Sushmita, Shanu (2012) *Study of result presentation and interaction for aggregated search.* PhD thesis.

http://theses.gla.ac.uk/3289/

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Glasgow Theses Service
http://theses.gla.ac.uk/
theses@gla.ac.uk

Study of Result Presentation and Interaction for Aggregated Search

Shanu Sushmita
School of Computing Science
University of Glasgow

A thesis submitted in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy
April 2012

Dedicated to

My Parents

Abstract

The World Wide Web has always attracted researchers and commercial search engine companies due to the enormous amount of information available on it. "Searching" on web has become an integral part of today's world, and many people rely on it when looking for information. The amount and the diversity of information available on the Web has also increased dramatically. Due to which, the researchers and the search engine companies are making constant efforts in order to make this information accessible to the people effectively.

Not only there is an increase in the amount and diversity of information available online, users are now often seeking information on broader topics. Users seeking information on broad topics, gather information from various information sources (e.g, image, video, news, blog, etc). For such information requests, not only web results but results from different document genre and multimedia contents are also becoming relevant. For instance, users' looking for information on *"Glasgow"* might be interested in web results about Glasgow, Map of Glasgow, Images of Glasgow, News of Glasgow, and so on. Aggregated search aims to provide access to this diverse information in a unified manner by aggregating results from different information sources on a single result page. Hence making information gathering process easier for broad topics.

This thesis aims to explore the aggregated search from the users' perspective. The thesis first and foremost focuses on understanding and describing the phenomena related to the users' search process in the context of the aggregated search. The goal is to participate in building theories and in understanding constraints, as well as providing insights into the interface design space. In building this understanding, the thesis focuses on the click-behavior, information need, source relevance, dynamics of search intents. The understanding comes partly from conducting users studies and, from analyzing search engine log data.

While the thematic (or topical) relevance of documents is important, this thesis argues that the "source type" *(source-orientation)* may also be an important dimension in the

relevance space for investigating in aggregated search. Therefore, relevance is multi-dimensional (topical and source-orientated) within the context of aggregated search. Results from the study suggest that the effect of the source-orientation was a significant factor in an aggregated search scenario. Hence adds another dimension to the relevance space within the aggregated search scenario. The thesis further presents an effective method which combines rule base and machine learning techniques to identify source-orientation behind a user query.

Furthermore, after analyzing log-data from a search engine company and conducting user study experiments, several design issues that may arise with respect to the aggregated search interface are identified. In order to address these issues, suitable design guidelines that can be beneficial from the interface perspective are also suggested.

To conclude, aim of this thesis is to explore the emerging aggregated search from users' perspective, since it is a very important for front-end technologies. An additional goal is to provide empirical evidence for influence of aggregated search on users searching behavior, and identify some of the key challenges of aggregated search. During this work several aspects of aggregated search will be uncovered. Furthermore, this thesis will provide a foundations for future research in aggregated search and will highlight the potential research directions.

Acknowledgements

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less" – Marie Curie. I started my PhD with similar belief, and it became a great journey — very enriching and very emotional, a journey of countless memories and experiences. I want to take this opportunity to mention and thank people who made this possible.

First, I would like to thank my PhD supervisors, Professor Mounia Lalmas and Joemon M. Jose, for their enthusiasm, guidance and support over the years. They were constant source of inspiration and I have benefitted greatly from their experience and wisdom. They were truly – a friend, philosopher and guide.

I want to thank the members of Information Retrieval team, I learned and laughed each day of my PhD because of them, and with them. Thank you Hideo, Robert, Benjamin, Martin and Ronan for valuable advice and encouragements. I also want to thank Peggy, Susanne, Beata and Punitha for being amazing friends. They were my family in Glasgow, and I will always cherish moments spent with them.

Thank you Sunil for making my life special and meaningful. Your love and support gave me strength to face the uncertainties of a PhD life. Thanks for being there.

Finally, my parents R.D. Singh and Meena Singh – words cannot truly express how much you mean to me. You gave me life and have done nothing but support me throughout it. Thank you for being the best parents a daughter could ever wish for, I love you with all my heart today and forever more. This one is for you–mom and dad!

Table of Contents

List of Figures

List of Tables

Part I

Introduction, Background and Problems

Preface

In this part an introduction to the thesis and the outline of the structure is provided (chapter 1) where, the concept of aggregated search will be introduced. Why investigating aggregated is important is also explained in the introduction chapter. The background material supporting the work presented in this thesis is presented next (chapter 2).

Chapter 1

Introduction

With the introduction of search engines, information access on the World Wide Web became easier and faster. Today, search engines offer sophisticated means to describe our information need, and allow us to search the vast amounts of online information that includes almost everything under the sun. As a result, increasingly people use search engines as a "primary means" to seek information. In the year 2009, it was estimated that Google received 34,000 searches per second, Yahoo! received 3,200 searches per second, and Bing received 927 searches per second[1-1]. These huge number of searches demonstrate increased usage of search engines; how people depend upon them; and hence their importance.

Not only is there an increase in the use of search engines to retrieve information on the World Wide Web, but also, the amount of information available on web continues to increase. The size of the Web was estimated to be 2.82 billion pages[1-2] in November 2010. The web is growing continuously, not only in terms of the amount of content, but also in the diversity of content. More and more multimedia (image, video, etc) and different types of documents (blogs, Wikipedia, etc) are being added every year.

In order to facilitate information access, search engines are now providing access to diverse data in a unified manner, called **aggregated search**. An aggregated search interface is designed to aggregate retrieval results from different information sources (image, video, maps, etc) into a single result page. An example of an aggregated result page can be seen in Figure 1.2 where, *map, image, Wikipedia and web* results are aggregated on one result page.

Prior to aggregated search, a conventional way of gathering relevant information from

[1-1] http://searchengineland.com/comscore-us-most-searches-china-slowest-34217
[1-2] http://www.worldwidewebsize.com/

several information sources (e.g., web, image, news, wiki) was to browse the search results from the individual sources separately, available through dedicated search engines (image search, news search, etc.). However, in aggregated search, users do not have to visit search engines separately to browse the search results to obtain a range of retrieved items. Therefore, aggregated search can be seen as an emerging paradigm that aims to facilitate information access from various sources (as shown in Figure 1.1).

This thesis is about the presentation of results from search when the search occurs across multiple physical sources or servers, each of them containing documents and media of a specific type (e.g., video, image, text). This is not specific to aggregated search, but a very important aspect of aggregated search. Such situation also arises in other web contexts, such as in digital libraries, where results from various sources and of different media can also be retrieved and thus need to be put together to form answers. This thesis contributes to the understanding of result presentation in the context of aggregated search, which is the focus of this thesis. For this purpose, this thesis looked at various aspects, depending on the area of investigation. For instance, for the interface related investigations, the effect of position of different types of results such as: image, video, and news on users' click-behavior is analyzed. Whereas, when determining suitable selection of sources, the relevance of the physical sources is taken into account.

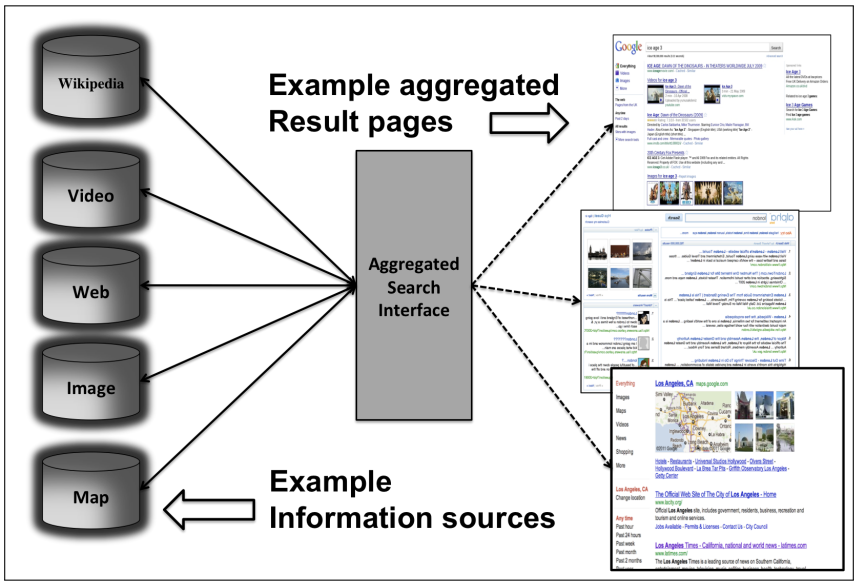


FIGURE 1.1: Example of an aggregated search, where results from various physical sources are retrieved and aggregated together to be presented on a single result page.

FIGURE 1.2: Example of an aggregated search result page by Google for the example query "Madonna". In this example, news, image, Wikipedia and web results are aggregated on a single result page.

1.1 Motivation

There were three main motivating factor for studying aggregated search: its increased usage, users often seeking information on general or broad topic, and the lack of rigorous or empirical studies of the effect of aggregated search on users.

1.1.1 Increased Usage

There are reported evidences that suggest the positive usage of the aggregate search, and three such evidences are discussed here.

First, iProspect[1-3], a search engine marketing firm reported the increased usage of image, video and news results after being aggregated within search result page. They conducted two surveys[1-4] with 2,218 US search engine users. During the first survey, users were asked how often they accessed image, video and news results using their respective verticals (for example, image search, news search, etc). The results showed

[1-3] http://www.iprospect.com/

[1-4] http://www.iprospect.com/about/researchstudy_2008_blendedsearchresults.htm

that the vertical search was not prevalently used (as shown in Figure 1.3). That is, a very low percentage of users (26% for image, 17% for news and 10% for video) actually used the dedicated verticals to access image, video or news results.

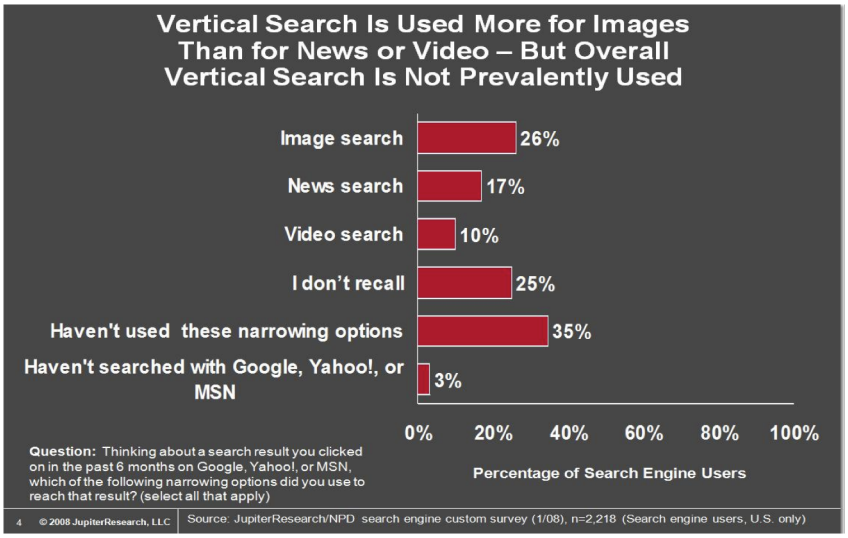


FIGURE 1.3: The figure is taken from the article[a]

[a]http://www.iprospect.com/about/researchstudy_2008_blendedsearchresults.htm

In the second survey, users were then further asked; *"within the last 6 months, when performing a general search within Google, Yahoo! or MSN (not using the narrowing options described in the previous question), which of the following types of results have you clicked on? "* That is, they were asked how often they clicked on image, video or news results, when provided on the aggregated result page.

Results from the second survey showed that with aggregated search, users accessed more image, video, etc results when compared to conventional way of searching results; that is, by using dedicated verticals. Thus, results show an increase in the click-rate on image, video and news result can be seen in Figure 1.4. It can be seen that, there is an increase in the percentage of image results accessed (from 26% to 31%). Similar observations were also made with respect to the video and news results.

Second evidence suggesting positive usage of aggregated search was that, Google Maps saw a 20% increase in its traffic after being included into its Universal search[1-5]. That is, more users were accessing maps after it was aggregated on the search result page.

Finally, third evidence suggesting usage of aggregated search results was reported dur-

[1-5]http://searchengineland.com/google-maps-gaining-on-market-leader-mapquest-13103

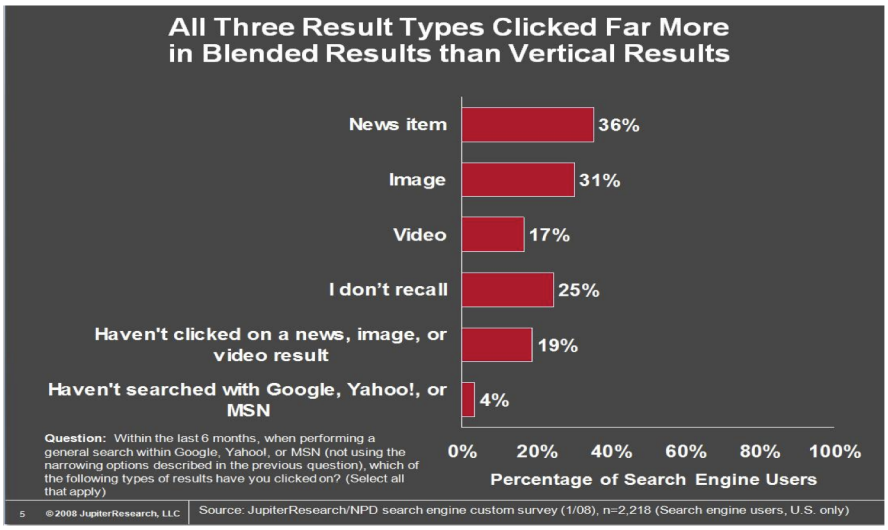


FIGURE 1.4: The figure is taken from the article[a]

[a]http://www.iprospect.com/about/researchstudy_2008_blendedsearchresults.htm

ing a joint Webinar[1-6] 2010 where, search engine market experts from comScore[1-7] and RankAbove[1-8]. The webinar looked at the growth of blended / universal search. The statistics reported during the seminar showed that the aggregated interface garner much higher click-through rates than regular results; 38% for aggregated, as compared to 12% on regular result pages, that is, when no aggregation was performed. Furthermore, it was reported that one-third of all the searches displayed an aggregated result page by most search engines.

The outcomes of the three reports discussed above suggest that aggregated search approach is now frequently followed by many search engine companies, and due to aggregation of rich media data on the search result page, results like image, video, maps and news are attracting users' click-through rates.

1.1.2 Users seeking information on broad topics

The second motivating factor for investigating aggregated search was that now people are often seeking general information on a broad topic such as "global warming" or "nutrition". [Jansen et al., 2008] reported that, 80% of the queries submitted on the web are informational in nature; and characterized informational queries as;

[1-6]http://www.slideshare.net/rankabove/com-score-rankabove-final
[1-7]www.comscore.com
[1-8]www.RankAbove.com

"The intent of informational searching is to locate content concerning a particular topic in order to address an information need of the searcher. The content can be in a variety of forms, including data, text, documents, and multimedia."

Information needs behind such informational queries are often satisfied by relevant information collected from multiple documents or media. Due to the increased quantity and diversity of multimedia contents available on the Web, images, audio, videos are also becoming relevant to many queries. An aggregated search approach facilitates access to multiple information sources (image, news, blog, etc) in a single interface, and without having to visit separate search engines. Therefore, aggregated search can be a suitable approach in accessing diverse information, and hence satisfying needs, which are informational in nature.

1.1.3 Lack of empirical evidences

Finally, the third motivation behind this work was the lack of empirical evidences investigating aspects of aggregated search reported in the research community. It is known that aggregated search is now being used by large number of users, and that there is need for such type of approach in order to satisfy users looking for information on broad topics. However, it is not know how and if at all this new approach of providing search results are influencing users' information seeking process. Apart from few reports released by search engine marketing companies, there is not much reported research investigating aggregated search. There are very limited (or none) scientific evidences reporting aspects of aggregated search and its influence. Therefore, the aim of this thesis was to establish an empirical understanding of aggregated search and its effect on users' behavior.

1.2 Problem

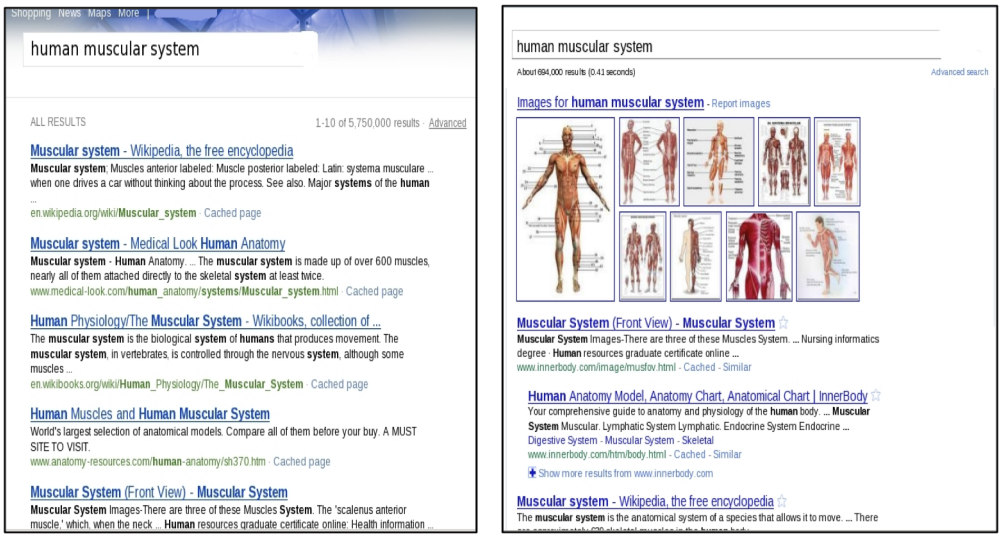
As discussed in previous sections, aggregated search provides a richer media experience to users, and by using only a small number of keywords or phrase. It does so by without mentioning the type of media, the results provided aims to cover everything relevant[1-9]. Furthermore, the surveys outlined in the previous Section suggest that, aggregated search seems to be a promising approach to facilitate access to diverse information; and is being used by large number of people.

[1-9]http://www.avangate.com/articles/universal-search-102.htm

However, influence of this *new* search approach on users' information seeking process is still not known. In-spite of its heavy usage, aggregated search still remains a *black-box* – in terms of its influence on search behavior, usefulness in task completion, and the overall user experience. In other words, users' interaction with this new search approach remains uncovered and there are many unexplored questions in this area.

In contrast, there have been a dedicated body of work through which, it has been possible to establish an understanding of various aspects of conventional web search; i.e., how users' information needs and searching behavior evolve, what factors affect click-behavior, architecture, interface design and issues [for example, Nielsen, 1993; Spink et al., 2002; Jansen and Spink, 2005; Hearst, 2009; Mateosian, 2010].

Although, aggregated search can be seen as an extension of conventional search approach. It is not clear if knowledge of conventional search engine implies within an aggregated search scenario as well. For instance it is known that, users' clicks are biased towards top ranking results on a conventional result page, but it is not known if it also the case for an aggregated result page. Furthermore, it is possible to expect that the design, implementation, and evaluation of aggregated search will be more complex and difficult when compared to the conventional search.



(a) Conventional search interface (b) Aggregated Search interface

FIGURE 1.5: Difference between conventional search interface and an aggregated search interface. Image results being added with web results in an aggregated interface.

For instance, before aggregated search was introduced, typing in a query like; *human muscular system* into a conventional search engine would have produced result page similar to Figure 1.5(a), where *only* web results were retrieved in response to the submitted query. Nevertheless, when the same query is submitted to an aggregated search,

it is highly likely that a result page similar to Figure 1.5(b) will be produced. From the Figure 1.5(b) it can be seen that there is an addition of image results with the web results, when compared to result page provided by conventional search engine. This addition of media data (image, video) or different document type (blog, twitter, etc) makes an aggregated search interface distinct from the conventional search interface, hence may cause users to interact differently.

Therefore, there is a need to explore aggregated search in terms of the effect of its various aspects such as: click-behavior, result presentation, search intent, etc. Broadly, four main aspects of a search system need to be explored in order build an overall understanding of it (as shown in Figure 1.6). (1) How useful is the system in finding relevant information? (2) How should relevance be determined or measured? (3) How users are expected to interact with the results when using the system? and (4) How can the interface be designed effectively. Furthermore, formal introduction to this emerging search approach is also missing from the literature; there is a lack of clear terminology associated with the aggregated search; and no framework in which the aggregated search can be described.

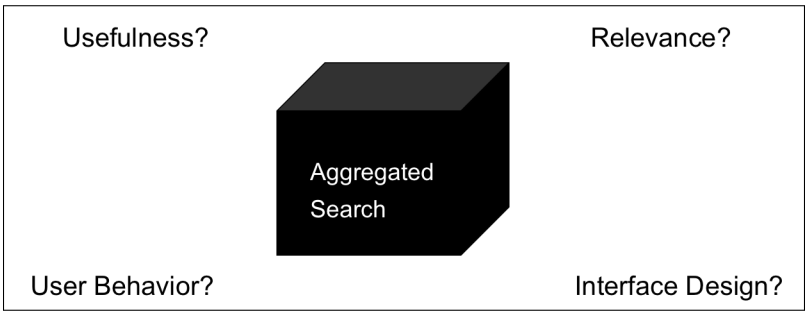


FIGURE 1.6: Aggregated Search remains a black-box, in terms of its effect on user behavior, usefulness, relevance, design space, etc.

1.3 Thesis Statement and Research Questions

This thesis aims to explore the aggregated search from the users' perspective. The thesis primarily focuses on understanding and describing the phenomena related to the users' search process in the context of the aggregated search. The goal is to participate in building theories and in understanding constraints, as well as providing insights into the interface design space. In building this understanding, the thesis focuses on the click-behavior, information need, source relevance, and dynamics of search intents. The understanding comes partly from conducting user studies and, from analyzing search

engine log data. In particular, the following five research questions are investigated in this thesis:

Res1: ***Does aggregation of results from different sources facilitate in task completion?***

For over a decade, web search engines has been increasingly used by number of users to satisfy their information need. In addition, it is well known that web search has proved to be beneficial in task completion by proving online information to its users. Through several dedicated studies [for example, Nekrestyanov and Panteleeva, 2002; Xu and Mease, 2009; Ali et al., 2009] attempts to determine the usefulness and effectiveness of web search have also been made. However, whether the aggregated search approach is also facilitates in task completion is not known, in particular for informational search tasks. The informational search tasks involve users to collect information from multiple source [Jansen et al., 2008]. Moreover, since aggregated search aims to provide unified access to multiple sources, therefore in this thesis, it is hypothesized that search result aggregation is most useful for supporting informational search tasks.

Res2: ***Which non-topical characteristics of results contribute to the relevance in the context of aggregated search?***

Although, relevance judgments are fundamental to the design and evaluation of all information retrieval systems, information scientists have not reached a consensus in defining the central concept of relevance [Schamber et al., 1990]. For some people, it is equivalent to retrieval effectiveness (i.e., precision and recall), for others such as IIR researchers, it is a multidimensional concept [e.g. Park, 1997; Mizzaro, 1998; Cosijn and Ingwersen, 2000; Borlund, 2003]. While, some also look at the characteristics of the document (type, format, etc) to determine the relevance [for example, Roussinov et al., 2001; Boese, 2005; Freund and Berzowska, 2010; Freund, 2010]. Therefore, it is important to recognize that it is difficult to have a universal interpretation of what relevance is. Within the context of the aggregated search, what attributes to the relevance of a result is not known. The aim of this research is to obtain an understanding of relevance within the context of aggregated search.

Res3: ***How does aggregated result presentation influence users' click-behavior (if any)?***

Search engine interface is an important component of a search engine, as it enables users in two major activities: formulating a query to represent an informa-

tion need, and finding needed information from a list of possibly relevant documents as retrieved by a search engine [Ali et al., 2009]. Therefore, in order to build effective search engine, understanding interface related factors that effect users' searching behavior is important. The goal of the third research question is to investigate factors effecting click-behavior in an aggregated interface.

Res4: ***How do users perform search when their information need requires results from multiple sources (e.g., image, Wikipedia, news, video)? Is there a pattern in searching behavior for such information needs?***

Understanding the users' searching behavior is one of the important goals in web search, and information on how users interact and view search results helps improving the user information seeking process [e.g., White and Morris, 2007b; White and Drucker, 2007]. This is because, having an insight on searching behavior of users have proved to be useful in building better information retrieval systems [e.g., Halvey and Jose, 2009; Agichtein and Zheng, 2006; Bilenko and White, 2008]. In context of aggregated search, understanding users' searching behavior when looking for information from different sources is important. As this would help in achieving suitable aggregation of different results. Furthermore, knowledge of associated searching patterns (if any) can prove to be useful in making the real-time aggregation effective. For instance, if known that users after viewing news results often click on video results. Then such information of result viewing pattern (news $\rightarrow$ video) would indicate that importance of including video results would increase as the search session progresses after the user has recently viewed news results. Recent users' actions can be easily captured through implicit feedback techniques[1-10] (e.g., click-through, skips, page reads, etc). The aim of the fourth research question is to uncover users' searching behavior with respect to underlying source related information need.

Res5: ***How does aggregation of different sources affect design of an aggregated interface. That is, what design issues exists pertaining to an aggregated search interface?***

There is a strong association between interface design and system satisfaction. For interface designers, it is challenging to incorporate the effective search technology into their design [Mateosian, 2010]. The aim of the last research question

[1-10]Implicit feedback algorithms utilize interaction between searchers and search systems to learn more about users' needs and interests than expressed in query statements alone. This additional information can be used to formulate improved queries or directly improve retrieval performance [Melucci and White, 2007].

is to identify some of the key design issues that may exits within an aggregated interface, so that suitable design guidelines could be proposed.

1.4 Contributions

This thesis makes following contribution in the context of the aggregated search:

- This thesis explores the new emerging aggregated search. At the time of writing, it is the first contribution towards the understanding this search approach. Therefore, this work provides a foundation and many future directions for research in this area.
- Aggregated interface related factors affecting click-though behavior are identified, which are crucial from an interface design perspective (Chapter 4).
- The notion of **source-orientation** is introduced, that identifies the multi-dimensional nature of relevance in the context of aggregated search. It refers to the degree to which documents *from* a specific source would be relevant in completing the corresponding search task (Section 4.1.2, Chapter 4).
- An effective **combination method**, using rule based and machine learning techniques is developed to identify the relevance of the source to the given information need (Chapter 6).
- A **random model** methodology is introduced to derive meaningful statistics from log-data analysis. The random model provides a novel evaluation methodology that allows the comparison of results obtained from log-data, in the absence of any previously computed statistics or baseline (Section 6.5, Chapter 6).
- The dynamics of users' searching behavior with respect to different information needs are investigated (Chapter 6).
- A set of **design guidelines** for an aggregated search interface are provided (Section 7.5, Chapter 7).

1.5 Thesis Outline

This thesis is structured in four main parts, which contains corresponding chapters:

- **Part I: Introduction, Background and Problems**
  This part comprises of three chapters. It introduces the concept of aggregated search and provides background material supporting this work. First, the overall aim and outline of the thesis is provided in Chapter 1. Next, related research is discussed in the Background chapter 2.

- **Part II: Result Presentation in Aggregated Interface**
  This part presents two user studies, which addressed result presentation issues in aggregated interfaces. Chapter 3 (study one) compares an aggregated interface to a non-aggregated interface during information access for informational queries. Another comparison between different types of aggregated interfaces is performed in the second study in Chapter 4, where factors affecting click-through behavior are analyzed.

- **Part III: Source-Orientation in Aggregated Search**
  In this part, the dynamics of user behavior with respect to different source intents are analyzed using large-scale log data from Microsoft. First in Chapter 5, effective ways to identify the source intent behind a user query are presented. Once the intents are correctly identified, they are then further exploited in Chapter 6 to uncover the behavioral patterns of users in an aggregated search scenario.

- **Part IV: Conclusion and Future Work**
  This part summarizes Part II and Part III, and discusses the overall findings in Chapter 7. The conclusions drawn from the user experiment in Part II and the results from the Log analysis in Part III are described in Chapter 8, where avenues for future works are also discussed.

1.6 Publications

This thesis also yielded several publications:

- **Factors Affecting the Click-Through Behavior on Aggregated Search Interfaces**, S.Sushmita, H. Joho, Robert Villa and M. Lalmas. *CIKM, The Conference on Information and Knowledge Management, 2010.*

- **Analyzing Domain and Genre Intents in Web Search**, S. Sushmita, B. Piwowarski and M. Lalmas. *AIRS, The Asia Information Retrieval Societies Conference, 2010.*

- **A Task Based Evaluation of an Aggregated Search Interface**. S. Sushmita, H. Joho and M. Lalmas. *SPIRE, 16th edition of the Symposium on String Processing and Information Retrieval 2009.*

- **Understanding Domain "Relevance" in Web Search**, S. Sushmita, H. Joho, M. Lalmas and J. Jose. *WSSP, 18th World Wide Web Conference workshop on Web Search Result Summarization and Presentation 2009.*

- **Using Digest Pages to Increase Users' Result Space : Preliminary Design Issues**, S. Sushmita, M. Lalmas and T. Tombros. *SIGIR, Special Interest Group on Information Retrieval, workshop on Aggregated Search, Singapore, 2008.*

1.7 Summary

This chapter introduces the emerging area of aggregated search and explained why it is an important topic to study. The research questions addressed in this thesis are also described in this chapter. Furthermore, several challenges are outlined and finally the main contributions of this thesis are discussed. In the following chapters various aspects of an aggregated search is investigated.

Chapter 2

Background and Related Work

The increase in the amount of available information on World Wide Web is remarkable. This plethora of information has continuously challenged researchers in devising suitable means for making them easily accessible to the end users. Providing easy access to this large amount of information involves three broad processes: retrieval, organization and presentation.

Given the collection of information, first, the relevant information needs to be fetched, second, the fetched information needs to be organized so that identifying topic of interest is easier; Then finally, an appropriate presentation of this organized information is required for them to be used by the users.

Once the means to organize and present these information is devised, it is then required to monitor on how effective or beneficial these methods are in information access. This helps to ensure the usability and effectiveness of the devised methods, and to identify limitations that may exits. To achieve this, users' response or feedback when using these support tools needs to be taken into account. In Web search, a common practice is to observe users' behavior during information seeking process and to understand benefits and limitations of the provided means.

The purpose of this Chapter is to present background material that supports this thesis. The aim of this work is to investigate various aspects of an aggregated search and to explore open issues or problems existing within it (if any). More specifically, the aspects of relevance, result organization, presentation and user behavior in an aggregated search scenario are investigated. In this Chapter, a background knowledge and related work on the dimensions of relevance, result organization, presentations and user behavior is provided.

This Chapter is structured as follows: Section 2.1 will discuss the concept of relevance

and various interpretations associated with it. While, standard approaches used by various information retrieval systems to aggregate results (or information) are presented in Section 2.2. How interface related issues have influenced users' information seeking behavior is discusses in Section 2.3. Furthermore, an overview of observed user behavior in context of various information retrieval systems is described in Section 2.4. Where appropriate the content of this chapter motivates, and is related directly to, the work presented in later chapters in this thesis.

2.1 The concept of Relevance

The concept of "relevance", sometimes also called "pertinence" or "aboutness", is central to the theory of information retrieval [Cooper, 1971]. In general, relevance can also be interpreted as, how "good" a retrieved result is with regard to the information need[2-1]. It is considered a key element in IR, and is often an important criterion in measuring the effectiveness of an IR system. Therefore, the aim of a retrieval system should be to retrieve all the relevant documents, and at the same time retrieving as few non-relevant documents as possible [Rijsbergen, 1979]. Therefore, estimating relevance is critical for IR systems.

Although, relevance judgments are fundamental to the design and evaluation of all information retrieval systems, information scientists have not reached a consensus in defining the central concept of relevance [Schamber et al., 1990]. For some people, it is equivalent to retrieval effectiveness (i.e., precision and recall), for others such as IIR researchers, it is a multidimensional concept [e.g. Park, 1997; Mizzaro, 1998; Cosijn and Ingwersen, 2000; Borlund, 2003]. While, some also look at the characteristics of the document (type, format, etc) to determine the relevance [for example, Roussinov et al., 2001; Boese, 2005; Freund and Berzowska, 2010; Freund, 2010]. Therefore, it is important to recognize that it is difficult to have a universal interpretation of what relevance is.

In the following sections, relevance and associated measures are described. Section 2.1.1 will discuss the system and the user based relevance measures, which are often used during standard IR evaluations. Section 2.1.2 will discuss how document properties can be utilized for the relevance assessment. Concept of relevance within the context of an aggregated search is explained in Section 2.1.3. Finally, Section 2.1.4 concludes the overall discussion.

[2-1] http://en.wikipedia.org/wiki/Relevance_(information_retrieval)

2.1.1 Relevance and IR Evaluations

In the early days of information retrieval, relevance expressed a criterion for assessing effectiveness in retrieval of information. Therefore in traditional IR models, relevance was considered a property of the system. That is, a system was judged how it acquired, represented, organized and matched the texts to the information need. Therefore, traditional IR systems focused on evaluating different approaches or algorithms based on how well they retrieved relevant results. Majority of IR evaluation studies, from Cranfield studies in the late 50's and early 60's to Text Retrieval Conference (TREC) evaluations in 1990s, are based on this framework for considering the nature of relevance [Saracevic, 1996b].

The Cranfield model [Cleverdon et al., 1966] is a classic example of the system-oriented approach to IR system performance and effectiveness evaluation [Hildreth, 2001]. Cranfield made two main assumptions: first, the users prefer to view results (documents) that are relevant to their search queries; and second, the document relevance to a query is perceived as the property of the document.

For decades, the notion of relevance was used as a measuring criteria for effectiveness of a retrieval system, and users' interactions with the system was not considered. However, interactive information retrieval (IIR) researchers highlighted the importance of users' role during information seeking process [for example, Ingwersen, 1996; Mizzaro, 1998; Robins, 2000; Borlund, 2003].

Cosijn and Ingwersen [2000] argued that the cognitive model for interactive IR comprises three of elements – systems, users, and the environment. The system involves documents or information objects, which are organized in different ways. The user typically has a problem or a work task to perform. The socio-organizational environment provides the context or situational framework influencing the activities of the user.

Since the traditional IR model did not reflect interaction, therefore in later research, a number of efforts were devoted to the development of IR models that incorporated the rich and complex nature of IR interactions. The prime weakness of the system-based relevance is that it was completely one-sided. It did not incorporate in any way anything from the users' interaction, except the query. It did not consider elements, variables, and context related to the user and their use, nor did it reflected the dynamic, interactive nature of IR as practiced. Therefore, situational, psychological, motivational relevance etc emerged as a reaction and challenge to the system-based relevance approach [Saracevic, 1996b].

In order to address the challenges of system-based relevance approach, Saracevic [1996b]

expressed relevance as a relation, and suggested that different theories (which he refers to as manifestations) of relevance encompass different relations. He listed five manifestations of relevance [also called as *"attributes"* of relevance by, Cosijn and Ingwersen, 2000] based on different relations as follows:

- **System or algorithmic relevance:** The relation between a query and information objects (texts) which are retrieved or failed to be retrieved by a given procedure or algorithm.This manifestation of relevance focuses on system or algorithm measures, whereas, the following four take users' interactions into account.
- **Topical or subject relevance:** The relation between the subject and topic expressed in a query, and topic or subject covered by retrieved texts. This relation is also system-oriented largely because the success of the relation depends on the system's input policy, as well as its indexing and searching ability to retrieve relevant objects. However, success also depends on the formulation of the request by the user, transformed into a query by the system [Cosijn and Ingwersen, 2000].
- **Cognitive relevance or pertinence:** The relation between the state of knowledge and cognitive information need of a user, and the information retrieved. Cognitive correspondence, informativeness, novelty and information quality are criteria by which cognitive relevance is inferred.
- **Situational relevance or utility:** The relation between the situation, task, or problem at hand, and texts retrieved by a systems or in the file of a system, or even in existence. Usefulness in decision-making, appropriateness of information in resolution of a problem, reduction of uncertainty, and the like are criteria by which situational relevance is inferred.
- **Motivational or affective relevance:** The relation between the intents, goals, and motivations of a user, and texts retrieved by a system or in the file of a system, or even in existence. Satisfaction, success, accomplishment, and the like are criteria for inferring motivational relevance.

Mizzaro [1998] further built upon Saracevic [1996b]'s concept of relevance as the manifestations of topical, cognitive and situational relevance, and suggested four-dimensions of relevance. He defined relevance as a four-dimensional relationship between an *information resource* (surrogate, document, and information) and a *representation of the user's problem* (query, request, real information need and perceived information need). Which is then judged according to one or more of the following *components*: topic,

task, or context, at a particular point in *time* [Cosijn and Ingwersen, 2000]. Graphical representation of Mizzaro [1998]'s four-dimensional representation of relevance can be seen in figure 2.1.

