen.	99.23	0.98
Scope	96.74	0.93
Objective	92.54	0.85
Authority Sen.	84.32	0.69
Topic	68.26	0.66
Task	75.71	0.63
Spatial Sen.	81.07	0.62
Genre	65.00	0.53
Specificity	55.44	0.33

Table 5.7: Inter-annotation Agreement and Kappa Coefficient for the facets

all the facets at the moment of classify it. Over the base of all the facets, the annotators had more elements to place the queries in a defined task category. From global results we see that this fact was useful to disambiguate *Not Informational* queries (since the proportion of *Informational* queries remains stable). Overall, we appreciate that the model of nine facets outperform the model of two facets, in terms of reduction of ambiguity. By combining the nine facets, annotators were able to identify a defined *task* for 91.8% of the queries.

With the purpose to determine the reliability and consistency of the manual classification, 10% of the 5249 processed queries (i.e., 523 queries) were selected to be classified by two annotators. Table 5.7 shows the overall agreement of judgements given by the experts, as well as the Kappa values for each facet. Results from the overall agreement indicate that the consistency of the manual classification is highly satisfactory. In average, the overall agreement is 80%. Eight out to the nine dimensions have reached an overall agreement higher than 65%, which is quite high if we consider the number of dimensions that were assessed, the number of possible values that each dimension can take, as well as the different criteria and the subjectivity of the judges.

5.3 Correlation among Dimensions

In order to establish the relations among the dimensions, we calculated the correlation coefficient among all of them. Further more, aiming to confirm the strength of the correlation that exist among the dimensions, we have divided the dataset in to three subsets, and calculated the correlation coefficient in each of them. The idea behind this schema is that if there is a correlation between a pair of dimensions, this correlation should be observable throughout different subsets. The subsets are the following: **s1**, queries classified by one judge –this set has 4,726 queries– (Table 5.8); **s2**, queries classified by two annotators –this set has 523 queries– (Table 5.9); and, **s3**, 578 queries from the set of the most popular ones according to the Zipf's Law distribution (Table 5.10).

The results show that, in general, genre and objective are the dimensions that are more highly correlated. From the three subsets of queries (*s1*, *s2*, *s3*) the average correlation is 0.47. The facets genre and topic also have positive correlations among the three subsets of data. The facet scope is related with the facet specificity, the trend is that a query tends to contain polysemic words if its specificity is broad. Typically, a query with low level of specificity is likely to be composed by general terms, i.e., it is likely that the query terms have more than one meaning (polysemy). Reversely, a query with high level of specificity is likely to be composed of terms with no more than one meaning. This fact is confirmed by the distribution of the queries for the specificity dimension; when scope=*Yes*, less than 1% of the queries are specificity=*Specific* while 38% of them are specificity=*Broad*. Some examples of queries with this configuration (i.e., specificity=*Broad*, scope=*Yes*) are: '*flat*', '*code*' and '*pascal*'.

In the remaining of this section we discuss with more detail the facets genre and topic, as they are the most representative facets. Since these facets have important associations with all the other facets, their analysis gives us an overall view of the behavior of the complete system of facets.

Facet	Aut. Sen.	Genre	Object.	Scope	Spat. S.	Specific.	Task	Time S.
Aut. Sen.	–	–	–	–	–	–	–	–
Genre	0.043	–	–	–	–	–	–	–
Object.	0.108	0.466	–	–	–	–	–	–
Scope	0.104	-0.085	-0.042	–	–	–	–	–
Spat. Sen.	-0.045	-0.11	0.066	-0.151	–	–	–	–
Specific.	-0.242	-0.007	-0.078	0.028	-0.207	–	–	–
Task	-0.043	-0.23	-0.357	-0.004	-0.23	-0.039	–	–
Time Sen.	-0.017	-0.082	-0.027	-0.008	-0.023	0	-0.022	–
Topic	0.018	0.288	0.111	0.025	-0.296	0.024	-0.177	0.013

Table 5.8: Correlation coefficient between the facets - Subset **s1**.

Facet	Aut. Sen.	Genre	Object.	Scope	Spat. S.	Specific.	Task	Time S.
Aut. Sen.	–	–	–	–	–	–	–	–
Genre	-0.13	–	–	–	–	–	–	–
Object.	-0.165	0.435	–	–	–	–	–	–
Scope	0.009	-0.078	-0.055	–	–	–	–	–
Spatial Sen.	-0.171	0.043	0.199	-0.076	–	–	–	–
Specificity	-0.462	0.009	-0.089	0.16	-0.043	–	–	–
Task	0.083	-0.358	-0.464	0.041	-0.17	-0.064	–	–
Time Sen.	-0.003	-0.075	-0.031	-0.01	0.061	-0.03	0.0293	–
Topic	0.165	0.181	-0.039	-0.076	-0.128	-0.119	0.011	-0.071

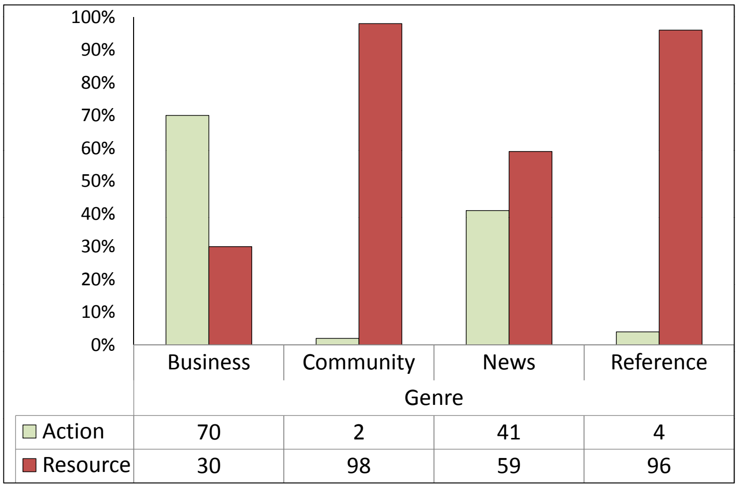
Table 5.9: Correlation coefficient between the facets - Subset **s2**.

Facet	Aut. Sen.	Genre	Object.	Scope	Spat. S.	Specific.	Task	Time S.
Aut. Sen.	–	–	–	–	–	–	–	–
Genre	0.088	–	–	–	–	–	–	–
Object.	0.115	0.473	–	–	–	–	–	–
Scope	-0.0007	-0.05	-0.049	–	–	–	–	–
Spatial Sen.	-0.024	-0.095	-0.010	-0.105	–	–	–	–
Specificity	-0.334	0.004	-0.038	0.288	-0.130	–	–	–
Task	-0.034	-0.199	-0.463	0.010	-0.013	-0.032	–	–
Time Sen.	-0.015	-0.013	-0.021	-0.019	0.035	-0.118	0.047	–
Topic	0.025	0.231	0.145	0.006	-0.211	-0.013	-0.196	-0.195

Table 5.10: Correlation coefficient between the facets - Subset **s3**.

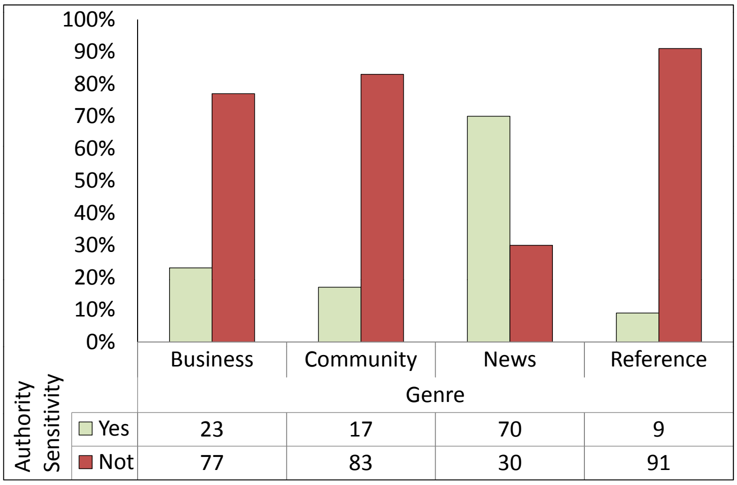
5.3.1 Genre Dimension

Genre is the most general facet from the complete set. Any of the four values that we defined for this facet, i.e., *Business, Community, News* and *Reference*, bring together queries that might belong to any topic, while are focused on performing a particular task, to reach a concrete objective, and have different features that distinguish them from the others.



	Business	Community	News	Reference
🟩 Action	70	2	41	4
🟥 Resource	30	98	59	96

Figure 5.4: Distribution of queries into the facets genre and objective.



Authority Sensitivity		Business	Community	News	Reference
	🟩 Yes	23	17	70	9
	🟥 Not	77	83	30	91

Figure 5.5: Distribution of queries into the facets genre and authority sensitivity.

As it was mentioned before, the facet with the highest correlation with genre is objective. Figure 5.4 shows the distribution of queries along these two facets. These distributions reveal specific relations. The strongest association is the one established between genre=*Community* and objective=*Resource*; genre=*Reference* and genre=*News* also exhibit clear associations with objective=*Resource*. The trend of the genres associated with the objective=*Resource* is to obtaining information instead of to performing a specific action. On the other hand, the genre=*Business* exhibits a different behavior compared to the others. *Business* is the genre with the highest number of queries with objective=*Action*. This genre concentrates the highest percentage of transactions –at commercial level–, this fact could explain the high number of queries with objective=*action*.

Genre is also correlated with the authority sensitivity dimension (ANOVA test: F=331.955, $p < 0.005$). Figure 5.5 shows the distribution of queries for genre facet with respect to the authority sensitivity. The distribution of the queries suggests a leaning of the genre=*News* towards queries that require trusted information. Queries that have genre=*News* represent a well defined interest which is, usually, associated with particular sources of information (authoritative). Some examples of queries with genre=*News* and authority sensitivity=*yes* are: '*official journal of chile*', '*job ads*', and '*Valparaiso University's radio station*'.