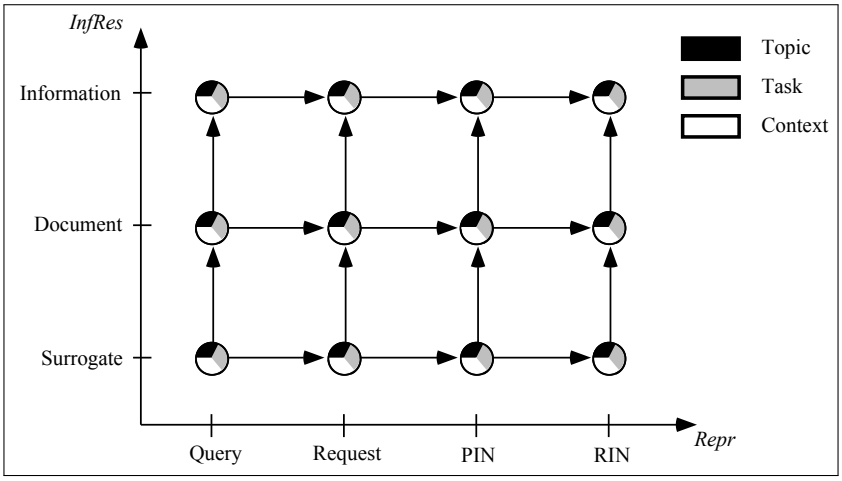


FIGURE 2.1: The Various kinds of relevance proposed by [Mizzaro, 1998]

Therefore, dimensions of relevance can be seen as the "attributes" based on which, the relevance of the result to the given information need is determined. So far in the web search, relevance is often estimated based on single attribute; topic, theme, time, category, and sometime based on the overlap of two or more attributes. For instance, location and theme as in Geographic Information Retrieval [Bucher et al., 2005]. However, within the context of the aggregated search, what attributes to the relevance of a result is not known. The aim of this research is to obtain an understanding of relevance within the context of aggregated search.

2.1.2 Relevance and Genre

Another important criteria for measuring the relevance are the genre of the document or the result. Genre is essentially a document type, which is judged based on purpose, form, and the content of the document [Rosso, 2008]. Kwaśnik and Crowston [2004] defines genre of a document as an information about its form and its purpose intended for a specific communication. Genre also includes a notion of social acceptance where, the genre defined for the community is recognized within a given discourse community.

A user's 'information need' is not only based on a particular topic, but it can also depend on many other features of a document retrieved. This is because, sometimes a user can be looking for a document for a specific purpose, for instance, school project or

technical review. At the same time the user might be looking for a document in a particular format, such as images (jpeg, png, etc), or a journal article (pdf). Therefore, the relevance of a document to a user's information need incorporates the topic area, as well as many other features that pertain to the genre of the document Boese [2005].

There have been dedicated bodies of work investigating the relation between the genre and the task relevance [For example, Freund et al., 2006; Rauen, 2009; Freund and Berzowska, 2010; Freund, 2010; Braslavski, 2011].

Freund performed several investigations to uncover the relation between relevance of a document type to the search task [e.g., Freund et al., 2006; Freund and Berzowska, 2010; Freund, 2010]. For instance, strong relation between appropriateness of a document and the task was suggested in [Freund et al., 2006], where, workplace search system that uses relationships among different tasks and genres to filter the search results were described. Freund [2010] further suggested that there is a need to find ways to strengthen real life communication among the cognitive actors. For this purpose, making use of genre in IR systems was proposed as a means to increase relevance.

While Rauen [2009] suggested that, for determining the relevance relations between text and context, genre structures provide a discursive context that cognitively focuses the attention of writer and reader both. This sets the relevance constraints and hence increases the communication efficiency. Therefore, a two-way influence between relevance and genre can be seen. Where, the relevance relations guide the generic structures; and generic structures guide relevance relations.

Furthermore, researchers like Roussinov et al. [2001] and Kwaśnik and Crowston [2004] explored the advantages of using genre features to improve web search retrieval algorithms. Roussinov et al. [2001] showed how recognition of the genre of a Web document can improve the quality of web searches. Through the study, Roussinov et al. [2001] identified genres which most/least frequently meet searchers' information needs. His study also suggested that certain genres are better suited for certain types of information needs.

While, Kwaśnik and Crowston [2004] examined whether it is possible to effectively incorporate document genre features into document relevance ranking. The evaluation of the results showed only moderate improvements. This was because some queries are very receptive to mixing genre ranks with traditional keyword relevance ranks, while other did not respond well to the genre information.

Despite the consistent efforts by several researchers, using genre information for determining relevance is still challenging. This is because, there is no clear and distinct

universal classification schema that covers all possible genre that is possibly available on the World Wide Web. Kwaśnik et al. [2006] pointed that, "one of the challenges of studying genre in general is that there does not seem to be a consensus on what a genre is, what qualifies for genre status, how genres "work," how we work with genres, how genres work with each other, or how best to identify, construe, or study genres. As a result, genres are recognized and used, but not so readily described and defined."

The issue gets profound in the context of web search, as new genres are still evolving. Due to the print document genres being adapted to the Web, and new electronic genres emerging frequently, the genres are continuously being shuffled, disassembled and then put together again. This results in a chaotic and ambiguous genre classification series. Furthermore, due to multimedia data being added and technologies converging, there arises a challenge of combining and recombining genres of many different kinds in inventive ways and for unexpected purpose [Kwaśnik and Crowston, 2005].

However, most researchers in this area, find genre to be a promising instrument for capturing the complexity of human interactions [Kwaśnik and Crowston, 2005]. This is because genres provide an efficient way of dealing with document starting from its creation, storage, retrieval and finally to its utilization. Knowledge about a document's genre, and therefore its intended utility. For instance, it helps a user in query formulation, or recognize the relevance of a document that is presented as the result of processing that query [Crowston et al., 2010].

Since an aggregated search aims to aggregate search results from multiple sources, this results in the combination of different types of documents and media data. As a result, genre of the result can also contribute to the relevance measurements in aggregated search. Therefore, genre is an important aspect and incorporating genre information within such search approaches can be beneficial. For instance, if known that a video and a map result is relevant to the topic of search (query), corresponding sources (video and map in this case) can be selected in order to fetch the desired results. This thesis investigates the relevance of the source to the search task, where the sources are the collection of specific genre (document or media). Furthermore, once the results of the relevant genres from their respective sources are fetched, their aggregation effect on users' searching behavior is also investigated. Next section will further elaborate on the relation between a genre and source, and which non-topical characteristics contribute to the relevance of a result within the context of aggregated search.

2.1.3 Relevance in Aggregated Search

An important task of an aggregated search system is to determine, for a given query, not only whether a document/surrogate/information is relevant, but also from *which* sources the relevant documents, should be retrieved. For an aggregated search system, finding the relevant results involves a three-stage process:

1. Selecting the relevant sources (image, news, video, maps, etc) from which the results are to be fetched;
2. Retrieving relevant results from the selected sources;
3. Merging the retrieved results and presenting them on a single result page.

This process is similar to the distributed information retrieval scenario, that is, *selection, retrieval* and *merging*. The concept of distributed information retrieval and its similarity to aggregated search will be discussed in Section 2.2.3. Here, we focus on the importance of source selection, for which, the relevance of the source to the given search task plays an important role.

In the context of aggregated search, a source is usually a collection of similar type of media or document. For instance, an image source will typically contain images, while a map source will possess maps only. The work presented in this thesis is based on the assumption that the genre of the result will also match to the genre of the source. For example, a news source will provide documents of genre news, a Wikipedia source will provide document of genre Wikipedia, and so on.

Due to the increase in diversity of content on the web, not only traditional html web results, but results like – twitter, blog, image, video, etc are also becoming relevant to the users' information need. An aggregated search aims to select the potential sources that would provide the required answer (in the form of document or media), and which will be judged by the users depending on their context or topic of search.

Therefore, for an aggregated search scenario, Mizzaro [1998]'s definition of relevance can be extended and restated as – "relevance is a four-dimensional relationship between an *information resource* (surrogate, document, information and **source**) and a *representation of the user's problem* (query, request, real information need and perceived information need). This relationship is then judged according to one or more of the following *components*: (topic, task, or context), at a particular point in *time*".

Thus, **source** can be incorporated as the fourth element to the *information resource* set. The graphical representation of the extended four-dimensional relevance space can be

seen in Figure 2.2. The values in each dimension set are as follows (as obtained from Mizzaro [1998]):

1. $InformationResource = \{surrogate, document, information, \textbf{source}\}$;
2. $Representation = \{Query, Request, PIN, RIN\}$;
3. $Time = \{t(rin_0), t(pin_0), t(r_0), t(q_0), t(q_1), t(r_1)..... .....t(r_m), ...t(q_n), t(f)\}$;
4. $Component = \{\{Topic\}, \{Task\}, \{Context\}, \{Topic, Task\}, \{Topic, Context\}, \{Task, Context\}, \{Topic, Task, Context\}\}$;

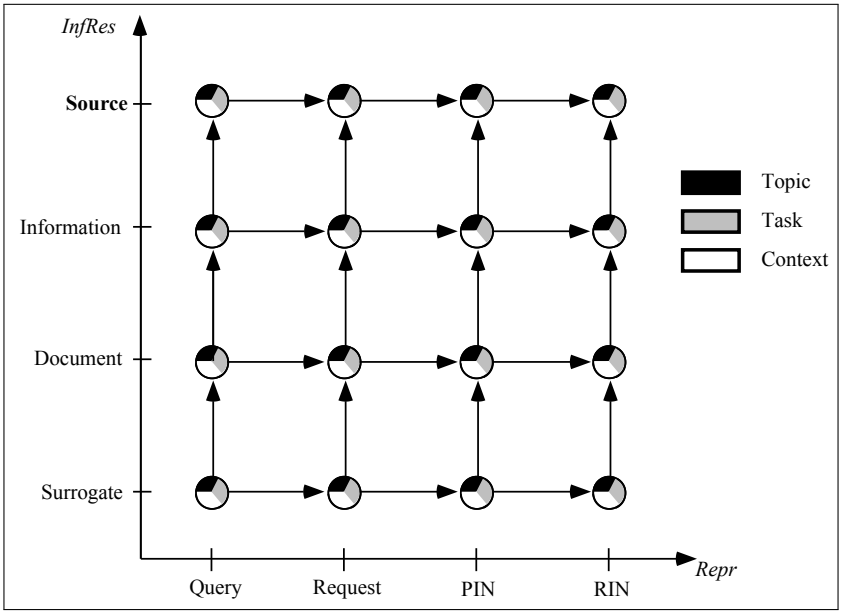


FIGURE 2.2: Extending Mizzaro [1998]'s dimensions of relevance. For an aggregated search scenario, element **source** is added to the information resource set, where **PIN** is the *Perceived Information Need* and **RIN** is the *Real Information Need*.

Consider two example search scenarios: (1) a school kid looking for information for a science project on "solar-system"; (2) an astronomical scientist looking for technical papers for current statistics related to the "solar-system". In both cases, it is possible to have similar search query – solar-system. However the task is different – preparing a school project and a technical review. For the first example, images of the planets and the solar systems might be very useful, but it is likely that, image results may not be required for the latter. Therefore for the query (solar-system) and the task (school

project), source (*image*) would be relevant. While, for the query (solar-system) and the task (technical review), source (*Wikipedia*) or a journal article might be more appropriate.

The research presented in this thesis is motivated to investigate the relevance of the source to the search task, and its effect on users' searching behavior. To capture this aspect, this thesis introduces the notion of ***"source orientation"*** of an information need, which refer to the degree to which results *from* a specific source would be relevant to complete the corresponding search task. In particular, the studies presented in this thesis investigates how and if relevance of the source to the search task affect users' interaction. Moreover, if so, how can the relevance of the source be determined.

2.1.4 Summary

In this section, background material on the concepts of relevance in information retrieval was provided. It was shown how different interpretations of relevance exist, and that there is no universal definition or understanding of the concept of relevance. However, the IR evaluation measures often use concepts like: system-based relevance – which lies in its pragmatism for systems applications and evaluation; and user-based relevance – where relevance is seen as the manifestations of topical, cognitive and situational relevance. This section also discussed how document genre can be used as an indicative for the relevance measurement, and that the source and genre are closely related in the context of aggregated search. Finally, estimating the relevance of the source to the search task was proposed. For this purpose, the notion of source-orientation is introduced which adds another element (source) to the relevance space in the context of an aggregated search. Next, various information aggregation approaches is described.

2.2 Result Organization and Aggregation

Once the relevant results are fetched, the next step is to effectively organize these relevant information, in order to facilitate the information access. A conventional search engine accepts queries from users and responds with a long list of ranked documents. In this paradigm, users have to manually select documents of interest from the returned list of documents. Such an approach of producing search results may be quick but it requires repeated effort of 'select' and 'click' from users to satisfy their information need.

Various studies (citep, [Spink et al., 2002; Jansen et al., 2000; Jansen and Spink, 2005, 2006]) showed that in the context of web search users access very small number of documents and are less tolerant to the long list of mixed documents. It was also suggested that selecting information of interest from a long list of mixed search results is a tiring and time consuming job and is generally not appreciated by users.

Therefore, it important to find methods to effectively organize and aggregate the retrieved results to facilitate identification of relevant information quickly, and with less effort. Document clustering and faceted browsing are existing examples, which are used to organize results in order to facilitate information access. Furthermore, federated search, distributed information retrieval and metasearch are other techniques that aim at aggregating results from various sources. The following sections discuss these techniques in more detail.

2.2.1 Clustering

Over the years there have been various organizing approaches (e.g. [Croft, October, 1978; Leouski and Allan, 1998; Tan et al., 2004]), which are based on visualization and presentation of relationship among the documents, users query terms, etc. Zamir in [Zamir and Etzioni, 1999] lists four major visualization techniques which attempts to visualize inter-document relationships: (a) document networks [Fowler et al., 1991], (b) spring embedding [Swan and Allan, 1998], (c) document clustering [Cutting et al., 1993; Osinski et al., 2004; Zeng et al., 2004; Wang and Kitsuregawa, 2002], and (d) self-organizing maps [Lagus et al., 2004]. However, studies in [Hearst and Pedersen, 1996; Zamir, 1998] showed that only document clustering appears to be both fast and intuitive and requires little training.

Document clustering ([Cutting et al., 1993; Hearst and Pedersen, 1996; Zamir, 1998; Osinski et al., 2004; Zeng et al., 2004; Wang and Kitsuregawa, 2002]) is a technique, which attempts to organize search results in clusters, where documents in a cluster focuses on some aspect of the query. Formally, clustering can be defined as an unsupervised learning method which organizes objects in groups based on certain similarity measures between objects. A cluster is therefore a collection of objects, which are 'similar' to each other and 'dissimilar' to objects belonging to the other clusters.

In web search, 'objects' correspond to individual search results (documents). Document clustering in IR is based on the 'Cluster Hypothesis' proposed by [Rijsbergen, 1979]. The hypothesis states that, *"closely associated documents tend to be relevant to the*

same request". The concept behind the hypothesis was that the relevant documents are more similar to each other than to non-relevant documents.

In the context of web search, clustering is a solution to reorganize search results (also called "snippets") in a more convenient way for browsing. In order to provide appropriate clusters of web search results, [Nguyen et al., 2009] suggests three key requirements for post-retrieval clustering systems: (1) the clustering algorithm should group similar documents together; (2) clusters should be labeled with descriptive phrases; and (3) the clustering system should provide high-quality clustering without downloading the contents of the whole web page.

Clustering of search results provides an insight into the different context of a query. By providing clusters, and means to grasp their content, identifying the focus of the search becomes easier for the web users. For instance, if for the query "nutrition", the clusters of the web results returned with the titles are: child nutrition, nutrition diet, etc. Then for the user looking for information related to 'child nutrition', locating suitable cluster (results) becomes easier, and hence the user can easily zero on his/her search goal.

Furthermore, clustering can also help to disambiguate queries having different meaning associated with it. For instance with the ambiguous query "java", users looking to find information about the 'java island', can more easily distinguish the set of results containing information about the island from the results discussing about the Java Programming Language. This is because the results for both the topics would be grouped in separate clusters: *Java Island* and *Java Programming Language*.

A number of document organizing approaches based on clustering technique have been developed over recent years. For instance, Yippy[2-2], Carrot[2-3], WebClust[2-4], iBoogie[2-5], etc are some of the existing clustering examples.

Clustering is useful when separating similar type of results (media or document) based on genre, topic, context, etc. However, in situations like aggregated search, where results are heterogeneous (i.e., media and document both), organizing results can be more challenging and complex. Because of the difference in the type of result (image, video, map, web, etc), there already exits distinct *sets* of results. That is, there is a set of image results, set of video results, set of web results, and so on.

[2-2] http://search.yippy.com/
[2-3] http://search.carrot2.org/stable/search
[2-4] http://www.webclust.com/
[2-5] http://www.iboogie.tv/

2.2.2 Faceted Browsing

In the history of search, there seem to exist two prominent search paradigm namely: navigational and direct search [Broder and Maarek, 2006]. Over the last few years, the direct search paradigm has gained dominance and the navigational approach became less and less popular. Recently a new approach has emerged, combining both paradigms, namely the faceted search approach. Faceted search enables users to navigate a multi-dimensional information space by combining text search with a progressive narrowing of choices in each dimension. It has become the prevailing user interaction mechanism in e-commerce sites and is being extended to deal with semi-structured data, continuous dimensions, and folksonomies [Broder and Maarek, 2006].

Unlike a simple hierarchical scheme, faceted classification gives the users the ability to find items based on more than one dimension[2-6]. For instance, while shopping for clothes from a store's online website, some users may be more interested in browsing by particular brand (e.g., Diesel, Forever Unique, etc), while others may be more interested in browsing by clothing type (e.g., dress, jacket, etc). Here, the "brand" and "type" are the example of facets; while Diesel, Forever Unique, dress and jacket are examples of facet values.

Similarly, aggregated search also allows users to search information in more than one information space (e.g., image, video, blog, etc). For instance, a user looking for information about "seven wonders of the world" might be interested in reading some *Wikipedia* articles, looking at pictures (*image*) of the monuments, reading any recent *news*, and so on.

[Polowinski, 2009] describes commonalities and general characteristics of faceted browsers. Among commonalities, it was suggested that most faceted browsers offer both, a visualization of the selection mechanism and a means to visualize the data itself. And that, faceted browsers have to often deal with instant updates and results. To some extent, aggregated interfaces also offer visualization and selection of multiple sources. Furthermore, in cases of 'news' aggregation, aggregated interfaces also need to deal with instant updates for news.

There have been several dedicated works addressing various aspects of faceted browsing. For instance, [Clarkson et al., 2009] describes formal models for faceted user interface; and [Stuart-Moore et al., 2006] established the components required to build a good browsing interface. Usefulness of a browsable interface utilizing the principle of faceted classification; and the users' preferred query submission methods in different

[2-6] http://www.webdesignpractices.com/navigation/facets.html

problematic situations were investigated by [Tang, 2007]. The study's findings showed that users preferred faceted search tools when their information needs were vague and the search topics were unfamiliar.

Furthermore, [Hopfgartner et al., 2010] explored the faceted browsing in video retrieval and demonstrated that, the faceted browser can potentially improve the search effectiveness of interactive video retrieval systems.

Finally, [Hearst, 2008] addressed the limitations and issues within faceted interface, and suggested that managing large scale subject space is challenging issue in faceted browsing, and acquisition of faceted subject metadata is also a problem. Furthermore, it is still an open question if faceted navigation is well suited for the small screen devices, for example in mobile computing. The study also highlighted that, although faceted navigation can be made more visually appealing with enhanced graphical displays, but to date it is not clear that these views enhance usability or substantially of the faceted interface.

Examples of existing search interfaces using faceted browsing approach are: iTunes[2-7], eBay[2-8], Amazon[2-9], etc. Usage of faceted browsing seem to be more prominent for e-commerce purposes, and have proved to be effective in facilitating information access. However, whether aggregated search can also facilitate information access is not clear. Since in aggregated search, facets might be of various media types and document genre.

2.2.3 Federated Search and Distributed Information Retrieval

Federated search and distributed information retrieval (DIR) systems provide a single user interface for searching information using multiple search engines [Avrahami et al., 2006]. The name federated search arose in the database research community; in the information retrieval (IR) research community the problem was usually described as distributed information retrieval [Avrahami et al., 2006]. Therefore federated search is also referred as distributed information retrieval [Si and Callan, 2005; Shokouhi et al., 2007].

[Craswell, May, 2000] describes the problem of distributed information retrieval as a situation when the documents are spread across many document servers; and an effective information retrieval system is required to access these distributed documents to

[2-7] http://www.apple.com/uk/itunes/
[2-8] http://www.ebay.co.uk/
[2-9] http://www.amazon.co.uk/

satisfy users' information need. In distributed information retrieval scenario, the information retrieval system available across the network is called a *search server*, and it is accessed using a *search client*.

In DIR a *search broker* is a sophisticated search client which when given a query and a set of search servers, selects the servers which are likely to provide relevant documents in response to the given query. The broker then sends queries to the selected servers, and finally produces a merged list of ranked documents from the set of documents provided by each server. Furthermore, [Craswell, May, 2000] defines three main tasks performed by the broker as: *selection, retrieval* and *merging*. The information flow during the three tasks performed by the broker are shown in the Figure 2.3.

During server *selection* the broker selects a subset S' of servers S which are best for answering the user's query q. Next, during *retrieval* the broker applies the query q at servers S' to obtain results lists $R_1, R_2 ..., R_{|S'|}$. Finally, during results *merging* the broker combines results $R1, R_2 ..., R_{|S'|}$ into a merged results list $RM = \langle D_M, o_M \rangle$, such that $D_M = D_1 \cup \ldots \cup D_{|S'|}$ and o_M is an effective ranking.

[Si and Callan, 2005] divides the problem of federated search and DIR into three sub-problems: *resource description, resource selection* and *result merging*; where resources corresponds to the severs defined by [Craswell, May, 2000]. There have been dedicated bodies of research focusing on either individual sub-problems of DIR and federated search (e.g., Luo-05,Shokouhi-07), or on the overall phenomenon of DIR (e.g., [Craswell, May, 2000]) and Federated search [Avrahami et al., 2006].

For instance, [Si and Callan, 2005] proposes a federated search technique that uses utility maximization to model the retrieval effectiveness of each search engine in a federated search environment. While [Avrahami et al., 2006] discusses a prototype federated search system developed for the U.S. government's FedStats Web portal, and the issues addressed in adapting research solutions to this operational environment. A collection-selection method based on the ranking of downloaded sample documents were proposed by [Shokouhi, 2007]. Furthermore, [Shokouhi et al., 2007] focused on the problem of maintaining representation sets for dynamically changing, uncooperative, distributed collections. It was suggested that as collections evolve over time, collection representations should also be updated to reflect any change.

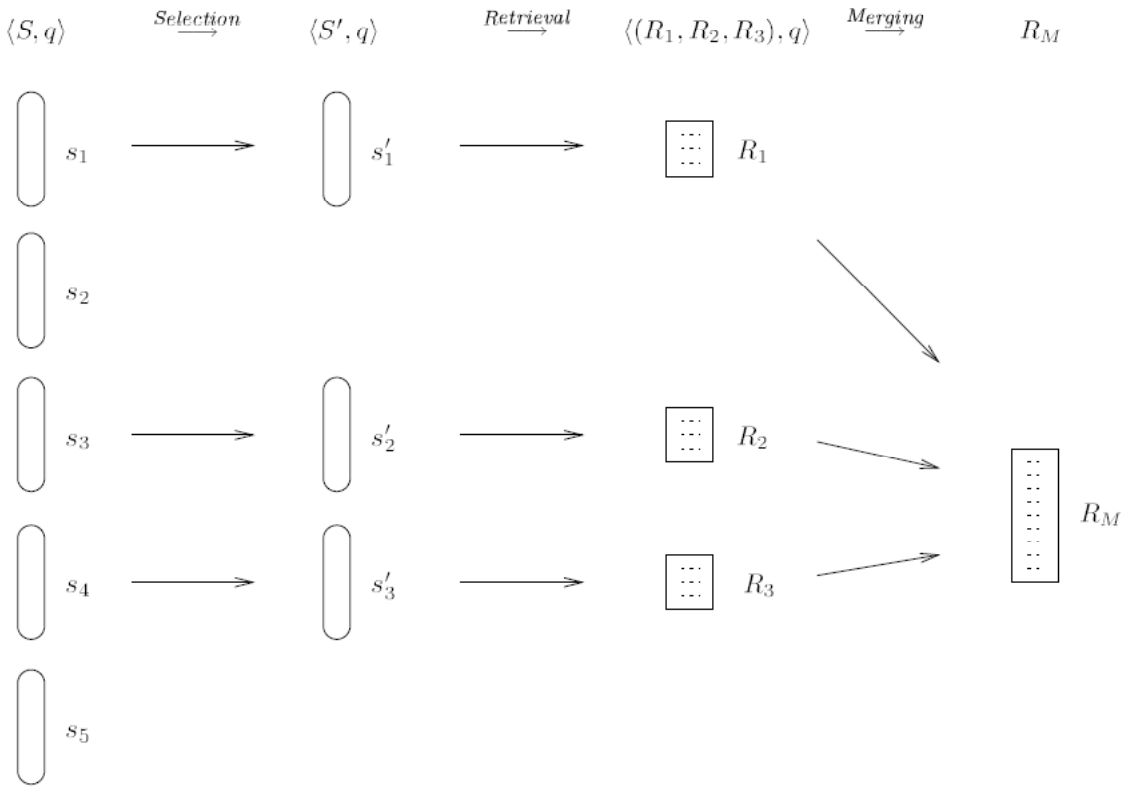


FIGURE 2.3: Search broker information flow. Given S and q the broker selects a subset $S' \subset S$, retrieves R_1, R_2, R_3 from those servers and builds the merged list RM. The query q usually guides each stage of the process, although in certain cases it may be ignored, for example if the broker's policy is to always select all its servers $S' = S$ (Figure taken from [Craswell, May, 2000]).

2.2.3.1 Metasearch

Metasearch can be considered an application of DIR and federated search[2-10]. Based on the underlying fundamentals of DIR and federated search, metasearch also aims to provide a unified access to information stored in the databases of multiple search engines. When a metasearch engine receives a query from a user, it invokes the underlying search engines to retrieve useful information for the user [Meng et al., 2002]. The similarity among the components of a metasearch and a federated search architecture can be seen in Figure 2.4 and Figure 2.5.

Primarily metasearch focuses on web collections, and aggregates results from various search engines, whereas, some of the DIR and Federated search systems may focus on specific database collection (e.g. FedLemur which was developed for the US govern-

[2-10]http://en.wikipedia.org/wiki/Federated_search#Implementation

ment's FedStats Web portal by [Avrahami et al., 2006]).

A metasearch engine sends a user query to several other search engines and/or databases and aggregates the results into a single list or display them according to their source. Metasearch engines enable users to enter search criteria once and access several search engines simultaneously. They operate on the premise that the web is too large for any one search engine to index it all. In addition, more comprehensive search results can be obtained by combining results from several search engines. This also may save the user from having to use multiple search engines separately[2-11].

There have been dedicated bodies of work addressing various aspects of metasearch (e.g., [Aslam and Montague, 2001; Wu et al., 2001; Meng et al., 2001, 2002; Thomas and Hawking, 2009; Thomas et al., 2010], etc.). For instance, [Thomas et al., 2010] investigated what user interfaces might be appropriate for presenting results from more than one source. Efficient ways for selecting search engines (servers) were suggested by [Dreilinger and Howe, 1997; Desai et al., 2006; Thomas and Hawking, 2009]. Furthermore, different models and Frameworks for metasearch were proposed by [Glover et al., 1999; Aslam and Montague, 2001; Aslam et al., 2003a].

The architecture proposed by [Glover et al., 1999] was designed to consider users' information need as well. [Aslam et al., 2003b] proposed a unified framework for simultaneously solving both the pooling problem (the construction of efficient document pools for the evaluation of retrieval systems) and metasearch (the fusion of ranked lists returned by retrieval systems in order to increase performance).

Techniques like clustering and distributed information retrieval have facilitated easier access to more information, but these approaches have been mostly limited to 'text' sources of similar genres or single media type (for example, image collections in faceted browsing).

2.2.4 Aggregated Search

In recent years, there has been a growing interest in finding ways to combine results from various vertical searches onto one result page, referred to as *aggregated search*. A vertical search engine searches a specific industry, topic, and type of content (e.g., travel, movies, images, blogs, live events), piece of data, geographical location, and so on[2-12].

[2-11] http://en.wikipedia.org/wiki/Metasearch_engine
[2-12] http://www.internettutorials.net/vertical-search.asp

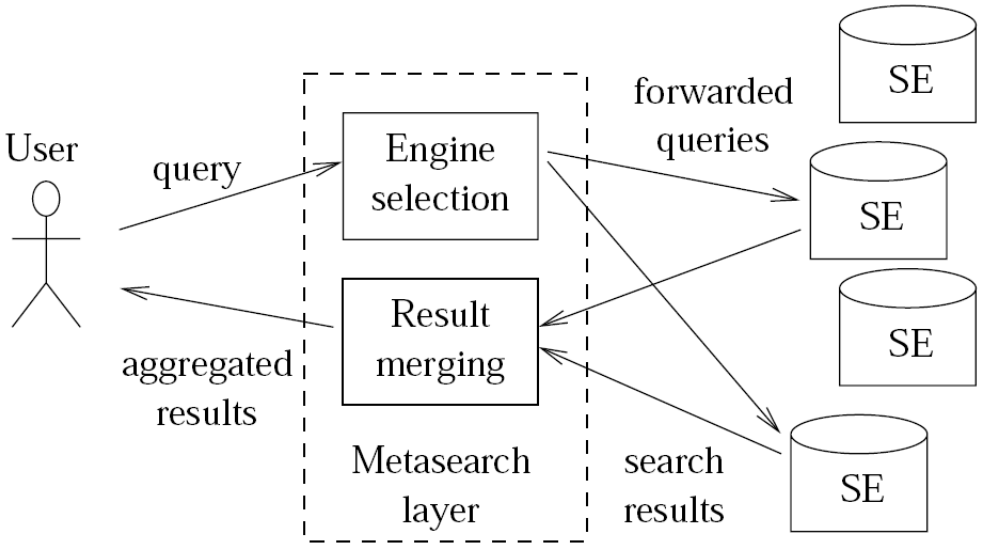


FIGURE 2.4: Federated search components architecture (The figure is taken from [Khoussainov and Kushmerick, 2004])

Aggregated search can be seen as an extension of metasearch, where, different dedicated search engines are used to fetch results from different verticals. When a user submits a query to a metasearch engine, it sends the query to its underlying collection of search engines (as shown in Figure 2.5). The metasearch engine returns a merged list of results provided by the different search engines to user in response to his/her query.

Similarly, an aggregated search engine sends the query to its vertical search engines; collects results from various sources (images, video, maps, etc); and finally aggregates the obtained results onto a search result page. The overall working architecture of an aggregated search is shown in Figure 2.6. It can be seen in the figure on how a user query is re-submitted to different collection of information sources. For instance, the video results from 'YouTube', maps results, Wikipedia pages and image results are returned to the aggregated search engine. These results are aggregated before returning to the user.

Although aggregated search is an extension of metasearch, the difference in the "type" of result obtained from different sources are more distinct in case of aggregated search. That is, aggregated search deals with more heterogeneous results when compared to earlier approaches like; federated search, distributed search and metasearch.

Therefore, aggregating results from different sources makes aggregated search paradigm more complex. Diversity in media and document content makes result organization

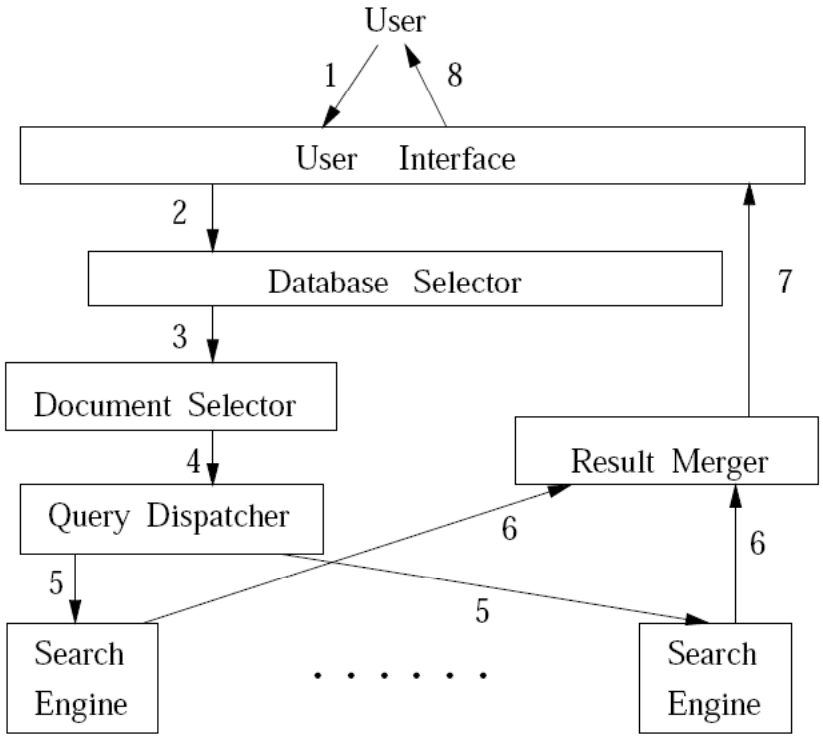


FIGURE 2.5: Metasearch components architecture (The figure is taken from [Meng et al., 2002])

in aggregated search more challenging. This thesis primarily investigates aggregated search, where results of different physical sources are aggregated.

2.2.5 Summary

This section has reviewed different approaches to facilitate information access followed by researchers in the past. All above approaches attempt to organize or aggregate information from various sources (collections, servers, etc) so that, information access can be made easier and faster. For instance, clustering organizes search results into groups where, documents containing information about a topic are grouped under one cluster. While federated search, distributed information retrieval and metasearch aims to facilitate information access by providing one interface using which, users can search through multiple information sources spread or distributed over network.

However, effects of aggregating information from different verticals (image, news, etc) within an aggregated interface is still not investigated. Dedicated studies have shown how information access have been facilitated by using clustering, faceted browsing,

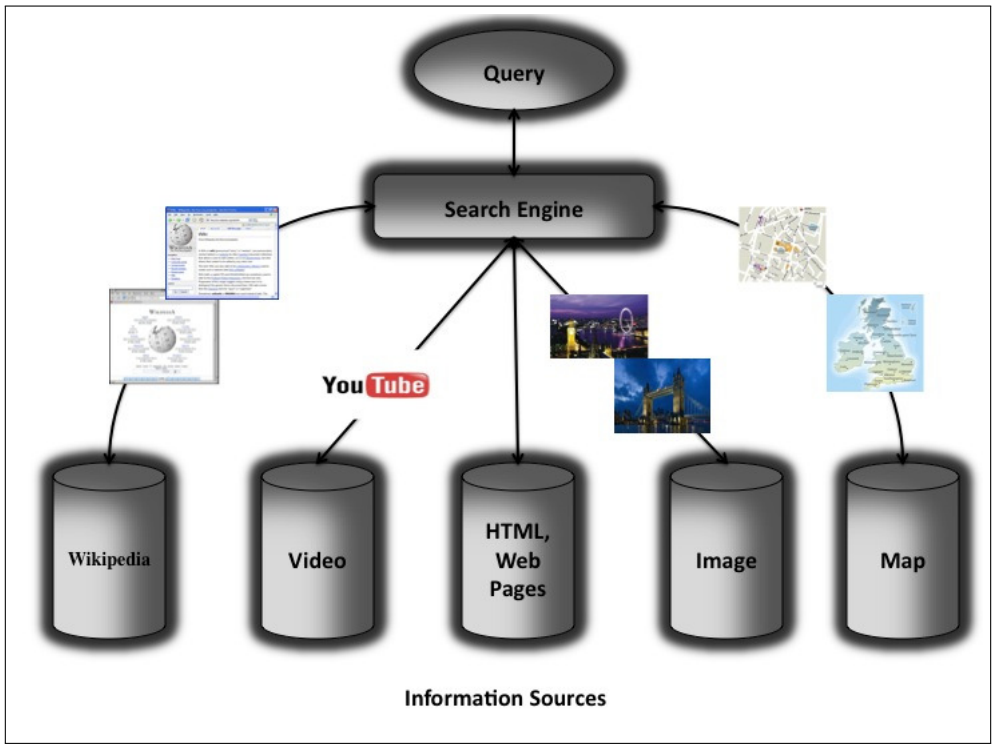


FIGURE 2.6: Architecture of an Aggregated Search Paradigm.

federated search, etc. Through studies it has been shown how these approaches have proved to be useful in users' search tasks, but usefulness of an aggregated approach in task completion is still not known.