The other facets have shown to have a significant statistical relation with genre, proving the general character of this dimension; each of the four genres encompass activities/actions that might be performed in all dimensions. Regardless the topic, a query has one genre as well, hence, introducing this new dimension allows the assignment of the appropriate topic to a query, without neglecting the more general dimension. To understand the user intent is also important to understand topic associations. The genre facet complements topical associations with an idea about what the user wants to do in a specific topic area. For example, it is different to know that the query *'hp printers'* has topic=*Technology and Electronic*, than to know that the same query belongs to topic=*Technology and Electronic* as well as genre=*Business*. Furthermore, following with

Topic	Business	Community	News	Reference
Adult	4.0%	**76.8%**	1.1%	18.1%
Arts	10.2%	**73.9%**	1.9%	14.0%
Beauty & Style	13.3%	**68.9%**	0.0%	17.8%
Cars & Trans.	**88.5%**	1.2%	0.0%	10.3%
Computers	**83.2%**	4.5%	0.0%	12.3%
Education	0.7%	**51.3%**	0.7%	**47.3%**
Entertainment	11.2%	**68.6%**	15.6%	4.6%
Finance	**74.2%**	21.0%	2.2%	2.6%
Food & Drink	7.9%	**85.0%**	0.0%	7.0%
Health	3.5%	87.7%	0.0%	8.8%
Home & Garden	**69.8%**	16.9%	2.7%	10.5%
Indust. Goods & Serv.	**70.3%**	8.1%	0.0%	21.6%
Politics	0.5%	**92.1%**	2.0%	5.4%
Religion	0.0%	**78.0%**	1.2%	20.7%
Science & Math.	3.2%	**48.1%**	0.2%	**48.5%**
Social Science	1.7%	**75.6%**	4.5%	18.2%
Sports	4.0%	**92.0%**	0.0%	4.0%
Technology	**75.0%**	3.1%	0.0%	21.9%
Travel	**36.3%**	10.8%	0.3%	**52.5%**
Work	15.4%	15.4%	**53.8%**	15.4%
Average	28.6%	49.0%	4.3%	18.1%
Sd Dev	33.54%	34.54%	12.17%	14.80%

Table 5.11: Distribution of queries into the facets topic and genre.

the example, if we find that the facet objective=*Action* then, there is a very high probability that the user intent is to perform a commercial transaction, in the topic=*Technology and Electronic*.

5.3.2 Topic Dimension

One of the most meaningful facets is topic. By contrasting this facet with the others, we observe the trends of the users' intents. Table 5.11 shows the distribution of the queries given the topic and genre. Since

Genre	Topics with highest percentage of queries	Topics with lowest percentage of queries
Business	Cars & Transportation, Computers, Technology & Electronic, Finance, Industrial Goods and Services	Religion & Belief systems, Politics & Government, Education
Community	Politics & Government, Sports, Health, Food & Drink, Religion & Belief systems	Cars & Transportation, Technology & Electronic, Computers & Internet
News	Work, Entertainment, Music & Games, Social Science	Beauty & Style, Cars & Trans., Computers, Food & Drink, Industrial Goods. (0.0%)
Reference	Travel, Science & Mathematics, Education, Industrial Goods, Religion & Belief systems	Finance, Sports, Entertainment

Table 5.12: Most important topics and less important topics for each genre.

the number of queries of genre=*News* is very low, the majority of queries are distributed among the other three genres: *Business, Community* and *Reference*. In this table we can see that some topics clearly belong to a particular genre, 13 of the 20 topic categories concentrated more than 70% of their queries in a specific genre.

Table 5.12 shows the most important topics and the less important topics (according to the accumulated percentage of queries) for each genre. While the number of queries grouped in the genre=*News* is not so high, topics with greatest concentration of queries have characteristics that are very representative of this genre. For example, some of the most important topics for the genre=*News* are *Work* and *Entertainment, Music, & Games*, which are topics that are continuously changing and generating new information to a wide community.

Figure 5.6 shows the combination of the facets topic and objective. While most of the queries are oriented to finding a *Resource*, queries that are oriented to performing an *Action* are related to few, but concrete topics, this is the case of *Cars & Transportation*. This specific relation among facets is useful to define the characteristics of each topic, and

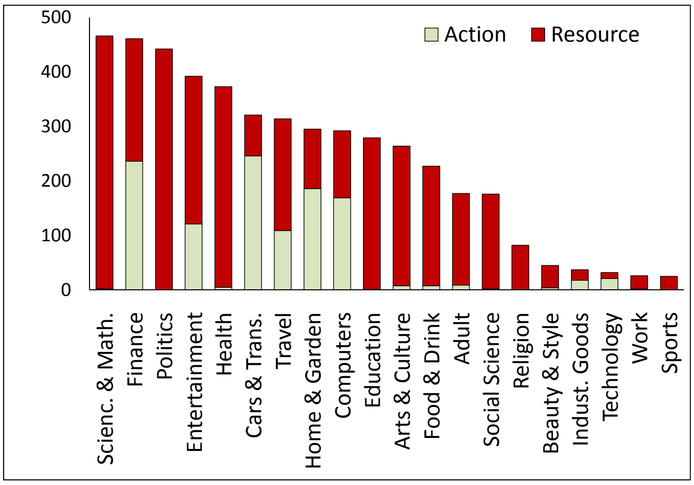


Figure 5.6: Distribution of queries into the facets topic and objective.

shows that depending on such characteristics, people can be interested in pages with different objectives. In addition to understanding the intent of the user query, it is important to provide insight into the types of result (Topic and Objective –*Action/Resource*–) that are associated with that intent.

5.4 Summary

Web Search Engines tend to view the query formulation and the retrieval process as a simple goal-focused activity. However, the intentions behind Web queries are more complex and the search process should be more oriented to a variety of characteristics than only to a simple goal. In this chapter we have introduced, analyzed and characterized a wide range of factors or facets that may be useful for user's intent identification when searching for information on the Web.

We have described the main features of each facet, its usefulness, and analyzed some relationships among them. From the analysis of the manual classification of queries we found that, to certain level, most facets plays an important role in the user's intent identification. We have confirmed the benefit to separate specific topics, from those that have a more general purpose. With the general purpose facets we can consider that genre is a meta–dimension. The use of this meta–dimension allows a fine-grained classification of queries at the topic level, as well as a more direct relation with the objective pursued by the user. For example, regardless the topic of a query, the *Business* genre, in most of the cases, is related to perform an *action*, which is one of the values of the facet objective.

In the next chapter, we use the faceted query intent model analyzed here for the automatic classification of Web queries. Through the use of the relationships between facets that we found in this work, we will apply supervised learning techniques, in order to determine the predictability of each facet. We will also evaluate the contribution of the combination of facets to the improvement of the quality of the prediction of the query intent.

Chapter 6

ADVANCED QUERY INTENT PREDICTION

As the Web continues to increase both in size and complexity, Web search is a ubiquitous service that allows users to find information, resources, and activities. However, as the Web evolves so do the needs of the users. Nowadays, users have more complex interests that go beyond to the traditional *informational queries*. For example, many users may want to perform a particular commercial transaction, locate a special service, etc. Thus, it is important for Web-search engines, not only to continue answering effectively informational and navigational queries, but also to be able to identify and provide accurate results for new types of queries.

Based on the premise above, Web-search engines try to improve the quality of their results by adopting a number of different strategies. For instance, diversification of search results aims at providing a list of diversified results that cover different interpretations of ambiguous queries [79]. The objective of diversification is to identify ambiguous queries and present the best results for each meaning of those queries. The first step towards this goal, is to identify the type of the query. In this respect, all the recent efforts to describe and identify the intent of the user's query are of great value [5, 49, 42]. Although there is a lot of work in the topic of identifying query intent, most of it is based on the analysis of only one possible facet of the query. The most common of these facets are the *topic category* and the *type of query intent*; mainly based on Broder's taxonomy [14]. However, one can argue that classifying a query with respect to one facet may improve the classification with respect to another facet. For example, knowing that a query topic is *Art* increases the prior that the query intent is *Informational*. Similarly, knowing that a query topic is *Electronics* increases the prior that the query intent is *Transactional*. Hence, we argue that identification of the query intent is a multi-faceted problem. We show that by treating the problem as such we can significantly improve the accuracy of the classification problem.

In the previous chapter we introduced and analyzed several facets that are useful in the task of identify the user's query intent. In this chapter we explore the feasibility of automatic faceted-classification. To accomplish this, we conduct several experiments. First, we automatically classify sets of queries into each facet in order to determine the individual predictability of the facets. Then we combine some of the most correlated facets and we perform multi-faceted classification. The results of these experiments as well as the discussion of the results are described in the next sections.

6.1 Predicting Individual Facets

In this section we use the labeled data-set described in Section 5.2 to train user intent prediction models. We trained a single support vector machine (SVM) classifier for each facet. The software used to implement SVM was LIBSVM [18]. We selected the one-against-one multi-class strategy given its good performance and because its training time is shorter compared with other multi-class strategies [45]. We used the radial basis function kernel for the SVM algorithm.

In order to validate and measure the performance of the prediction, we split the data into training and unseen test sets by using percentage-fractions of 50/50 and 70/30. We believe that this data split gives us an indication of how the classifiers are able to generalize over their inputs. For the experiments, three metrics were considered: the recall, the pre-

Training/Testing	50%/50%			70%/30%		
Task	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.
Informat.	0.7037	**0.9889**	**0.8223**	0.7227	**0.9915**	**0.8360**
Not Informat.	0.8408	0.2670	0.4053	**0.8917**	0.2948	0.4431
Ambiguous	**0.9167**	0.0550	0.1038	0.8571	0.0526	0.0992
Average	0.8204	0.4370	0.4438	0.8238	0.4463	0.4594

Table 6.1: Performance evaluation of automatic prediction for task.

Training/Testing	50%/50%			70%/30%		
Objective	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.
Action	**0.9451**	0.1673	0.2843	**0.9375**	0.2007	0.3306
Resource	0.8116	**0.9973**	**0.8949**	0.8235	**0.9964**	**0.9017**
Average	0.8783	0.5823	0.5896	0.8805	0.5985	0.6162

Table 6.2: Performance evaluation of automatic prediction for objective.

Training/Testing	50%/50%			70%/30%		
Genre	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.
Business	**0.9146**	0.2884	0.4386	**0.8984**	0.3159	0.4675
Community	0.5649	**0.9867**	**0.7185**	0.5765	**0.9836**	**0.7269**
Reference	0.8033	0.1038	0.1839	0.8537	0.1203	0.2108
Average	0.7609	0.4597	0.4470	0.7762	0.4733	0.4684

Table 6.3: Performance evaluation of automatic prediction for genre.

cision and the F-measure. For the recall and precision we use the same definition introduced in Chapter 4 (see Section 4.4.2). F-measure combines recall and precision with an equal weight, using the harmonic mean of precision and recall [53].