There remain many unanswered questions in regard to the aggregated search approach. For example, Does aggregated search facilitates information access? Is there a need for an aggregated search? Does aggregated search improve in the diversity of information accessed? The aim of this thesis is to investigate the usefulness of an aggregated interface in task completion, and to see if such approach facilitates information access by providing diverse results.

The next Section provides the background knowledge on result presentation and interface design. The aim of the Section is to discuss the importance of *presentation* of results, and to discuss some of the attempts in providing useful interface by information retrieval and its related communities.

2.3 Result Presentation and Interface Design

The problem of effective information presentation has received substantial attention from researchers in the past (For example, [Aloia et al., 1996; Kang, 2005; Joho and Jose, 2008; Hearst, 2009; Mateosian, 2010]). Success of a retrieval system not only depends on a good retrieval mechanism; but also on how the retrieved items are presented to the end users. Having a perfect retrieval mechanism but a poor result presentation may lead to user dissatisfaction and result in the failure of the retrieval system [Joho and Jose, 2008]. Therefore, different web interface designs may determine whether the company (or search engine) keeps or loses their customers [Kang, 2005].

There is a strong association between interface design and web site system satisfaction and self-efficacy. For interface designers, it?s challenging to incorporate the effective search technology into their designs [Mateosian, 2010]. For every query submitted, it is required to generate most effective and expressive presentation of results, so that the amount of user cognitive is reduced [Aloia et al., 1996].

the art of search interface design, both in academic research and in deployment in commercial systems. In his book, he discusses on how the worldwide reach of the Web has brought with it a new realization among computer scientists and laypeople of the enormous importance of usability and user interface design.

Furthermore, through several dedicated efforts over a decade in the field of interface design, it has been possible to achieve certain understanding on what works in search interfaces from a usability perspective, and what does not. Researchers and practitioners have developed a wide range of innovative interface ideas, but only the most broadly acceptable make their way into major web search engines [Hearst, 2009].

In interface design, usability is one the key aspects which contributes to a successful user interface. Usability can be defined as, a quality attribute that assesses how easy user interfaces are to use[2-13]. Optimized user interface design requires a systematic approach to the design process. But, to ensure optimum performance, usability testing is required[2-14].

[Nielsen, 1993] suggested ten general principle guidelines for user interface design, which are followed by many interface designers while designing an interface. They were called; “usability heuristics" which describes basic characteristics of usable interface. The ten proposed heuristics for usability were as follows:

[2-13] http://www.useit.com/alertbox/20030825.html
[2-14] http://www.usernomics.com/user-interface-design.html

- Simple and natural dialog: User interface should be simplified as much as possible.
- Graphics design and color: The principles of graphics structure should be used to help the users understand the structure of the interface.
- Speak users' language: The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
- Minimize users' memory load: The system should take over the load of memory as much as possible. In general, users find it easier to identify something from what is shown to them, than having to recall the same information from memory without help.
- Consistency: Consistency is one the most basic usability principle. The same information should be presented in the same location on all screens. If users know that same command will always have the same effect; they feel more confident in using the system.
- Clearly marked exits: users do not like being trapped by the computer. In order to increase the users' feeling of being in control, the system should offer the user an easy way out of as many situations as possible.
- Shortcuts: Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
- Good error messages: Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution[2-15].
- Prevent Errors: System should be designed to avoid error situations as much as possible.
- Help and documentation: Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large[2-16].

[2-15] http://www.useit.com/papers/heuristic/heuristic_list.html
[2-16] http://www.useit.com/papers/heuristic/heuristic_list.html

In the following Sections, various user interfaces for information presentation is discussed. Attempts by different information retrieval and its related communities in regard to result presentation are described. Section 2.3.1 describes attempts made by database community. Section 2.3.2 shows how result interface for digital libraries were designed and investigated. Finally, Section 2.3.3 would survey various interface related studies dedicated to web search interface and design.

2.3.1 Database Interface

In nineties, importance of usability and design of interface was recognized within the database community (For example, [Aloia et al., 1996; Mitchell et al., 1996; Curl et al., 1997], etc). For instance, a framework for user-interfaces to databases (IDSs) was proposed by [Mitchell et al., 1996], where the framework was described in terms of a classification of the characteristic components of an IDS.

[Curl et al., 1997] made attempts to understand to what extent leveraging the graphical user interface's ability to convey spatial information can improve a user's ability to write effective database queries. A laboratory experiment was conducted to explore the influence of interface style and the spatial visualization ability of the user on the performance of the query development process.

[Voisard, 1994] investigated the problem of designing graphical geographic database user interfaces (GDUIs) and have integrating them into a database management system (DBMS). The study proposed functionalities that should be provided by a graphical user interfaces for display, editing and querying the database.

[Aloia et al., 1996] developed a new approach to the precise definition of effective presentations in a specific class of systems: query results of multimedia databases. The traditional entity-relationship schema was exploited to identify semantically related data and give suggestions for their presentations.

Another entity-based database interface was presented by [Cattell, 1980]. Experience with the proposed prototype system suggested that the entity-based presentation was appropriate for types of casual interactive use that existing database interfaces did not address, such as browsing. It was proposed that such an interface could be used to supplement a query language or other interface to allow users both kinds of views of the data.

[Mantei and Cattell, 1982] investigated data entry, information retrieval, and browsing behavior with respect to a database interface. The study further examined and contrasted the behavior with more conventional storage media.

Interface studies in database community provide insight on various interface related factors. Through several dedicated studies it was possible to establish suitable result presentation approaches and design guidelines. Findings from these studies have proved to beneficial in presenting results on a database interface. However, if the similar design guidelines could be applicable or suitable for presenting results in an aggregated interface is not clear.

2.3.2 Digital Library Interface

A digital library is a collection of documents in organized electronic form, available on the Internet or on CD-ROM (compact-disk read-only memory) disks. Depending on the specific library, a user may be able to access magazine articles, books, papers, images, sound files, and videos [2-17]. Digital libraries are organized collections of information and are focused on a particular topic or theme [Witten, 2003].

Several attempts have been made to generate effective and usable interface for accessing collections within digital libraries (For example, [Yoon and Kim, 1998; Wang Baldonado, 2000; Davis, 2006; Dubinsky et al., 2007; Shiri, 2008], etc).

For instance, [Shiri, 2008] showed how metadata can be utilized to enhance visual user interfaces to digital libraries. The study was conducted to investigate and analyze a specific category of digital library visual interface that supports information seeking, exploration and retrieval based on metadata representations, namely metadata-enhanced visual interfaces. Results from the study demonstrated that the combined use of visualization techniques and metaphors is becoming increasingly prevalent as a design strategy to support users' information exploration. The results also suggest that visual interfaces enhanced with metadata are becoming more widespread to provide a richer representation of digital collections.

[Wang Baldonado, 2000] proposed "SenseMaker", a user-centered interface named for information exploration in a heterogeneous digital library. The proposed interface unified citations and articles from heterogeneous sources by presenting them in a common schema with affordances for quick comparisons of properties.

Catalog browsing approach was used to devise a speech-based mobile interface to a digital library by [Dubinsky et al., 2007]. A physical to digital relationship was examined when using vocal commands to control the search activity. Finally, challenges, decisions, and processes that affected the design and development of the search user interface in a digital library were identified by [Davis, 2006].

[2-17] http://www.cesa8.k12.wi.us/media/digital_dictionary.htm

Similar to database, the information collection in digital library is stored in a specific format depending on the collection type. Results within digital library is focused on one topic, theme or type of information. Therefore, the interface design or result presentation approach followed in digital library is often tailored to fit to a specific collection. In case of aggregated search, the information collection is heterogeneous; exits in many different formats; and contains information on diverse topics and themes. Therefore, it is not clear if the design and result presentation approach of digital interface could be adapted to fit within an aggregated search scenario.

2.3.3 Web Search Interface

Significant efforts have been devoted to the design and evaluation of conventional web search interfaces (For example, [Dziadosz and Chandrasekar, 2002; Vaughan et al., 2003; Nakarada-Kordic and Lobb, 2005; Kammerer and Gerjets, 2010; Mateosian, 2010]).

For instance, [Kang, 2005] examined the effect of information search guidance in a web site on the understanding of the product and the acceptance of the Web information system. [Wildemuth, 2006] proposed an evidence-based practice to search interface design, with the goal of designing interfaces that adequately support search strategy formulation and reformulation. Furthermore, a hierarchical category browser interface for children was proposed by [Hutchinson, 2003].

[Villa et al., 2009] presented a search interface, which was designed to support broad search tasks, allowing a user to create search aspects, each of which models an independent subtask of some larger task. The interface was built on the intuition that users should be able to structure their searching environment when engaged on complex search tasks, where the act of structuring and organization may aid the user in understanding his or her task.

[Dziadosz and Chandrasekar, 2002] described an empirical evaluation of the utility of thumbnail previews in web search results. Results pages were constructed to show text-only summaries, thumbnail previews only, or the combination of text summaries and thumbnail previews. It was found that in the combination case, users were able to make more accurate decisions about the potential relevance of results than in either of the other versions, with hardly any increase in speed of processing the page as a whole.

"Advanced" factors such as interface attractiveness [Nakarada-Kordic and Lobb, 2005] and result trustworthiness [Kammerer and Gerjets, 2010] have also been evaluated. A

taxonomy of result presentation techniques has been proposed as a reference for designers of web search systems [Wilson et al., 2010]. [Hearst, 2009] extended the design guidelines proposed by [Nielsen, 1993] (as discussed early in this Section) to search interface design. She further explained some of the difficulties with search interface design and provided a set of design guidelines tailored specifically to search user interfaces. These guidelines were:

- Offer efficient and informative feedback,
- Balance user control with automated actions,
- Reduce short-term memory load,
- Provide shortcuts,
- Reduce errors,
- Recognize the importance of small details, and
- Recognize the importance of aesthetics.

Search results presentation plays a critical role in the typical web search cycle [Hearst, 2009]. Appropriate representation of retrieved items is an important aspect of web search interface [Joho and Jose, 2008].

There is a large body of work on effective document surrogates for web search results (For example, [Marchionini and White, 2007; Clarke et al., 2007; Joho and Jose, 2008; Hearst, 2009]). In information retrieval, an alternative representation of the retrieved objects is sometimes referred to as a *surrogate* [Joho and Jose, 2008]. In other words, document surrogates are summary information intended to help the user understand the primary object, as opposed to metadata more broadly construed, which can also serve this purpose but is often more tailored towards use by computer programs [Hearst, 2009].

For instance, [Joho and Jose, 2008] presented a comparative evaluation of textual and visual forms of document representation as additional components of document surrogates. The experimental results suggested that an increased level of document representation available in the search results can facilitate users' interaction with a search interface.

While [Clarke et al., 2007] suggested on how understanding the influence of features of the captions (comprising of title, snippet, and URL) in web search may provide

guidelines for improving web search interface. Findings from the study showed that the occurrences of certain terms like; *official, attraction, sexy, etc*, attracted more users' clicks; while terms like *free, encyclopedia, etc* had negative influence on users' clicks.

Although, there has been significant amount of research devoted to the effective design of web search interface, yet there is very limited understanding of the design and result presentation in an aggregated interface. For instance, what is a suitable "surrogate" for representing different types of results (image, video, map, etc) on an aggregated interface is not known. What makes a well-designed, attractive, trustworthy and engaging aggregated search interface is not clear. In addition, how different design and visualization approach can be used to improve information access in an aggregated search interface is still unexplored. What is an optimum or suitable combination of results (web+image, web+image+video, etc) to be presented on an aggregated search interface.

2.3.4 Summary

This Section has shown that there has been many research dedicated to understand interface design and result presentation among various information retrieval related communities namely; database, digital library and web search. Several efforts have been made to generate an effective and usable interface. Many design guidelines have also been proposed in order to build a suitable result interface.

These design guidelines or result presentation approach have shown to be effective when presenting results from focused or homogeneous information collection. For instance, information in database or digital libraries is focused on one topic, theme or information type (document, media). But, in aggregated search, results are often heterogeneous and includes combination of both document and media results (e.g., image, video, map, web, etc). This combination of different information "types" makes the result presentation and design of aggregated search interface more complex.

There is established understanding of what makes a usable and effective interface to present results from a database or a digital library or web search. Findings from the dedicated studies have shown on how interface design can be tailored appropriately for result presentation from the respective information collection. However, if these guidelines can also be tailored to fit into aggregated search scenario is not known.

There exits many unanswered questions with respect to an aggregated search interface design and result presentation. For instance, does existing interface design guidelines

apply to aggregated search interface. Is there a need to incorporate new design guidelines for an aggregated search interface? Does aggregation of different types of result cause cognitive load on users? What is a good (optimum) number of results that should be presented on a single aggregated result page? The goal of this thesis is to investigate aspects of result presentation, and to uncover some of the design issues that may exists within an aggregated search paradigm. This thesis also aims to suggest some design guidelines that could be useful in building an aggregated search interface.

In the following Section, the background knowledge on user information seeking behavior is provided. So far in this Chapter, it was shown that result organization and aggregation used by various retrieval system have proved to be effective in making information access easy. Furthermore, it was also shown that how appropriate presentation of these organized results is required in order to make these results useful for the users.

To ensure if the organization or presentation of results were effective during information seeking process; it is important to see how users interact or respond to these information organized and presented in a certain way. Analyzing user behavior with the results provides an insight on the usefulness and effectiveness of the organization and presentation approach.

2.4 User Behavior

[Borgman, 1987] suggested possible definitions for the "study of user behavior" . For instance, studying ways to optimize human efficiency at retrieval; studying ways to make systems more "user friendly"; studying "satisfaction" with the system and it?s output; or study of the human characteristics and human problem solving activities that bear some relationship to the retrieval performance of the user.

With the emergence of information retrieval systems, an alternative means to access large data collections were required. Information retrieval systems are rapidly cultivating an end user market in a variety of domains [Borgman, 1983]. The overall process of an information retrieval can be divided into two aspects: *the system*, and *the end-user*. The system can be further divided into back-end and front-end system aspects. The back-end aspect of the system focuses on retrieval mechanisms, optimizations, etc. While, the front-end of the system deals with interface design, result presentation, etc. The user aspect involves the information seeking behavior associated with the retrieval system.

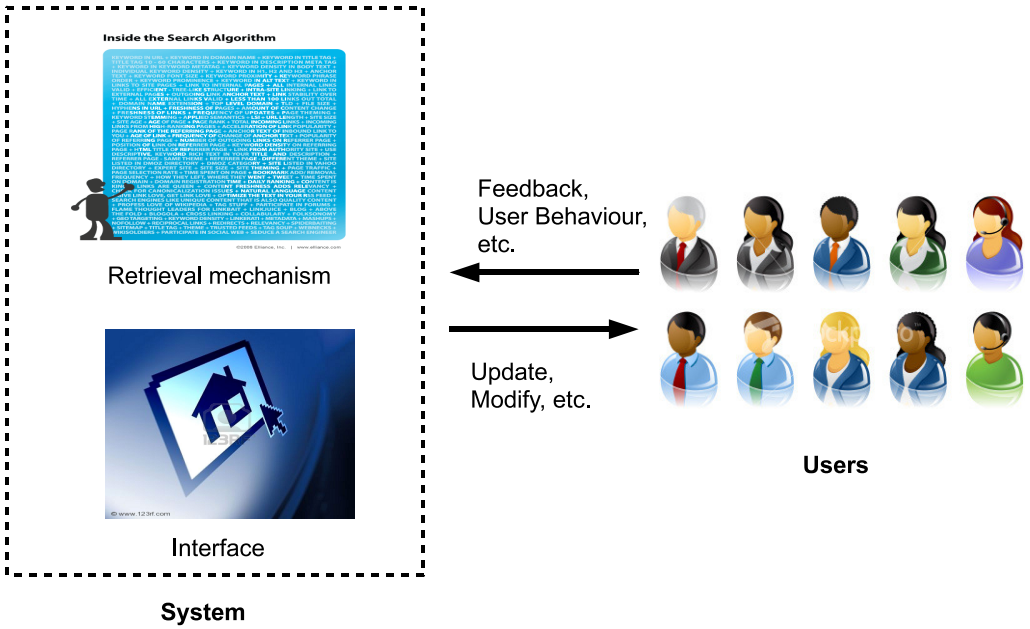


FIGURE 2.7: Information Retrieval Cycle

The process of information retrieval evolves by iterating repeatedly over user and system aspects (as seen in Figure 2.7). That is, first a system is devised which is capable of retrieving information from a data collection. Then the system is used by the end-user for retrieving desired information from the data collection. Finally, based on users' feedback, or after analyzing users' information seeking behavior, the retrieval system is further updated or modified to overcome the limitations (if any) observed in the previous versions of the system, or to suit domain specific tasks. Therefore, in order to build successful information retrieval systems, understanding the user aspect is also critical [Jeon et al., 2008]. For instance, incorporating user behavior data can significantly improve retrieval mechanism ordering of top results in real web search setting [Agichtein and Zheng, 2006].

2.4.1 Methodology

To understand users' response for a given retrieval system, it requires monitoring users' information seeking behavior while performing a search task. There seem to exist two prominent methods used for analyzing user behavior: (1) Experimental method and (2) Log analysis. An experimental method is a lab based user study, which involves users (participants) to perform specific search tasks with the retrieval system being tested. While during log analysis, users' interaction with the system over a time period is

recorded for analyses.

2.4.1.1 Experimental method

Answers to certain research questions can be obtained through experimental method. For instance: Do people in Glasgow drink more during winters than in summers? Do people become crazy after writing PhD thesis? Experimental methods or laboratory studies allows researchers to examine relationship between two or more variables. During an experimental method, environment is deliberately manipulated to ascertain what effect one variable has on the other [Field and Hole, 2003].

In experimental method, answers to research questions are obtained by observing the "cause" and "effect" of certain factors. For instance, to answer research question: Does listening to music reduces stress? Here, the cause (*music*) can be measured by observing its effect (*stress level*).

Experimental studies are often conducted in labs so that the researcher has the maximum control over the experimental situations. Advantage of conducting a laboratory studies is that, it offers explanation for the observed behavior [Kelly, 2009].

In past, many researchers have conducted experimental studies to observe effects of certain features of retrieval system on users' information seeking behavior (For example, [Joho and Jose, 2008; Villa et al., 2009; Thomas et al., 2010; Hopfgartner et al., 2010]).

However, the experimental methods also has certain limitations, and they are primarily due to the fact that these methods typically involve subjective elements, small number of participants, and have large variance in the results due to sampling bias. It is hard to generalize the results from experimental methods due to the small number of participant [Zhang et al., 2009].

Furthermore, experimental methods have limited benefits because it does not provide intervention-free studies of user-initiated searches, where no observers were present to monitor end-user searching, no researchers assigned tasks to end users, and no interviewers questioned users before, during, or after the search [Markey, 2007a].

Nevertheless, experimental studies are good with respect to the amount of control researchers have over the study situation. This is useful when trying to isolate the impact of one or more variables [Kelly, 2009]. Next, a naturalistic method (log analysis) to investigate users' information seeking behavior is described.

2.4.1.2 Log analysis

Search engine companies and the researchers often use transactional log to understand the user behavior during information seeking process (For example, [Spink et al., 2002; Jansen and Spink, 2005; Agichtein and Zheng, 2006; Jansen, 2006; White and Morris, 2007a; Duarte Torres et al., 2010]).

A search engine's transactional log is an electronic record of interactions that have occurred during a searching episode between a Web search engine and searchers who are seeking information on that Web search engine. Therefore, the use of data stored in transaction logs of Web search engines, Intranets, and Web sites can provide valuable insight into understanding the information-searching process of online searchers [Jansen, 2006].

Although, transactional log data continues to be useful in uncovering search behavior patterns and is used by many researchers and search engine companies, there exist some limitations and social issues associated with it.

[Kurth, 1993] identified the limitations of the transactional log analysis and suggested that, although transaction log data effectively describe what searches patrons enter and when they enter them, but they do not reflect, except through inference, who enters the searches, why they enter them, and how satisfied they are with their results. Furthermore, [Murray and Teevan, 2007] reported following social and technical challenges of transactional log analysis:

- Can query log data be safely collected and analyzed? Should it?
- Can query log data be anonymized and shared? How shall it be done?
- Can we establish standards of practice for query log analysis?

Key points and suggestions from the discussions for the above questions were that, using transactional log data is crucial for training researchers in this field. In addition, there is a strong potential for search engine companies to benefit from the findings of the studies from these log data. In order to make fair use of these data, setting up review boards and clear guidelines were suggested. Furthermore, preserving logs was suggested as it provides a reflection of the character, values, fears, hopes and desires of the people who issue them.

In-spite of the above limitations and challenges, transactional logs continues to be a valuable resource for analyzing user behavior and search patterns. It allows to learn

how end users search online IR systems left to their own devices and unaffected by potential biases such as the presence of an observer, their knowledge that a reviewer would scrutinize their search at a later time, their aptitude for or the potential biases of a researcher-assigned task [Markey, 2007a]. In the following Sections findings of user behavior analysis from experimental methods and log analysis are presented.

2.4.2 User Behavior in General IR

There has been a large body of dedicated research in user behavior analysis for various online information retrieval systems (for example, [Das Neves and Fox, 2000; Borgman, 1983; Park, 2000; Mongy et al., 2005; Mat-Hassan and Levene, 2005; Yi et al., 2006; Markey, 2007a,b; Dupret and Piwowarski, 2010], etc).

For example, [Borgman, 1983] reported a computer monitoring study of users of the Ohio State University Libraries' online catalog. The study characterizes user behavior in terms of types of searches done, patterns of use, time spent on searching, errors, and system problems. Results from the study suggested that users have much shorter sessions than on other types of retrieval systems (in this study, information desk area and the card catalog area of the same library). Patterns of use vary between campus libraries, academic quarters, and between short and long sessions.

[Das Neves and Fox, 2000] investigated the effect of clustering techniques and query highlighting on search strategy users develop in the virtual environment for digital libraries. One of the key findings of the study was that users did not search until they found the best match, but until they found a match "good enough".

[Mongy et al., 2005] suggested two levels of user modeling approach based on browsing behavior of users in video search engine. The first level aims at modeling and clustering user behavior on a single video sequence (intra video behavior), the second one aims at modeling and clustering user behavior on a set of video sequences (inter video behavior).

Furthermore, [Yi et al., 2006] compared the differences in the nature of the queries from history and psychology databases. The aim was to see if difference in the subject specific database searches can aid in understanding of user search behavior and the information needs of their respective users. In other words, the research focused on how the particular subject (history and psychology in this case) for which users need to find information influences their selection of terms for queries.

[Park, 2000] investigated issues on how to support effective interaction of users with heterogeneous and distributed information resources. In the Text REtrieval Conference

(TREC) environment a study was conducted to compare the usability, user preference, effectiveness, and searching behaviors in systems that implement interaction with multiple databases through a common interface. There were four main findings from the study; the general assumption of the information retrieval (IR) literature that an integrated interaction is best needs to be revisited; it is important to allow for more user control in the distributed environment; for digital library purposes, it is important to characterize different databases to support user choice for integration; and last, certain users prefer control over database selection while still opting for results to be merged.

Finally, [Markey, 2007a] observed that, most of the IR users accept default values when using advanced features provided by the retrieval systems. Also, most studies suggest that, when users use advanced search features in their queries, they use them incorrectly about a third of the time. The results from the past studies demonstrate that most of the users are satisfied by their online searches. Although users are not conducting very sophisticated online searches, yet, the vast majorities were satisfied with their searches.

Studying user behavior in general IR have helped in understanding the effect of the retrieval systems during information seeking process. That is, when do users provide long and short queries; does users' use different searching mechanisms when searching different information collection. Furthermore, effectiveness of the retrieval systems (database, digital library) in task completion was also evaluated.

However, results from the user behavior studies in database and digital libraries do not provide insight on how users interact with an aggregated interface. For instance, what is the average length of a query in an aggregated search? Does users of an aggregated interface also prefer to have control over verticals? Are users of aggregated search interfaces satisfied by their search performances? etc.

2.4.3 User Behavior in Web Search

In context of web search, several studies attempted to understand the dynamics of the search behavior of users, i.e. how users information need and searching behavior evolve (For example, [Spink et al., 2002; Jansen and Spink, 2005; Agichtein and Zheng, 2006; White and Morris, 2007a], etc). These studies analyzed web search engine click data to gather information about the queries being submitted to the search engine, their length, query frequency per session, click frequency per session, etc.

[Markey, 2007a] in a meta-study compared several researches done in past twenty-five years which analyzed user behavior. The study outlined the overall user behavior observed in most online information retrieval systems. Summing up the overall findings

from various studies, the author addressed the following research questions: What characterizes the queries that end users submit to online IR systems? What search features do people use? What features would enable them to improve on the retrievals they have in hand? What features are hardly ever used? What do end users do in response to the system' retrievals?

While comparing the observations regarding queries submitted to the IR systems, the author ([Markey, 2007a]) made two important notes; users submit few queries, and users' queries contain few terms. That is, most studies reported means between two and four queries per session. These two to four query sessions are not limited to searches of Web search engines, but include online catalogs and digital libraries. Furthermore, queries submitted to online catalogs are relatively shorter than web queries; where approximately one-two terms mean terms are contained in catalog queries, while web queries contained two-three terms.

[Agichtein and Zheng, 2006] showed that incorporating user behavior data can significantly improve ordering of top results in real web search setting. Various alternatives for incorporating feedback into the ranking process were explore and were then compared to other common web search features. Results from the study showed that, incorporating implicit feedback can augment other features, improving the accuracy of a competitive web search ranking algorithms.

[Liu et al., 2008] analyzed different user behavior pattern for accessing ordinary web pages and web spam pages, and proposed features to identify spam pages. The analyses showed that spam pages are mostly designed to attract users attention, they do not provide any valuable contents. Therefore, Web spam page receives most of its user visiting from search engines instead of from non-spam pages or bookmark lists.

Furthermore, [Jansen et al., 2010] analyzed user interactions with a real time search engine to investigate the query frequency, terms and its structure during information seeking in real time searching. The findings from the study showed that, 60% of the traffic comes from the engine's application program interface, indicating that real time search is heavily leveraged by other applications. Of the queries, 30% were unique (used only once in the entire dataset). The search topics associated with the query terms were often related to technology, entertainment, and politics.

Studies like above have shown how users' interact with web search results, and how these information can be further utilized in improving the retrieval systems. However, in context of an aggregated search, there is very limited understanding of user behavior with respect to the aggregated results. That is, not much is known on how users seek

information when different types of results are provided on a single result page, and how user behavior can be modeled within an aggregated retrieval system.

2.4.4 User Behavior in Aggregated Search

Through informal studies, interactions with Google's Universal search results were looked at[2-18]. Eye tracking techniques were used to capture users' visual attention on the aggregated result page.

Results from the study showed that, due to the addition of pictures on the result page, the orientation point of viewing results seem to change, leading to a distinctly different experience. In pre-aggregated search interfaces, the common observed trend was to orient in the upper left corner (indicated by A) and to start the scanning from there, first vertically (the down arrow) and then scanning across when a title catches your attention, as seen in heat map on right in Figure 2.8.

While in the aggregated results (left heat map of Figure 2.8), it was found that while there still was some scanning in the upper left (B), it does not appear that the scanning starts there. Instead, the orientation appeared to happen by the graphic thumbnail in the results (C), and then started from there. Scanning seems to be predominantly to the side and below (D). Furthermore, "F" pattern [Nielsen, 2009] for scanning results in conventional web interface seem to change to "E" pattern in an aggregated interface, as seen in Figure 2.9.

Through another eye-tracking study[2-19] performed by Google on their Universal interface (aggregated interface) it was observed that, non-web results (image, video, etc) within a result page did not affect the order of scanning the results and does not disrupt the information seeking process of users. Furthermore, [Ostergren et al., 2010] also performed an eye-tracking study to investigated how searchers interact with graphical, non-textual search engine results page user interfaces in order to reveal the potential value of these alternative display strategies.

So far studies related to aggregated interface have primarily been investigated through eye-tracking studies. These studies provide insight on how users' view results in an aggregated interface. That is, there is some understanding on where on the result page does user looks at or gives attention; and where does the user start viewing the web and the non-web results. However, whether and how these non-web search results affect

[2-18] http://searchengineland.com/eye-tracking-on-universal-and-personalized-search-12 233
[2-19] http://googleblog.blogspot.com/2009/02/eye-tracking-studies-more-than-meets.html

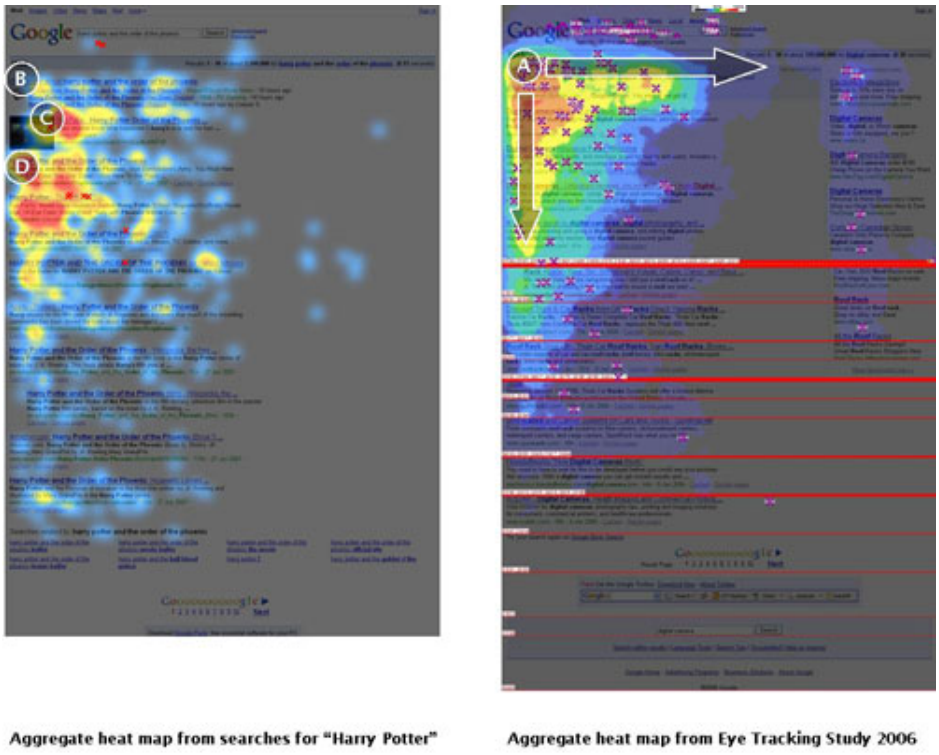


FIGURE 2.8: Heat Map of an Aggregated Interface. The figure is taken from the article[a]

[a]http://searchengineland.com/eye-tracking-on-universal-and-personalized-search-12233

users click-through ratio is not very clear from these studies. Also, if result presentation in an aggregated interface is useful and effective in task completion is not known. Despite some initial attempts (as mentioned above), there lies a big gap in understanding of overall information seeking process of users in an aggregated scenario. Above studies provide insight on how users "view" an aggregated interface, but how users "use" an aggregated interface remains unexplored; This is the aim of the research presented in this thesis.

2.4.5 Summary

In this Section a background material on user behavior analysis was provided. How analysis of user behavior have influenced various information retrieval systems were also discussed. Through studies various behavioral observations were made during information seeking process, such as: length of queries submitted, difference in information access with respect to particular subject, etc.

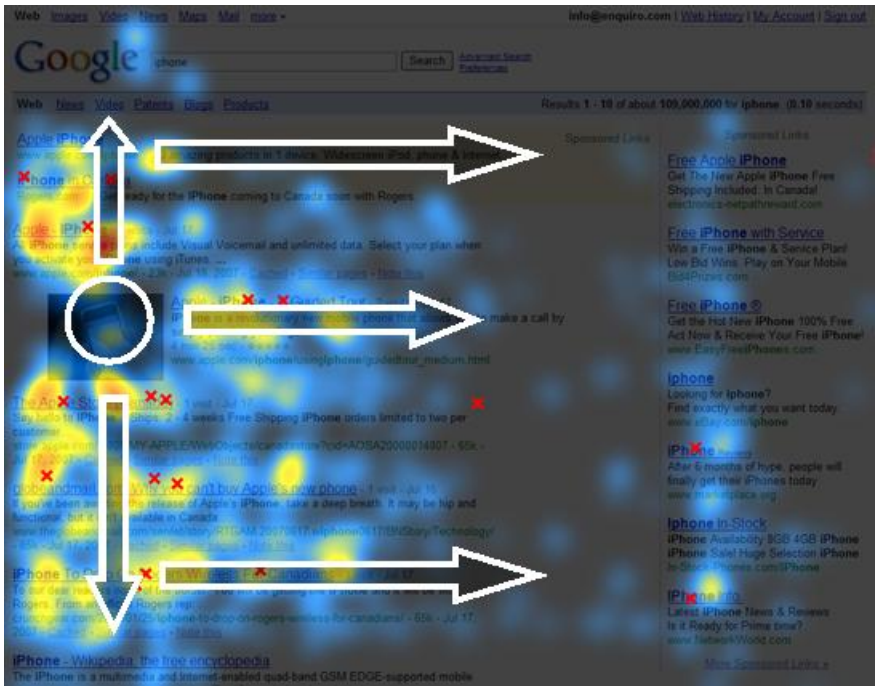


FIGURE 2.9: The "E" shape result-viewing pattern in aggregated interface. The figure is taken from the article[b]

[a]http://searchengineland.com/eye-tracking-on-universal-and-personalized-search-12233
[b]http://searchengineland.com/eye-tracking-on-universal-and-personalized-search-12233

Above studies provide insight on users' searching behavior with respect to various IR systems. There exist certain similarities as well as differences among observed user behaviors pertaining to different IR systems. However, the studies in the past did not take into account the underlying source need. That is, understanding users' searching behavior when their information need requires results from different sources is not explored. For instance, how does session length correspond to the type of information (e.g., image, video, map) accessed? How often users combine results from different information sources? Which are the most frequent combination of information sources accessed? The aim of this thesis is to obtain an understanding of the users' searching behavior with respect to different sources, and to investigate if there exits any patterns in such searching behaviors. Furthermore, the thesis also aims to investigate the interaction of users with the aggregated interface.

2.5 Summary

The main contribution of this Chapter is to present background and motivation for the work presented in this thesis. The work presented in this thesis is related to four main research aspects; relevance in IR, result aggregation, result presentation and user behavior. It is shown that there have been significant efforts in past to address these aspects in context of various information retrieval systems, for example, in digital libraries, databases, and web search.

Relevance has been a key aspect in evaluating the IR systems. Researchers in past have described various attributes and dimensions of relevance depending on the context of search. However, what relevance means within the context of aggregated search in not clear. Furthermore, what dimensions of relevance exits within the relevance space in an aggregated search scenario is also not known. The aim of this thesis to address these issues so that an estimation of relevance can be done effectively.

Through several dedicated studies in past, understanding of how result aggregation or result presentation influence users' behavior in different retrieval systems have been achieved. However, effect of these aspects still remains uncovered in context of aggregated search. Techniques like clustering and distributed IR have shown to be effective in making information access easier. But, whether aggregation of different types (image, video, map, etc) on one result page also improves information access is not known.

There is also limited understanding of result presentation within an aggregated interface. For instance, it is not clear if the result presentation of an aggregated interface has similar effects in task completions as compared to conventional search interface. How (if any) does aggregation of different search results effect users information seeking process. Also, how different design and visualization approach can be used to improve information access in an aggregated search interface is also not known. Answers to these questions are still missing in an aggregated search scenario.

There is also a significant amount of work dedicated in analyzing user behavior with respect to various information retrieval systems. These studies have uncovered various patterns of user behavior dynamics, which were further utilized in improving the retrieval systems. Whereas in aggregated search there is limited body of work addressing user behavior dynamics, which do not provide sufficient insight on information seeking behavior of users in an aggregated search.

Furthermore, a formal introduction to this emerging paradigm is also missing from the literature. There is no clear definition for terminologies associated to aggregated search

paradigm. In addition, framework of an aggregated paradigm is also not provided.

This PhD work aims to address the issues associated to aggregated search, and to lay foundation for the future research in this field. Effect of result aggregation and result presentation in aggregated search is uncovered in Chapter 3 and Chapter 4. The focus of Chapter 5 and 6 is to explore dynamics of user behavior in an aggregated scenario. Finally, overall findings from this thesis work is presented in Chapter 7, which provided several directions for future research.