The results for the automatic prediction of the facets are reported in Tables 6.1 to 6.5. As we can see from these tables, in general we obtain good results for estimating the facets. There is not too much difference between the results obtained with the automatic classifiers based on training-sets of 50% of queries and those classifiers based on training-sets of 70% of queries. That is, the prediction's performance is good for different quantities of training data.

The best results are for the facets task and objective (see Tables 6.1 and 6.2). The average precision for these facets is 0.822 and 0.879 respectively. In the case of task, the classifier was most useful for predicting *Informational* and *Not Informational* queries, where we can observe a good balance of precision and recall (see F-measure values). For queries in the *Ambiguous* category, we maintain high precision (0.88 on average) at the expense of recall. In the case of objective, the classifier is very effective, specially to distinguish *Resource* queries (F-measure is 0.89 on average). For *Action* queries the precision is also very good (0.94 on average), although the F-measure is not as high as for *Resource* queries. In general, the automatic classification results for task and objective show the feasibility of the prediction of these facets. This is important, as being able to correctly identify the task and the objective of queries give us insight into the query intent and the type of resource associated to it. In the Figure 6.1 we show the distribution of queries along the task and the objective. As we can observe, *Informational* queries have a clear orientation towards *Resource*-objective and *Not Informational* queries are more oriented towards *Action*-objective. An interesting point is that *Ambiguous* queries are also oriented towards *Resource*-objective. This finding suggest that queries with an *Action*-objective are less ambiguous than queries with a *Resource*- objective. Since most of the queries with an *Action*-objective belong to the *Not Informational*- task we can say that the ambiguity of *Not Informational* queries is low.

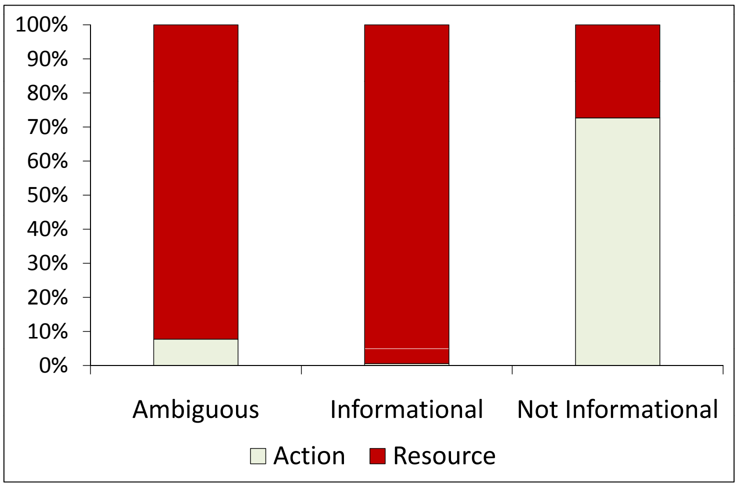


Figure 6.1: Distribution of queries into facets task and objective.

Table 6.3 shows the performance results for the facet genre. We obtain good genre identification performance for *Business*, *Community* and *Reference*, as SVM classifiers yielded good overall precision, around 0.76 (F-measure is 0.46 on average). These three genre categories are the most representative categories of the facet, they group together 97.4% of the total of queries. For the category *News* the classifier did not provide predictions, which may be caused by the small number of queries belonging to this category (56 queries in the training set, 122 in total). Apart of the *News* category, the performance of prediction for the facet genre is good. As we mentioned in the Chapter 5, genre is an important facet because it brings a general context for the intent of the user query. Overall, the automatic prediction results of this facet show that most of the queries can be correctly classified into one of the genre categories.

For the facet topic, the classifier provides predictions for fifteen of the twenty topics that were considered for the classification, as shown in Table 6.4. Overall, the precision is good, and over 0.6 for most of the topics. The best precision values are for the topics: *Adult, Cars & Trans-*

Training/Testing	50%/50%			70%/30%		
Topic	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.
Adult	0.7692	0.2000	0.3175	0.7000	0.2414	0.3590
Arts	0.5769	0.0938	0.1613	0.5556	0.1020	0.1724
Cars & Trans.	**0.9107**	0.3778	0.5340	**0.8800**	0.2857	0.4314
Computers	**0.8600**	0.3308	0.4778	**0.8261**	0.3167	0.4578
Education	0.8043	0.2483	0.3795	**0.8529**	0.3222	0.4677
Entertainment	0.1502	**0.8743**	0.2564	0.5474	0.4685	0.5049
Finance	0.6622	0.3858	0.4876	0.2149	**0.8052**	0.3393
Food & Drink	**0.8667**	0.2364	0.3714	**0.9500**	0.2923	0.4471
Health	0.6383	0.6667	0.6522	0.7347	0.4800	0.5806
Home & Garden	0.7692	0.1538	0.2564	0.5600	0.1667	0.2569
Politics	0.4483	0.4437	0.4460	0.4903	0.5549	0.5206
Religion	**1**	0.1224	0.2182	**1**	0.1515	0.2632
Science & Math.	0.4279	0.6394	0.5127	0.3846	0.6765	0.4904
Social Science	0.4375	0.0753	0.1284	0.3636	0.0690	0.1159
Travel	0.7083	0.2252	0.3417	0.7333	0.2472	0.3697
Average	0.6687	0.3382	0.3694	0.6529	0.3453	0.3851

Table 6.4: Performance evaluation of automatic prediction for topic.

Training/Testing	50%/50%			70%/30%		
Facet	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.
Authority Sen.	**0.8570**	0.5193	0.4865	**0.8441**	0.5245	0.4961
Spatial Sen.	0.7526	**0.6140**	**0.5471**	0.7618	**0.6494**	**0.5913**
Time Sen.	0.4884	0.50	0.4941	0.4873	0.50	0.4936
Scope	0.49450	0.50	0.49723	0.4958	0.50	0.4979
Specificity	0.5126	0.3412	0.3101	0.5060	0.3444	0.3160

Table 6.5: Performance evaluation of automatic prediction for authority sensitivity, spatial sensitivity, time sensitivity, scope and specificity.

portation, Computers & Internet, Education, Food & Drink and *Health*. These topics group an important number of queries and have interesting connections with the other facets (see Figure 6.2). The topics for which

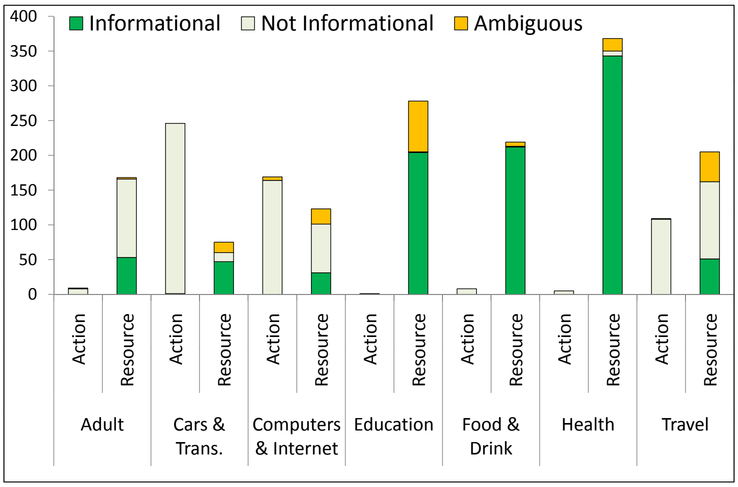


Figure 6.2: Distribution of queries into facets topic, task and objective.

the classifier does not provide predictions are: *Beauty & Style, Industrial Goods & Services, Sports, Technology & Electronic* and *Work*. All these topics have a very small concentration of queries (less of 50 queries each one, see Table 5.6 in the Chapter 5). This small concentration of queries suggest that these topics should be reconsidered in the list of topics. For example, the topic *Work* could be contained in the topic *Finance*.

Table 6.5 shows the performance of the prediction for the facets authority sensitivity, spatial sensitivity, time sensitivity, scope and specificity. The results for this group of facets is good. Some of the facets obtain better results than the others, but overall results are balanced. The facets spatial sensitivity and authority sensitivity show the best precision results. To be able to identify correctly these two facets is important because the search results for spatially-sensitive and authority-sensitive queries, must be both relevant to the query and valid for the associated location and for the authoritative requirement. This is the typical case while searching in a mobile device. The performance for time sensitivity, scope and specificity is similar in average.

6.2 Combining Multiple Facets

Although learning the exact intent behind a user query is a complicated task, it is not impossible. In the previous chapter we saw how through a multi-faceted model it is possible to obtain a better description of the query intent of the user and in the last section we show that it is possible to automatically classify the queries into our multi-faceted model.

The idea behind a multi-faceted model to the query intent of the user is to provide a richest context for the query, and has two main implications. First, in the absence of a rich context, the intent of a user query is often ambiguous. The use of multiple facets to describe the query intent of the user might help to reduce the ambiguity in the classification process. Second, a multi-faceted model would allow search engines to provide more accurate results to the users, according to the specific characteristics of the different dimensions of the query intent.

Although the set of facets that we are studying here are different dimensions of the user's intent and not all of them are necessarily correlated, we are interested to study how the combination of multiple facets in the classification process can improve the performance of the prediction of the facets. Based on the query log analysis we made in last chapter, we selected two groups of related facets and performed a multi-label classification of the queries with these facets. We now present the results of the multi-label classification and then we make a comparison of these results with the results of the single-label classification of the facets. Our hypothesis is that the performance of the single-label classifiers can be improved by introducing the information of multi-label training samples into the learning procedure effectively.

6.2.1 Multi-label Classification

We address the problem of classifying a query into a set of relevant facets as a multi-label classification problem. The traditional *single-label* classification, also known as multi-class classification, is the common machine learning task where an instance is assigned a single label l, that is chosen from a previously known finite set of labels L. A data-set D of n instances is composed of instance-classification pairs $(x_0, l_0), (x_1, l_1), ..., (x_n, l_n)$. The *multi-label* classification task is an extension of this problem, where each instance is associated with a subset of labels $S \subseteq L$. A multi-label data-set D of n instances is composed of instance-classification pairs $(x_0, S_0), (x1, S_1), ..., (x_n, S_n)$.