Part II

Result Presentation in Aggregated Search Interface

Preface

This part presents two user studies, where outcomes of research questions addressing result presentation issues in aggregated interfaces are described. Chapter 3 (study one) compares an aggregated interface to a non-aggregated interface during information access for informational queries. Another comparison between different types of aggregated interfaces is performed during the second study in Chapter 4, where factors affecting click-through behavior are analyzed.

Chapter 3

Aggregated verses Non-Aggregated Search Interface

3.1 Introduction

This chapter presents a user study that evaluated the effectiveness of an aggregated search interface in the context of informational search tasks Sushmita et al. [2009a]. An experimental system was developed to present search results aggregated from multiple information sources, and was compared to a conventional tabbed interface. Sixteen participants were recruited to evaluate the performance of the two interfaces. The results from the study suggest that the aggregated search interface is a promising way of supporting informational search tasks. In particular, the quantity and diversity of the retrieved items, which participants accessed to complete a task, increased in the aggregated interface. Participants also found that the aggregated presentation provided easier access to relevant items when compared to conventional interface.

3.1.1 Structure

The outline of this chapter is as follows: In Section 3.1.2 the study presented in this chapter is motivated. The experimental design is described in Section 3.2. Section 3.3 presents the results and analysis of the study. Finally, Section 3.4 discusses the findings and limitations of the study.

3.1.2 Motivation

This work is motivated by two observations: first, the positive evidence in favor of the usage of aggregated interface; and second, the growth in the percentage of informational queries issued to the search engines.

Starting with the first observation, which motivated this study. Introduction Chapter 1 discussed that there have been evidence on how aggregating results have influenced users' information gathering process. After search engines introduced the aggregation approach in their result pages, an increase in access of non-web[3-1] results (image, video, etc) were observed. People started to access non-web results more often, as compared to when they were accessed through their dedicated vertical, that is, via the conventional tabbed interface. Furthermore, the results of a log analysis study Sushmita et al. [2009b] suggested a potential need for aggregated search interfaces. The study reported in Sushmita et al. [2009b] collected evidence if users actually accessed non-web results during their search. It was found that, although by large standard web pages are being accessed (clicked), the percentage of non-standard web pages that are clicked (news, blogs, images, etc) is not negligible.

Although, there is much evidence to support the need for an aggregated search, yet the effectiveness of an aggregated search interface is not known. There is very little understanding of how effective the aggregation of results are in information gathering process. It was clear that more and more people were accessing non-web results; when provided on an aggregated interface. However, it was not very clear if it was also effective in task completion. This motivated the study (described in this chapter), where effectiveness of an aggregated interface was empirically evaluated.

Next, moving to our second observation mentioned earlier in this section. The increase in the percentage of informational queries issued to search engines further motivated this study. In a study, Broder [2002a] classified the "need behind the query" into three intents namely: *informational*, *navigational* and *transactional*. He explained that, for navigational queries, the immediate intent is to reach a particular site (e.g., BBC Homepage). While for queries which are informational in nature, the intent is to acquire some information assumed present on one or more web pages. Finally, in case of transactional queries, the intent is to perform some web-mediated activity (e.g., download, purchase, etc).

[3-1] The term "non-web" is used to distinguish these media and other document genre to the standard web documents. In current search engine terminology, these "non-web" results could be retrieved directly from the respective verticals.

In the context of aggregated search, the idea of combining results from different sources (images, web, video, etc) seems to be more useful for informational queries. This is because, the aim behind such queries is to gather information from multiple documents or sources. Whereas, for navigational and transactional queries, aggregating results may not be useful, since the aim is to reach a particular website, and not to gather information from different website.

For instance, for the navigational query, "British airways", where the user might aim to visit the British airways website (URLs of the airline website); showing image or video results with the web results in this case may not be useful. Similarly, for an example transactional query, such as; "amazon.com", the intention might be to purchase some product from the Amazon website. It is again not obvious if the aggregation of results would be necessary for satisfying information need behind this query. Therefore, returning an aggregated result page can be considered more appropriate for queries, which are informational in nature.

A recent study reported that 80% of queries submitted to search engines are informational Jansen et al. [2008]. The growth in the percentage of informational queries submitted to search engines suggests that more people are now often seeking general information on a broad topic such as "global warming" or "nutrition". Therefore, there is a need to find effective way of combining results from different sources in order to make information access easier and faster. Aggregated search is one such approach, which aims to make search results more comprehensive and useful [3-2].

Although it is already known that, there is an acceptance among users for information being presented in an aggregated manner, and that now many users are seeking information on general topics (informational queries). However, there still lies a gap in understanding how effectively an aggregated search approach helps in supporting informational search tasks. In this chapter, a task-based user study that compares the performance of an aggregated search interface to a conventional interface for informational queries is presented.

3.2 Experimental Design

A within-subject[3-3] experiment design was used in this study, where two search interfaces (controlled and experimental) were tested by sixteen participants, performing

[3-2] http://googleblog.blogspot.com/2007/05/behind-scenes-with-universal-search.html

[3-3] In a within-subjects (also called 'repeated measures') design, each participant is exposed to all of the conditions of the experiment Field and Hole [2003].

two search tasks with each interface. In the following sections the research hypotheses of this study is defined (section 3.2.1), and the experiment designed to investigate the hypotheses is discussed in detail (section 3.2.2).

3.2.1 Research Hypotheses

The overall hypothesis of this study is that an aggregated result presentation can facilitate informational search tasks by offering diversified search results. More specifically, the following sub-hypotheses were formulated to investigate:

H1 *An aggregated presentation can increase the quantity and diversity of documents viewed by users to complete a task (Section 3.3.1).*

H2 *An aggregated presentation can increase the quantity and diversity of relevant information collected by users to complete a task (Section 3.3.2).*

H3 *An aggregated presentation can improve users' perceptions of the search system (Section 3.3.3).*

While an increased number of clicks can be seen as a sign of confusion in navigational queries, informational search tasks often require viewing a range of documents to complete the task. Therefore, an effective interface should be able to facilitate the browsing of retrieved documents **[H1]**. This should also affect the relevant information collected to complete a task **[H2]**. Finally, participants were expected to have a positive perception on the system that enabled them to perform a task successfully **[H3]**.

3.2.2 Search interfaces

Two search interfaces within a system, called DIGEST system, were devised to address the research hypotheses. Both interfaces used the same back-end search engine (Yahoo! search API[3-4]). For a given query, the API was set to retrieve the top 30 items from four information sources. In this study, web, image, news, and Wikipedia were the four information sources used. The difference between the two interfaces was the presentation of retrieved items.

The interface representing the conventional search approach is the controlled system whereas, the interface representing the aggregated search approach is the experimental

[3-4]http://developer.yahoo.com/search/boss/

system. Tiled aggregation approach was used in the experimental system. Since the aim of the study was to see if aggregated approach improves information access by providing results from different sources, therefore, focus was not on selecting the aggregated design. The comparison between the type of aggregations: vertical and tiled will be shown in the next chapter.

Figure 3.1 shows the controlled system (for the example query "London") where the results from the four sources were presented in separate tabs. The default source was set to the web tab, and the participants could click other tabs at the top of the interface to view the results from other sources. This represented a conventional vertical presentation of search results available in major search engines. The controlled system presented the first 10 results for every selected information source with an option of "more results" at the bottom to view the remaining 20 results (in sets of 10).

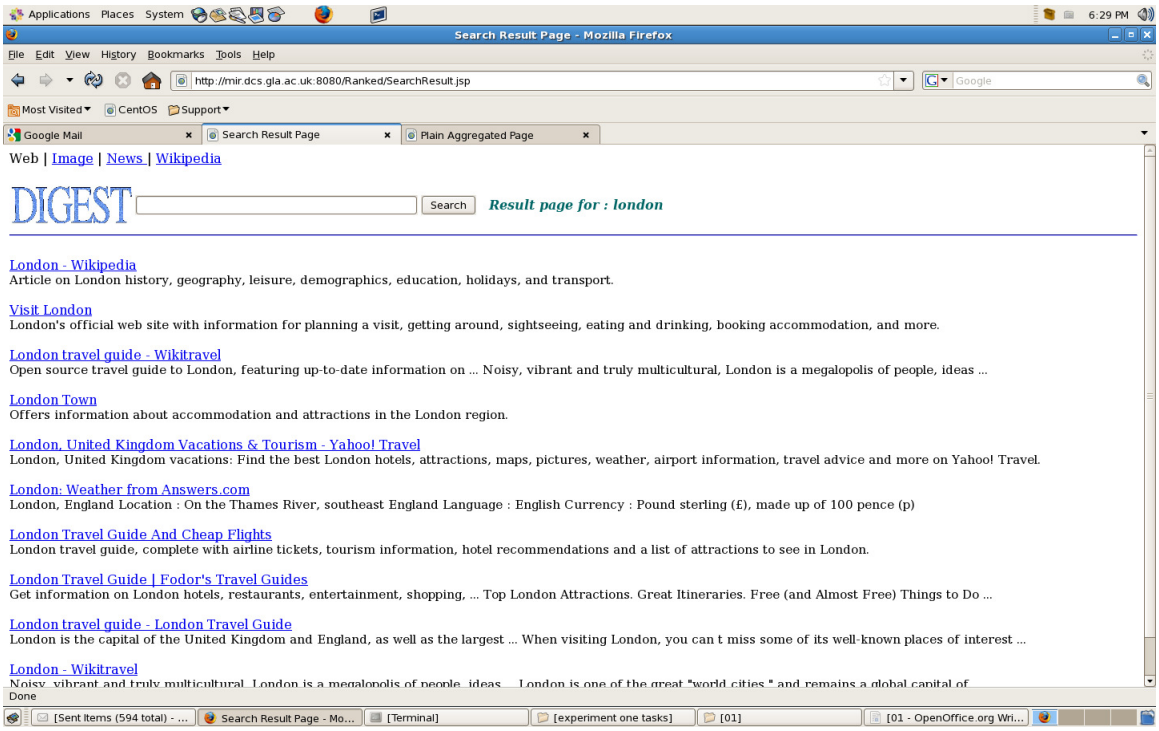


FIGURE 3.1: Controlled System (Tabbed)

Figure 3.2 presents the experimental system (for the example query "London") where the results from the four sources were integrated into a single page. This represents an aggregated presentation of search results. The first 10 web results, 12 image results, 10 wiki results and 5 news results, were shown, in each corresponding panel. Every information source on the experimental system also had an option of "more results" (similar to the controlled system) in order to view the remaining results.

The layout of the four sources was synthesized from one of the existing examples of

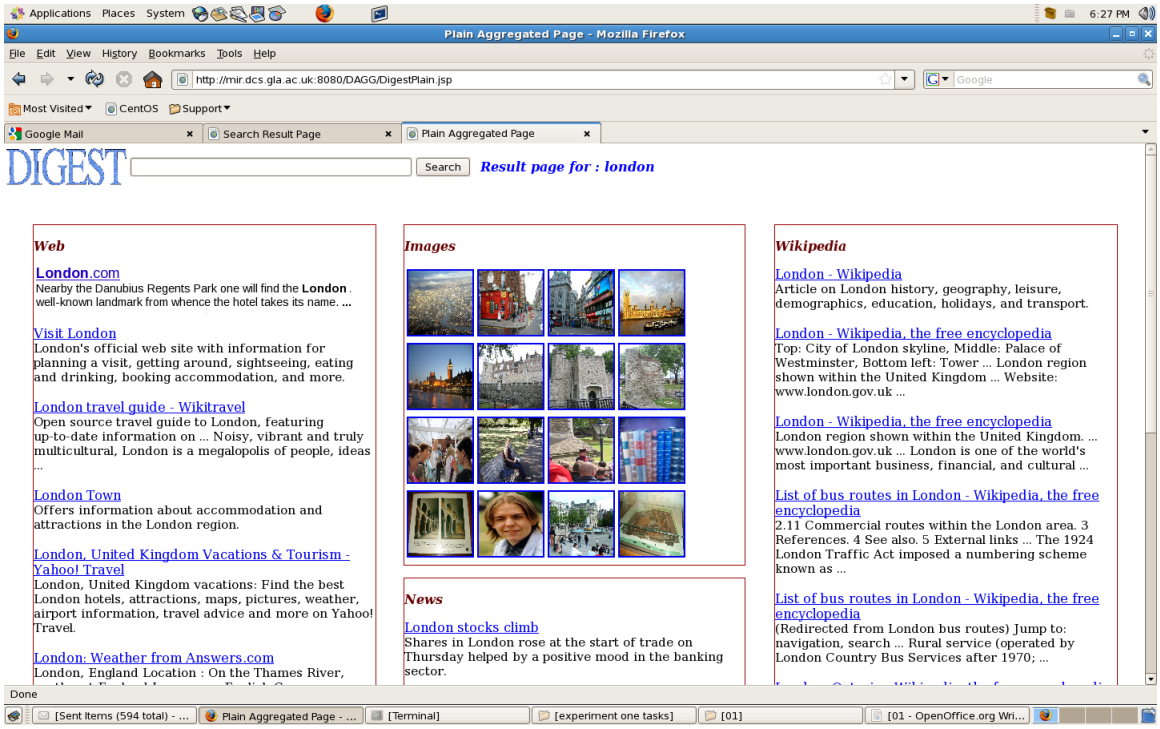


FIGURE 3.2: Experimental System (Aggregated)

aggregated interface[3-5], and was fixed throughout the experiment. Since the focus of the experiment was not to determine an optimal layout, therefore, one design layout was approximated and used in the study. Furthermore, since a single layout was used throughout the study, it minimized the layout effect on user searching behavior during the study; which otherwise might have been the case, if different layouts were used instead. A formal study to determine an optimal layout is suggested for future work in section 8.4.5 of Chapter 8.

3.2.3 Task

Participants of the user study were asked to perform informational search tasks using the interfaces described earlier. Each search task was based on the simulated work task[3-6] situation framework proposed by Borlund [2000]. The framework was designed to encourage participants to engage with an artificial task by giving a situational background scenario for the task. Figure 3.3 shows an example of the search scenario. As can be seen, our search tasks required browsing several documents and collecting relevant information from multiple sources.

[3-5] http://au.alpha.yahoo.com/search/web?p=london

[3-6] A simulated work task situation is a short 'cover story' that describes a situation that leads to an individual requiring to use an IR system Borlund [2000].

Participants were asked to copy and paste relevant texts, URLs, and images to word processing software during the task. Microsoft word software was used as an electronic notebook. Examples of notebooks made by participants are shown in Figure 3.4. Six search scenarios (provided in Appendix) were prepared so that participants could choose the scenarios based on their interest. This design aimed to facilitate participantsÕ engagement with the artificial search tasks. Participants were given fifteen minutes to complete each task. Each participant performed four search tasks, two with the experimental system and two with the controlled system. The order of the systems was rotated to reduce learning effects.

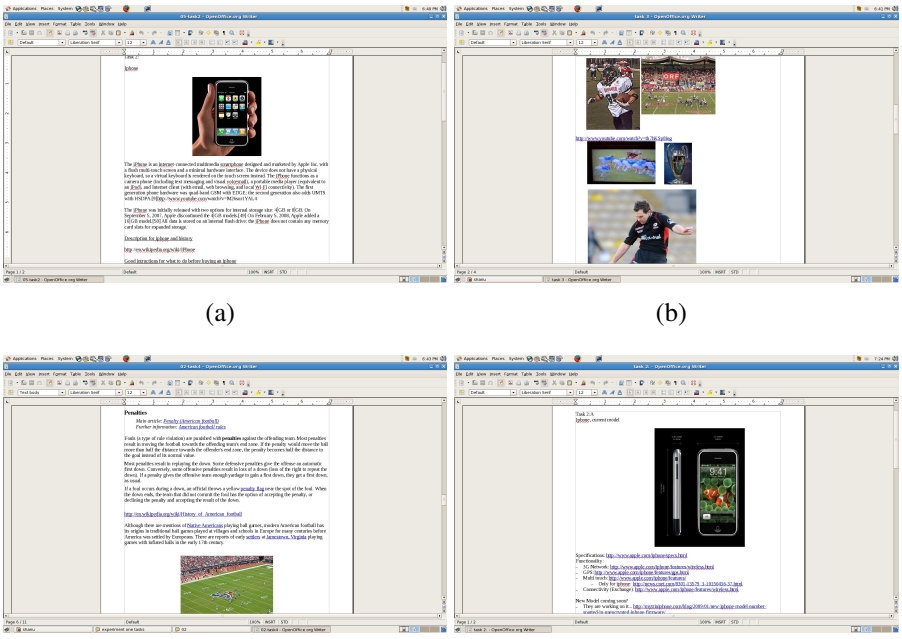
Topic: European verses American football

Task: Your friend shared her experience of her recent visit to a state in US, and mentioned about an American football match to which she was invited to watch. Having only watched European soccer in her life, she was surprised to see that American football was quite different from European football matches she had watched. She found rules and the game to be completely different from what she expected and was quite confused during the match. You decided to help her by explaining the difference between them. Your task is to collect information on American and European football, their rules, the game, their origin, etc. You may also show her some pictures of both the games, pictures of some popular sports person from both, etc. It would be nice to update her about any recent event or news about these games.

FIGURE 3.3: An example simulated work task used during the experiment.

3.2.4 Participants

The experiment was carried out with 11 males and 5 females from University of Glasgow. Out of 16 participants, 7 were undergraduate students, 2 postgraduate students, 3 PhD students, and 4 were research staff members. The participants were from various educational fields, namely: computing, business management, arts and commerce. The participants were recruited through a formal call for participation email distributed to several lists. An entry questionnaire established that 82% of participants stated that they had accessed more than one information source to complete a search task. Therefore, the participants of the study were not unfamiliar with search tasks that require multiple sources. However, none had used the Digest interfaces or the tasks before.



(a) (b)

(c) (d)

FIGURE 3.4: Sample information collected during search tasks by participants

3.2.5 Procedure

For each participant, the experiment was performed in the following manner. When they arrived at the experiment site, they were welcomed and explained the overall aim of the experiment. When they agreed to participate, a consent form was signed. Then, they were asked to fill in an entry questionnaire to capture their profile and search background. Next, they had a training session with both interfaces using a sample search task. The training session typically lasted for five minutes

Then, they were asked to perform the first search task by selecting the most interesting scenario from the six scenarios. During the task, the system automatically logged participants' interaction with the interface. When the first task was completed, they were asked to fill in a post-task questionnaire to capture their subjective assessments on the system and task. Then, participants were informed of the change of the interface, and the second scenario was selected. This was repeated four times. After the completion of the four tasks, they were asked to fill in an exit questionnaire to capture their perceptions of systems and tasks as a whole. Participants were rewarded fifteen pounds for their participation after the experiment.

3.3 Results

This section presents the results of the experiment based on the research hypotheses stated in Section 3.2.1. A total of thirty-two search sessions per system were used in the analysis. To measure the statistical significance of the results, both t-test (parametric) and Wilcoxon signed rank test (non-parametric) were applied to the difference between the controlled and experimental systems. All tests were paired and two-sided, and critical value was set to 0.05, unless otherwise stated.

3.3.1 Quantity and diversity of information accessed

The first hypothesis [**H1**] looked at the effect of an aggregated presentation on the quantity and diversity of documents participants viewed to complete a task. To examine this hypothesis, first the participants' click-through data on different information sources were analyzed. The results are shown in Table 4.12.

The bottom row of the table shows the average number of retrieved items viewed to complete a task. As can be seen, participants viewed a significantly larger number of items in the experimental system when compared to the controlled system, where $t(31) = 4.2, p = 0.000$, with a 95% confidence interval.

The breakdown of the information sources suggests that the difference was due to the significantly different frequency in the wiki $(t(31) = 4.2, p = 0.000)$ and image $(t(31) = 3.5, p = 0.0013)$ sources. These results provide a support for that the aggregated presentation increased the quantity of retrieved items viewed.

The combinations of information sources accessed by participants to complete a task were also looked at during the analysis. The results are shown in Table 3.2. As can be seen, in five more sessions (10th row in Table 3.2), participants accessed all four information sources in the experimental system (13) when compared to the controlled system (8). In addition, more sessions were completed by a single source (Web) in the controlled system. This suggests that the aggregated presentation encouraged participants to view more diversified sources from search results. It was also noticed that the frequent source was different in the two systems. When the diversity score of 3 was considered, the sources of web, image, and news was the most popular combination in the controlled system while the web, image, and wiki were the most common combination in the experimental system. This aspect will be discussed later. Overall, the results provided some evidence to support **[H1]**.

TABLE 3.1: Frequency of participants' clicks per information sources for Controlled System (Cont. Sys) and Experimental System (Exp. Sys). Where Mean (SD) = mean and SD of click through frequency on augmented elements, and, t= t value, df = degree of freedom and p_1 = p value from T-test. The z = z statistics and p_2 = the p value from the Wilcoxon test. The number of sessions analyzed = 64, 32 for each interface.

Source	Mean (SD) of Cont. Sys	Mean (SD) of Exp. Sys	t	*df*	p_1	z	p_2
Web	7.7 (6.7)	8.6 (5.9)	0.9	31	0.3700	180	0.1847
News	1.7 (1.7)	1.0 (1.2)	1.9	31	0.0663	256	0.1039
Wiki	1.0 (1.5)	2.7 (2.2)	4.2	31	**0.0002**	39	**0.0005**
Image	2.4 (3.2)	6.8 (7.7)	3.5	31	**0.0013**	47.5	**0.0004**
All	12.8 (8.8)	19.1 (12.1)	4.2	31	**0.0002**	60.5	**0.0002**

TABLE 3.2: Combination of information sources, where W=web, I= image, N=news and Wi= wiki

Diversity	Sources	Controlled system	Experimental system
1	W	4	0
2	W+I	2	3
2	W+N	1	2
2	W+Wi	1	2
2	I+Wi	0	1
3	W+I+N	12	1
3	W+N+Wi	3	1
3	W+I+Wi	1	9
4	W+I+N+Wi	8	13
	Total	32	32

3.3.2 Quantity and diversity of information collected

The second hypothesis examined whether or not an aggregated presentation increased the quantity and diversity of relevant information collected by participants to complete a task. To answer this hypothesis, a similar analysis to the previous section was performed but on the number of texts, images, and URLs collected in the notebook. The number of texts was counted based on the number of paragraphs. The results of the analysis are shown in Table 3.3. Again, the bottom row of the table shows the average number of collected items to complete a task. As can be seen, participants collected five more items in the experimental system when compared to the controlled system. The difference was found to be significant by the Wilcoxon test (where, $z = 122.5, p - value = 0.0409$). The breakdown of collected items shows that participants tended to collect more items in all three types (Texts, Images, and URLs) when they used the experimental system.

TABLE 3.3: Information collection using Controlled (Cont. Sys) and Experimental systems (Exp. Sys). Where Mean (SD) = mean and standard deviation of information collected, and, t= t value, df = degree of freedom and p_1 = p value from T-test. The z = z statistics and p_2 = the p value from the Wilcoxon test.

Information	Mean (SD) of Cont. Sys	Mean (SD) of Exp. Sys	t	*df*	p_1	z	p_2
Text	7.8 (13.2)	10.8 (21.4)	0.9	31	0.3657	95.5	0.3211
Images	3.3 (2.8)	4.6 (3.4)	1.6	31	0.114	137.0	0.0815
URLs	6.1 (5.3)	7.4 (7.1)	1.3	31	0.1956	113.5	0.3025
All	17.3 (12.7)	22.7 (18.6)	1.6	31	0.1173	122.5	**0.0409**

TABLE 3.4: Information collected using Controlled and Experimental systems for text, image and ulr combinations. Here, I=image, T= text and U = url

Diversity	Sources	Controlled system	Experimental system
1	I	0	2
1	U	2	0
2	I+T	9	6
2	I+U	11	1
2	T+U	0	12
3	I+T+U	10	11
	Total	32	32

However, no difference was found to be significant. Table 3.4 shows the combination of the collected items. As can be seen, the number of sessions where all three types were collected (diversity score 3) was similar across the systems.

The frequency in the other two diversity scores (diversity score 1 and 2) was also found to be comparable. However, there was some noticeable difference in the combinations. More specifically, the combination of Image and Text (I+U) and combination of Text and URLs (T+U) had a very different frequency across the systems. Further examinations of log files could not provide any clear answer to the cause of this difference. To summarize, the results provided partial evidence to support the quantity aspect of **[H2]**, but no obvious evidence was found to support the diversity aspect of the hypothesis.

3.3.2.1 Quality of Information Collected:

In order to ensure that the information collected by users were relevant to the given task, the information collected by them, that is, images, the text and the URLs of the web pages were also analyzed. The aim was to ensure that the collected information by the participants were not only higher in quantity but were also relevant to the given

task. After analyzing the collected information, it was observed that the images, text and URLs collected by the participants were relevant to the respective task in most cases. This was observed with the experimental as well as the controlled system, and the participants were able to collect relevant information using both the systems. However, since the quantity of the collected information was higher when the experimental system was used (Section 3.3.2), therefore, the amount of relevant information collected by using aggregated approach was also higher. This suggests that, aggregated search approach can be seen as an effective medium, which facilitated information access.

3.3.3 User perceptions

The last hypothesis [**H3**] looked at the effect of the aggregated presentation on participants' perceptions of the systems. To answer this hypothesis, participants' subjective assessments on the systems were analyzed, which were captured by a 5-point Likert scale in the exit questionnaire. More specifically, their agreement on the following two statements for each of the two systems was asked.

Q1 The system was useful to complete my search tasks (1 = Strongly agree; 5 = Strongly disagree).

Q2 It was easy to find relevant information with the system (1 = Strongly agree; 5 = Strongly disagree).

TABLE 3.5: Users' perceptions on the Controlled (Cont. Sys) and Experimental systems (Exp. Sys) for N=16. Where Mean (SD) = mean and standard deviation, t= t value, df = degree of freedom and p_1 = p value from T-test. The z = z statistics and p_2 = the p value from the Wilcoxon test.

Question	Mean (SD) of Cont. Sys	Mean (SD) of Exp. Sys	t	*df*	p_1	z	p_2
Q1	2.4 (1.1)	1.9 (1.1)	1.13	15	0.2622	37	0.3542
Q2	2.4 (1.0)	1.8 (0.9)	1.84	15	0.0859	52	0.0932

Since the third hypotheses (**[H3]**) expected the experimental system to have a better assessment than the controlled system, the statistical tests were applied with paired but one-tailed where an alternative was set to be greater. Note that a lower value represented a higher degree of agreement in our analysis. The results are shown in Table 3.5. As can be seen, participants tended to find the experimental system easier to find relevant information to complete a task. Although participants tended to give a better score on

the experimental system regarding the usefulness, the difference was not found to be significant. We also asked participants which system was easier to access search results in the exit questionnaire. 75% of participants selected the experimental system for the question. Overall, these results provide partial evidence to support third hypothesis **[H3]**.

3.4 Discussion

This section first discusses the implications of the results on the design of aggregated search interfaces, followed by the limitation of this study. Beaulieu [1997] observed the trade-off between the complexity of search interfaces and cognitive load of the users. This applies to the design of aggregated search interfaces, too. The experimental system of this study used a more complex presentation than the controlled system to integrate multiple information sources in a single page. Therefore, the aggregated search interface could increase the cognitive load of the end-users. However, the experimental results suggested that participants were capable of interacting with an aggregated presentation, and tended to find the experimental system easier to find relevant information when compared to the controlled system. This might be because the controlled system still required extra effort to select information sources to access a range of retrieved items.

Another implication was that the layout of aggregation was likely to affect people's selection of information sources. In Section 3.3, it was found that the combination of the web, image, and news was the most common selection in the controlled system while the web, image, and wiki were the popular selection in the experimental system. They were exactly the same order of the sources in the interfaces. The tab on the top of the controlled interface listed the sources in the order of web $\rightarrow$ image $\rightarrow$ news $\rightarrow$ wiki. The top three panels of the aggregated interface were the web, image, and wiki. This suggests that people's browsing of information sources can be sequential, and their attention moves horizontally rather than scrolling down the result page vertically. This also implies that an aggregated search interface might be able to offer an effective support by optimizing the order of information sources for different tasks or queries.

Furthermore, when the participants were asked for their overall preference for aggregated and non-aggregated interface during study 62% preferred aggregated, 18.75% preferred non-aggregate approach, and 18.75% preferred both. Participants preferring non-aggregated approach suggested that because of their acquaintance with similar system (Google, yahoo, etc) over the years, they were more comfortable while using the

system. Participants who preferred both the system suggested that they would prefer to have a combination of both systems where they could have an option to switch between both the systems depending upon their search task. This was an important feedback since it was also hypothesized that aggregation of search results would be beneficial for informational queries.

Therefore, it would be interesting to have a system capable of dealing (or switching) between aggregated and non-aggregated paradigm, depending on the query intent. This would be an interesting research for future work. Some users also suggested that having results aggregated on one page helped them in comparing the various results together and hence provided a broader overview of the search task topic.

3.4.1 Limitations

There are some limitations in this study; First, only one back-end search engine (Yahoo! API) to test the effectiveness of the interfaces was used. Although this made the comparison fair, the implication of the results is limited to this particular engine. Second, the systems were tested with a small number of topics compared to a system-centered evaluation. Other types of tasks such as a decision-making task would also give a better understanding of the effect of aggregated presentation. Third, the collected items were based on perceived relevance and the quality of collected items was not assessed. Finally, the layout of aggregation was fixed in our experiment. This seems to have an implication on participants' information seeking behavior, which will be discussed next.

3.5 Summary

In this chapter, a task-based user study was described to compare the effectiveness of an aggregated presentation to a conventional presentation. The study presented investigated the effect of the aggregated presentation on the quantity and diversity of information objects accessed by users during informational search tasks.

In conclusion, the study provided empirical evidence to support that an aggregated presentation of information sources can increase the quantity and diversify of the retrieved items accessed to complete informational search tasks. Participants tended to find the aggregated presentation easier to access retrieved items and to find relevant information. Although these positive effects were not strong enough to increase the amount

of relevant information collected, it can be speculated that an intelligent way of organizing information sources is a key to achieve such a goal. Despite its limitations, the presented study is the first step into a new and unexplored domain and a contribution to the exploration of the understanding of aggregated interface.

Another interesting observation, which came up during the analysis of this study, was that the layout of an aggregated page might have implications on users' click behavior. In the next Chapter 4, study presenting factors affecting click behavior in aggregated interface is described, which would help in further understanding of aggregated interface, and its design concerns.

Chapter 4

Factors Affecting Click-Through Behavior in Aggregated Search Interfaces

4.1 Introduction

In the pervious chapter, the effectiveness of an aggregated search result presentation approach was evaluated. It was observed that an aggregated interface seems to be a promising approach in satisfying information-gathering process for informational queries. This chapter presents two user studies investigating factors affecting users click-through behavior on aggregated search interfaces Sushmita et al. [2010a]. The aim is to gain an insight into the implications of result aggregation with respect to their position, relevance of source to the given search tasks, etc.

In order to achieve this, two aggregated search interfaces is tested: one where results from the different sources are presented into a single list, that is **vertical representation** of search results (Section 4.3.1), and another, where results from each source are presented in a separate panel, that is **tiled representation** of search results (Section 4.3.2). A total of 1,296 search sessions performed by 48 participants in two separate studies are analyzed in this chapter. Results from the study suggest that: 1) the position of augmented search results is only significant in the vertical design and not in the tiled design; 2) participants' click-through behavior for videos is different from other sources; and finally 3) capturing a task's orientation towards particular sources is an important factor for further investigation and research.

4.1.1 Structure

The rest of the chapter is structured as follows: The study presented in this chapter is motivated in Section 4.1.2. The research questions investigated during the study are described in Section 4.1.3. The experimental design to address the research questions is described in Section 4.2. Section 4.3.1 reports the results from the study investigating factors on vertical approach of aggregating results while, Section 4.3.2 reports the results from the second study investigating factors on tiled approach of aggregating search results. Section 4.5 contains a discussion of findings. Finally, the chapter finishes with conclusions in Section 4.6.

4.1.2 Motivation

Aggregated search interfaces are now a common paradigm for search result presentation. An aggregated search interface is designed to integrate search results from different sources (web, image, video, news, blog, tweet, etc) into a single result page. An objective of aggregated search is to facilitate the access to the increasingly diverse content available on the web.

There appears to be at least two main types of aggregation approaches; **vertical** and **tiled**. A vertical aggregation (as initiated by Google Universal Search[4-1], and now used by many other search engines) presents results from different sources within a single ranked list. Vertical aggregation is one where results from different sources are blended in together to form a single ranked list. Such type of aggregation is also referred to as merged or slotted aggregation, since results from image, video, etc are merged in with the web results or slotted between web results. An example of vertical aggregation is shown in Figure 4.1. The Figure shows aggregation of image and video results with web results on the result page.

Whereas a tiled aggregation (e.g., Alpha Yahoo![4-2] or Naver[4-3]) presents results from each source in a separate panel. There is no blending or merging of results in such type of aggregation, and the distinction among sources is more evident. Example of a tiled like aggregation followed by Yahoo! alpha is shown in Figure 4.2. It can be seen that for the example query *iceage-3*, results from image, web and video are provided in separate panels.

[4-1] http://googleblog.blogspot.com/2007/05/universal-search-best-answer-is-still.html
[4-2] au.alpha.yahoo.com/
[4-3] www.naver.com/

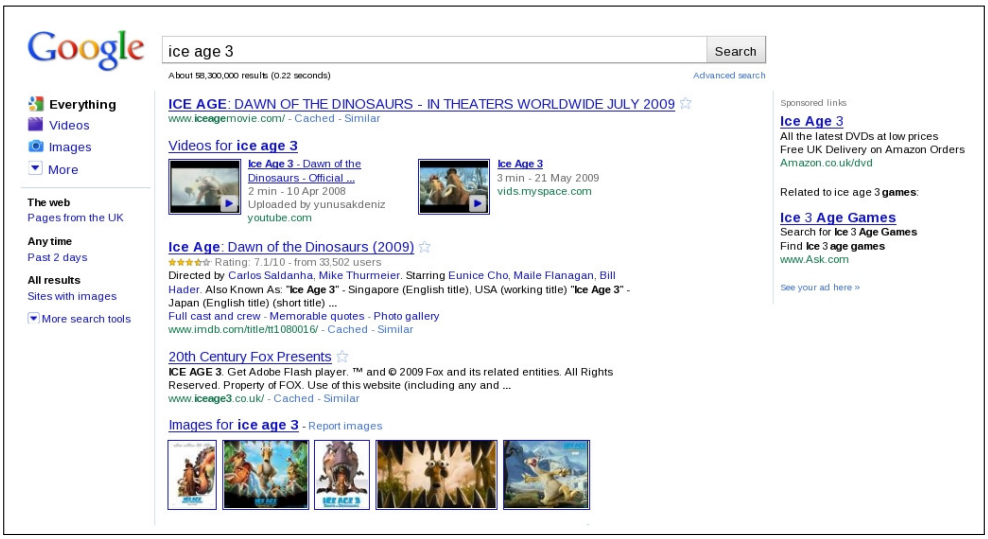


FIGURE 4.1: Example of a vertical aggregated search result page by Google for the query "ice age 3".

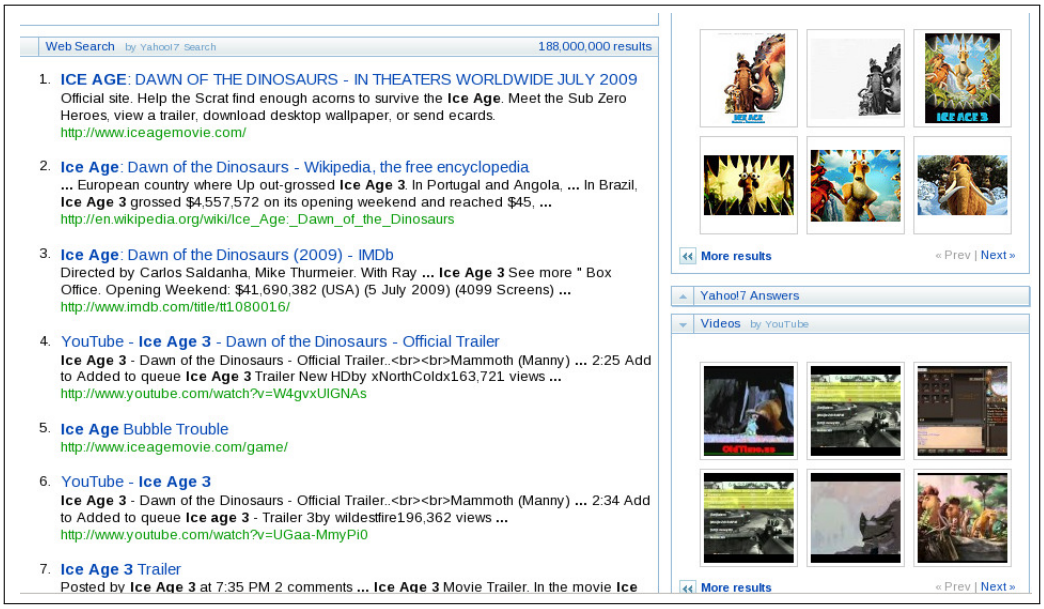


FIGURE 4.2: Example of a tiled like aggregated search result page by Yahoo! alpha for the query "ice age 3".