We can find many examples of multi-label classification problems in the real world. Some of the most popular applications are text categorization [23, 29, 46] and automatic image annotation [26, 71, 99]. Although it has been traditionally managed as a single-label classification problem, automatic query intent classification can be seen as a multi-label problem. In [49] Jansen, Booth and Spink concluded that there is an important portion of queries that are vague or multi-faceted that cannot be managed through single-label classification methods.

Our goal is to induce multi-label models from the labeled data set with a supervised machine learning method. Learning a multi-label model can be achieved through one of two approaches: problem transformation methods, and algorithm adaptation methods. The problem transformation methods, transform the multi-label classification problem into one or more single-label classification or regression problems. The adaptation methods, extend specific learning algorithms in order to handle multi-label data directly [96].

In this work we explore multi-class transformation, where each combination of facet-values is labeled with a single distinct label. This transformation explicitly captures overlaps between facets. Since Support Vector Machine have shown good generalization ability in different single-label multi-class problems, is also one of the most used techniques to resolve multi-label classification problem [78]. We used the multi-label classification tool of LIBSVM[1] to build multi-label classifiers for our group of facets. We selected the label combination option as the transformation method.

[1] Available at http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/multilabel/

6.2.2 Genre-Objective Combination

As we show in Chapter 5 (see Section 5.3), two of the most correlated and descriptive facets are genre and objective. We are interested to explore how combined classification of these facets can improve the quality of the classification results. In order to obtain comparable results, the training and test sets used for multi-label classification are the same data-sets used to perform single-label classification. That is, we used the same group of queries and the only variation are the training-set labels. In this case, each query was marked with two values, genre and objective, respectively.

The quality of multi-label classification is assessed through either partial or complete class label matching. First, we present the results for complete label matching and then we present the results for partial class label matching.

Complete Class Label Matching

Complete class label matching is often referred to as exact matching. The exact matching measure is in the context of multi-label classification a very strict measure as it requires the predicted set of labels to be an exact match to the true label set (Eq. 6.1) [87]. This measure is, in a way, similar to accuracy in the case of the single-label classification. However, it does not account for e.g. two out of three correctly predicted labels. This criterion gives the "automatic" ability of the predicting models. It shows the proportion of the data that do not need any human check.

$$\text{Exact match ratio} = \frac{\sum_{i=1}^{M} I[Y_i^{predicted} = Y_i^{actual}]}{M} \tag{6.1}$$

Where $I[S]$ is 1 if the statement S is true and 0 otherwise, and M represents the number of classified examples.

The multi-label classifier correctly predicted two of the two facets (genre and objective) for 1,517 queries, that is, the exact match ratio is 64.19%. This accuracy to predict simultaneously genre and objective is a fairly good success rate in the context of multi-label classification.

	Objective	
Genre	Action	Resource
Business	**67.82%**	33.33%
Community	3.70%	**77.23%**
News	**37.50%**	10.00%
Reference	0	**49.45%**

Table 6.6: Exact match ratio for multi-label classification based on the combination of the facets genre and objective.

It is also interesting to evaluate the exact match ratio for each pair of values of the combination genre-objective (see Table 6.6). As we can observe, there are clear relations between some pair of values. The best results are for the combination *Business-Action* and *Community-Resource*, the average exact match ratio for these two pair of values is 72.52%. The results of this automatic multi-label classification show that it is possible to predict simultaneously, values of multiple facets for a query. The association of multiple facets could improve the understanding of the intent of the query.

Partial Class Label Matching

To evaluate if multi-label classification obtains better recognition of single facets with respect to single-label classification, we also measure the performance prediction of the multi-label classifier for each individual facet (*partial class label matching*).

Table 6.7 shows the multi-label classification results for the facet genre. Overall, multi-label classification outperforms the single-label classification for all categories of genre. Specially, we note important improvements in recall for small categories (i.e., categories with less representation of queries in the sample data) like *News* and *Reference*. For *Reference*, single-label classification yielded good precision result (0.82 on average) but the recall value is low (0.11 on average). With the multi-label classification the results for *Reference* are more balanced, the recall

Genre	Precision	Recall	F-Measure
Business	**0.7620**	0.7052	0.7325
Community	0.7204	**0.7874**	**0.7524**
News	0.2727	0.2143	0.2400
Reference	0.5561	0.4936	0.5230
Average	0.5778	0.5501	0.5620

Table 6.7: Multi-label classification results based on the combination genre-objective for the facet genre (50% training and 50% testing).

Objective	Precision	Recall	F-Measure
Action	0.6762	0.6420	0.6587
Resource	**0.9019**	**0.9145**	**0.9082**
Average	0.7890	0.7783	0.7834

Table 6.8: Multi-label classification results based on the combination genre-objective for the facet objective (50% training and 50% testing).

improves to near 0.5 and the precision is still good (above 0.55). In the case of *News*, the overall performance is not so good, but it is noticeable that the multi-label classifier provides predictions for this category, given that the single-label classifier did not report any results for this case, because was too small. This is because when we use multi-label classification, the additional information provided with the multi-label training set allows the classifier to predict small categories. For the large categories the multi-label classification also improved the results. In Figure 6.3 we can see the comparison of the performance (F-measure values) of single-label and multi-label classification for the facet genre.

The results of the multi-label-classification for the facet objective are shown in Table 6.8, being the results for objective even better. With respect to the single-label classification results, the major improvements are for the category *Action*, as seen in Figure 6.4, where recall for *Action* improves dramatically from 0.18 to 0.64.

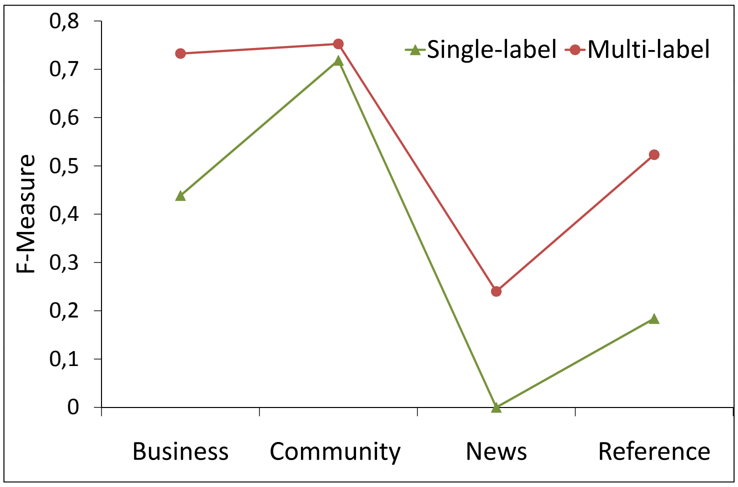


Figure 6.3: Comparison of single-label and multi-label F-Measure results for the facet Genre.

For the category *Resource* the multi-label classifier maintains the high recall and precision obtained with the single-label classification (F-measure is 0.908).

Overall, the combination genre-objective is positive for multi-label classification. From the results, obtaining better performance than the traditional single-label classification.

6.2.3 Genre-Task-Topic Combination

Three of the most important facets we are evaluating are genre, task and topic. The combination of these facets might influence the prediction results of the queries. In the Section 5.3, we show that there are some topics that are oriented to specific genre categories. For instance, the topic *Cars & Transportation* is related with the genre-category *Business* and the topic *Politics & Government* is related with the genre-category *Community*. These relations suggest that knowing that a query belongs to a particular topic, could also indicate that the query belongs to a partic-

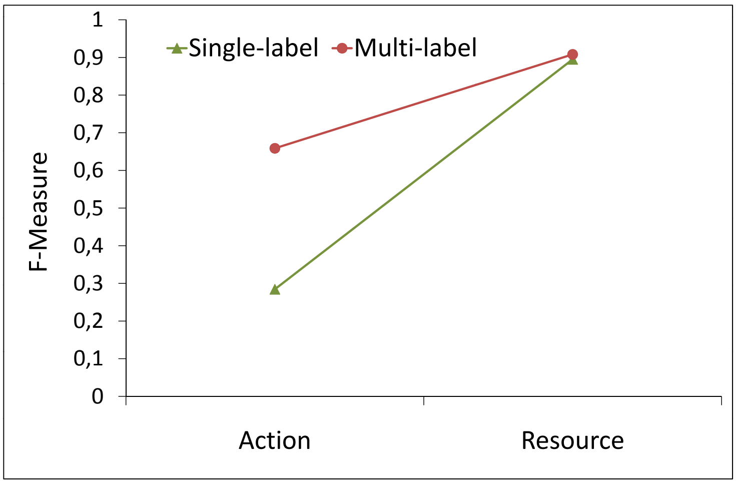


Figure 6.4: Comparison of single-label and multi-label F-Measure results for the facet Objective.

ular genre-category. Hence, we test this hypothesis through performing multi-label classification with these three correlated facets.

Complete Class Label Matching

Being able to predict correctly three facet-values (genre, task and topic) for a query is quite valuable to identify the query intent. This is an important but difficult task as predicting three labels is substantially more difficult than predicting one single label. In this case, where the number of facets increase, the correlation between the set of evaluated facets is a relevant factor that influences the quality of the multi-label classification results. The exact match ratio for the multi-label classification of genre, task and topic is 43.58%. This accuracy is not as high as the accuracy obtained with the combination genre-objective; however it is still good if we note that now we are considering one more facet. In the Figure 6.5 we show the distribution of queries that were correctly classified into the three facets.

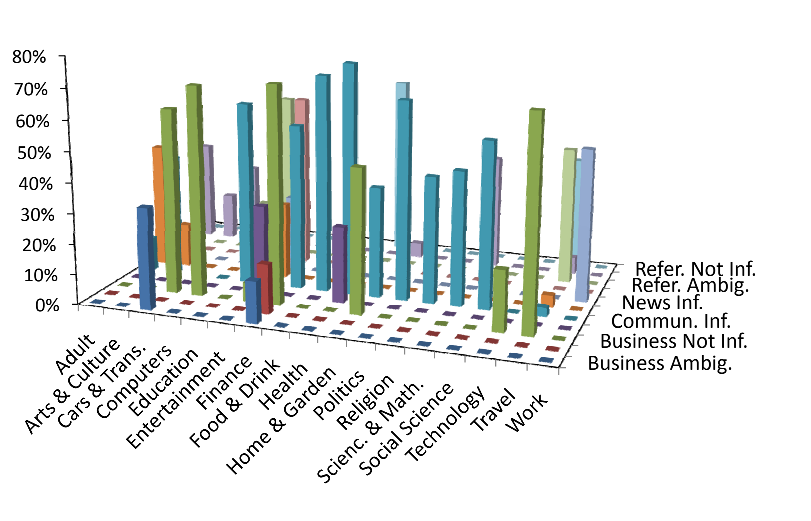


Figure 6.5: Distribution of queries correctly classified (exact matching) for the facets genre, task and topic.

As we can observe, the combinations of facet-values with highest percentage of correct queries are:

- genre-*Business*, task-*Not Informational*, topic-{*Cars & Transportation, Computers & Internet, Finance, Home & Garden, Travel*}.
- genre-*Community*, task-*Informational*, topic-{*Education, Entertainment, Music & Games, Finance, Food & Drink, Health, Home & Garden, Politics & Government, Religion & Belief Systems, Science & Mathematics, Social Science*}.

The first combination of facets represents queries which intent is oriented to perform commercial actions in one of the associated topics. Search results for queries with these characteristics should include service-oriented Web pages. The second combination of facets represents

Genre	Task		
	Informational	Not Informational	Ambiguous
Business	8.08%	**72.08%**	11.53%
Community	**72.94%**	28.91%	24.52%
News	18.91%	**57.14%**	0.0%
Reference	27.30	33.92%	**42.22%**

Table 6.9: Exact match ratio for genre and task based on the results of the multi-label classification genre-task-topic.

Topic	Genre			
	Business	Community	News	Reference
Adult	0.0%	**34.38%**	0.0%	30.77%
Arts & Culture	0.0%	**39.53%**	0.0%	11.76%
Cars & Trans.	**59.83%**	0.0%	n/a	26.67%
Computers	**73.15%**	0.0%	n/a	6.67%
Education	n/a	**60.32%**	0.0%	43.53%
Entertainment	5.00%	**44.12%**	40.91%	0.0%
Finance	**56.65%**	**54.41%**	0.0%	0.0%
Food & Drink	0.0%	**70.79%**	n/a	0.0%
Health	0.0%	**75.22%**	n/a	0.0%
Home & Garden	**42.86%**	33.33%	0.0%	4.76%
Indust. Goods & Serv.	7.14%	0.00%	n/a	0.00%
Politics	0.0%	**64.31%**	0.0%	0.0%
Religion	n/a	**40.00%**	0.0%	0.0%
Science & Math.	0.0%	**44.35%**	0.0%	37.78%
Social Science	0.0%	**55.26%**	0.0%	0.0%
Technology	**18.18%**	0.0%	n/a	0.0%
Travel	**66.67%**	33.33%	0.0%	37.04%
Work	n/a	**100.00%**	**75.00%**	0.0%

Table 6.10: Exact match ratio for topic and genre based on the results of the multi-label classification genre-task-topic.

queries with informational intent in community interest topics. Although in this case the multi-label classifier is based on the combination of three facets, it is also interesting to observe the results for each pair of facets involved in the classification process. The results of exact matching among the values of each pair of facets are reported in Tables 6.9 to Table 6.11. The best combination is the one integrated by the facets genre-task, that has an exact match ratio of 56.58%. For these facets the best combination of pair of values are genre-*Business* and community-*Informational*, the average exact match ratio for the values of the two facets is 72% (see Table 6.9). If we observe the results for the combination of the three facets, genre-*Business* and community-*Informational* are also included in the group of best combinations. In the case of two facets, the exact match ratio is higher considering that the number of labels is lower than in the case of multi-label classification with three facets.

The values of the exact match ratio for the other two combinations of facets, genre-topic and task-topic, are 46.97% and 48.15% respectively. According to the facet which is combined, topic shows good results for different group of values. Each facet influences in a different way to the other facet. For example, the best combination of values for genre-topic is *Community-Work* and for task-topic the best combinations of values is *Not Informational-Finance*. Although the two combinations include the facet topic, the topic with the highest accuracy is different for each combination.

Partial Class Label Matching

In the same way that in Section 6.2.2, we report now the results of the multi-label classification for the individual facets. We compare these results with the results obtained with single-label classification.

Table 6.13 shows the performance evaluation of the multi-label classification for the facet genre. In general, we obtain very similar results to the multi-label classification based on the combination genre-objective (see Table 6.7). The difference between the results of the two multi-label classifications is not high, the combination genre-objective has slightly

	Task		
Topic	Informational	Not Informational	Ambiguous
Adult	39,13%	**44,00%**	0.0%
Arts & Culture	**47,26%**	12,50%	16,67%
Cars & Trans.	18.18%	**60.91%**	33.33%
Computers	0.0%	**65,38%**	9.09%
Education	**48,48%**	0.0%	**55.10%**
Entertainment	22,08%	**28,33%**	26.09%
Finance	42,37%	**70,87%**	12.12%
Food & Drink	**65,31%**	0.0%	0.0%
Health	**70,73%**	**50.00%**	16.67%
Home & Garden	20,83%	**46.84%**	0.0%
Indust. Goods & Serv.	8,33%	0.0%	0.0%
Politics	**66,43%**	0.0%	0.0%
Religion	**33.33%**	0.0%	n/a
Science & Math.	**48,50%**	0.0%	0.0%
Social Science	**50,00%**	0.0%	n/a
Technology	0.0%	20.00%	n/a
Travel	13.33%	**58,16%**	43,48%
Work	**28.57%**	n/a	n/a

Table 6.11: Exact match ratio for topic and task based on the results of the multi-label classification genre-task-topic.

better results, but in general the improvements are the same in both cases. Since genre is the facet that has more positive correlations with other facets, it could be combined in different ways with several facets.

The results of the multi-label classification for the facet task are shown in the Table 6.12. In the same way that in Section 6.2.2, the multi-label classification maintain the good results of the large categories and improve the results of the small categories. In this case, the large category is *Informational*. For this category, the multi-label classifier yields similar results than the single label classifier (F-measure=0.82). For the

Genre	Precision	Recall	F-Measure
Business	**0.7170**	0.7068	0.7119
Community	**0.7116**	**0.7807**	**0.7446**
News	0.1538	0.3214	0.2081
Reference	0.5908	0.3792	0.4619
Average	0.5433	0.5471	0.5316

Table 6.12: Multi-label classification results based on the combination genre-task-topic for the facet task (50% training and 50% testing).

Task	Precision	Recall	F-Measure
Informational	**0.8235**	**0.8170**	**0.8202**
Not Informational	0.6923	0.6682	0.6801
Ambiguous	0.2692	0.3150	0.2903
Average	0.5950	0.6001	0.5969

Table 6.13: Multi-label classification results based on the combination genre-task-topic for the facet genre (50% training and 50% testing).

other two categories, *Not Informational* and *Ambiguous*, the multi-label classifier yields much better results than the single-label classifier.

In Figure 6.6 we can see a comparison between the overall results (F-measure values) of the multi-label classification and the single-label classification for the facet task. In this case, we included in the comparison the results obtained with the single-label classification of the Chapter 4, where we test a two facets model for the query intent identification. The idea of this comparison is to appreciate that what make the difference in the improvement on the prediction results is not the number of facets that we consider (two or nine), but the combination of these facets. Some previous works have analyzed several facets, but they have not shown how these facets can be combined to automatically identify the intent of the query. As we can see in the Figure 6.6, the best results are obtained with multi-label classification (combination of facets).

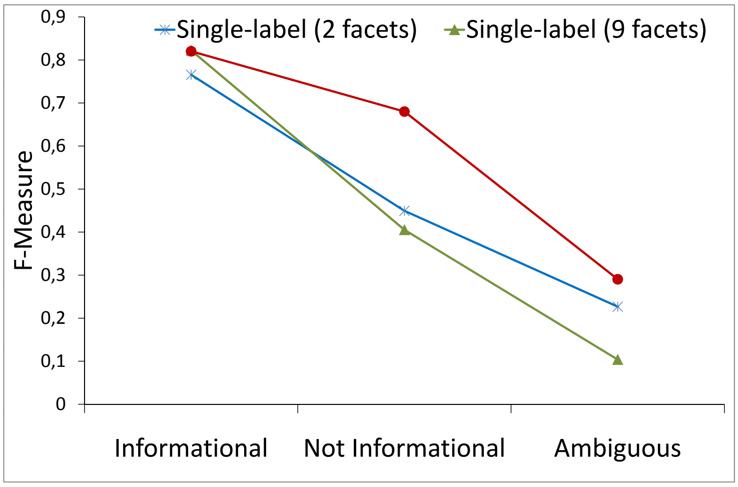


Figure 6.6: Comparison of single-label and multi-label F-Measure results for the facet task.

Finally, in Table 6.14 we show the multi-label classification results for the facet topic. The multi-label classifier provides predictions for eighteen of the twenty considered topics, while with the single-label classification, five topics obtained F-measure values of zero. Hence, when we use combinations of facets to train automatic classifiers, the probability to predict small categories increases. We observe this effect of multi-label classification in facets like topic and genre. In general, the coverage (recall) of topics is substantially increased. The results of the multi-label classification are more balanced (precision-recall) than the results of the single-label classification. Figure 6.7 shows the F-measure values for the two types of classifications. As we can observe, multi-label classification outperforms single classification in all the topics. In some of them with remarkable improvements, like *Food & Drink* and *Politics & Government* categories.

In summary, the combination of the facets genre, task and topic is good for the prediction of each facet. Specially, we show that the coverage of the facets increases considerably and the precision is more balanced.

Topic	Precision	Recall	F-Measure
Adult	0.5500	0.4400	0.4889
Arts & Culture	0.2313	0.4625	0.3083
Cars & Trans.	0.7333	0.5704	0.6417
Computers	0.6214	0.6692	0.6444
Education	0.6560	0.5503	0.5985
Entertainment	0.4944	0.4863	0.4904
Finance	0.5927	0.5787	0.5857
Food & Drink	**0.7742**	0.6545	**0.7094**
Health	0.6194	**0.7111**	0.6621
Home & Garden	0.5455	0.4154	0.4716
Indust. Goods & Serv.	0.5000	0.0870	0.1481
Politics	0.7245	0.6553	0.6882
Religion	0.5455	0.3673	0.4390
Science & Math.	0.6293	0.4796	0.5443
Social Science	0.3431	0.5054	0.4087
Technology	0.2000	0.1333	0.1600
Travel	0.6074	0.5430	0.5734
Work	0.0606	0.5714	0.1096
Average	0.5238	0.4934	0.4818

Table 6.14: Multi-label classification results based on the combination genre-task-topic for the facet topic (50% training and 50% testing).