The main aspect *distinguishing* an aggregated search interface from a conventional search interface is that results from different sources are aggregated within the default web search results. This increases the complexity in the design of search interfaces and result presentation. This distinction raises two main issues:

1. How should we determine the relevance of a source to a search task when given

a query?

2. How should we organize the search results with multiple sources?

There have been several studies looking at the first issue (e.g. Arguello et al. [2009a]; Diaz [2009]; Arguello et al. [2009b]; Sushmita et al. [2009b]). On the other hand, research on the second issue in the context of aggregated search interfaces is still limited when compared to the first issue. Therefore, to examine the second issue, two user studies are carried out (described in this chapter): one using a vertical representation and another using a tiled approach. The objective of the studies is to investigate the impact of factors on people's click-through behavior on the aggregated search result presentation. The factors studied include:

1. position of search results,
2. source types and,
3. strength of search task's orientation towards a particular source type (Source-Orientation).

As already discussed in the Background chapter (Section 2.1), an important task of an aggregated search system is to determine, for a given query, not only whether a document is relevant, but from *which* sources relevant documents should be retrieved.

It should be noted that it was not the aim of this work to study how to select the sources from which documents should be retrieved. Instead, given that the source orientation of an information need is known, the aim is to investigate the effect of this on users click-through behavior.

In other words, the presented study is motivated to investigate the relevance of the source to the search task, and its effect on users' searching behavior. To capture this aspect, this study measures the ***"source orientation"*** of an information need, which refers to the degree to which documents *from* a specific source would be relevant to complete the corresponding search task.

To summarize, the work reported in this Chapter is concerned with: the effect of the source orientation of an information need, the positioning of search results, and source types on users click-through behavior. These effects are studied for both vertical and tiled aggregated search interfaces separately (Sections 4.3.1 and 4.3.2), and then compared in Section 4.4.

4.1.3 Research Questions

This subsection defines the research questions investigated in this study. In this chapter, results from conventional web search will be referred as the *"base elements"* of aggregated result presentation, and those from all other sources such as images, videos, and news, as *"augmented elements"*.

R1 *How does the position of augmented elements affect users' click-through behavior in aggregated search interfaces?*

It has been shown in previous studies that click-through behavior is strongly affected by the position of results in the ranked list (e.g., Joachims et al. [2005]; Agichtein and Zheng [2006]; Guan and Cutrell [2007]; Keane et al. [2008]). However, less is known for aggregated result presentation. Therefore, the aim here is to investigate how the positions of augmented elements affect user behavior, in both aggregation designs, to gain a further insight into this problem.

R2 *How does the source orientation of an information need affect users' click-through behavior in aggregated search interfaces?*

While the thematic (or topical) relevance of documents is important, it can be argued that the source type may also be an important dimension in the relevance space for investigating aggregated search. This can be thought of in a similar manner to how structural relevance is considered as another dimension of relevance in XML retrieval research Lalmas and Tombros [2007]. Therefore, it was decided to incorporate the source orientation factor into the experimental design of the study presented in this chapter.

R3 *How does the effect of position and source orientation differ over the source types of augmented elements?*

The complexity of aggregated search interfaces partly stems from the range of potential sources that can be aggregated, and because of this, it is important to understand the influence the different types of source can have on the click-through rates of the augmented elements. Therefore, the last research question is concerned with the interaction effect between the augmented source types and the other two factors described in R1 and R2 (position and source orientation).

4.2 Experimental Design

To investigate the factors affecting click-through behavior in aggregated search interfaces, two separate within subject studies with 48 participants (24 using a vertical interface design and 24 using a tiled design approach) were carried out.

It should be emphasized that it was not the aim of this work to compare the performance of the two interfaces, but instead to make separate observations for each. Since these two main approaches to aggregating results exist (vertical and tiled), the factors for both approaches were tested.

Research reported in Sushmita et al. [2010b] and Arguello et al. [2009b] showed that the typical pattern with multiple-sources (in terms of click-through and relevance) was that at most two sources are involved, with the conventional web source being highly predominant. Consequently, in this study only the aggregation of two sources (conventional web and one other source) is considered at any one time. In addition, image, news and video were chosen as the non-web sources, as these were some of the most frequently used sources in existing aggregated search interfaces.

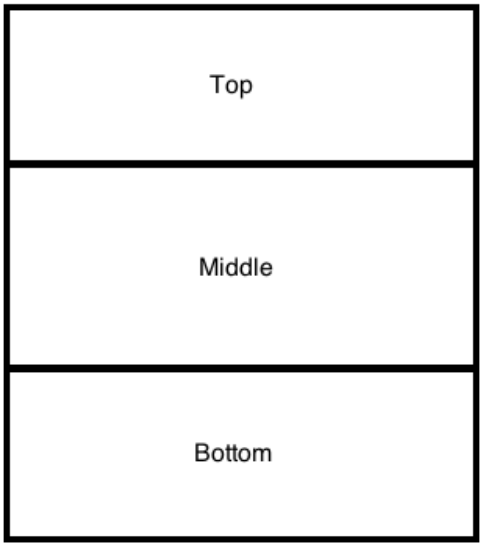
The factors examined in the study were represented by three independent variables: position, source of the augmented elements, and the source orientation of the search task. Each variable had three levels, as listed below:

1. **Position (P):**
   In vertical representation: Top, Middle and Bottom (Figure 4.3(a)).
   In tiled Format: Left, Top-Right and Bottom-Right (Figure 4.3(b)).
2. **Augmented Element (AE):** Image, News and Video.
3. **Source-Orientation (SO):** High, Medium and Low.

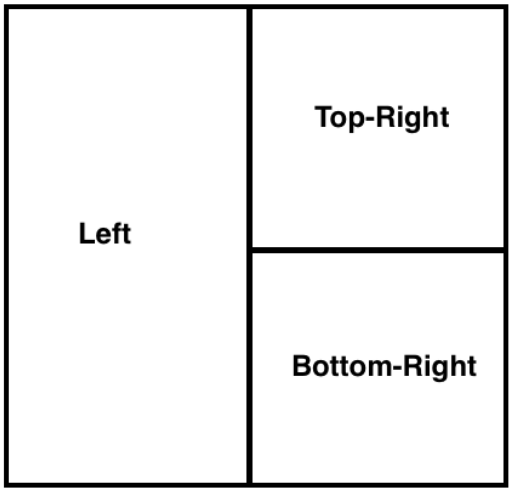
To measure the effect of the independent variables on users click-through behavior, two dependent variables were proposed: the frequency of clicks on augmented elements, and the ratio of those clicks that are bookmarked by participants. Participants of the study were asked to bookmark the search results perceived to be needed to complete the search tasks.

Deciding which interface positions augmented elements could be placed and thus examined in this study was not trivial. There are many possible combinations in both the vertical and tiled interface designs. Based on the observations of existing aggregated search interfaces (as shown in Figure 4.1 and 4.2), it was decided to focus on the three

positions listed above. These positions can be considered as reasonable approximations for both designs. The estimations of positions are shown in Figure 4.3.



(a) Vertical Representation



(b) Tiled Format

FIGURE 4.3: Three positions that were considered for each type of aggregation approach.

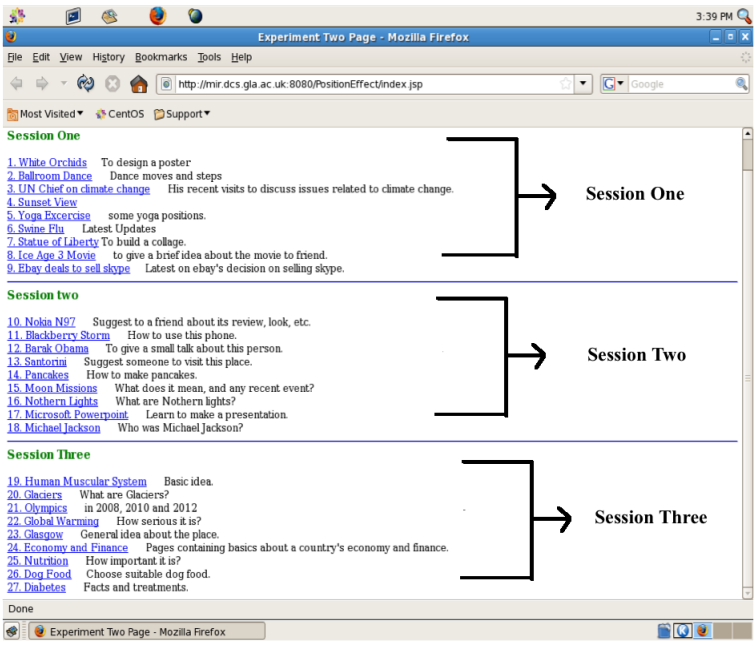
For both the studies (vertical representation and the tiled representation), the same ex-

perimental design was used, unless otherwise stated. This means that, the number of participants, the search tasks and topics, the duration of experiments, and the questionnaires were same. The same search results were shown on both interfaces. The rest of this section contains the details of the experimental design

4.2.1 Interfaces

Two separate aggregated search interfaces were designed, one for each type of aggregation. The first interface (Figure 4.6) was used to study the factors' effect in vertical representation for aggregating results, and the second interface (Figure 4.7) was used to investigate the factors for the tiled format for the aggregation.

FIGURE 4.4: List of twenty-seven topics as shown to the participants. The topics were divided into three sessions, therefore each session contained nine topics. For each topic, a fixed result page was shown to the participant, as shown in Figure 4.5



Each interface contained a fixed set of search results, one set for each of the twenty-seven topics experimented. Participants were shown one result page at a time, one for each topic. List of topics as displayed to the participants can be seen in Figure 4.4, and an example result page for a given topic is shown in Figure 4.5. The search tasks and topics are described in the next Section 4.2.2.

All the search results were fetched prior to the experiment, so that all participants were presented with the same set of results for a given topic. Yahoo! search API was used

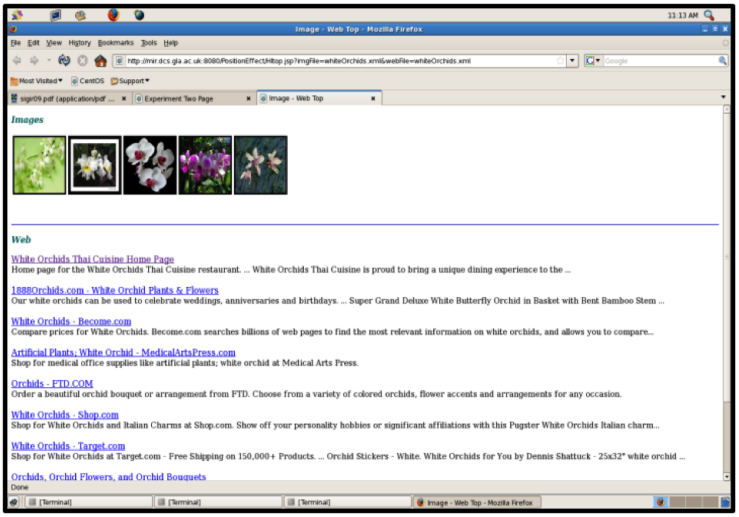


FIGURE 4.5: An example result page. The result page was fixed for all participants, and for each topic one such result page was displayed.

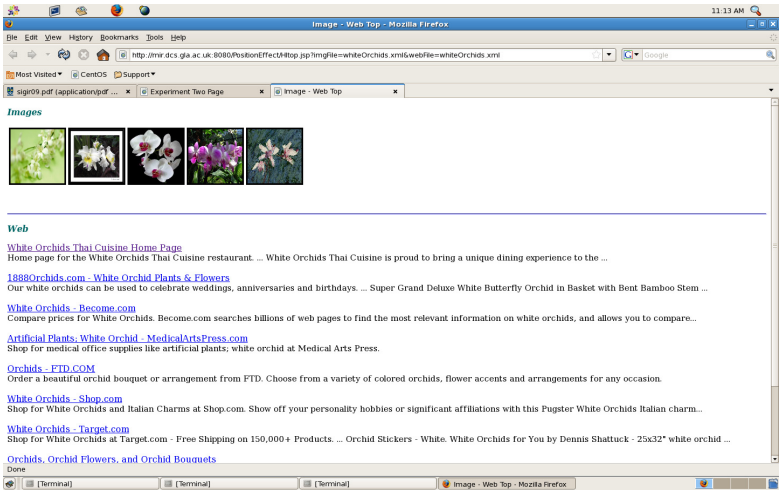
to fetch the results using the topic texts as queries. For each query, the results fetched corresponded to the top 10 results from the conventional web, and the top 5 from each of the other three sources. The ranking of results within each source was preserved (no re-ranking was performed). The only difference between the vertical and tiled interfaces was the organization of the results (i.e. their positions on the interface).

For a given topic, the interface was set to display the corresponding result page showing 15 items from two selected sources. These items were fixed and were the same for all participants. The result page showed 10 results from the conventional web (base elements) and five from the non-web source (augmented elements). These sizes are similar to typical aggregated interfaces.

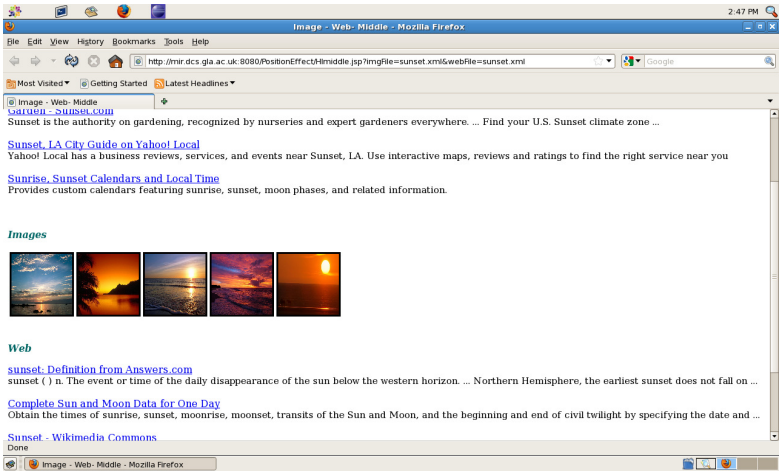
The positions of the augmented elements were varied; at the top, the middle and the bottom of the web results in the interface representing vertical aggregation, and left, top-right and bottom-right in the tiled like aggregated interface. The base and the augmented elements were not merged[4-4] in the vertical design. This is the strategy adopted by most current search engines (with the sources used in this study), and avoids the need to devise a merging algorithm, and consider its effect on users click-through behavior.

Since three positions were to be tested in both interfaces, this resulted in three panels for the tiled layout design. To use the same number (two) of sources in both studies, two panels were used to display, respectively, the base elements and the augmented

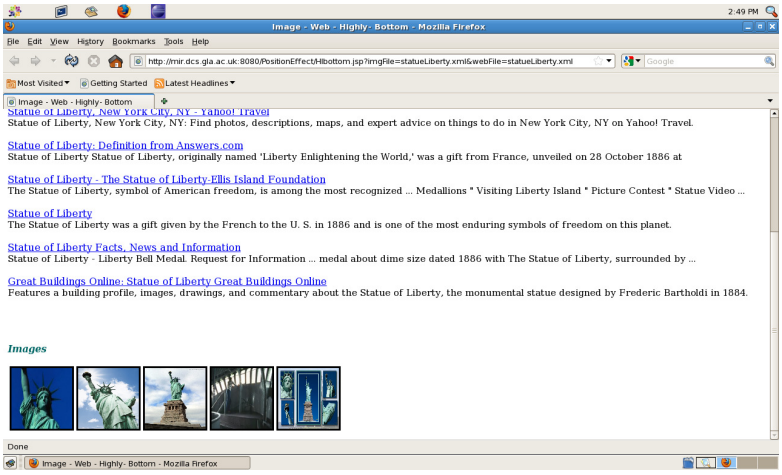
[4-4]In search engine terminology, the non-web (image, video, etc) results were slotted at three different positions on the result page (top, middle and bottom).



(a) Image on top

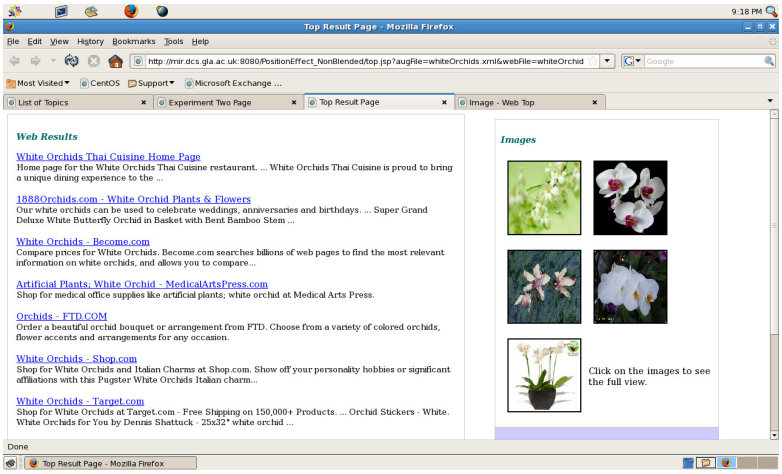


(b) Image in the middle

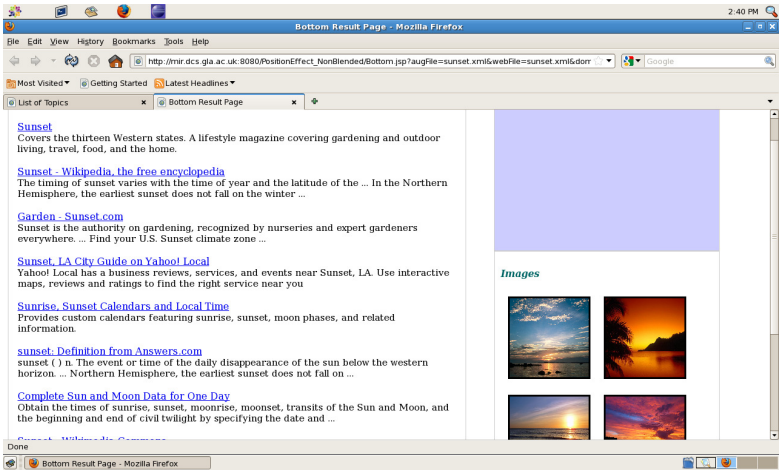


(c) Image at the bottom

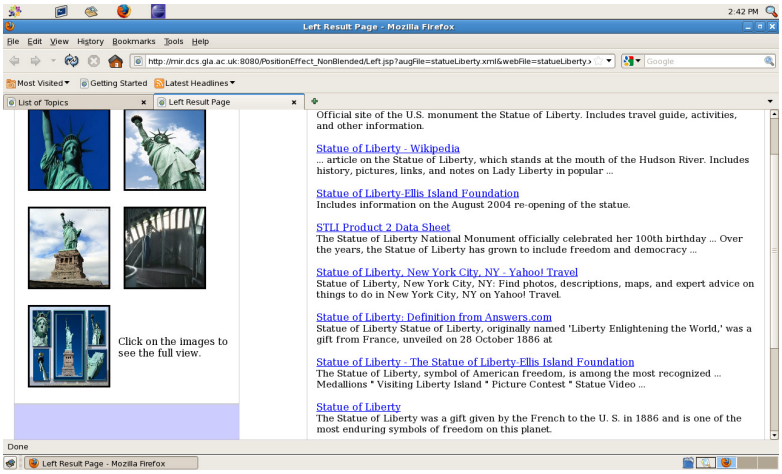
FIGURE 4.6: Examples of result pages in the vertical form of aggregation approach.



(a) Image on right-top



(b) Image at the right-bottom



(c) Image in the left

FIGURE 4.7: Examples of result pages in the tiled format approach.

elements; the third panel was kept empty. To distinguish between the three panels, the empty panel had a light background color (see Figure 4.7).

Questions in the exist questionnaire checked whether participants were distracted or disturbed by the presence of the colored panel. It was found that 98% of the participants did not notice the colored panel. Those who did (2%) confirmed that the color did not distract them. Therefore, we can assume that the colored panel did not lead to any biased observations for the tiled design.

4.2.2 Tasks

Since the experimental design had three independent variables (position, source of the augmented elements, and source orientation of the search task) and each variable had three levels, one search task (a topic and a task description) for each combination of these variables ($3 \times 3 \times 3 = 27$) were formulated, as shown in Table 4.1. This led to a total of twenty-seven search tasks.

The degree of source orientation of a task was determined according to intuition. For instance, it was speculated that a highly video-oriented search task was one where some visual learning would be highly desirable (e.g. dance steps, aerobic movements). A medium video-oriented search task was one where it was assumed that video results may or may not be required; e.g. for the topic "baking pancakes", some participants might prefer reading a recipe rather than watching a video.

For each topic, participants were shown the search results associated with that topic, displayed in the vertical or the tiled design (as shown in figure 4.5). Participants were asked to bookmark those results they judged useful in completing the task (see Figure 4.8). A brief task description for each topic was provided, which did not refer to the source orientation of the corresponding task (e.g. words such as photo, image, video, were not used). For instance, for the topic *salsa dance*, the task description provided was *"provide examples of dance steps"*.

The twenty-seven topics were distributed into three sessions. Each session was composed of all search tasks with the same level of source orientation: high, medium or low. What then varied in each session were the source of the augmented elements (i.e., three; image, news, video) and the position of the augmented elements (another three, for example in case of vertical interface; top, middle and bottom), leading to (3×3) nine topics per session (as seen in Table 4.1). For instance, for the "high" oriented session, at position p_1, c_1 included image, video and news topics. Similarly, at position p_2,

TABLE 4.1: Position and source orientation combinations for each source. For vertical representation P_1=top, P_2=middle and P_3=bottom. For the tiled approach, P_1=left, P_2=right-top and P_3=right-bottom. Here, each cell (combination c_n) is tested for three sources, image, video and news. Thus leading to twenty-seven combinations.

Position	High	Medium	Low
P_1	c_1	c_2	c_3
P_2	c_4	c_5	c_6
P_3	c_7	c_8	c_9

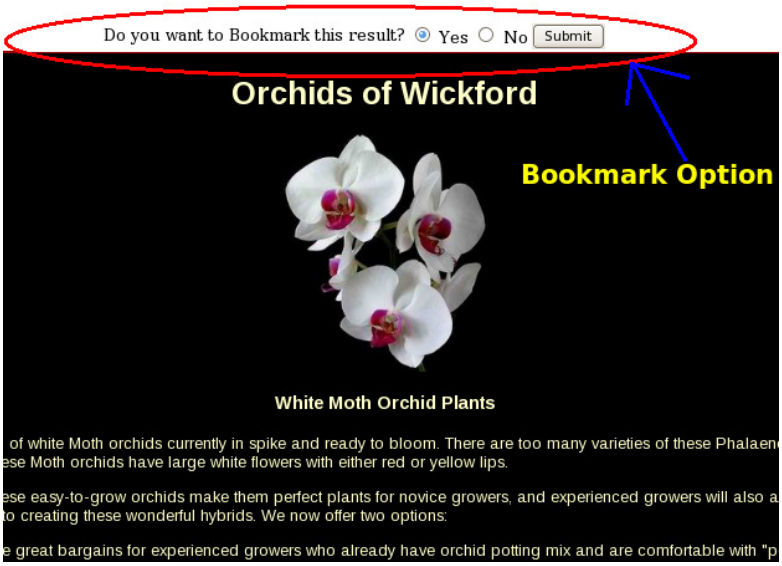


FIGURE 4.8: Example of a bookmarked page, which shows how web pages were augmented with an extra feedback bar at the top of the page, which allowed the users to mark the result as relevant or not relevant.

c_4 also included image, video and news; and at position p_3, c_7 again included image, video and news related topics. Thus resulting in total nine topics for the session "high".

Every result page was timed for two minutes and participants were asked to spend it in their usual search manner. That is, participants could view all 'fifteen' results if they wished or just 'one' result within the two minute time. On average, each session lasted for 18-20 minutes (2 minutes each for 9 topics). The sessions were rotated to minimize learning effects, and allowing participants to select the topics in random order within each session reduced ordering effects.

4.2.3 Participants

The experiments for the vertical aggregation was carried out by 18 males and 6 females from our university, of which 16 were graduates, 7 postgraduates and 1 an undergraduate. The experiment for the tiled aggregation was carried out by 18 males and 6 females, of which 1 was a graduate, 21 were postgraduates and 2 were undergraduates. To prevent any learning effect or biases, participants from one study were not involved in the second study. Participants in each study did all twenty-seven topics, thus leading to a total of $(27 \times 24 \times 2)$ 1296 sessions.

The participants were from various educational fields, e.g. computing, law, life sciences, real estate, aerospace, business management, arts and commerce. Participants were recruited through an email distributed to several mailing lists. An entry questionnaire was used to capture participants' profile and search background. As the experiment was divided into three sessions, post-session questionnaires were used to capture subjective assessments on the sessions and tasks. Exit-questionnaires provided participants' perceptions of sessions and tasks as a whole. Participants were rewarded fifteen pounds for their participation.

4.3 Results

This section presents the results from the experiments. First, results from the study investigating factors effect on the vertical representation of an aggregated interface will be presented in section 4.3.1. Next, the results from the second study where factors effect on the tiled approach of aggregation will be presented in section 4.3.2.

4.3.1 Study One: Vertical Aggregation

In Section 4.3.1.2, the overall results of the main and interaction effects of position, task source orientation, and augmented element type on users click-through behavior is presented. Then the details of each of the independent variables is described in Sections 4.3.1.3, 4.3.1.4, and 4.3.1.5, respectively. Finally, the result of participants' perceptions of the augmented elements is presented in Section 4.3.1.6.

TABLE 4.2: Distribution of clicks. Total = total number of clicks on augmented elements, % click = percentage of clicks on augmented elements (augmented element/(web+augmented element)), Mean (SD) = mean and SD of click through frequency on augmented elements, and % Bookmark = percentage of bookmarked results (number of bookmarked augmented elements/number of clicks on augmented elements).

Augmented Element	Total	Click %	Mean (SD)	% Bookmark
Image	360	33.7	1.7 (1.72)	53.6
Video	324	34.4	1.6 (1.48)	63.6
News	294	30.1	1.4 (1.59)	46.9
All	978	32.8	1.6 (1.60)	54.9

4.3.1.1 Click-Through Frequency

First, the overall frequencies with which augmented elements are clicked were analyzed. The results are shown in Table 4.2, which shows the total number of clicks on each augmented element, the percentage with which augmented elements were clicked compared to the web results, the mean number of augmented click through per session, and the percentage of clicked augmented elements which were also bookmarked.

The results in the table show similar distribution of the clicks across the three augmented elements. That is, participants accessed similar number of image, video and news results during the task. Furthermore, it can be seen that most of the data fall within ≈ 1.5 standard deviation of the mean, which suggests that the data are concentrated about the mean.

4.3.1.2 Anova Test

The next analysis performed was the significance of the effect of the three independent variables (IVs) such as Position (P), Source-Orientation (SO), and Augmented Element type (AE), on participants' click-through behavior. ANOVA tests[4-5] were used to examine the significance of three main effects (P, SO, AE), three interaction effects of two factors (P:SO, P:AE and SO:AE), and one interaction effect of three factors (P:SO:AE).

A main effect is the simple effect of a factor (independent variable) on a dependent variable. It is the effect of the factor alone averaged across the levels of other factors[4-6]. Whereas, an interaction effect is the variation among the differences between means for

[4-5] ANOVA tests (analysis of variance) is a statistical technique for comparing means for multiple (usually ≥ 3) independent population Field and Hole [2003].

[4-6] http://www.stats.gla.ac.uk/steps/glossary/anova.html#maineff

TABLE 4.3: Anova test results of the main and interaction effects of the position (P), Source-Orientation (SO), and Augmented Element type (AE) on the click-through ratio. Where, Df_1 = The first degree of freedom (Lower bound), Df_2= The Second degree of freedom (Upper bound), Pr = Associated p value. The Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.'

Factors	Df_1	Df_2	**F value**	**Pr(>F)**
P	2	35	12.67	0.0000 ***
SO	2	35	28.99	0.0000 ***
AE	2	34	4.60	0.0171 *
P : SO	4	78	2.26	0.0704
P : AE	4	77	9.53	0.0000 ***
SO : AE	4	78	5.92	0.0003 ***
P : AE : SO	8	161	10.18	0.0000 ***

different levels of one factor over different levels of the other factor[4-7].

The raw click-through frequencies were first converted into ratios (click-through frequency on augmented element/total click-through frequency) before performing the ANOVA test. When the difference was found to be significant among the levels of independent variables, further multiple post-hoc tests were performed to find significant pairs among the factor levels. Multiple t-tests were run in the post-hoc analysis while adjusting the p-values using the Bonferroni method[4-8], where the critical value was set to = (0.05/number of multiple comparisons).

The overall ANOVA results are given in Table 4.3, which shows the F-value from the test. The significant effects (Pr(>F)) are highlighted with asterisk symbols (*) . The results from the anova tests suggest that the mains and the interaction of the factors in most cases affected participants click behavior.

It can be seen that the main effects of position (F(2,35) = 12.67, p = 0.0000), source-orientation (F(2,35) = 28.99, p = 0.0000) and augmented elements (F(2,34) = 4.60, p = 0.0171) was found to have significant effect on participants' click-through ratios. In particular, the F ratio for the factor source-orientation (Third row in Table 4.3) was found to be significantly large, thus suggesting greater differences among groups.

Furthermore, interaction between the position: augment element (F(4,77) = 9.53, p = 0.0000) and, interaction between source-orientation and augment elements (F(4,78) = 5.92, p = 0.0003) was found to be significant too. This indicates that participants'

[4-7] http://www.stats.gla.ac.uk/steps/glossary/anova.html#intern

[4-8] The Bonferroni method is a simple method that allows many comparison statements to be made (or confidence intervals to be constructed) while still assuring an overall confidence coefficient is maintained Prins et al. [2003].

TABLE 4.4: Click-through frequency across different positions.

	Top	**Middle**	**Bottom**
	Mean (SD)	Mean (SD)	Mean (SD)
Image	2.2 (1.90)	1.5 (1.43)	1.5 (1.73)
Video	1.8 (1.53)	1.6 (1.41)	1.3 (1.46)
News	1.8 (1.71)	1.1 (1.25)	1.4 (1.71)
All	1.9 (1.72)	1.4 (1.38)	1.4 (1.63)

click-through behavior was not affected by a single factor but by a combination of multiple factors. Finally, the bottom row of Table 4.3 shows that there was a significant interaction of the three factors on participants' click-through behavior.

These results are helpful in presenting a high-level overview of the ANOVA tests. The following subsections will discuss these results in detail.

4.3.1.3 Position Effect

The first research question [**R1**] considered the effect of the augmented elements' position on users' click-through behavior. To examine this research question, the participants' click-through data was analyzed on results from different augmented elements (image, video and news) at three positions on the vertical aggregation design: top, middle and bottom (as shown in Figure 4.6). Results from click-distribution at three positions are shown first, then results from successful click distribution are presented.

As seen in the previous section (Section 4.3.1.2), the position had a significant effect on participants' click-through behavior. Table 4.4 shows the breakdown of the click-through frequency based on the three levels of positions (i.e., top, middle, and bottom). The overall mean click frequency (row "All" in Table 4.4) shows that there is a position effect between the top and middle/bottom positions, but less so between the middle and bottom positions. Multiple comparisons were then preformed to test that the differences observed were significant. These post-hoc tests revealed that differences between the top and middle, and top and bottom positions were significant (Table 4.5). This suggests that participants clicked the augmented elements more frequently when they were located at the top of the results than the other positions.

Next, the aim was to ensure that the relevance of results shown to the participants was consistent across all three positions, and that the observed click behaviors were because of position effect and not due to the difference in relevance of results. In order to achieve this, the number of successful clicks (bookmarked results/clicked results) on

TABLE 4.5: Multiple t-tests (p-value) on position combinations

Position	p-value
Top-Middle	**0.0010**
Top-Bottom	**0.0000**
Middle-Bottom	0.1059

TABLE 4.6: Success rates across different positions.

	Top	Middle	Bottom
	Mean (SD)	Mean (SD)	Mean (SD)
Image	0.5 (0.40)	0.5 (0.46)	0.5 (0.41)
Video	0.6 (0.42)	0.6 (0.42)	0.8 (0.38)
News	0.4 (0.43)	0.4 (0.44)	0.5 (0.48)
All	0.5 (0.42)	0.5 (0.45)	0.6 (0.44)

image, video and news results at the three different positions were computed. The averaged distribution of successful clicks is shown in Table 4.6. It can be seen that overall the successful clicks were found to be similar across three positions ("All" row in Table 4.6). Although, the mean success rate at the bottom position was higher than top and middle, post-hoc analysis revealed that these differences were not significant. This ensures that the higher click-through rates at top position were not due to the fact that, the results presented at top were more relevant than the results presented at the bottom or middle position.

4.3.1.4 Source Orientation Effect

The second research question [**R2**] looks at the effect of the source orientation of an information need on users' click-through behavior. To address this, we analyzed the click-data of the participants with respect to the different levels of source orientation (e.g., highly image oriented search, medium image oriented search, low image oriented search). The order of the experimental results is the same as Section 4.3.1.3.

As already seen in Table 4.3, the source orientation was found to be significant in vertical aggregation design. The breakdown of the results based on the three levels of the orientation is shown in Table 4.7. It can be seen from the bottom of the table, the click-through frequency decreases as the level of source orientation weakens. The post-hoc tests show that the differences between the high level and middle level, and between the

TABLE 4.7: Click-through frequency across different source orientation levels.

	High	**Medium**	**Low**
	Mean (SD)	Mean (SD)	Mean (SD)
Image	2.5 (1.92)	1.4 (1.47)	1.3 (1.46)
Video	1.7 (1.66)	1.9 (1.43)	1.2 (1.26)
News	2.2 (1.89)	1.0 (1.20)	1.0 (1.25)
All	2.1 (1.85)	1.4 (1.41)	1.2 (1.33)

high level and low level, are significant (2nd column of Table 4.8). This suggests that participants clicked the augmented results more frequently when an information need has a strong orientation towards a particular information source type, in the vertical aggregation design.

TABLE 4.8: Multiple comparisons for source orientation combinations on click-through frequency.

Orientation	**p-value**
High-Medium	**0.0000**
High-Low	**0.0000**
Medium-Low	0.4268

Furthermore, analysis of the successful clicks (bookmarked results/clicked results) show that more results were bookmarked when the orientation of the task at hand was high or medium, than when task orientation was low (as seen in Table 4.9). Results from post-hoc analysis suggest that there was significant difference in successful click rates between the high-low and medium-low levels of source orientation. This reflects to the findings above, from the click data, where users clicked more often on source results when the source orientation of the task at hand is high.

TABLE 4.9: Success rate across different source orientations.

	High	**Medium**	**Low**
	Mean (SD)	Mean (SD)	Mean (SD)
Image	0.6 (0.40)	0.5 (0.42)	0.5 (0.44)
Video	0.6 (0.40)	0.7 (0.37)	0.6 (0.46)
News	0.6 (0.40)	0.4 (0.48)	0.2 (0.36)
All	0.6 (0.40)	0.5 (0.44)	0.4 (0.45)

4.3.1.5 Interaction across Augmented Elements

The last research question [**R3**] aims to investigate the interaction effect between the augmented source types and the position or source orientation. The interest was to see if the interaction effect of factors such as position and source orientation vary, or are similar across augmented elements. To answer this question, the click-through data from image, video and news results across different levels of position and source orientation were compared. For the vertical aggregation interface, first, results for the interaction effect of position and augmented elements (P:AE) is reported, and then, results for the interaction of source orientation and augmented element (SO:AE) is presented.

- **Interaction of position and augmented elements:**

  Returning to Table 4.3 again, it can be seen that the interaction effect of position and augmented element (PxAE) is significant in the vertical interface (5th row of Table 4.3). This suggests that participants' click-through behavior was affected by a combination of the two factors. To understand this effect better, an interaction plot using the position and augmented elements were plotted. The result is shown in Figure 4.9.

  The interaction plot shows that the pattern of videos is different from images and news. The post-hoc tests show that, at the middle level of positions, participants clicked video results more frequently than images or news results (As shown in the Table 4.10).

- **Interaction of Source-Orientation and Augmented Elements:**

  Next, results from the ANOVA tests suggest that the interaction effect of source orientation and augmented element (SO:AE) is another significant factor affecting user's click-through behavior in the vertical design (6th row of Table 4.3). The interaction plot of the two factors is shown in Figure 4.10. Again, the pattern of video results appears to be different from images and news results. The post-hoc tests show that, at the middle level of source orientation, participants clicked the

TABLE 4.10: Results of post-hoc pair wise comparisons for the interaction effect of position and augmented elements (PxAE) in Figure 4.9.

	Top	Middle	Bottom
Image-Video	0.6759	**0.0023**	0.9853
Image-News	0.5643	**0.0159**	0.4724
News-Video	0.8619	**0.0000**	0.4786

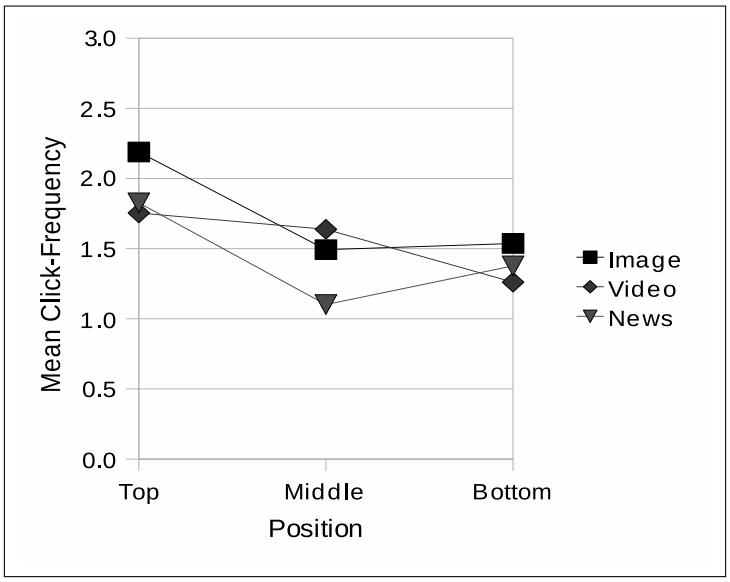


FIGURE 4.9: Interaction effect of position and augmented elements (PxAE) in the vertical representation design. Results of post-hoc pair wise comparison among augmented elements at three positions are shown in Table 4.10

video results more frequently than the other sources in the vertical design (See Table 4.11).

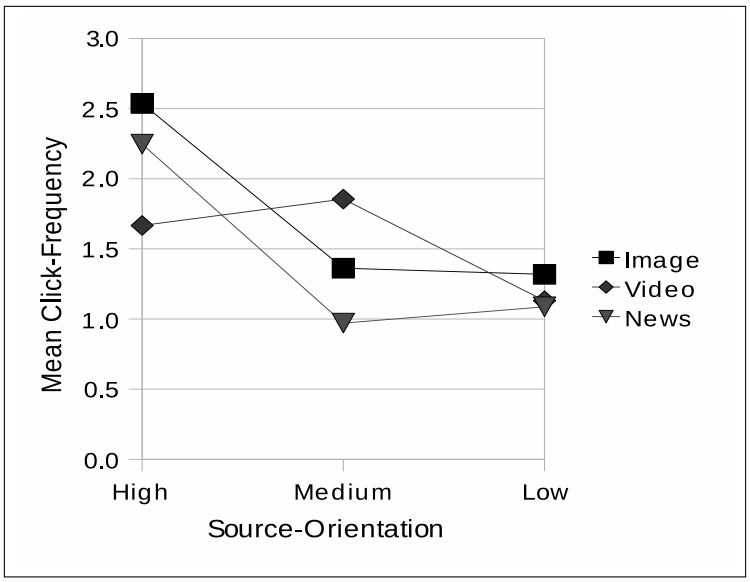


FIGURE 4.10: Interaction effect of source-orientation and augmented elements (SOxAE) in the vertical representation design. Results of post-hoc pair wise comparison among augmented elements at three positions are shown in Table 4.11

TABLE 4.11: Results of post-hoc pair wise comparisons for the interaction effect of position and augmented elements (SOxAE) in Figure 4.10.

	High	Medium	Low
Image-Video	0.0975	**0.0010**	0.1611
Image-News	0.8643	**0.0175**	0.6302
News-Video	0.1099	**0.0000**	0.0708

Next Section presents the results from the questionnaires where the overall users' preference regarding position of augmented elements depending on the level of source-orientation is shown.

4.3.1.6 User Perception

Recall that the experimental design had three sessions (Section 4.2) with each session corresponding to one level of source orientation (high, medium and low). After each session, participants were asked to provide their preferred position for the augmented elements on the result page. More specifically, participants were asked if they prefer to view non-web results (image, news, etc.) at the top, middle or bottom of the result page for the vertical interface, and left, top-right or bottom-right of the result page for the tiled interface. The results from the questionnaires for the study are now presented.

The questionnaire results are shown in Figure 4.11. Results from the vertical interface study suggest that for highly oriented sessions users prefer to have augmented results on the top. Users tended to get less specific and their preferences becomes mixed when the source orientation decreases. This is consistent with the findings from the click data where a clear position effect was observed on users' click-through behavior (Section 4.3.1.3)

So far, results for position, source-orientation and interaction effect on participants' click-through ratio was presented for the interface with vertical aggregation approach. Next, results for the factors' effect within the titled aggregated interface will be presented (Section 4.3.2). First, position effect on click-behavior will be presented in Section 4.3.2.3. Effect of levels of source-orientation will be shown in Section 4.3.2.4. Finally, interaction effect of position and source-orientation across augmented elements will be presented in Section 4.3.2.5.

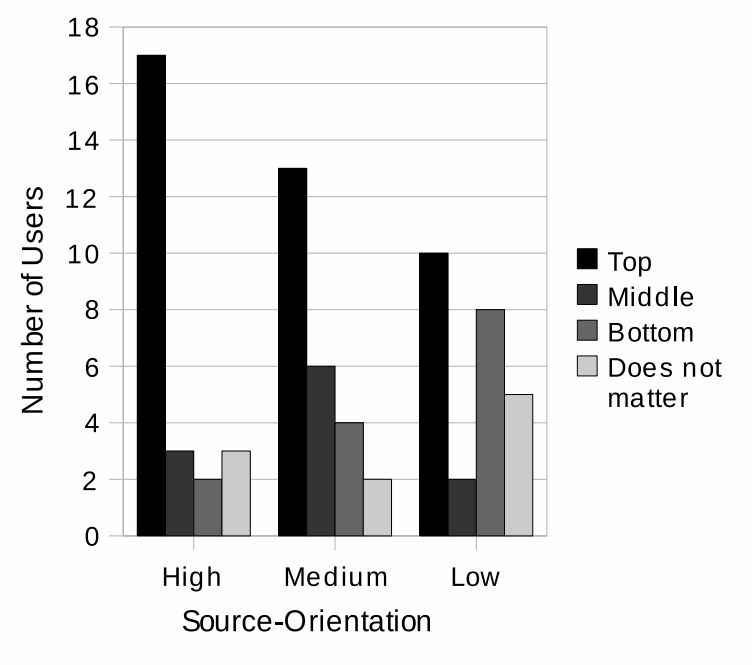


FIGURE 4.11: User preference on position of augmented elements across source orientations.

4.3.2 Study Two: Tiled Aggregation

The three research questions stated in the beginning of this chapter (Section 4.1.3) were investigated again in this study. Recall that there exist two main types of aggregation approaches: vertical and tile like format. In study one, the research questions were investigated for the first type of aggregation (Section 4.3.1). While the second study (presented here) explored the second form of aggregation.

The structure for the result from the second study will be similar to that of the first study . First, click-distribution of participants during the study is presented in Section 4.3.2.1. Next, in Section 4.3.2.2, the overall results of the main and interaction effects of position, task source orientation, and augmented element type on users click-through behavior is presented. Then the details of each of the independent variables is described in Sections 4.3.2.3, 4.3.2.4, and 4.3.2.5, respectively. Finally, the result of participants' perceptions of the augmented elements is presented in Section 4.3.2.6.

4.3.2.1 Click-Through Frequency

First, the overall frequencies with which augmented elements are clicked were analyzed. The results are shown in Table 4.12, which shows the total number of clicks on each augmented element, the percentage with which augmented elements were clicked

TABLE 4.12: Distribution of clicks. Total = total number of clicks on augmented elements, % click = percentage of clicks on augmented elements (augmented element/(web+augmented element)), Mean (SD) = mean and SD of click through frequency on augmented elements, and % Bookmark = percentage of bookmarked results (number of bookmarked augmented elements/number of clicks on augmented elements).

Source	Total	Click %	Mean (SD)	% Bookmark
Image	410	37.3	1.9 (1.28)	41.5
Video	339	33.0	1.6 (1.21)	58.4
News	290	28.6	1.3 (1.39)	33.4
All	1039	33.1	1.6 (1.31)	44.8

compared to the web results, the mean number of augmented click through per session, and the percentage of clicked augmented elements which were also bookmarked.

The results in the table show that the participants bookmarked more video results when compared to image and news results (Column *% Bookmark* in Table 4.12). However, participants seemed to access image results slightly higher than the news and video results. On an average, participants viewed ≈ 2 image results from the list of five images shown to them during a session. Next, the results from the Anova test are presented.

4.3.2.2 Anova Test

Analysis of variance test (Anova) was also preformed to test the significance of the effect of the three independent variables (IVs) such as Position (P), Source-Orientation (SO), and Augmented Element type (AE), on participants' click-through behavior. For which, the raw click-through frequencies were first converted into ratios (click-through frequency on augmented element/total click-through frequency) before performing the ANOVA test. When the difference was found to be significant among the levels of independent variables, further multiple post-hoc tests were performed to find significant pairs among the factor levels. Multiple t-tests were run in the post-hoc analysis while adjusting the p-values using the Bonferroni method, where the critical value was set to = (0.05/number of multiple comparisons). Results from the Anova test are shown in Table 4.13.

From the results shown in the Table 4.13 we conclude that, on average, position of the augmented elements had no significant effect ($F(2,41)= 0.62$, $p= 0.5437$)on the click-behavior of the participants. On the other hand, source-orientation ($F(2,43) = 24.37$, $p = 0.0000$)) and the type of the augmented element ($F(2,42) = 9.22$, $p = 0.0005$) affected the participants' click-behavior significantly (second and third row in Table 4.13).

TABLE 4.13: Anova test results of the main and interaction effects of the position (P), Source-Orientation (SO), and Augmented Element type (AE) on the click-through ratio. Where, Df_1 = The first degree of freedom (Lower bound), Df_2= The Second degree of freedom (Upper bound), Pr = Associated p value. The Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.'

Factors	Df_1	Df_2	**F value**	**Pr (> F)**
P	2	41	0.62	0.5437
SO	2	43	24.37	0.0000 ***
AE	2	42	9.22	0.0005 ***
P : SO	4	87	0.37	0.8314
P : AE	4	86	0.66	0.6228
SO : AE	4	88	14.16	0.0000 ***
P : SO : AE	8	178	3.37	0.0013 **

For the tiled aggregated interface, the interaction between position and source-orientation (F(4,87) = 0.37, p = 0.8314) was not significant suggesting that, the effect of position on participants click-behavior did not depend on the level of source-orientation. Furthermore, the interaction between position and augmented element (F (4,86) = 0.66, p= 0.6228) was also not significant. Thus, effect of position did not change with the type of the augmented element either. Finally, there was a significant effect on participants' click-behavior by source-orientation and the type of augment element interaction (F(4,88) = 14.16, p = 0.0000). Thus indicating that the participants' click-ratio was affected with the levels of source-orientation and the type of the augmented element presented.

4.3.2.3 Position Effect

The first research question [**R1**] considered the effect of the augmented elements' position on users' click-through behavior. To examine this research question, the participants' click-through data was analyzed on results from different augmented elements (image, video and news) at three positions on the tiled aggregation design: left, top-right and bottom-right (as shown in Figure 4.7). Results from click-distribution at three positions are shown first, then results from successful click distribution are presented.

Unlike the vertical aggregation design, the position effect was found to be insignificant in the tiled aggregation design. The breakdown of the click-through frequency based on the position of the second design can be found in Table 4.14. As can be seen from the bottom row of the table, the overall frequency was similar across the position levels. This echoes our findings from the ANOVA test, where overall, position did not affect

TABLE 4.14: Click-through frequency across different positions.

	Left	**Top-Right**	**Bottom-Right**
	Mean (SD)	Mean (SD)	Mean (SD)
Image	2.0 (1.36)	1.8 (1.21)	1.9 (1.27)
Video	1.6 (1.17)	1.5 (1.21)	1.6 (1.26)
News	1.4 (1.48)	1.3 (1.47)	1.3 (1.21)
All	1.7 (1.37)	1.6 (1.32)	1.6 (1.24)

TABLE 4.15: Success rates across different positions.

	Left	**Top-Right**	**Bottom-Right**
	Mean (SD)	Mean (SD)	Mean (SD)
Image	0.4 (0.39)	0.4 (0.41)	0.5 (0.40)
Video	0.6 (0.45)	0.5 (0.45)	0.6 (0.43)
News	0.2 (0.33)	0.4 (0.41)	0.3 (0.40)
All	0.4 (0.42)	0.4 (0.42)	0.5 (0.43)

the users' click-behavior. On the other hand, there appears to be some difference among the augmented source types, which will be discussed in Section 4.3.2.5.

Next, the analysis of successful clicks across the three different positions in the tiled interface suggests that users bookmarked approximately the same number of results at the three positions (row "All' in Table 4.15). This again confirmed that the results shown at all three positions in the interface were of similar relevance.

4.3.2.4 Source Orientation Effect

The second research question [**R2**] looks at the effect of the source orientation of an information need on users' click-through behavior. To address this, we analyzed the click-data of the participants with respect to the different levels of source orientation (e.g., highly image oriented search, medium image oriented search, low image oriented search). The order of the experimental results is the same as Section 4.3.2.3.

Similar to the vertical design, participants' click-through frequency was also significantly affected by the source orientation in the tiled design. The breakdown of the results based on the three levels of orientation in the tiled design is shown in Table 4.16. Again, the frequency appears to decrease as the level of orientation weakens. The post-hoc tests show that the differences between the high level and medium level, and between the high level and low level, are significant (3rd column of Table 4.8). This

TABLE 4.16: Click-through frequency across different source orientation levels.

	High	**Medium**	**Low**
	Mean (SD)	Mean (SD)	Mean (SD)
Image	2.5 (1.31)	1.6 (1.17)	1.6 (1.16)
Video	1.6 (1.14)	1.9 (1.30)	1.2 (1.11)
News	2.3 (1.44)	0.9 (1.17)	0.9 (1.03)
All	2.1 (1.35)	1.5 (1.28)	1.2 (1.14)

TABLE 4.17: Success rates across different levels of source orientation.

	High	**Medium**	**Low**
	Mean (SD)	Mean (SD)	Mean (SD)
Image	0.5 (0.36)	0.3 (0.39)	0.5 (0.41)
Video	0.6 (0.45)	0.7 (0.40)	0.5 (0.46)
News	0.5 (0.37)	0.2 (0.38)	0.1 (0.21)
All	0.5 (0.40)	0.4 (0.44)	0.4 (0.45)

suggests that participants clicked the augmented results more frequently when the task's source orientation was high.

Analysis of the successful click rates show that the numbers of bookmarked results were higher when the source orientation was high. Results from post-hoc analysis suggest that the bookmarking rate significantly decreases when the orientation of the task drops from high to medium or high to low.

4.3.2.5 Interaction across Augmented Elements

The last research question [**R3**] aims to investigate the interaction effect between the augmented source types and the position or source orientation. The interest was to see if the interaction effect of factors such as position and source orientation vary, or are similar across augmented elements. To answer this question, the click-through data from image, video and news results across different levels of position and source orientation were compared. For the vertical aggregation interface, first, results for the interaction effect of position and augmented elements (P:AE) is reported, and then, results for the interaction of source orientation and augmented element (SO:AE) is presented.

- **Interaction of Position and Augmented Elements:**

  Recall that for the tiled aggregation design, an interaction effect of position and augmented element type (P:AE) was not found to be significant during Anova

TABLE 4.18: Results of post-hoc pair wise comparisons for the interaction effect of position and augmented elements (PxAE) in Figure 4.12..

	Left	Right-Top	Right-Bottom
Image-Video	0.5597	1.0000	1.0000
Image-News	0.1507	0.1490	**0.0147**
News-Video	0.3677	0.5451	0.1238

test (Sixth row in Table 4.13). In order to get a better understanding of this an interaction plot for clicks across position and augmented element type was plotted. As can be seen from the interaction plot shown in Figure 4.12, all augmented elements show a similar pattern indicating that the participants clicked similar number of results (image, news or video) irrespective of their position on the result page. For instance, there was not much difference in the amount of image results accessed when placed on left side of the result page, or when placed at the top-right position, or when it was placed on the bottom-right panel of the result page.

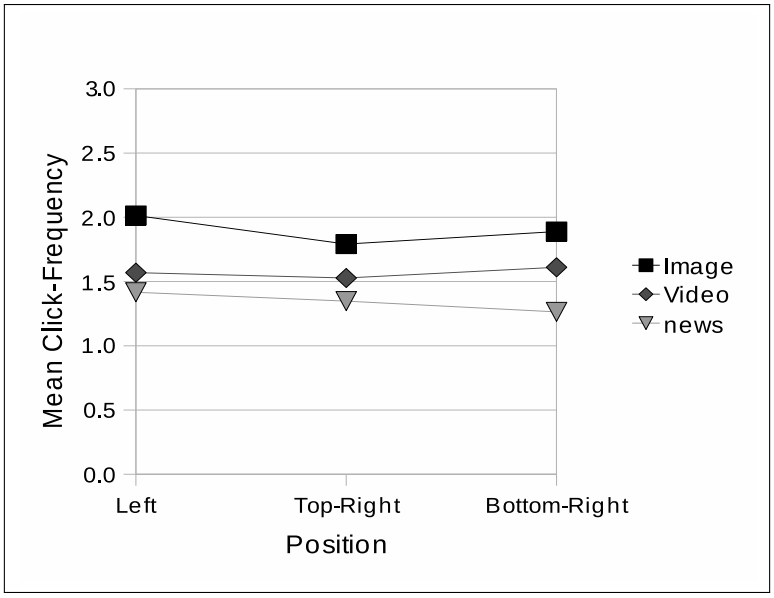


FIGURE 4.12: Interaction effect of position and augmented elements (PxAE) in the vertical representation design. Results of post-hoc pair wise comparison among augmented elements at three positions are shown in Table 4.18

- **Interaction of Source-Orientation and Augmented Elements:**

  On the other hand, participants' click-through frequency was affected significantly by the interaction effect of the source orientation and augmented element

TABLE 4.19: Results of post-hoc pair wise comparisons for the interaction effect of source and augmented elements (SOxAE) in Figure 4.13..

	High	Medium	Low
Image-Video	**0.0001**	**0.0028**	0.2394
Image-News	0.8904	**0.0006**	**0.0000**
News-Video	**0.0011**	**0.0000**	**0.0045**

types (SO:AE). The interaction plot of the two variables is shown in Figure 4.19. As can be seen from the plot, the video results again have a different pattern from the image and news results. While the significant effect was found only in the middle level of source orientation in the vertical design, the tiled design has significant effects in the high and low levels of orientation, too (See Table 4.19).

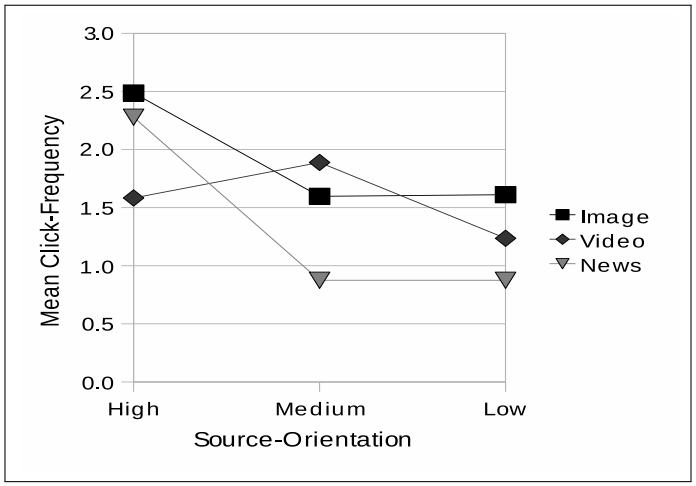


FIGURE 4.13: Interaction effect of source-orientation and augmented elements (SOxAE) in the vertical representation design. Results of post-hoc pair wise comparison among augmented elements at three levels of orientations are shown in Table 4.19

4.3.2.6 User Perception

For the tiled like interface, results show that user preferences are mixed when considering the position of augmented results for different orientations (Figure 4.14). In addition, many users suggested that since they were able to see all the panels (base and augment elements) on the screen without having to scroll, the position of the panels did not matter. Again, this is consistent with the findings from the click data analysis, where position was not significant for the tiled interface (Section 4.3.2.3), although there is a trend for users to prefer the top-right position for the display of non-web results.

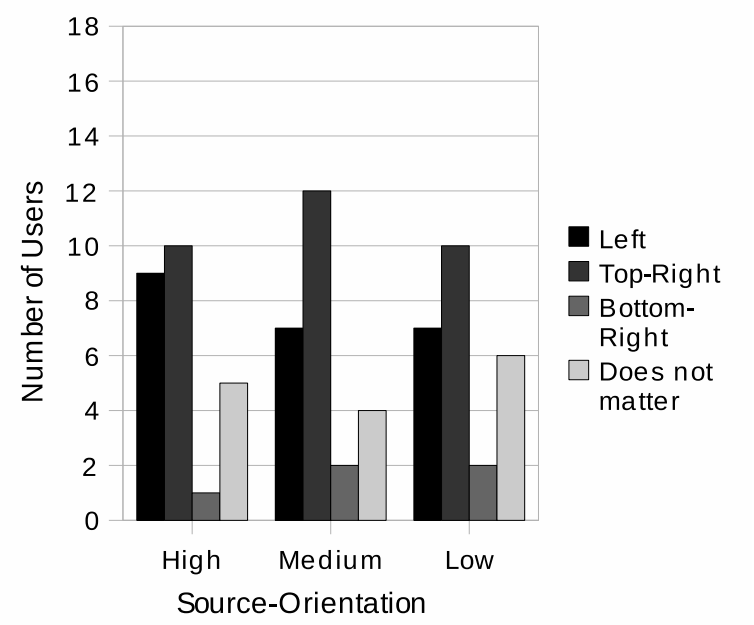


FIGURE 4.14: User preference on position of augmented elements across source orientations.

4.4 Result Analyses

This section summarizes the overall findings from both the studies. As stated earlier, it should be noted that the aim of the study was not to compare the two designs, but instead to obtain an overall understanding of the effect of the factors on different ways of aggregating results.

4.4.1 Position Effect

To summarize, the results for research question one [**R1**] suggest that there is a significant position effect on users' click frequency in the vertical aggregation, whereas in the tiled-like interface, the position of augmented elements does not seem to affect users' click frequency.

4.4.2 Source-Orientation Effect

As already seen, the effect of the source orientation was found to be equally significant in both the vertical and tiled aggregation designs. The standard deviation in the bottom row of Tables 4.7 and 4.16 suggests that this trend is slightly more consistent in the tiled design than the vertical design. However, in terms of the frequency, the results

are comparable. This suggests that the effect of the source orientation is an important factor to be investigated in aggregated search research.

4.4.3 Interaction across Augmented Elements

A highlight of the above analysis is the difference in the click behavior for video results, compared to news and image. Multiple cases were observed where the frequency pattern of videos differed from that of images and news. Videos tended to have a low frequency even when the position was high or source orientation was high. On the other hand, participants tended to click video results more frequently than the other sources when the position was middle or the orientation was middle. Implications of this finding will be discussed in the next Section.

4.5 Discussion

The work presented in this chapter was motivated by the prominent complexity of aggregated search interfaces compared to the conventional single-source design. Despite the complexity, there has been a limited number of studies looking at the effects of factors such as augmented elements' positions, source types, and the search tasks source orientation, on people's search behavior.

As mentioned above, the intention was not to decide which type of aggregation is more effective than the other. Instead, the interest was in understanding the characteristics of both designs so that we can leverage their advantages depending on the context of system use. The following discusses the main findings of this study and their implications on the design of aggregated search interfaces.

The first finding is that the factors that affect participants' click-through behavior differ between the vertical and tiled designs. This may sound obvious, however, this should not be underestimated, since it suggests that the way in which we present results from different sources indeed matters. For example, participants' click-through behavior was significantly affected by the position of augmented elements in the vertical design, echoing the findings of previous studies Joachims et al. [2005]; Agichtein and Zheng [2006]; Guan and Cutrell [2007]; Keane et al. [2008], yet not in the tiled design. This suggests that we need a careful estimation of the position of the augmented elements with respect to the base elements when the vertical design is employed. Such information can be useful when the suitable aggregation approach is required. For instance, when it is not possible to afford to measure the position of augmented elements, the

tiled-like approach might be more appropriate since participants' click-through behavior was not affected by the position in this type of aggregation. Such a situation may arise in digital libraries and elsewhere. These results address the first research question [**R1**] of this study.

The second finding is that videos resulted in a different click-through pattern from news and images. This trend was common in both the vertical and the tiled designs. This suggests that, when deciding to retrieve videos, different behavior from other sources may be observed.

While it is not entirely clear why videos are different, it is possible to suggest some possible reasons. That is, we can speculate on a couple of potential factors for the trend. First, videos are multimodal media Halvey et al. [2009] combining text, images, and audio in a dynamic way. The dynamism and multimodality of the information source might cause a user to give a different priority to videos during a search task. The different priority may cause the different click-through pattern of videos when compared to news and images. Secondly, it may be due to the type of surrogate used to represent video results, being less informative of the contents of the video than the equivalent image or text representations. The title of news articles and the thumbnail of an image can provide a good indication of the respective content of the documents. On the other hand, although basic metadata of the videos were presented, this may not have been as informative of the contents. This difficulty in getting the preview of videos might cause the different click-through pattern Song and Marchionini [2007]. It should be noted that, the task time limit of 2 minutes did not seem to discourage users from viewing video results.

Although, it was not clear why the difference in the click-behavior across results from different sources were observed, yet, the overall message that we get from this finding is that depending on the type of source aggregated on the result page, it might be possible to observe difference in users' searching behavior. Alternatively, more generally, this study suggests that participants' click-through behavior can be different across source types. These observations addresses the third research question [**R3**]

The third finding was that a search task's orientation towards a particular source could affect participants' click-through behavior. This trend was common to both the vertical and tiled designs. Traditional information retrieval research has been focused on the modeling of thematic (or topical) relevance of documents. However, research on XML document retrieval Lalmas and Tombros [2007] and geographic information retrieval (GIR) Purves and Jones [2004] has demonstrated that relevance can be multidimensional, there being a structural relevance in XML retrieval, and geographic relevance in

GIR, which can be considered apart from the thematic relevance. In a similar way, the experimental design of this study controlled the level of orientation towards a particular source (i.e., news, images, and videos). The significant effect of the source orientation observed in our experiments suggests that the task's source orientation is an important factor to investigate in research on aggregated search. On the design level, it suggests that devising a means of capturing a searcher's intent about the source is an important problem to tackle. These observations address the second research question [**R2**].

4.5.1 Limitations

The study described in this chapter has some limitations. First, the results to be displayed on the interfaces were generated a priori. Although this made the investigation fair, the implications of these results are limited to this particular set of search results. Second, to reduce the complexity and duration of the experiment, only three sources were tested; image, video and news. Finally, only a single search engine results were used throughout the experiments. Therefore, the findings may or may not apply to different environments. Further studies should be carried out to deepen our understanding of aggregated search interfaces. In addition, no attempts were made to ensure that the documents retrieved by the search engine were relevant to the tasks, to ensure that the presented results were representative of real world search engine performance.

However, as the ratio of bookmarked documents over all clicked documents suggests, participants did perceive some of the retrieved documents as relevant in order to complete the search tasks. Furthermore, careful attention was paid to deal with any technical problems during the experiment, and no cases were observed where participants were deliberately bookmarking clearly irrelevant documents. Therefore, it can be asserted that the backend engine did retrieve relevant documents and participants were able to find some of them although any quality controls of retrieval were not performed.

4.6 Summary

In this chapter, a study testing two aggregated search interfaces was presented. Two separate experiments were carried out during the study; one where results from the different sources were blended into a single list, and the other, where results from each source were presented in a separate panel. A total of 1,296 search sessions performed by 48 participants were analyzed.

The study reported in the chapter led to three main findings. First, the position of search results was only significant when the results are aggregated in a vertically ranked list. Second, participants' click-through behavior on videos was different compared to other sources. Finally capturing a task's orientation towards particular sources is an important factor to consider when considering the use of an aggregated search design.

These findings provide further insight to aggregated interface and its result presentation issues. Results from the study provide initial guidelines to the designers of aggregated interfaces and their associated concerns. Furthermore, the study triggers many potential research directions to be explored in aggregated search paradigm.

To conclude, overall aggregated search, as already pushed forwards by major search companies, is a useful paradigm. Producing aggregated result page in response to an informational query improves information access to the users and hence makes task completion quicker. Furthermore, results suggest that the designers of aggregated search interfaces need to concentrate on different aspects over the aggregation styles. In a vertical aggregation, one needs an accurate estimation of the best position of augmented elements and that relevance of the source is a key element in both form of aggregation. The outcomes from the above studies not only provide insight to some initial issues of result presentation associated with aggregated interface, but also suggest many future directions for research in aggregated search.

In the next part, methods to identify the source-orientation behind a user's query are described in Chapter 5. Furthermore, large scale log data is analyzed to uncover the dynamics of users' searching behavior when their information need is oriented towards multiple sources in Chapter 6.

Part III

Source-Orientation in Aggregated Search

Part III

So far in this thesis, different aspects related to the result presentation in an aggregated page were discussed. The effectiveness of an aggregated interface, and the factors affecting click behavior in aggregated interface were investigated. Motivated by the results from the previous studies, the work was further extended to gain an insight into users' searching behavior with respect to the information need oriented towards different sources. In this part, the dynamics of user behavior with respect to different sources are analyzed using large scale log data from Microsoft. First in Chapter 5, effective methods to identify the source-orientation behind a user query are presented. Once the orientations are correctly identified, they are then further exploited in Chapter 6 to uncover the patterns and behavioral aspect of users for such information needs.

Chapter 5

Identifying Source-Orientation

5.1 Introduction

In the previous Chapter [4], results from the analysis of click-behavior on aggregated interfaces suggested that source-orientation is a key element in an aggregated interface. In this chapter, ways to identify the source orientation behind users' queries are presented. Two methods are tested for this purpose; Rule Based and Combination (Rule base + Machine Learning). For the first method (described in Section 5.5), a simple rule based technique is applied on the clicked URLs of the log entries. For the second method (described in Section 5.6), a combination of rule base and machine learning techniques are applied on the submitted query, clicked URL, and the title of the clicked documents.

In this work, Microsoft 2006 RFP search click data were analyzed to understand the source-orientation behind the queries. More specifically, six sources were looked at, namely, image, video, map, news, blog and Wikipedia. All other sources were viewed as standard "web", i.e. the typical web search result.

These six sources were chosen based on a survey[5-1], which showed that images, news, and videos were the three most frequently accessed sources. Map and Wikipedia were chosen because results of these types are now frequently included within the top ten result list by major search engines.

[5-1] http://www.iprospect.com/about/researchstudy_2008_blendedsearchresults.htm

5.2 Structure

The structure of the chapter is as follows. The motivation for identifying source-orientation is presented in Section 5.3. The rule-based method is described in Section 5.5. While Section 5.6, presents the combination method used in an attempt to identify the source-orientation. Finally, the chapter finishes with the summary in Section 5.8.

5.3 Motivation

The diversity of the content available on the web has dramatically increased in recent years. Multimedia content such as images, videos, maps, voice recordings has been published more often than before. Document genres have also been diversified, for instance, news, blogs, FAQs, Wikipedia. Therefore, a particular challenge is to properly determine the type(s) and source of information sought by a user, in order to better support the user information seeking process.

As observed in previous Chapter 4, in the context of aggregated search, source-orientation seem to be a key element in designing of an aggregated interface, therefore on the design level, devising a means of capturing a searcher's intent about the source is an important problem to tackle.

The result should not only be relevant to the search topic but also relevant to the type of information (image, news, etc) required by the user. Therefore, the challenge in aggregated search is to identify the information need in terms of topic as well as in terms of sources. The work described in this chapter aims to identify information need in terms of source specific results.

In web search, information need of a user is often identified by the query submitted to the search engines or some retrieval system (Zazo et al. [2005]; Stojanovic [2005]; Vechtomova and Zhang [2009]; Stamou and Kozanidis [2009]). Understanding the information need behind a user query, i.e. *query intent* (Guo and Agichtein [2008]; Brenes et al. [2009]; Hu et al. [2009]) is one of the important goals in web search. When the query intent is correctly identified, the user information need can be better satisfied Lee et al. [2005]. In context of aggregated search, identifying the information need in terms of source is further required, so that the results from the appropriate sources could be provided.

An information need may be of different nature. Figure 5.1 shows how user expresses his/her information need through a query, and how it can be identified in various ways.

It can be based on the type of desired interaction, e.g. navigational, transactional or informational Broder [2002b]. It can relate to some predefined general topics, e.g. travel, sport, shopping Beitzel et al. [2005]. Finally, it can also apply to the information from specific source being sought, e.g. image, video, blog Diaz [2009], as in case of aggregated search.

In this chapter, we are concerned with the latter that is source-orientation. More precisely, the interest here is in identifying the source from which the result is being sought. Therefore, in this chapter ways to identify the source-orientation of the user query in terms of 'type' source is proposed. In the following sections two methods to identify source-orientation are described, namely, rule based method (Section 5.5) and combination method (Section 5.6).

.

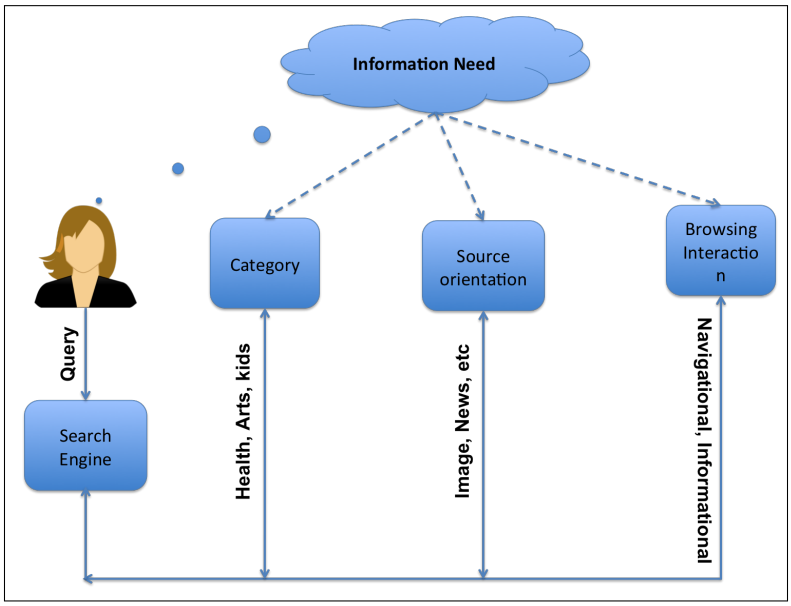


FIGURE 5.1: Various aspect of users' information need.

5.4 Data-Set

The data set used in this work consists of the Microsoft 2006 RFP Dataset, which is made of 15 million log entries Conference Chair-Craswell et al. [2009]. The data is a subset of web search logs from US users spanning over a month. Each log entry corresponds to a click and contains the following information:

1. The **timestamp** (time and date), which is used to order the clicks within a session.

2. A unique **session id** representing a search session. A session is the entire series of queries, one or more, submitted to the search engine by a user over some given time.

3. A unique **query id** is given to every distinct query submitted during a search session.

4. The **query terms** used by the search engine to display results.

5. The **URL** of the clicked result.

Next, rule based method to identify the source-orientation is presented.

5.5 Rule Based Method

During the rule based method, a simple rule based technique was applied on the clicked URLs of the log entries. This work was the first step towards identification of the "source relevance" to a given query Sushmita et al. [2009b]. In other words, given a query, the challenge was to identify the intended source, from which the result would be required in order to satisfy the information need. In the following sub-sections, the methodology followed is described in Section 5.5.1, and the outcomes are presented in Section 5.5.3. Finally, the overall outcomes are discussed as summary in Section 5.5.4.