6.3 Comparison with other Approaches

Given that in the comparison of machine learning techniques done in Section 4.5, supervised and semi-supervised learning obtained close results, we compare now the multi-label classification results with the semi-supervised classification results in order to analyze the effect of to incorporate multiple facets into the learning process.

We compare the automatic classification results for the facets genre (Table 6.15), task (Table 6.16) and objective (Table 6.17). As we can observe from the tables, the multi-label classification outperforms the semi-

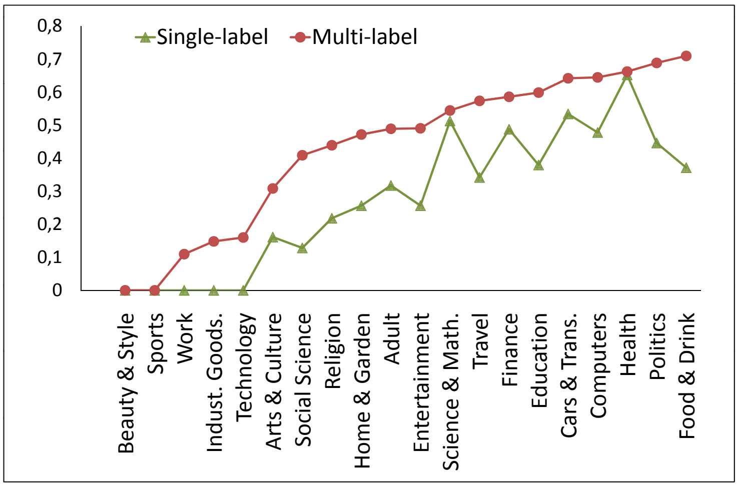


Figure 6.7: Comparison of single-label and multi-label F-Measure results for the facet topic.

supervised classification in most of the categories of each facet. The difference is especially large for the facet objective, where the F-measure value obtained by multi-label classification is nearly double than the value obtained by semi-supervised classification.

A particular case are the *Ambiguous* queries. Although that the results of multi-label classification for this type of queries improved substantially with respect to the initial results shown in Chapter 4, specially for the recall value, the F-measure value is not higher than 0.3. For this type of queries semi-supervised learning obtains better results than the multi-label classifier. Given that semi-supervised learning combine labeled with unlabeled data, this could suggest that this kind of learning is better to identify queries with an ambiguous nature.

Overall, we can appreciate that using combination of facets into the learning process, allows to obtain better classification results than those of the traditional machine learning techniques, like semi-supervised learning.

Genre	Supervised			Semi-supervised		
	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.
Business	0.7620	0.7052	**0.7325**	0.678	0.582	0.6263
Community	0.7204	0.7874	**0.7524**	0.621	0.543	0.5794
News	0.2727	0.2143	0.2400	0.543	0.483	**0.5112**
Reference	0.5561	0.4936	0.5230	0.665	0.576	**0.6173**

Table 6.15: Comparison of performance of multi-label classification and semi-supervised learning for the facet genre

Task	Supervised			Semi-supervised		
	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.
Informational	0.8235	0.8170	**0.8202**	0.801	0.738	0.7682
Not Informat.	0.6923	0.6682	**0.6801**	0.506	0.648	0.5683
Ambiguous	0.2692	0.3150	0.2903	0.473	0.411	**0.4398**

Table 6.16: Comparison of performance of multi-label classification and semi-supervised learning for the facet task.

Objective	Supervised			Semi-supervised		
	Prec.	Recall	F-Meas.	Prec.	Recall	F-Meas.
Action	0.6762	0.6420	**0.6587**	0.543	0.385	0.4505
Resource	0.9019	0.9145	**0.9082**	0.537	0.418	0.4701

Table 6.17: Comparison of performance of multi-label classification and semi-supervised learning for the facet objective.

6.4 Summary

In this chapter we explore the feasibility of multi-faceted query intent prediction. We first perform individual classification processes for nine facets and then we perform multi-label classification of the facets based on the combination of correlated facets.

The results of the classification of the individual facets are very promising. The results show that the predictability of the evaluated facets is good. Since most of the previous research in query intent identification is oriented to the study of single facets, these results offer a wider spectrum of facets that can be used for automatic user intent identification.

We also study the combination of multiple facets to improve the performance of the automatic classification process. Our experimental evaluation shows that the combination of facets can effectively improve the prediction of the query intent. Some previous works have analyzed several facets, but they have not shown how these facets can be combined to automatically identify the intent of the query. As we can see from the results of our experiments, the best performance is obtained with multi-label classification.

There are several potential areas for future work, including the study of optimal combination of facets that could be implemented through multi-labeled classification.

Chapter 7

CONCLUSIONS AND FUTURE WORK

In order that Web search engines continue evolving together with its dynamic context, they must provide elements that increase the knowledge of the behavior of the user and of the needs underlying their queries. In this thesis we have shown that determining the intent behind the user query has the potential to highly improve the performance of a Web search engine. The intent of the query has a direct impact in the user as is shown with the eye tracking analysis of user search behavior that we presented in Chapter 3. With this study we show that the query intent of the user affects all the decision processes in the SERP. Different types of query intent have different implications on users' search behavior. As a consequence, knowing the intent of the user Web queries could help Web search engines in the task of associate resources available on the Web with user goals.

There has been a lot of work in the literature attempting to characterize the intent of Web-search queries. Much of this work has characterized query intent in single facets, such as geographical locality, topic, type of resources required to answer the query, and so on. Although all the early approaches to identify the intent of the user query are valid, in this thesis we hypothesize that query intent is a complex phenomena, which can be modeled more accurately by considering a multi-dimensional set of facets that interact with each-other in non-trivial ways. In order to prove our hypothesis of the multidimensionality of the intent of the user query, we carry out several experiments that allow us to determine if it is possible to build a model for accurate identification of query intent based on multiple facets.

7.1 Main Results

We first present an initial automatic classification of query intent approach using two facets: the query intent and the topic categories in which the queries within these intents can be located. Our results show that query intent can be automatically predicted with accuracy 60%. We observe that the distribution of the query intent along the topics varies for each intent, that is, the intent of the query influences the associated topic of that query. For instance, topics such as *Education, News, Science* and *Society* are mostly informational topics given that more than 80% of the queries of these topics are associated with the *Informational* intent. In the same way, the topics *Adult* (72%) and *Games* (58%) show clear association with the *Not Informational* intent. These findings are important, because these kind of relationships between types of intent and topic categories can help to Web search engines to map incoming search queries to specific content. Overall, the results of this study are promising because they show that the combination of several facets can offer more interesting use-cases driven by the process of query intent identification. A natural next step was to explore new combinations of facets and new methods that would be able to take advantage of the combination of multiple facets.

In a more advanced approach to the identification of the multiple facets of the query intent, we analyze and characterize a wide range of factors or facets that may be useful for user's intent identification when searching for information on the Web. We study some of the most used facets to characterize user intent, like the two we have studied before, the task (type of resource) and the topic, we also include other popular facets like spatial sensitivity and time sensitivity. As a result of the query log analysis and of the initial automatic classification results, we also introduce some new facets that we considered useful to the identification of the query intent like genre (*Business, Community, News, Reference*) and objective (*Action, Resource*). We describe the main features of each facet, their usefulness, and analyze the relationships among them. From the analysis of the nine facets, we found more interesting relationships between the facets, than those relationships that we have found when we analyzed only two facets. For instance, most of the queries associated with genre=*Business* and objective=*Action* were also related with the topics *Cars & Transportation* and *Computers & Internet*. This association of values of facets describe queries with a clear commercial intention. Nowadays, where the users are not only interested in obtaining what they are looking for, but also in not be disturbed with what they are not looking for, this kind of fine-grained description of the query intent would be very appreciated by the users.

Considering the importance of the relationships between the facets, we use the analyzed facets for the automatic classification of Web queries. We first classify the query intent along the nine facets independently and then we perform multi-label classification of the facets based on the combination of correlated facets. The objective of this procedure is to evaluate if the automatic classification of queries based on a multi-faceted method can yield more insights on the query intent of the user, than the insights obtained with the query classification based on mutually exclusive facets. The results of the classification of queries using combination of facets are very promising. They show that the predictability of the evaluated facets is good and the combination of correlated facets can effectively improve the prediction of the query intent. Some previous works have analyzed several facets, but they have not shown how these facets can be combined to automatically identify the query intent. As we can see from the results of our experiments, the best performance is obtained with multi-faceted classification. Given the variety and complexity of the information seeking behaviors, our findings are a first step to a more comprehensive description of the intents underlying the real needs of the users.

7.2 Future Work

There are many areas of potential interest for future work, however we highlight three of them:

- *Identification of optimal combinations of facets and multi-faceted classification methods*: This thesis is a first step to a functional multi-faceted query intent classification, however, there is still work to do. More study about the best facets and their combination is needed. The facets explored in this work are useful to the identification of query intent, but it does not mean that other facets can not be considered for multi-faceted query intent classification. In the same way, it is important to experiment with different multi-label classification methods in order to determine which technique is more suitable, in terms of effectiveness and scalability.

- *Multi-faceted classification of Web content*: All the work presented in this research is concerned with the user queries, however, in order to be able to give answer to the intent behind the query is necessary to count with the adequate description of the Web content. If the query is described through multiple facets, would be desirable to match the right documents for those facets.

- *Personalizing of search engine result page regarding the intent of the user query*: As we show with the eye tracking study of the behavior of the users in SERPs, the intent influence the browsing behavior of the users and also the areas of interest in the SERP. According to this, it is very important to adapt the search results to the needs of the users. The SERP should be completely dynamic and adapted to each query intent.

Bibliography

[1] Eugene Agichtein, Eric Brill, Susan Dumais, and Robert Ragno. Learning user interaction models for predicting Web search result preferences. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '06, pages 3–10, New York, NY, USA, 2006. ACM.

[2] Azin Ashkan and Charles L.A. Clarke. Characterizing commercial intent. In *CIKM '09*, pages 67–76, New York, NY, USA, 2009. ACM.

[3] Ch. Aswanikumar, Ankush Gupta, Mahmooda Batool, and Shagun Trehan. An information retrieval model based on latent semantic indexing with intelligent preprocessing. *JIKM*, pages 279–285, 2005.