5.5.1 Methodology

In web search, it is possible to breakdown search queries into at-least two broad categories[5-2]: navigational and informational, because informational and transactional queries have similar characteristic Lee et al. [2005]. It was hypothesized that search result aggregation was most useful for supporting informational queries (as also discussed in chapter 3).

To focus on the informational queries in the dataset, the data was separated into two sets based on the number of clicks made within a single session. More specifically, two click sets were made; a single-click set, which had only one click in a session, and multiple-click set, which had more than one click in a session. Although this was a very simple method to separate navigational queries from the others, single clicks are one of the main properties of the navigational query Yuan et al. [2008]. As a result, 3,218,588

[5-2]http://www.seobythesea.com/?p=1021

single-click sessions (27%) and 8,932,479 multiple-click sessions (73%) were obtained. The following analyses mainly focuses on the multiple-click data-set, and zero-click queries were not included in the analysis. That is, the sessions where no clicks were made in response to the submitted queries were not used in this analysis.

5.5.2 URL analysis

The following set of pattern rules were used to identify the sources of click-through documents. For example, if a click-through URL contained a string *movies*, it was assumed that the main content of the clicked document was a movie.

Image: /img/ /images/ /image/ /pictures/ /picture/ /photo/ /photos/

Video: /vid/ /video/ /videos/ /movie/ /movies/

Wikipedia: /wiki/

News: /news/

Blogs: /blogs/ /blog/

Audio: /audio/ /audios/

Map: /map/ /maps/

Web + Others: URLs that did not match any of above

While the patterns may not be exhaustive to identify all pages belonging to a source, it can be considered as a reasonable approximation of the distribution of different orientations in the dataset. It should also be noted that the overlap of orientations were not considered during these methods (Rule based and Combination both). Although, it is possible that a query will be associated with multiple orientations, but this study focused on identifying only single orientation for a given query. The aim was to first build a suitable single orientation classifier and to explore its possibilities and outcomes. For future research direction, it should be possible to further elaborate these methods to be able to identify overlap of orientations.

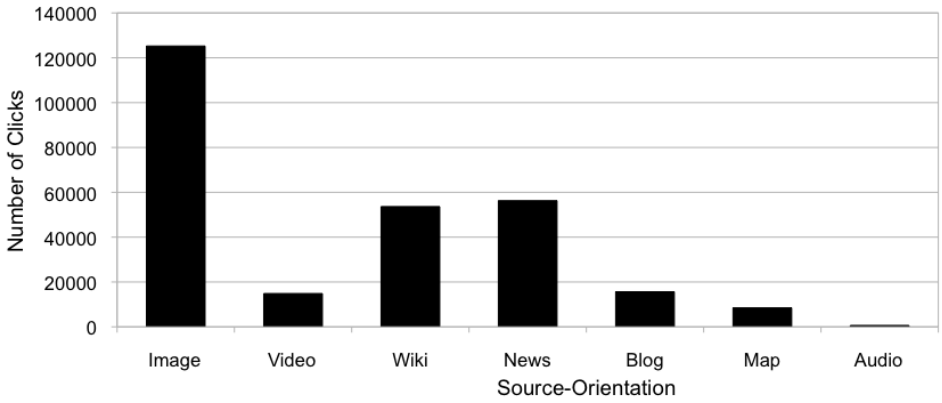


FIGURE 5.2: Distribution of Source-Orientations.

5.5.3 Results

Once the URLs were classified using the rule base technique, the number of clicked URLs that matched with any of the sources (as shown in section 5.5.2) in the multiple-click set was counted. There were 2,74,755 click-through URLs (3%) that matched with one of the six source orientations, image, video, blogs audio, Wikipedia, news and map, the rest was classified as Web + Others. Figure 5.2 shows the distribution of the seven orientations based on the matched URLs. The results show that, the images were the most frequent orientation followed by Wikipedia, news and video. Furthermore, the percentage of queries oriented to audio source was found to be very low.

Similar findings were also reported in the survey[5-3]. The findings were based on users' feedback on a set of questionnaires asked during the survey. The survey selected online consumers randomly from the NPD U.S. online consumer panel. A total of 2,404 individuals responded to the survey. Respondents received an email invitation to participate in the survey with an attached URL linked to a WebÐbased survey form. A result presented after an image search is clicked by 26% of users (the most frequently clicked "vertical search" category). The second most commonly clicked vertical search category was found to be news search at 17%, followed by video search at just 10%. Since Wikipedia was not considered in this survey, therefore percentage of clicks for Wikipedia were not reported. The findings from log analysis reported in this chapter are therefore in consistent with the findings of the online survey, hence the estimations can be assumed reasonable.

[5-3]http://www.iprospect.com/about/researchstudy_2008_blendedsearchresults.htm

5.5.4 Summary

During *rule based* method, one possible way to identify the source-orientation of the given query was shown. The method used a rule base technique on the clicked URLs, which showed promising results towards identification of associated source orientation with respect to the user query. Although the method was simple, approximately 3% of the total log data were identified to have one of the predefined orientations, which considering the large size of the log was not inconsiderable. Furthermore, it was found that although by large number of standard web pages were being accessed (clicked), the percentage of pages from other sources (e.g., news, blogs, images) that were clicked was not negligible.

The results from the above method show that it is possible to identify the source-orientation of a query in a log-data using techniques like rule based classification. Therefore, the above method was then further formalized in order to obtain stronger evidences for identifying the source-orientations. For this purpose, the combination method described in the next Section 5.6 is used. Furthermore, since the percentage of identified audio orientation was very low, it was not considered in the next method.

5.6 Combination Method: Rule and Machine Learning

Using only rule based method, to some extent it was possible to identify the source-orientation based on the URLs of the document. While, the results were promising. yet, a more sound and formal method was required in order to have stronger evidences for the identified source-orientations. Therefore, the previous method was further modified, where a combination of rule base and machine learning technique was used in order to identify the source-orientations Sushmita et al. [2010b].

For this purpose, three different evidences were used; the submitted query, the clicked URL in response to the query, and the title text of the clicked document. This made it possible to have higher degree of confidence for the results obtained as compared to the previous method, where only one source of evidence was used; the URL of the clicked document. In the following sub-sections, the methodology, results, and summary are presented.

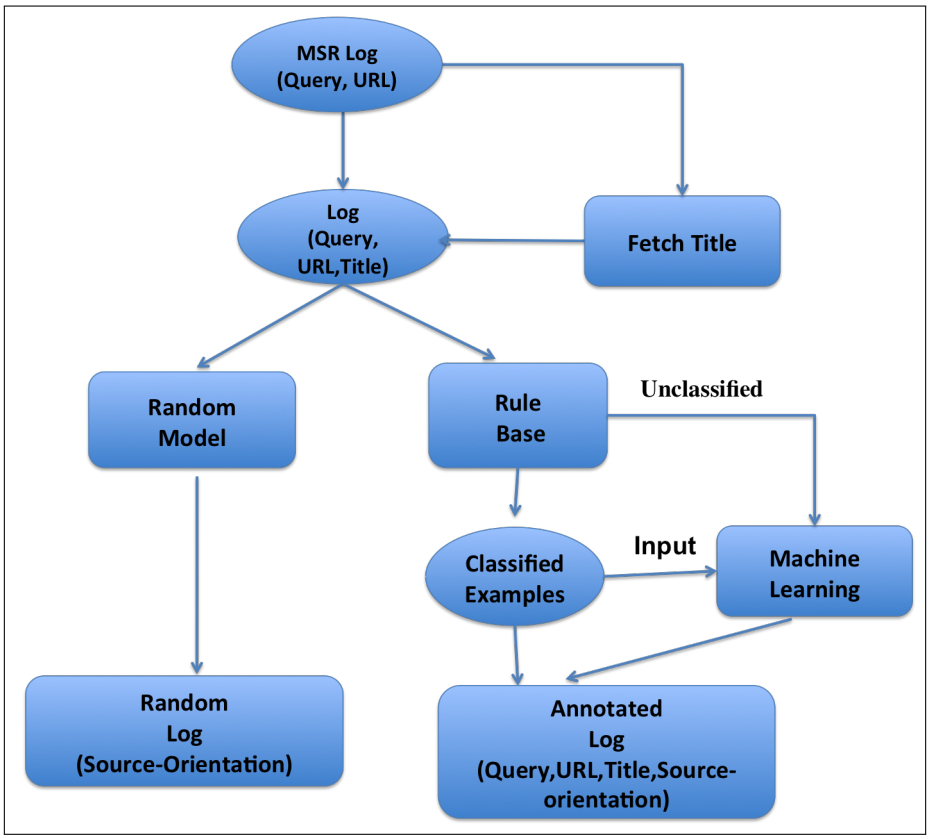


FIGURE 5.3: Overall process for classifying the log-data used for the combination method.

5.6.1 Methodology

The overall process followed during the combination method is depicted in the diagram shown in Figure 5.3. To determine the source-orientation behind a query, three supporting evidence was used. The first one is the query itself, more precisely the use of source-specific terms in the query were looked at, such as "video", "map", etc (as also used during rule based method in Section 5.5). The second and third evidences were the clicked URLs and the titles of the corresponding web documents, respectively. The reason for choosing URL and the title as supporting evidences was because in previous work, e.g. He et al. [2007] has shown that result snippets (title, URL and excerpt) of the clicked web documents could be used to determine the information need behind the query (in a different context though). For instance, one example query in the Microsoft 2006 RFP Dataset is "muscular systems", and one clicked document has as title "images and diagrams of human muscular system"; here the term image in the document title, and the fact that the user has accessed that document, may be a good indication that the user is looking for images. Therefore, the information need is oriented towards

Total number of entries	6,637,590
Total number of sessions	3,960,541
Number of sessions with one click	2,654,794
Number of sessions with two clicks	721,223
Number of sessions with three clicks	282,980
Number of sessions with four clicks	132,834
Number of sessions with five clicks	68,403
Number of sessions with more than five clicks	100,307
Average length of a session (hh:mm:ss)	00:03:44

TABLE 5.1: Statistics about the click data set used in the combination method.

the source *image*.

Document excerpts were not used since access to them was not provided, and it was not evident how these could be reconstituted. As the data set did not contain the titles of the clicked documents, they had to be retrieved. Therefore, the clicked documents were downloaded. Note that given the time lag between when the data set was created and time of download of the clicked documents (approximately 3 years), it was possible to obtain the titles for only 50% of the clicked documents. In the rest of this chapter, only those sessions are used for which we could download the titles for all clicked documents. The statistics reported in Table 5.1 correspond to these log entries.

Using these three evidences, query terms, URL and title, each click was then classified using two different strategies, the first using hand crafted rules and the second using machine learning techniques (as can be seen in the Figure 5.3). A *random model* is also built using the log-data, which is used to validate the statistics obtained from the classified log-data. The details of random model will be explained in the next chapter (Section 6.5).

The first stage of this work is to determine the intended source-orientation (in this study, image, video, map, blog, Wikipedia and news) for given queries. More precisely, from a given query of the log-data, the intention is to identify whether a user was looking for an image, blog entries, etc. Therefore, first, a rule-based classifier was used, output of which was used to build the features used by a machine learning classifier. These two steps are discussed next, including their outcomes.

5.6.2 Rule-based classifier

A rule-based classifier was used to identify the source orientations. The aim here was to classify as accurately as possible some of the log entries. The interest here was

Orientation	Rule (%)	SVM (%)	Combined(%)
Image	1.1	0.32	1.42
Video	0.64	0.86	1.5
News	0.71	1.43	2.14
Map	1.5	0.03	1.54
Blog	0.03	0.15	0.17
Wikipedia	0.07	0.89	0.96
Web	0	92.28	92.28

TABLE 5.2: Percentages of classified log entries into the different source-orientations with the rule-based (Rule) and the machine learning (SVM) methods. Note that the rule-based method did not consider "web" source as a class. The total percentages of classified log entries is given in the last column (Combined).These percentages are used as a basis to investigate the dynamics of query orientations in the in next Chapter 6 .

more in precision (properly classified) than recall (all identified). This was important as these classifications were to be used as inputs to calculate the features for the machine learning approach. These classified log entries are further used in our analysis of the dynamics of source-orientations (described in next Chapter 6).

To build a high precision classifier, the most reliable evidence was used, namely the query terms. For example, if a query contains the term *photo*, i.e. the user has explicitly typed *photo* in his or her query, it seems correct to deduce that the user is looking for image, i.e. has an information need oriented towards result from an image source. Here *photo* is referred to as an "source-specific" term. As also used in the previous method (Section 5.5), following source-specific terms were used to classify the clicks (log entries) into the three domains and three genres investigated in this part of the work.

Image: *image, images, photo, photos, picture, pictures*.

Video: *video, videos, movie, movies*

Blog: *blog,blogs*

Map: *map, maps*

Wikipedia: *wiki, Wikipedia*

News: *news*

The source-specific terms were manually selected from the previous method. Although not necessarily exhaustive, they are a reasonable approximation of how users would

search, in terms of the queries they would submit to a search engine, for documents of particular domains or genres.

With this approach, 268,491 log entries (i.e. clicked URLs), that are 4% of the total logs, were identified to be of one of the six orientations. A closer inspection however shows incorrect classifications (false positives). For example, a query "video games" was classified as having a video orientation. The percentage of such false positives is nonetheless low enough (e.g., 0.4% for the case of video games, and no obvious false positives could be observed for the other source-orientations).

5.6.3 Machine learning classifier

Since a small set of source-specific terms were used, and that the source may often not be explicitly specified by the user (through the use of an source-specific term), it is likely that the percentage of such information need is higher. To identify additional percentage of clicks oriented to these sources, machine learning techniques is used. Using machine learning requires the following four main steps:

1. A manually labeled set of log entries (section 5.6.3.1).
2. Designing features correlated with the possible source-orientations that can be computed for every log entry (section 5.6.3.2).
3. Training a classifier to predict the source-orientation given the training examples from (1) and the features defined in (2) (section 5.6.3.3).
4. Predicting the source-orientation of the non-manually classified log entries using the classifier (section 5.6.3.4).

In the following sections, each step is described in detail.

5.6.3.1 Manual Classification:

For the first step (1), 3800 unique clicks were sampled randomly and were manually classified into one of the following six *categories* of specific source-orientations – image, video, blog, map, Wikipedia, and news. A click was classified as having a web orientation (which corresponds to our unclassified category) when it could not be classified into any of the other six categories. A web orientation corresponds mostly to the typical web search result, and we expect it to be the predominant source.

Orientation	N	image	video	blog	map	wiki	news	web
image	98	**0.30**	**0.18**	0.01	0.00	0.00	0.01	**0.49**
video	131	0.02	**0.46**	0.02	0.00	0.00	0.03	**0.47**
blog	62	**0.08**	**0.12**	**0.27**	0.00	0.01	0.03	**0.49**
map	25	0.00	0.00	0.00	**0.63**	0.00	0.00	**0.37**
wiki	66	0.04	0.02	0.04	0.00	**0.74**	0.00	**0.17**
news	75	**0.08**	0.04	0.02	0.02	0.01	**0.32**	**0.52**
web	3354	0.04	0.04	0.02	0.01	0.00	0.02	**0.87**

TABLE 5.3: Statistics for the machine learning approach. The first column gives the source-orientation, the second gives the number of training examples for this source, In addition, the remaining columns present the percentage of time the predicted. source-orientation was correctly classified or mis-classified as other source-orientations (values above 0.05 are in bold).

In order to manually classify the log entries three evidences were looked at:

1. The query terms,
2. The URL terms and,
3. The document title.

That is, the three supporting evidences described in Section 5.6.1 were analyzed before assigning a predicted source-orientation. For instance, for the query "Ice Age3", if the corresponding URLs contained terms like: *video, movie, etc*, and if the document titles contained terms like: *preview, trailer, etc*. Then these combined observations from the three evidences indicate that users typing "Ice Age 3" was looking for video results and therefore had information need oriented towards video results. Hence, **video-orientation** was manually assigned to the corresponding log entry.

It should be noted that although efforts were made to ensure accuracy and that there is no researcher's biases in predicting the source-orientation, it is possible to have different predictions for the given query. However, since three evidences were used to predict the orientation, overlap in decisions is likely to occur. The outcomes of the manual classification is shown in Table 5.3, the **N** column. As expected, web orientation corresponds to the predominant source of information need.

5.6.3.2 Feature Generation:

As explained at the beginning of this section, the second phase (2) consists in defining a set of features associated with each log entry, where each feature should be a good predictor for identifying a subset of source-orientation (ideally one). In this case,

the set of features were based on the language models computed from the data set (classified source-orientations) obtained through the rule-based approach described in Section 5.6.2.

A language model for each evidence (i.e., query terms, clicked URL and document title) and each category was built. A background language model was also built, which estimates the probability of a term to appear in the given category regardless of the source. It was chosen to model separately the three evidences since they are of very different nature (query, URL and title of the web page).

The intuition and the working hypothesis here is that each category uses a specific vocabulary that is often associated with explicit source-specific terms – see Section 5.6.2. For instance, in a query, if "Aniston" is often associated with the term "photo", the term "Aniston" will be associated to a large number of log entries classified as an image orientation by the rule-based classifier. As a result, the query language model for the orientation "image" will give a higher probability to "Aniston" than the background language model, and thus comparing both probabilities gives the classifier an indication of how likely a term (or a set of terms) is generated by an source specific language model rather than by the background one. The parameters of each language model[5-4] were estimated, one for each evidence **e** and orientations **o** (21 in total, i.e. one for each of the 3 evidences and for each of the 6 source-orientations plus the background model). We estimated the probability that a term t occurs using the standard maximum likelihood estimate, and smoothed it using the background language model for a given evidence.

$$P(t/\mathbf{o},\mathbf{e}) = \lambda P_{ML}(t/\mathbf{o},\mathbf{e}) + (1-\lambda)P_{ML}(t/\mathbf{e}) \quad (5.1)$$

$$= \lambda \frac{c_{\mathbf{o},\mathbf{e}}(t)}{\sum_{t'} c_{\mathbf{o},\mathbf{e}}(t')} + (1-\lambda)\frac{c_{\mathbf{e}}(t)}{\sum_{t'} c_{\mathbf{e}}(t')} \quad (5.2)$$

In the above equation, the probability P_{ML} is the maximum likelihood estimate of the probability of a term occurring in the selected evidence **e**, and if given, for the source-orientation **o** (otherwise, it is the background language model). These probabilities are estimated with the following statistics: $c_{\mathbf{o},\mathbf{e}}(t)$ denotes the number of times the term t appeared for the evidence **e** with the orientation **o**; $c_{\mathbf{e}}(t)$ is the number of times term t appeared for source **e**. These statistics were computed from the set of automatically classified clicks using the rule-based classifier defined in the previous section. The

[5-4]For the URLs, we considered that terms were any maximal sequence of alphanumeric characters. For example, `http://www.abc.com/video` has four terms, www, abc, com and video

smoothing parameter λ was heuristically set to 0.95, since the intension to emphasize the importance of the orientation.

The probability that a sequence of terms T is generated by any of the language models can be computed by

$$P(T/\mathbf{e},\mathbf{o}) = \prod_{t \in T} P(t/\mathbf{e}) \tag{5.3}$$

Finally, the probability $P(T/\mathbf{e},\mathbf{o})$ is not a perfect indication on whether the text T was generated since some set of terms can have a low probability because they do not occur frequently, and not because they are not specific to the source. A more useful value is the ratio[5-5] of the probability (for a given evidence) of observing T given the orientation to the probability of observing T, *i.e.*, the ratio $R_{\mathbf{o},\mathbf{e}}(T)$ is defined as:

$$R_{\mathbf{o},\mathbf{e}}(T) = \log \frac{P(T/\mathbf{e},\mathbf{o})}{P(T/\mathbf{e})} \tag{5.4}$$

whose value is above zero if it is more likely that the text was generated given the orientation than in general, and below zero in the opposite case. This gives rise to a set of 18 features (one for each of the 6 categories and 3 sources) that are used as an input to build a multi-class classifier.

5.6.3.3 Training:

The third step of machine learning process is to train the classifier using the example training data, obtained from manual classification described in section 5.6.3.1; and the features defined, as obtained during feature generation in section 5.6.3.2). In this step, an SVM classifier[5-6] Crammer and Singer [2002] were used because it performed the best for the task at hand, when evaluated with the 10-fold cross-validation was performed (using nine tenth of the manually classified data to learn, and one tenth to compute the performance, and repeating this operation 10 times). During the selection process, models that predicted either the correct or the web orientations were preferred, over those with better performance that predicted an incorrect orientation, i.e. the aim was to achieve a high precision rather than a high recall.

[5-5] To compare two models for a data set, we can calculate their likelihood ratio, which results from simply dividing the probability of the data according to one model by the probability of the data according to the other model Manning et al. [2008].

[5-6] The implementation of Karatzoglou et al. [2006].

The SVM classifier was then trained using a three cross validation and a Gaussian radial basis function. With respect to the manually labeled data, to give less importance to the web orientation, the number of corresponding examples were down-sampled and only 20% of those were chosen, which gives a total of 817 manually labeled examples. Given this low number of manually classified log entries, an equal number of automatically labeled log entries were added (randomly sampled among the 224,241 classified log entries), using the same rules as in Section 5.6.2 but using the title and URL as sources of evidence rather than the query terns. Since the output of rule-based classifier based on query terms (Section 5.6.2) were used to generate the features for building the SVM classifier, they were not used during the training of the classifier (as it would have resulted in biased features). Furthermore, it is possible to expect a high enough precision from the URL and title rule-based classifier, which thus could be used for learning.

5.6.3.4 Prediction:

Table 5.3 shows the confusion matrix[5-7] with the final settings that were used for classification. It can be seen that the correct classification rate is low (between 0.27 and 0.87), but that most of the time, when a click is misclassified its predicted orientation is web. The only exceptions are for image (18% are classified as video) and blog (20% are classified as either image or video). Further manually classified data and better features would be necessary to increase the classification performance, but given that web is the unclassified category, the results show that it was possible to have improved recall without hurting too much precision.

Overall, Table 5.2 shows the statistics about the classification (both rule-based and using SVM as our machine learning approach) of all the log entries. It is this labeled log that is used in the analysis for the next Chapter [6]. In this table, it can be seen that approximately 8% of the total log entries were identified to have a source orientation other than web. This is not negligible considering the large size of the log data analyzed, and is compatible (although not directly comparable) to the results reported in Arguello et al. [2009b].

It can also be seen that using a machine learning classifier doubles (from 4% to 8%) the total number of log entries classified as having one of our six orientations of interest. Given the low level of noise (false positives) or said otherwise the high precision, the

[5-7] A confusion matrix shows which classes were classified properly and which were misclassified/confused with other classes and in what degree Klopotek et al. [2006].

analysis presented here is thus more sound than previous work relying on a rule-based classifier only (Section 5.5).

5.6.4 Evaluation

Finally, once the log-data was classified to have one of the pre-defined source-orientations, the next step was to judge the accuracy of the classifier. Apart from the ten-fold cross validation (as discussed in the pervious section), two other evaluation measures were also carried out. First, the classified source-orientations were compared with the source-orientations provided by human annotators (section 5.6.4.1). Second, the overall agreement of the predicted orientation with the existing search engine companies was obtained (section 5.6.4.2). The following sections describe these measures and their outcomes are provided.

5.6.4.1 Human Judgment:

First, the classified log-data were compared with the manually classified examples provided by human annotators. For this purpose, 1,000 unique queries were randomly selected. These queries were then manually annotated by ten human judges, providing source-orientation for 100 queries each. That is, for the given query, the annotators were asked to provide an expected source result (from the six sources used in this work), which they would wish to see in response to the given query. For instance if for a given query, the annotator wished an *image* result, then the expected source "image" was indicated.

The manually classified queries by annotators were then compared with the predicted orientations by the classifier (Section 5.6). The result obtained from the comparison is shown in Figure 5.4, where percentage of predicted source-orientation by the classifier that matched with the expected source provided by the annotators is shown. It can be seen that, approximately for the 60% of the queries, the source-orientation were correctly identified. While approximately 40% of query orientations did not match with the expected orientations provided by the human annotators.

Although, the percentage of correctly identified source-orientation that matched with the expected source provide by the annotators is not very high, yet, considering the fact that the log-data was very skewed and noisy, the results seem fairly reasonable. Furthermore, there were very limited examples for each orientation (as discussed in section 5.6.3.1) to train the classifier. Therefore, given more training data for each category

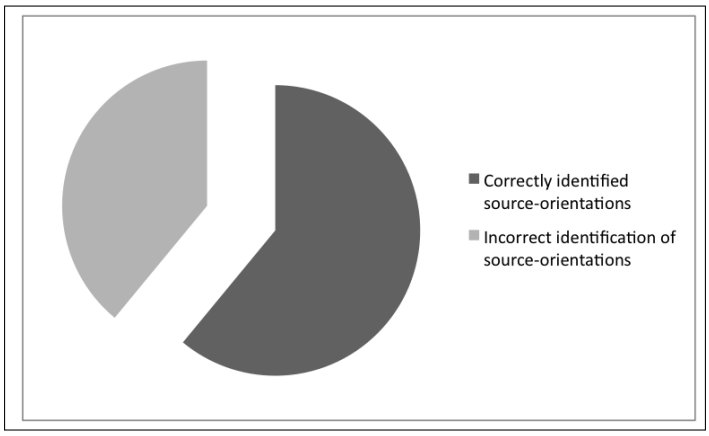


FIGURE 5.4: Percentage of sample queries for which the predicted source-orientation matched with the expected source provided by the human annotators. Approximately 60% of query's information need was correctly identified by the classifier, while approximately 40% of the time the predicted orientation did not match with the expected orientation provided by the annotators.

(orientation), it should be possible to obtain better classifications. Next, accuracy of the classifier evaluated based on the agreement with three search engines is presented.

5.6.4.2 Agreement with Search Engines:

Finally, in order to estimate the accuracy of the classifier, the results from the classified log-data were also compared with the results from top three[5-8] commercial search engines; Google[5-9], Bing[5-10] and Yahoo![5-11]. For this purpose, six hundred (200 for each of six sources) unique queries from the classified log were randomly sampled. These queries were then submitted to the three search engines (Google,Bing and Yahoo!).

For each query, a match was considered if, the results (among top ten) provided by the search engine included results from one of the source (image, video, map, etc), and if the source-orientation of the same query as identified by the classifier was similar. For instance, if for a sample query, the predicted orientation by the classifier was "image", and for the same query if the search engines included image results, it was then assumed that the source-orientation was correctly identified.

It should be noted that, the aim of this comparison was not to compare the performance of the search engines; among themselves or with the classifier described in this

[5-8] http://www.seoconsultants.com/search-engines/
[5-9] https://www.google.com/
[5-10] http://www.bing.com/
[5-11] http://uk.yahoo.com/?p=us

chapter. Instead, the comparison was made to ensure that the results obtained from the classifier were not random and that the predicted source-orientation for the query was in-consistent with the current search engine results. That is, if for a query the classifier predicts relevance of an image result, and if search engines also include image result, then it can be said that the search engines also predicted the relevance of an image result; how the search engines made this prediction was not the aim of this work. Therefore, although the underlying mechanism of judging relevance may not be comparable, but overall agreement of predicted information need can be compared. The outcome of the comparison is shown in the Figure 5.5.

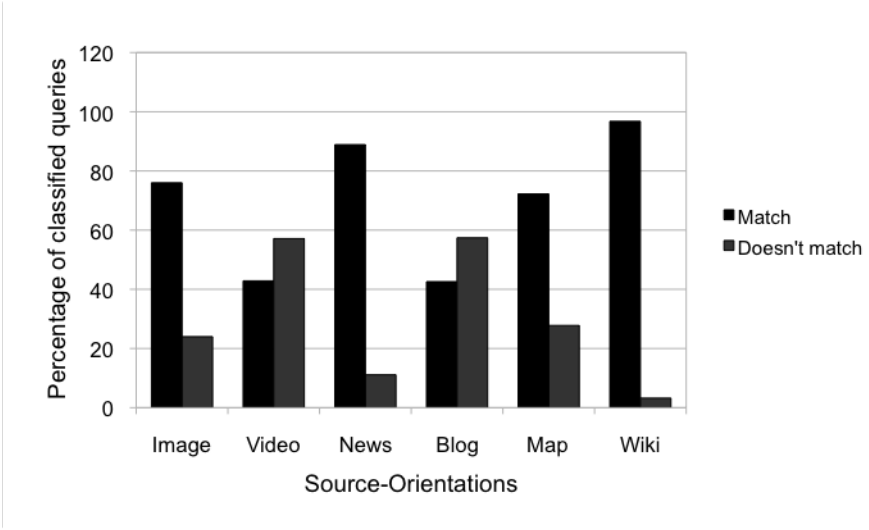


FIGURE 5.5: Percentage of sample queries for which the predicted orientation matched with at-least one of the search engines.

The results show that predicted source-orientation for the sample queries matched with the search engine results in most cases (over 60%). It was found that, the search engines among their top results also included the results from the predicted source. An exception was observed in case of video and blog, where the predicted source-orientation did not match the results of the search engines in many cases.

When the queries for video-orientations were further analyzed, it was found that among the predicted video-oriented queries, which did not match with the results of the search engines, approximately 50% of the queries were adult queries, and therefore were probably filtered by the search engines. Consequently, the result page did not include any video result for those queries. Although, no obvious reason was observed in case of blog queries.

The breakdown of the percentage of matched orientations across three search engines

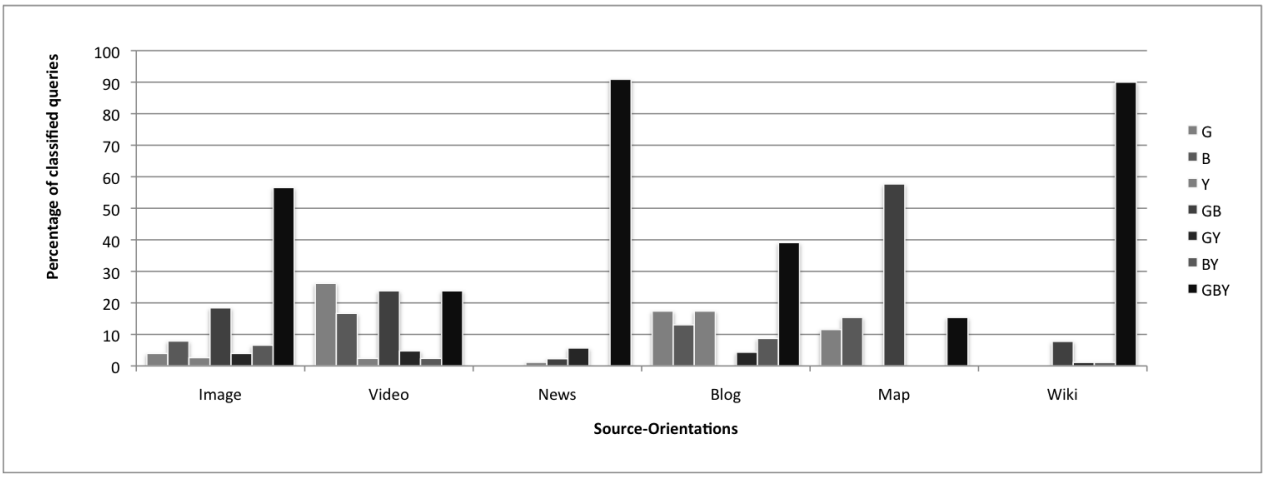


FIGURE 5.6: Comparison of predicted source-orientation with search engines for the sample queries from the classified log. Here, G= Google, B= Bing and Y = Yahoo! search engine respectively. Predicted source-orientations for sample queries, which matched more than one search engine, are also shown. For instance, GB represents percentage of those that matched with the results of Google and Bing both. It can be seen that for most orientations, the match was observed across all three search engines (GBY).

are shown in Figure 5.6. The results show that in most cases, the predicted orientation for the sample queries matched all three search engines (GBY; Google, Bing and Yahoo!). Otherwise, match with at-least one or two search engines was observed. Finally, overall results after comparing the predicted orientations with the existing search engine results, and by comparing the results with the classified queries provided by the human annotators, it can be said that the classifier performed with reasonable accuracy. Therefore, analyses performed on these classified data in the next chapter can be assumed to have reasonable estimations.

5.7 Limitations

There are some limitations of this work, and this section will outline the main limitations of the work presented in this chapter. First, the log data used in the study was from the year 2006, when aggregation of results from different sources was not performed, and the web results were the prime search results provided. As a consequences of which, the log data was highly dominated by the source category "Web". This limited the number of samples from the other sources (e.g., image, map, video, etc) to train the classifier. As a result, the accuracy of the classifier was low due to the lack of training examples. However, using the same classification model, it should be possible to

achieve high accuracy when more example from each source is provided. The overall aim of this work was to suggest possible ways in which the relevance of a source to the users' query can be identified. For this purpose, the Microsoft log-data was used to show a working example for such classification purposes.

Second, although careful attention was paid during manual classification of the training data by the researcher, and also by the ten human annotators during evaluation, it is still possible to observe different percentage of the classified queries per source, if performed by other people. This is because, source-orientation can change with an individual's interpretation of the query meaning. It is known that the quantity and quality of training data is critical for most classification problems. However, it is hard to collect large scale manually labeled Web queries. Therefore, some previous work manually labels Web pages or queries as the training data [as done by several researchers in the past. e.g., Li et al., 2008; Arguello et al., 2009b; Ashkan and Clarke, 2009]. Furthermore, manual classification also provides a high-precision element [Beitzel et al., 2007]. Therefore, for the classifier described in this chapter, it was considered appropriate to manually classify the data for training and for the evaluation purposes.

Finally, the third limitation of this works is that only six source-orientations were considered, so there is a need to extent this work to include more sources, as for example, now provided by search engine verticals (shopping, jobs, local, finance, books, etc). This work can been as an initiation for finding ways to identify source-orientation for a given query.

5.8 Summary

In this chapter, ways to identify the source-orientation behind a user query were presented, more specifically, two methods were describe; *rule based* and *combination*. The first method applied a simple rule based technique, while a combination of rule based and machine learning techniques were used in the second method.

Although simple, yet the rule based method was the first step towards the development of aggregating results from source "relevance" to a given query, and suggested promising results. This method was then further formalized (combination method) in order to obtain stronger evidences of source-orientation. Using only rule based method approximately 3% of log data was identified to have one of the predefined orientations, while using the combination method, it was possible to identify approximately 8% of the total log data to have one of the source-orientation.

Finally, agreement on the predicted orientation for the queries was observed when compared with the results from the search engines. It was found that, for sample queries, the search engines included the results from the sources, which were identified as the intended orientation by the classifier. Furthermore, predicted orientations were also found to be comparable over 60% of the times with the expected orientations provided by the human annotators.

To conclude, this work contributes in understanding the information need behind a user query with respect to different source-orientations. In the next chapter 6, research questions would be investigated to uncover the patterns in which users access results from different sources, i.e. the dynamics of source-orientation.

Chapter 6

Dynamics of Source-Orientation

6.1 Introduction

In this chapter, the classified Microsoft 2006 RFP click dataset is analyzed to obtain an understanding of user behavior with respect to different source-orientations. In particular, the interest is to understand their dynamics (i.e. how information need with respect to different sources evolve within search sessions and their effect on query reformulation). It will be observed that users have information need which requires results from at most two sources within a session (Section 6.6), and these often evolve according to some patterns (Section 6.7). It will also be observed that source-specific terms (e.g. video, map) are often added to or removed from the query whenever there is a change of source-orientation within a session (Section 6.8).

In the previous Chapter, methods to identify the source-orientations behind a query were presented. Source-orientations with respect to six sources were identified. For this purpose, a combination of rule-based and an SVM classifier was used to classify log entries, using which, approximately 8% of the total click dataset were classified to have one of the six domain or genre intents. The remaining (92%) were identified to have as web orientation. In this chapter, the classified log entries (obtained from the previous chapter) are investigated to address research questions to uncover the patterns in which users access results when they have information need for results of different types, i.e. the dynamics of source-orientations.

A second contribution of the work presented in this chapter is the methodology adopted to derive meaningful statistics regarding the dynamics of source-orientations. Indeed, as will be discussed later in the chapter, the statistics computed from the click data are heavily biased by both the skewed distribution over the length (number of clicks) of a

search session, and the different orientations (the web being by far the most predominant source). Therefore the notion of a *random user behavior* is introduced with respect to the object of this study, i.e. the source-orientation dynamics. Any meaningful results then come from comparing the statistics calculated from the original data set and from that generated by a random user. Random model is described in detail in Section 6.5.