[4] Ricardo Baeza-Yates. Applications of Web query mining. In *Advances in Information Retrieval*, volume 3408 of *Lecture Notes in Computer Science*, pages 7–22. Springer Berlin / Heidelberg, 2005.

[5] Ricardo Baeza-Yates, Liliana Calderón-Benavides, and Cristina González-Caro. The intention behind Web queries. In *Proc. of the 13th SPIRE*, volume 4209 of *LNCS*, pages 98–109. Springer, 2006.

[6] Ricardo Baeza-Yates, Carlos Hurtado, and Marcelo Mendoza. Improving search engines by query clustering. In *JASIST, Special*

Issue on Mining Web Resources for Enhancing Information Retrieval, 58(12), 2007.

[7] Ricardo Baeza-Yates, Carlos Hurtado, Marcelo Mendoza, and Georges Dupret. Modeling user search behavior. *Latin American Web Congress*, 0:242–251, 2005.

[8] Bing Bai, Jason Weston, David Grangier, Ronan Collobert, Kunihiko Sadamasa, Yanjun Qi, Olivier Chapelle, and Kilian Weinberger. Supervised semantic indexing. In *Proceeding of the 18th ACM conference on Information and knowledge management*, CIKM '09, pages 187–196, New York, NY, USA, 2009. ACM.

[9] A. Basu, C. Watters, and M. Shepherd. Support vector machines for text categorization. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03) - Track 4 - Volume 4*, HICSS '03, pages 103.3–, Washington, DC, USA, 2003. IEEE Computer Society.

[10] Steven M. Beitzel, Eric C. Jensen, Ophir Frieder, David D. Lewis, Abdur Chowdhury, and Aleksander Kolcz. Improving automatic query classification via semi-supervised learning. In *ICDM'05*, pages 42–49. IEEE, 2005.

[11] Steven M. Beitzel, Eric C. Jensen, David D. Lewis, Abdur Chowdhury, and Ophir Frieder. Automatic classification of web queries using very large unlabeled query logs. *ACM Trans. Inf. Syst.*, 25, April 2007.

[12] N. J. Belkin, R. N. Oddy, and H. M. Brooks. Ask for information retrieval: Part i. background and theory. *Journal of Documentation*, 38(2):61–71, 1982.

[13] Nicholas J. Belkin and W. Bruce Croft. Information filtering and information retrieval: two sides of the same coin? *Commun. ACM*, 35:29–38, December 1992.

[14] Andrei Broder. A taxonomy of Web search. *SIGIR Forum*, 36:3–10, September 2002.

[15] Andrei Broder. The next generation Web search and the demise of the classic IR model. In *Proceedings of the 29th European conference on IR research*, ECIR'07, pages 1–1. Springer-Verlag, 2007.

[16] Georg Buscher, Susan T. Dumais, and Edward Cutrell. The good, the bad, and the random: an eye-tracking study of ad quality in Web search. In *SIGIR '10*, pages 42–49, New York, NY, USA, 2010. ACM.

[17] Liliana Calderón-Benavides, Cristina González-Caro, and Ricardo Baeza-Yates. Towards a deeper understanding of the user's query intent. In W. Bruce Croft, Michael Bendersky, Hang Li, and Gu Xu, editors, *Query representation and understanding. A workshop at SIGIR'2010*.

[18] Chih chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines, 2001.

[19] Colleen Cool and Nicholas J. Belkin. A classification of interactions with information. In *Proceedings of the Fourth International Conference on Conceptions of Library and Information Science*, pages 1–15. Greenwood Village, CO: Libraries Unlimited, 2002.

[20] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, 2000.

[21] Edward Cutrell and Zhiwei Guan. What are you looking for?: an eye-tracking study of information usage in Web search. In *CHI '07*, pages 407–416, New York, NY, USA, 2007. ACM.

[22] Cristian Danescu-Niculescu-Mizil, Andrei Z. Broder, Evgeniy Gabrilovich, Vanja Josifovski, and Bo Pang. Competing for users' attention: on the interplay between organic and sponsored search

results. In *WWW '10*, pages 291–300, New York, NY, USA, 2010. ACM.

[23] Sareewan Dendamrongvit and Miroslav Kubat. Undersampling approach for imbalanced training sets and induction from multi-label text-categorization domains. In *Proceedings of the 13th Pacific-Asia international conference on Knowledge discovery and data mining: new frontiers in applied data mining*, PAKDD'09, pages 40–52. Springer-Verlag, 2010.

[24] Beniamino Di Martino. An approach to semantic information retrieval based on natural language query understanding. In *Proceedings of the 10th international conference on Current trends in Web engineering*, ICWE'10, pages 211–222. Springer-Verlag, 2010.

[25] Thomas G. Dietterich and Ghulum Bakiri. Solving multiclass learning problems via error-correcting output codes. *J. Artif. Int. Res.*, 2:263–286, January 1995.

[26] Ivica Dimitrovski, Dragi Kocev, Suzana Loskovska, and Sašo Džeroski. Detection of visual concepts and annotation of images using ensembles of trees for hierarchical multi-label classification. In *Proceedings of the 20th International conference on Recognizing patterns in signals, speech, images, and videos*, ICPR'10, pages 152–161, Berlin, Heidelberg, 2010. Springer-Verlag.

[27] Michael J. Driver and Theodore J. Mock. Human information processing, decision style theory, and accounting information systems. *The Accounting Review*, 50(3):490–508, 1975.

[28] Katrin Erk and Sebastian Padó. A structured vector space model for word meaning in context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 897–906, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[29] Andrea Esuli, Tiziano Fagni, and Fabrizio Sebastiani. Boosting multi-label hierarchical text categorization. *Inf. Retr.*, 11:287–313, August 2008.

[30] Ali F. Farhoomand and Don H. Drury. Managerial information overload. *Commun. ACM*, 45:127–131, October 2002.

[31] Joseph H. Goldberg, Mark J. Stimson, Marion Lewenstein, Neil Scott, and Anna M. Wichansky. Eye tracking in Web search tasks: design implications. In *ETRA '02*, pages 51–58, New York, NY, USA, 2002. ACM.

[32] Cristina González-Caro and Ricardo Baeza-Yates. A multi-faceted approach to user's query intent classification. In *Proceedings of the 18th International Symposium on String Processing and Information Retrieval*, SPIRE '11. Springer-Verlag, 2011.

[33] Cristina González-Caro, Liliana Calderón-Benavides, and Ricardo Baeza-Yates. Web queries: The tip of the iceberg of user's intent. In David Carmel, Vanja Josifovski, and Yoelle Maarek, editors, *User modeling for Web applications. A workshop at WSDM'2011.*

[34] Cristina González-Caro and Mari-Carmen Marcos. Different users and intents: An eye-tracking analysis of Web search. In David Carmel, Vanja Josifovski, and Yoelle Maarek, editors, *User modeling for Web applications. A workshop at WSDM'2011.*

[35] Laura A. Granka, Thorsten Joachims, and Geri Gay. Eye-tracking analysis of user behavior in WWW search. In *SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 478–479, New York, NY, USA, 2004. ACM.

[36] Luis Gravano, Vasileios Hatzivassiloglou, and Richard Lichtenstein. Categorizing Web queries according to geographical locality. In *Proc. of the CIKM '03, pages 325–333, New York, NY, USA, 2003. ACM.*

[37] Qi Guo and Eugene Agichtein. Exploring mouse movements for inferring query intent. In *SIGIR '08*, pages 707–708, New York, NY, USA, 2008. ACM.

[38] Qi Guo and Eugene Agichtein. Exploring searcher interactions for distinguishing types of commercial intent. In *WWW '10: Proceedings of the 19th international conference on World Wide Web*, pages 1107–1108, New York, NY, USA, 2010. ACM.

[39] Qi Guo and Eugene Agichtein. Ready to buy or just browsing?: detecting Web searcher goals from interaction data. In *SIGIR '10*, pages 130–137, New York, NY, USA, 2010. ACM.

[40] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11:10–18, November 2009.

[41] B. S. Harish, D. S. Guru, S. Manjunath, and R. Dinesh. Cluster based symbolic representation and feature selection for text classification. In *Proceedings of the 6th international conference on Advanced data mining and applications - Volume Part II*, ADMA'10, pages 158–166, Berlin, Heidelberg, 2010. Springer-Verlag.

[42] Mauro Rojas Herrera, Edleno Silva de Moura, Marco Cristo, Thomaz Philippe Silva, and Altigran Soares da Silva. Exploring features for the automatic identification of user goals in Web search. *Inf. Process. Manage.*, 46:131–142, 2010.

[43] Thomas Hofmann. Probabilistic latent semantic analysis. In *Proc. of Uncertainty in Artificial Intelligence, UAI'99*, Stockholm, 1999.

[44] G. Hotchkiss. Enquiro eye tracking report ii: Google, MSN and Yahoo! compared, 2006.

[45] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, 13(2):415 –425, March 2002.

[46] Liu Hua. Research on multi-classification and multi-label in text categorization. In *Proceedings of the 2009 International Conference on Intelligent Human-Machine Systems and Cybernetics - Volume 02*, IHMSC '09, pages 86–89, Washington, DC, USA, 2009. IEEE Computer Society.

[47] Mark I. Hwang and Jerry W. Lin. Information dissemination, information overload and decision quality. *Journal of Information Science*, 25(3):213–218, 1999.

[48] Bernard J. Jansen. The comparative effectiveness of sponsored and nonsponsored links for Web e-commerce queries. *ACM Trans. Web*, 1(1):3, 2007.

[49] Bernard J. Jansen, Danielle L. Booth, and Amanda Spink. Determining the informational, navigational, and transactional intent of Web queries. *Inf. Process. Manage.*, 44(3):1251–1266, 2008.

[50] Bernard J. Jansen and Marc Resnick. An examination of searcher's perceptions of nonsponsored and sponsored links during ecommerce Web searching. *J. Am. Soc. Inf. Sci. Technol.*, 57(14):1949–1961, 2006.

[51] Bernard J. Jansen and Amanda Spink. An analysis of Web searching by European alltheweb.com users. *Inf. Process. Manage.*, 41:361–381, 2005.

[52] Bernard J. Jansen and Amanda Spink. How are we searching the World Wide Web? a comparison of nine search engine transaction logs. *Inf. Process. Manage.*, 42(1):248–263, 2006.

[53] N. Jardine and C. J. van Rijsbergen. The use of hierarchical clustering in information retrieval. *Information Storage and Retrieval*, 7:217–240, 1971.

[54] Wei Jin and Rohini K. Srihari. Graph-based text representation and knowledge discovery. In *Proceedings of the 2007 ACM symposium on Applied computing*, SAC '07, pages 807–811. ACM, 2007.

[55] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. Accurately interpreting clickthrough data as implicit feedback. In *In Proceedings of SIGIR*, pages 154–161, 2005.

[56] Rosie Jones and Kristina Lisa Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management*, pages 699–708, New York, NY, USA, 2008. ACM.

[57] Rosie Jones, Wei V. Zhang, Benjamin Rey, Pradhuman Jhala, and Eugene Stipp. Geographic intention and modification in Web search. *Int. J. Geogr. Inf. Sci.*, 22:229–246, 2008.

[58] In-Ho Kang and GilChang Kim. Query type classification for Web document retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, SIGIR '03, pages 64–71. ACM, 2003.

[59] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine Piatko, Ruth Silverman, and Angela Y. Wu. The analysis of a simple k-means clustering algorithm. In *Proceedings of the sixteenth annual symposium on Computational geometry*, SCG '00, pages 100–109, New York, NY, USA, 2000. ACM.

[60] Robert Krovetz and W. Bruce Croft. Lexical ambiguity and information retrieval. *ACM Trans. Inf. Syst.*, 10:115–141, April 1992.

[61] Uichin Lee, Zhenyu Liu, and Junghoo Cho. Automatic identification of user goals in web search. In *Proceedings of the 14th international conference on World Wide Web*, WWW '05, pages 391–400, New York, NY, USA, 2005. ACM.

[62] David D. Lewis. *Text representation for intelligent text retrieval: a classification-oriented view*, pages 179–197. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1992.

[63] Ying Li, Zijian Zheng, and Honghua (Kathy) Dai. KDD CUP-2005 report: facing a great challenge. *SIGKDD Explor. Newsl.*, 7:91–99, December 2005.

[64] Yiqun Liu, Min Zhang, Liyun Ru, and Shaoping Ma. Automatic query type identification based on click through information. In *AIRS'06*, pages 593–600. Springer-Verlag, 2006.

[65] Reitbauer M. Keep an eye on information processing: Eye tracking evidence for the influence of hypertext structures on navigational behaviour and textual complexity. *LSP and professional communication*, 8(2):14–38, 2008.

[66] Mari-Carmen Marcos and Cristina González-Caro. Comportamiento de los usuarios en la página de resultados de los buscadores. *El profesional de la información*, 19:348–358, Julio-Agosto 2010.

[67] Mazlita Mat-Hassan and Mark Levene. Associating search and navigation behavior through log analysis: Research articles. *J. Am. Soc. Inf. Sci. Technol.*, 56(9):913–934, 2005.

[68] Marcelo Mendoza and Juan Zamora. Identifying the intent of a user query using support vector machines. In *Proceedings of the 16th International Symposium on String Processing and Information Retrieval*, SPIRE '09, pages 131–142. Springer-Verlag, 2009.

[69] Donald Metzler, Rosie Jones, Fuchun Peng, and Ruiqiang Zhang. Improving search relevance for implicitly temporal queries. In *SIGIR'09*, pages 700–701. ACM, 2009.

[70] Dan I. Moldovan and Mihai Surdeanu. On the role of information retrieval and information extraction in question answering systems. In *SCIE'02*, pages 129–147, 2002.

[71] Gulisong Nasierding and Abbas Z. Kouzani. Empirical study of multi-label classification methods for image annotation and retrieval. In *Proceedings of the 2010 International Conference on*

Digital Image Computing: Techniques and Applications, DICTA '10, pages 617–622, Washington, DC, USA, 2010. IEEE Computer Society.

[72] Viet B. Nguyen and Min-Yen Kan. Functional faceted Web query analysis. In *Query Log Analysis: Social And Technological Challenges. A workshop at the WWW' 07, May 2007.*

[73] Shahrul Azman Noah, Lailatulqadri Zakaria, and Arifah Che Alhadi. Extracting and modeling the semantic information content of Web documents to support semantic document retrieval. In *Proceedings of the Sixth Asia-Pacific Conference on Conceptual Modeling - Volume 96*, APCCM '09, pages 79–86. Australian Computer Society, Inc., 2009.

[74] M. Oussalah, S. Khan, and S. Nefti. Personalized information retrieval system in the framework of fuzzy logic. *Expert Syst. Appl.*, 35:423–433, July 2008.

[75] Bing Pan, Helene A. Hembrooke, Geri K. Gay, Laura A. Granka, Matthew K. Feusner, and Jill K. Newman. The determinants of Web page viewing behavior: an eye-tracking study. In *ETRA '04: Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 147–154, New York, NY, USA, 2004. ACM.

[76] Stefan Pohl, Justin Zobel, and Alistair Moffat. Extended boolean retrieval for systematic biomedical reviews. In *Proceedings of the Thirty-Third Australasian Conferenc on Computer Science - Volume 102*, ACSC '10, pages 117–126. Australian Computer Society, Inc., 2010.

[77] Alex Poole and Linden J. Ball. Eye tracking in human-computer interaction and usability research: Current status and future. In *C. Ghaoui (Ed.): Encyclopedia of Human-Computer Interaction.*, pages 211–219. Hershey, PA: Idea Group, 2005.

[78] Yu-ping Qin and Xiu-kun Wang. Study on multi-label text classification based on SVM. In *Proceedings of the 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery - Volume 01*, FSKD '09, pages 300–304, Washington, DC, USA, 2009. IEEE Computer Society.

[79] Davood Rafiei, Krishna Bharat, and Anand Shukla. Diversifying Web search results. In *WWW'10*, pages 781–790. ACM, 2010.

[80] Keith Rayner. Eye movements in reading and information processing. *Psychological Bulletin*, 85(3):618 – 660, 1978.

[81] Daniel E. Rose and Danny Levinson. Understanding user goals in Web search. In *WWW '04*, pages 13–19, New York, NY, USA, 2004. ACM.

[82] Khaled Shaban. *A semantic graph model for text representation and matching in document mining*. PhD thesis, Waterloo, Ont., Canada, Canada, 2006. AAINR23537.

[83] Dou Shen, Rong Pan, Jian-Tao Sun, Jeffrey Junfeng Pan, Kangheng Wu, Jie Yin, and Qiang Yang. Q2c@ust: our winning solution to query classification in KDDCUP 2005. *SIGKDD Explor. Newsl.*, 7:100–110, December 2005.

[84] Dou Shen, Jian-Tao Sun, Qiang Yang, and Zheng Chen. Building bridges for Web query classification. In *SIGIR'06*, pages 131–138. ACM, 2006.

[85] Sajad Shirali-Shahreza. Compact representation of multimedia files for indexing, classification and retrieval. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems*, MEDES '09, pages 81:489–81:492, New York, NY, USA, 2009. ACM.

[86] Vikas Sindhwani and S. Sathiya Keerthi. Large scale semi-supervised linear svms. In *Proc. of the 29th ACM SIGIR conference*, pages 477–484. ACM Press, 2006.

[87] Marina Sokolova and Lapal Guy. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manage.*, 45:427–437, July 2009.

[88] Paul Solomon. Looking for information—a survey of research on information seeking, needs, and behavior. *Inf. Retr.*, 6(2):284–288, 2003.

[89] Fengxi Song, Shuhai Liu, and Jing yu Yang. A comparative study on text representation schemes in text categorization. *Pattern Analysis and Applications*, 8:199–209, 2005.

[90] Vassilis Spiliopoulos, Konstantinos Kotis, and George A. Vouros. Semantic retrieval and ranking of semantic web documents using free-form queries. *Int. J. Metadata Semant. Ontologies*, 3:95–108, December 2008.

[91] Amanda Spink, Dietmar Wolfram, Major B. J. Jansen, and Tefko Saracevic. Searching the Web: the public and their queries. *J. Am. Soc. Inf. Sci. Technol.*, 52:226–234, 2001.

[92] Xiaoying Tai, Fuji Ren, and Kenji Kita. An information retrieval model based on vector space method by supervised learning. *Inf. Process. Manage.*, 38:749–764, November 2002.

[93] Jaime Teevan, Christine Alvarado, Mark S. Ackerman, and David R. Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '04, pages 415–422. ACM, 2004.

[94] Jaime Teevan, Susan T. Dumais, and Daniel J. Liebling. To personalize or not to personalize: modeling queries with variation in user intent. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, pages 163–170, New York, NY, USA, 2008. ACM.

[95] Hitoshi Terai, Hitomi Saito, Yuka Egusa, Masao Takaku, Makiko Miwa, and Noriko Kando. Differences between informational and transactional tasks in information seeking on the Web. In *IIiX '08: Proceedings of the second international symposium on Information interaction in context*, pages 152–159, New York, NY, USA, 2008. ACM.

[96] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *Int J Data Warehousing and Mining*, 2007:1–13, 2007.

[97] Marnix S. van Gisbergen, Jeroen van der Most, and Paul Aelen. Visual attention to online search engine results. *Market Research Agency De Vos & Jansen*, pages 1–13, 2007.

[98] Vladimir N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[99] Zhengxiang Wang, Yiqun Hu, and Liang-Tien Chia. Multi-label learning by image-to-class distance for scene classification and image annotation. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, CIVR '10, pages 105–112, New York, NY, USA, 2010. ACM.

[100] Steve Wedig and Omid Madani. A large-scale analysis of query logs for assessing personalization opportunities. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 742–747, New York, NY, USA, 2006. ACM.

[101] Wilson, T. D. Models in information behaviour research. *Journal of Documentation*, pages 249–270, 1999.

[102] Zuobing Xu and Ram Akella. Improving probabilistic information retrieval by modeling burstiness of words. *Inf. Process. Manage.*, 46:143–158, March 2010.

[103] Qiang Yang, Hai-Feng Wang, Ji-Rong Wen, Gao Zhang, Ye Lu, Kai-Fu Lee, and Hong-Jiang Zhang. Towards a next-generation search engine. In *Proceedings of the 6th Pacific Rim international conference on Artificial intelligence*, PRICAI'00, pages 5–15. Springer-Verlag, 2000.