6.1.1 Structure

The outline of the Chapter is as follows; The work presented in this Chapter is motivated in Section 6.1.2. The research questions are discussed in details in Section 6.2. They are then investigated in Section 6.3 Section 6.4, Sections 6.6, 6.7 6.8 and Section 6.9, respectively. The random model is described in Section 6.5. Finally section 6.11 concludes this chapter.

6.1.2 Motivation

Understanding the users' search behavior is one of the important goals in web search, and information on how users interact and view search results helps improving the user information seeking process (e.g., White and Morris [2007b]; White and Drucker [2007]). This is because, having an insight on searching behavior of users have proved to be useful in building better information retrieval systems (e.g., Halvey and Jose [2009]; Agichtein and Zheng [2006]; Agichtein et al. [2006]; Bilenko and White [2008]). For instance, Agichtein et al. [2006] show that incorporating user behavior data can significantly improve ordering of top results in real web search setting.

Several area have benefited from exploiting the knowledge of user behavior, for instance, health care (e.g., Anagnostopoulos and Maglogiannis [2006]; Mahoui et al. [2009]), commerce (cite,Lin and Wang [2005]) , social network (e.g., Maia et al. [2008]; Benevenuto et al. [2009]; Kwak et al. [2010]), etc have shown to benefit from the knowledge of the user behavior. For example, Benevenuto et al. [2009] state that, understanding how users behave when they connect to social networking sites creates opportunities for better interface design, richer studies of social interactions, and improved design of content distribution systems.

Furthermore, understanding dynamics of user behavior can also be useful in improving personalized search experiences (e.g., Sendhilkumar and Geetha [2007]; Sieg et al. [2007]; Kumar et al. [2008]). For instance, Kumar et al. [2008] suggests a personalized re-ranking of URLs returned by a search engine using user's browsing behaviors, while

Sieg et al. [2007] presents an approach to personalized search that involves building models of user context as ontological profiles.

Therefore, understanding of user behavior is vital for building better systems. In context of web search, many dedicated research have analyzed user behavior during information seeking process, and have further used this knowledge in building sophisticated search engines. For instance, studies reported in, (e.g., Spink et al. [2002]; Granka et al. [2004]; Jansen and Spink [2005]; Rose [2006]) aimed to understand the dynamics of the search behavior of users, i.e. how users' information need and searching behavior evolve, query length, etc. These studies analyzed web search engine click data to gather information about the queries being submitted to the search engine, their length, query frequency per session, click frequency per session, etc.

The example studies stated above show that the knowledge of user behavior has proved to be beneficial in many ways. These studies have provided insight on how user perform search or interact with search results. However, there is very limited (if any) knowledge of user behavior with respect to different source-orientations. That is, it is not known how users interact with results when they have information need which desires results from different sources.

Furthermore, there are not many evidences, which can reflect similar findings (if any) of user behavior with underlying source-orientations; as observed in different search scenarios mentioned above. For example, it is known how users modify their queries when looking for "only" web results or image results, and so on. It is not clear how users modify their queries, when their information need changes in context to different sources. Although it is known on how users perform searches when looking for video results (e.g., Albertson [2010]; Vrochidis et al. [2010]), or image results (e.g., Villa et al. [2010]), news result (e.g., Hu et al. [2008]). However, it is not know (to the best of knowledge) that how users perform searches when the information need is oriented towards different sources, or combination of sources.

In an aggregated search scenario, understanding users' information need with respect to different sources is important, as aggregated search requires aggregation of results from different sources. Therefore, information of users' expectations and searching behavior when looking for results from different sources can prove to be useful in building better aggregated systems. For instance if known, results from which sources users look for when issuing health related queries, then this information can be used by the aggregated system in selection of appropriate sources for the aggregated result page.

Finally, as already discussed in Chapter 2, web search engine click data is often used by search engine companies and the researchers to understand the user behavior dur-

ing information seeking process. A search engine's click-data is an electronic record of interactions that have occurred during a searching process between a Web search engine and searchers who are seeking information on that Web search engine. Therefore, the use of data stored in logs of Web search engines, Intranets, and Web sites can provide valuable insight into understanding the information-searching process of online searchers Jansen [2006]. Therefore, in this chapter, the Microsoft 2006 RFP click dataset is analyzed to obtain an understanding of user search dynamics with respect to different source-orientations. Next, we discuss the research questions addressing dynamics of users' searching behavior with respect to different source-orientations.

6.2 Research Questions

Having classified the clicks into different source-orientations (domains and document genres) in pervious Chapter [5], it was then possible to study the dynamics of these classified source-orientations. The following research questions were therefore formulated for this purpose:

R1: *Does rank position affect user click behavior for different source-orientations?*
The first research question looked at the frequency of click-through at different rank positions (Section 6.3). It was hypothesized that click-through frequency would be skewed with respect to results positioned at the top of the result page, since studies for instance by Keane et al. [2008], reported people being biased towards items listed at the top of the result list.

R2: *Does there exit an effect of click order on the access pattern of sources?*
The objective here was to examine if certain source relevant results tended to be accessed earlier or later in the session (Section 6.4). For instance, do users tend to click a news related result earlier or later in the search session. The question is further examined by research question **R4**, where patterns among the click order for frequent domain and genre combinations within a session were analyzed.

R3: *What are the frequent combinations of source-orientations within a search session?*
By looking at how source-orientations co-occur within a search session, we gain insights on what are common combinations of sources in web search. For example, and not surprisingly, as it will be seen in Section 6.6, the combination

web+image is more common than video+blog. Answers to this research question also allow us to restrict the options to consider when investigating the other two research questions.

R4: *Does source-orientations evolve according to some patterns?*
In sessions with several source-orientations, do these follow any pattern? For example, as will be seeing in Section 6.7, a common pattern is to start with a web orientation (a typical web result), and then to switch towards map results.

R5: *Is there a relation between query reformulation and a change of source-orientation?*
When the source-orientations changes, e.g. switching from a news results to a blog results, is the query re-formulated? As discussed in Section 6.8, a particular re-formulation strategy is to add a source-specific term.

R6: *Is there an association between the query category and the source-orientation?*
This aim is to see if there exits any relation between category (art, business, entertainment, etc) of the submitted query and, the intended source specific result. For instance, do "health" related queries often fetch results from image and Wikipedia? In that case it might also be possible to suggest a way of aggregating results based on users' query. It would be seen in Section 6.9 that, there is an indication of the existence of associations between query categories and sources.

In order to obtain answer for above research questions (**R1, R2, R3, R4, R5** and **R6**), the classified log data obtained from Chapter 5 were used. In following sections, outcomes for the above research questions are presented.

6.3 Effect of Rank Positions

The aim in this section is to investigate if there is any rank effect on users' click behavior. To address research question one [**R1**], the rank positions of clicked results, classified to have one of the predefined source-orientations were analyzed. While there is a general trend of higher ranked documents being clicked more frequently than lower ranked documents Guan and Cutrell [2007]; Keane et al. [2008], there is very limited understanding of the effect of domains and genre on this behavior. Figure 6.1 shows the distribution of ranked positions (one to five) within each source. Related findings were obtained in a survey[6-1].

[6-1] http://www.iprospect.com/about/researchstudy_2008_blendedsearchresults.htm

As one can see, most source-orientations follow the general click-through pattern, which is a high frequency of the top at rank one and then with monotonically decreasing frequency. This suggests that selection of sources may strongly be affected by the ranking positions. This observation is also consistent with results obtained from user study performed in previous Chapter 4, where a clear position effect was observed on users' click-through behavior (Section **??**). Therefore results from the rank affect further backs up the findings from the click data from the user study.

The effect is found to be strong in case of news and map orientations. Users seem to click on news and map results more often when positioned on top of the result page. Therefore, in aggregated scenario, when including news or map results with web results, it might be necessary to place them at top position, in order to get them accessed by users.

Furthermore, in case of image, Wikipedia and blog orientations, a similar level of click-through rate across the rank positions is observed. Although they still follow the general decreasing trend. This suggests that the influence of ranking positions on the selection of the image, Wikipedia and blog results was weaker than with the other intents.

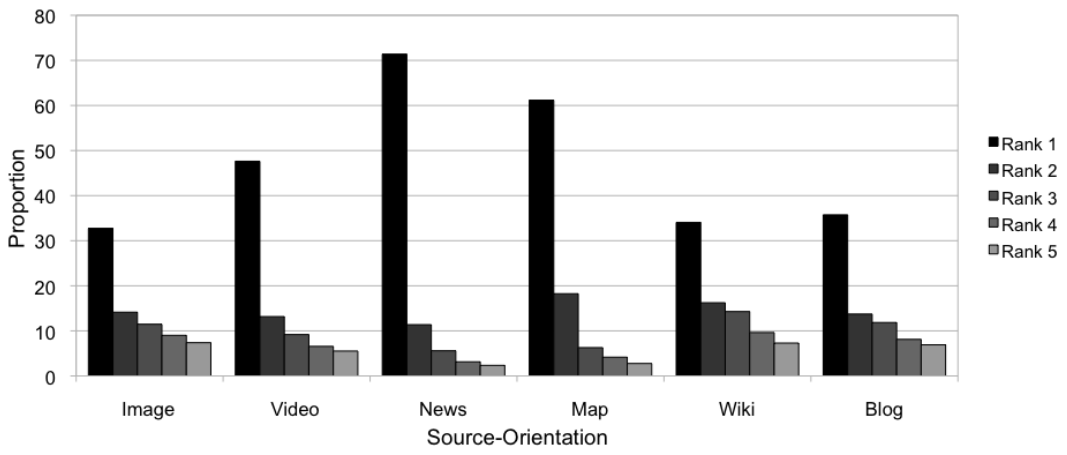


FIGURE 6.1: Effect of rank positions (Rank 1-5).

6.4 Effect of Click Order

Next, the effect of click orders on the access of results with different source-orientations were analyzed, research question two [**R2**]. The objective was to examine whether a certain source relevant result tended to be accessed earlier or later in search sessions.

For this analysis, the numbers of clicks were counted and the click order for those sessions that had at least three click-through documents were recorded. It should be noted that for this analysis, log data classified using only rule base method was used (Section 5.5, Chapter 5). 6788 sessions having three or more than three clicks were analyzed and the result of the analysis are shown in Figure 6.2.

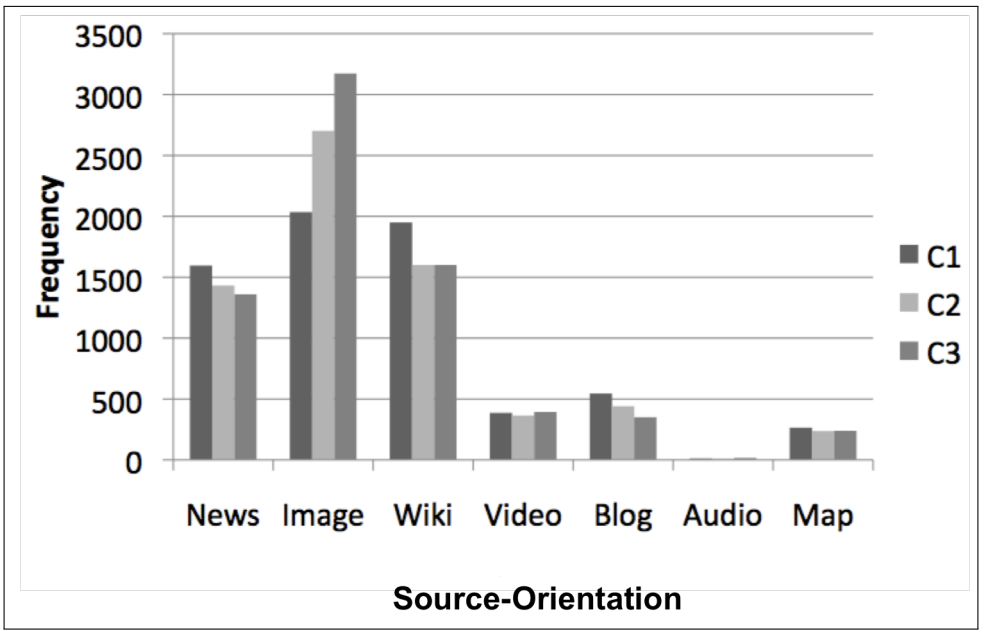


FIGURE 6.2: Effect of click order (Click 1-3).

As can be seen, most results had a similar level of frequency of clicks across the click order in sessions. There was a trend of decreasing frequency as the click number increased. Image domain, however, had a very different result. People seem to click the image related results more frequently as the search session progresses. This suggests that the value of image results can increase as the search progresses in aggregated search results. The above research question is further examined by research question four [**R4**], where patterns of accessed intents within the session are analyzed.

In the following sections, research questions three, four, five and six (**R3, R4, R5** and **R6**) are discussed. Nevertheless, before discussing the remaining research questions, the random model methodology adopted to compare the answers is described first.

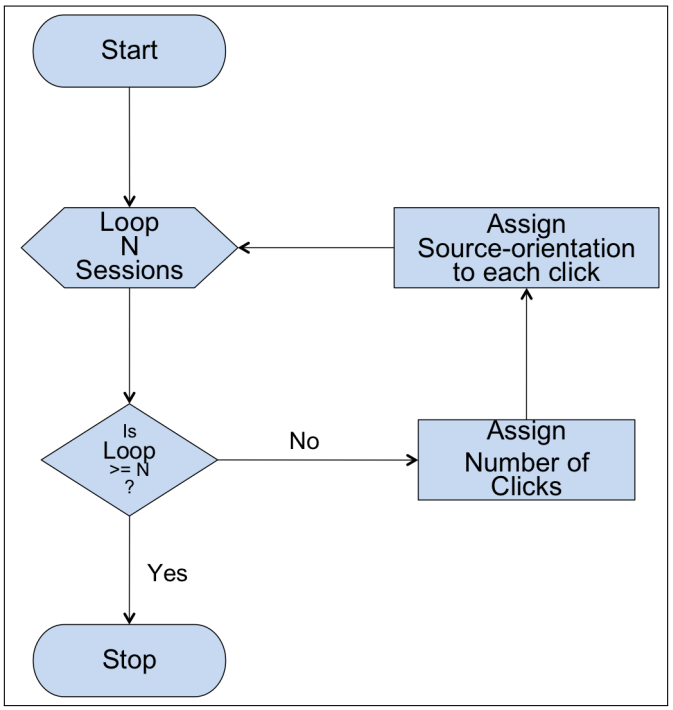


FIGURE 6.3: Process for generating random log. Random log is used to overcome bias from predominance of single and standard web clicks.

6.5 Random Model

To overcome the skewness of the observed distributions in the log data, a methodology was needed. Indeed, from Table 6.1, it can be seen that around 96% of the sessions have only one underlying source-orientation. However, most sessions are composed of one or two clicks (85%, from Table 5.1 in Chapter 5), and that web is the most likely source (92%, from Table 5.2 in Chapter 5). In such situation, it can then be expected that a high percentage of sessions will be one or two clicks sessions with a web orientation, and hence will be single-orientation sessions. It is therefore not possible to know whether the observed statistic of 96% is because users do not have multiple source-orientations, or that the click and the source-orientation distributions are skewed.

To overcome this, statistics with those obtained from the log data were compared from a so-called *random log*. This random log is exactly similar to the real log, but instead of using a classifier to assign an orientation to each log entry, one source-orientation is selected by random, in accordance with the source-orientation distributions presented in Table 5.1. The random log is thus a log where the source-orientation would be independent of what the user is searching for, and of his or her search history.

To compute statistics for this random log, all the possible random source-orientation

assignments have to be averaged. In practice, the different formulas were formally derived which were needed to compute the statistics from the source-orientation and click distributions, so the number reported are probabilistic expectation of the different statistics.

Process for generating random log is shown in Figure 6.3 where, the random log of size equal to the size of the original log is generated, that is having equal number of sessions. For each session, number of clicks (random) is predicted, and then for each click a random source-orientation (from seven source-orientations) is assigned. This is repeated for all the sessions in the random log. Ten random logs are generated using this process, and finally their assignments are averaged to obtain a single random log.

In the rest of this chapter, statistics computed for both the real and the random logs are reported. Going back to the example of the 96% of single source-orientation sessions in Table 6.1, it can be seen (Table 6.1, column "Random") that we would expect 93% of the sessions to be single source-orientation in the random log. This sets the following limit: if the real number was below this limit (even with a percentage as high as 90%), then it would be possible to say that users tend to combine more than one source-orientations within a session; instead, it is observed that the statistic is higher (96%), which means that sessions are indeed generally single source-oriented.

6.6 Combination of Source-Orientations

Here the third research question [**R3**] is investigated, that is the existence of frequent combinations of source-orientations in search sessions. To this end, how often source-orientations co-occur within a same search session is computed. This was done on 3,957,888 sessions (the number of sessions with less than twenty clicks). The sessions up to twenty clicks were considered because while generating the random log, sessions up-to twenty clicks were generated. This was done in order to limit the size of sessions and to have comparable session size between random log and the real log. Since generating a random log required two steps, first generating a session with random number of clicks, then assigning a random source-orientation to each click. It was necessary to restrict the number of session in order to have fair comparisons between the two log-data. It should be noted that limiting the session size to maximum of twenty clicks did not cause discarding large log data samples. When analyzed, there were very few sessions that contained over twenty clicks (less than 200).

First, in Table 6.1 the percentage of sessions that contain two, three, ... , to seven different source-orientations are reported. This was done for the original log data, and the

TABLE 6.1: Percentage of sessions with one, two, three, four, five, six and seven different source-orientations. Note that the number of distinct orientation does not correspond to the number of clicks.

Distinct orientations in the session	Log (%)	Random (%)
Seven	0.000	0.000
Six	0.000	0.000
Five	0.000	0.002
Four	0.005	0.034
Three	0.121	0.470
Two	3.385	6.354
One	96.489	93.140

randomly generated log data as well. From the table, it can be observed that there are very few sessions with more than two different source-orientations. This is in accordance with the study reported in Arguello et al. [2009b] where, through manual assessments, only 30% of the queries had more than one source-orientation (it should be noted that, in the study, verticals were assigned to the queries, some of which corresponding to the domains and genres studied here).

The difference between the original and the random log statistics show that the fact that a session is associated with a low number of source-orientations is not due to chance. Moreover, the difference increases as the number of sources increase (from 2 times higher with two intents to 7 times higher with four intents), which shows that when users have diverse orientations, it is generally restricted to at most two. Therefore, in the rest of the study, the analyses are restricted to sessions having two source-orientations.

Returning to the research question [**R3**], which source-orientations co-occurred frequently is now investigated. The percentage of sessions where at least two source-orientations appeared were computed and reported the values in Table 6.2. For instance, in this table, the value 0.01% for "nb" means that there are very few sessions with a blog and a news orientations. Given that the different source-orientations have distinct probabilities of occurrence, it is also interesting to look at the conditional probability that a source-orientation can be present (third and fourth series of columns). For instance, in the line labeled "bb", it can be seen that if the first observed orientation is "blog" in the session, then there is a 0.189 probability that a second "blog" orientation would be observed in the same session. The results are further analyzed in the rest of this section.

The most important observation from the results is that most users do not mix sources. Indeed, the first and last series of rows of the table show that users are less likely to combine two different sources in the same session that what would be expected by

TABLE 6.2: Pairs of intents. Column % report the percentage of sessions that had the corresponding pair of intents. Column % among 1st (respectively 2nd) reports the percentage of sessions with the first intent of the pair (respectively second) that also had the second (respectively first). L stands for log, and R for random. For the intents, we use n= News, m= Map, i= Image, v=Video, w (lowercase) = Wikipedia, b=Blog and W= Web.

	%		% among 1st		% among 2nd	
Combination	**L**	**R**	**L**	**R**	**L**	**R**
bm	0.00	0.02	0.5	4.6	0.1	0.5
nb	0.01	0.03	0.3	0.5	2.6	6.2
vm	0.01	0.2	0.3	4.5	0.4	4.4
bw	0.02	0.01	4	2.8	0.7	0.5
ib	0.02	0.02	0.8	0.5	4.1	4.3
im	0.02	0.19	0.9	4.5	1.1	4.1
in	0.04	0.26	1.5	6.2	1.1	4.2
nm	0.04	0.28	1	4.5	1.8	6.2
nw	0.04	0.18	1.1	2.8	1.6	6.2
vb	0.04	0.02	1.4	0.5	7.8	4.5
wm	0.04	0.13	1.5	4.5	1.9	2.8
vn	0.05	0.27	1.7	6.2	1.3	4.4
iw	0.09	0.12	3.3	2.8	3.4	4.1
iv	0.10	0.18	3.6	4.4	3.5	4.2
vw	0.06	0.12	2.1	2.8	2.2	4.4
bb	0.09	0.00	18.9	0.2	18.9	0.2
ww	0.50	0.04	18.6	1.5	18.6	1.5
mm	1.11	0.1	51.8	2.3	51.8	2.3
vv	1.17	0.1	43.1	2.2	43.1	2.2
nn	1.33	0.2	35.5	3.2	35.5	3.2
ii	1.40	0.09	52.5	2.1	52.5	2.1
bW	0.43	0.51	88.6	97.1	0.4	0.5
mW	1.35	4.42	62.7	97.3	1.4	4.4
iW	1.85	4.09	69.2	97.3	1.9	4.1
vW	1.99	4.31	73.1	97.3	2.1	4.3
wW	2.48	2.77	92.9	97.2	2.6	2.8
nW	2.87	6.1	76.5	97.4	3	6.1
WW	91.77	91.41	94.9	91.7	94.9	91.7

random (around 3 times less likely in average).

Moreover, looking at the second series of rows it is on average around ten times more likely that users repeat a click on the same source (rows "bb" to "ii") than what would be expected by random. In sessions made of two or more clicks, when one orientation is map, video, image or news, then there is above 35% of chance to observe a second click with the same orientation (third and fourth group of columns). For blog and

Wikipedia, the probability is lower although still high (around 19%). This could be because users might consider Wikipedia and blog result pages as web pages and hence do not differentiate them as coming from separate sources.

We however observe some potential exceptions for the pairs blog/blog ("bw"), image/blog ("ib"), and video/blog ("vb"). These three pairs occur more often together that would be expected by random. However, the percentages of sessions are so low that this is likely due to noise. More data, and eventually, a more effective classifier, is needed to obtain a real understanding of these potential exceptions.

A last observation is that when there are two source-orientations, these are often a web orientation and any of the non-web one, as shown in the last series of rows in the table. This is not surprising, and means that web search results should continue to contain mostly web results, and when appropriate, images, videos, blogs, in addition. This is nowadays the approach followed by all major search engines.

6.7 Patterns of Source-Orientations

In the Section 6.4, it was observed that there is a pattern in click order of accessed source related results (research question two [**R2**]), and that, certain results were accessed more at the beginning of the search session, while some tended to be accessed more as the session progressed.

In this section, how source-orientation evolve within search sessions (research question [**R4**]) is investigated. Exploiting the findings from the previous section, the analysis is restricted to sessions with two source-orientations, as most sessions have at most two source-orientations. In addition, the five most frequent co-occurrences of two source-orientations from Table 6.2 are considered during this analysis (other co-occurrences were too low to derive any significant findings). The frequent pairs were; image+web, wiki+web, video+web, news+web, map+web. For each such pair, all possible sequences of changes of orientations were looked at. Let us consider five source-orientations; image, video, news, Wikipedia, map as **NW** (non-web), and web as **W**, then the four main sequences observed were of the form:

1. W $\rightarrow$ NW,
2. NW $\rightarrow$ W,
3. NW $\rightarrow$ W $\rightarrow$ NW,

4. W → NW → W.

In Table 6.3, for each pair, the percentages of sessions containing each of the identified four sequences are reported. All the others possible sequences are reported under "Other" in the table. In this calculation, sessions with less than three clicks were excluded. This was done to avoid results biased towards the large number of sessions with two clicks or less. Finally, the calculations are done for the original log and the random log. Three observations were made from the results.

First, with the Wikipedia orientation, users do not follow any particular pattern. Indeed, the sequences obtained are close to what would be expected by random. This confirms the findings of the previous section, where we made the hypothesis that users do not differentiate between Wikipedia and web documents (Section 6.6).

Second, when the orientation is towards news and map results, users switch from one orientations to the another, but do not switch back to the previous source-orientation. By random we would expect more users to move back and forth between source-orientations. This can be seen in the difference between the random and real logs for the four sequences (Table 6.3).

Third, in case of an image and video orientation, more users first view a web result and then view an image or video related result. This further explains the different pattern observed for image results when addressing research question two [**R2**], where the click order for image result increased as the session progresses (Section 6.4). Therefore, including more image results as the session progresses might be beneficial in an aggregated scenario.

However, unlike video, map and news orientations, users with image orientation also move back and forth between web and image related results (web → image → web), but this would be expected by random too, hence it does not contribute to the observed sequence for the image orientation.

The above results show that there is a clear sequence in the source-orientations for all except Wikipedia intent. Users have a tendency to change orientation from source to another, and then to end the session, rather than switching several times between source-orientations within the same session. Such observation could be useful in improving the search result pages when using implicit feedback (clicks) or in real-time aggregations. For instance, when users change from a web to an image result, it might be useful to include more image results when user returns to the result page, since it is less likely that user would switch back to web results again. This in turn may help in improving

TABLE 6.3: Sequence of intents in search sessions, for each pair web+non-web. The rows labeled "Other" give the percentage of sessions with none of the listed sequences for that pair. Percentage numbers for the most frequent sequences are in bold.

Sequence	**Log** (%)	**Random** (%)
wiki → web	25	26
web → wiki	30	26
web → wiki → web	**41**	**46**
wiki → web → wiki	1	0
Other	3	1
video → web	32	26
web → video	**34**	26
web → video → web	31	**46**
video → web → video	1	0
Other	2	1
map → web	**36**	26
web → map	34	26
web → map → web	28	**46**
map → web → map	1	0
Other	1	1
news → web	**34**	26
web → news	31	26
web → news → web	32	**45**
news → web → news	1	1
Other	2	1
image → web	27	26
web → image	**37**	26
web → image → web	31	**46**
image → web → image	1	0
Other	4	1

information seeking process of users by appropriately returning the results from the most likely domain or genre.

6.8 Source-Orientation and Query Re-formulation

In this section, both quantitatively (how many queries are modified?) and qualitatively (how were they modified) the effect of a change of source-orientation on a user query are analyzed (research question **R5**). This was done by comparing pairs of consecutive queries with two different source-orientations, within the same session. That is, when there was change in source-orientation within a session, query with the initial

TABLE 6.4: Most frequently added and removed terms during a change from web → non-web orientation.

Orientation change	Terms added	Terms removed
web → Wikipedia	Wikipedia, what, how, history, states	world, does, what, state
web → blog	blog(s), how, women, song, love	with, teens
web → news	news, newspaper, press, daily, channel	post, new, times
web → map	map(s), mapquest, state, city	estate, southwest, hotels, airlines
web → video	video(s), movie(s), porn, free, funny	lyrics, free, myspace
web → image	picture, photo, image, free, gallery	what, hair, home, ideas

TABLE 6.5: Most frequently added and removed terms during a change from non-web → web orientation.

Orientation change	Terms added	Terms removed
Wikipedia → web	world, does, what, state, dictionary	Wikipedia, wiki, what, how, history, states
blog →web	lyrics, world, teens	blog(s), how, have, women, song
news → web	post, new, times, york, county	news, newspaper, press, daily, channel
map → web	estate, southwest, hotels, airlines	map(s), mapquest, state,city
video → web	lyrics, free, myspace	video(s), movie(s), porn, free, funny
image → web	what, hair, home, ideas	picture, photo, image, free, gallery

source-orientation was compared to the query with the new source-orientation.

First, for each pair of consecutive and different source-orientation within a session, the numbers of queries that were exactly the same, modified or completely different were computed. Similar statistics were calculated by Jansen et al. [2000], but on a different log dataset, and without taking into account any underlying source-orientation. Results are shown in Table 6.8. For blog and Wikipedia, over 50% of the users did not change their query (orienting from a blog or Wikipedia results to web results or vice versa). It is likely that this happens because both types of results are present in the top ranked documents for the same query. Users do not have to change their queries to obtain results from blog or Wikipedia and then web sources (and vice versa).

The situation is reversed for news, image, video, and map orientations. Most of the time, users did change their query (over 65%). Furthermore, it can be noted that there is a slight difference between news/map/video and image orientation. In the former case, users issued different queries, whereas in the latter, in half of the cases, the users modified their queries by adding or removing terms.

TABLE 6.6: Percentage of sessions where a query was not modified, was modified (i.e. by adding or removing terms), and was different (no terms in common).

Sequence	Exact (%)	Modified (%)	Different (%)
web → news	21	21	**58**
news → web	18	19	**63**
web → wiki	**59**	18	23
wiki → web	**54**	20	26
web → video	35	25	**40**
video → web	33	24	**43**
web → image	30	**40**	29
image → web	30	34	**37**
web → map	4	24	**73**
map → web	3	21	**76**
web → blog	**56**	24	19
blog → web	**52**	23	25

In the case where a query was modified, we also looked at which terms were added or removed (qualitative analysis). Tables 6.8 and 6.5 show the most frequent terms that have been added and removed during a change of source-orientation, respectively. When a user went from a web orientation to any other of the other six orientations (Table 6.8) or vice versa (Table 6.5).

It is possible to easily identify terms that are linked with a source, which were referred in previous Chapter 5, as source-specific terms (Section 5.6.2). Namely for,

- Wikipedia → "Wikipedia", "what", "how"
- Blogs → "blog", "how", "song"
- News → ' 'news", "newspaper", and "press"
- Maps → "map", "mapquest", "state" and "city"
- Video → "video", "movie", "porn", "free" and "funny"
- Image → "picture", "photo", "image", "gallery", and "free"

It should be noted that some of these terms were present in the rule-based classifier described in Section 5.6.2, but some were rightly (re)identified here.

TABLE 6.7: Categories from Open Directory projects (OPD), and associated example topics.

Category	Example Topics
arts	Movies,Television,Music, Photography, Body Art
games	Video Games, RPGs, Gambling
kids	Arts, School Time, Teen Life
reference	Maps, Education, Libraries
shopping	Clothing, Food, Gifts
world	Dansk, Deutsch, Italiano
business	Jobs, Real Estate, Investing
health	Fitness, Medicine, Alternative
news	Media, Newspapers, Weather
regional	US, Canada, UK, Europe
society	People, Religion, Issues
computers	Internet, Software, Hardware
home	Family, Consumers, Cooking
recreation	Travel, Food, Outdoors, Humor.
science	Biology, Psychology, Physics
sports	Baseball, Soccer, Basketball.

6.9 Query Category and Source-Orientation

The aim of last research question [**R6**] is to investigate if there is any association between query category and the selected source-orientation. That is, the aim was to see if for a given query, the identified category (health, arts, etc.) and the source (image, map, etc) have any association. For example, for queries belonging to 'health' category, do users generally access 'image' Wikipedia' and web results?

In order to achieve this, the 200 most frequent unique queries were collected for the six source-orientations, and for the six most frequent combinations (as observed from Section 6.6). Therefore in total, twenty-four hundred (2,400) queries were categorized for this purpose. 1,200 with single source-orientation, and 1,200 for six combinations of source-orientations. Each query from the selected source-orientation and the their combination was then manually assigned to one of the categories (e.g., health, business, etc), as defined in the Open Directory Project ODP [2010].

Some of the example topics for the categories defined in ODP are shown in Table 6.7. These pre defined topics help in selecting suitable categories for the sample queries. Any query that did not fit into an existing category was assigned to an 'others' category. The categorization of queries was carried out to identify associations between query

categories and the source-orientation, and for the combinations of source-orientations. The percentage of clicks per category with respect to single source-orientation and combinations of source-orientations are shown in Table 6.8 and 6.9 respectively. Two observations can be made from the results.

TABLE 6.8: Association between query category and source-orientations. The percentage values are shown in the table

Category	Blog	image	map	video	news	wiki
arts	17.5	29.0	0.0	50.0	1.5	10.5
games	0.0	0.0	0.0	6.0	0.0	0.0
kids	0.0	17.5	0.0	0.5	0.0	3.0
reference	2.0	1.0	96.5	0.0	0.0	3.5
shopping	4.0	5.0	0.0	1.0	0.0	1.5
world	0.5	0.0	0.0	0.0	0.0	1.0
business	5.0	0.0	0.0	0.5	2.0	0.0
health	6.5	4.0	0.0	0.0	0.5	7.0
news	7.5	0.0	0.0	1.5	86.5	0.0
regional	2.5	1.0	0.0	0.0	1.0	6.5
society	2.5	2.5	0.0	1.0	3.5	22.5
computers	6.0	0.5	0.0	3.0	0.0	1.5
home	7.0	3.5	0.0	0.0	0.0	0.5
recreation	1.5	4.5	0.5	0.0	0.0	4.0
science	4.5	9.5	0.5	0.5	0.0	16.5
sports	1.0	0.0	0.0	0.5	1.5	2.0
adult	23.0	9.0	0.0	28.0	0.0	1.0
others	9.0	13.0	2.5	7.5	3.5	19.0
Total	100.0	100.0	100.0	100.0	100.0	100.0

First, there is evidence of associations between the category of a query and source-orientation. For queries belonging to a certain category, there exists preference for results from specific sources. For example, the percentages of clicks for the map are higher for "reference" category (fourth row in Table 6.8), and for category "business", the combinations of news and web results were preferred more (row seven, Table 6.9).

Second, users do not have many source-orientations, that is, users do not access results from many sources for a single information need. For instance, for category "kids" (third row Table 6.8), intent for image results was higher than other intents. In other words, for queries belonging to category "kids", image results are mostly preferred. Similar observations are also made when combination of two intents is analyzed with respect to different categories. For example, results from Wikipedia and web are preferred most for queries belonging to category "society" (row eleven, Table 6.9), while

image and web results are preferred for category "home" (row thirteen, Table 6.9).

However, the associations for category "arts" (first, Table 6.9) is found to be more diverse than other categories, and results from image, video, blog and web are intended. Although some dominance of preference for the 'video' domain is observed for this category yet, other intents are also preferred. One of the possible reason for this observation could be that, the category like "arts" includes many visual topics, such as, movies, photography, entertainment, animation, etc. Therefore, it is possible to have a distribution of clicks across different forms of visual results, such as image and video in this case.

Above results also backs up the findings from Section 6.6, where frequent combinations of source-orientations were looked at, and it was found that users often do not desire results from multiple sources.

TABLE 6.9: Association between query category and source-orientation combination

Category	Blog-Web	Image-Web	Map-Web	Video-Web	News-Web	Wiki-Web
arts	18.0	18.0	1.0	50.5	2.0	10.5
games	1.0	0.0	0.5	0.0	0.0	0.5
kids	0.0	0.0	0.0	0.0	0.0	0.0
reference	3.0	3.5	45.0	0.0	0.0	7.0
shopping	15.0	6.0	6.0	2.0	2.0	2.0
world	0.0	0.0	0.0	0.0	0.0	0.0
business	10.0	0.5	1.5	1.0	34.0	1.0
health	5.0	5.0	0.5	0.5	1.0	6.5
news	4.5	0.5	3.0	2.0	45.0	0.0
regional	0.0	1.0	8.5	0.5	1.0	7.5
society	4.0	3.5	1.5	3.5	9.5	32.0
computers	5.0	4.0	0.0	0.5	0.0	2.0
home	0.0	15.0	0.5	1.5	0.0	5.0
recreation	1.5	21.5	12.5	0.0	0.0	0.5
science	13.0	1.0	0.5	0.0	0.0	19.5
sports	0.0	0.5	0.0	0.0	0.0	2.0
adult	16.5	10.5	0.0	27.5	0.5	0.5
others	3.5	9.5	19.0	10.5	5.0	3.5
Total	100.0	100.0	100.0	100.0	100.0	100.0

To conclude, observations from this analysis give an indication that there are associations between a query category and sources. For queries from certain categories, results from specific sources are more relevant. Therefore, information like query category

can be further exploited in order to suggest suitable aggregation of results from appropriate sources. The analysis hence provides some insight into the information seeking behavior of web users in terms of the sources and the query category.

6.10 Limitations

There are some limitations of this study. First, it makes use of data generated by a search engine where general web results were predominant. This has consequences in the analyses, since this lowers the probability of observing more specific orientations (e.g., image, video, etc.) and also forces users to use source specific terms (e.g., photo, movie) in order to retrieve source specific results. Second, due to this low probability, more data (sessions) would have been necessary to obtain more reliable estimates of the different statistics. Third, only six non-web intents were considered, therefore, for future research, there is a need to extent this work to include more sources.

Finally, the log-data used in this study was from the year 2006 when the aggregation approach was not used by the search engine companies, and hence users were not provided with diverse results. Therefore, it is possible that the users' source related searching patterns observed during this study might differ to how users' perform search today. This is because, the aggregation of results from different sources are now followed by the most search engines and for most of the queries submitted to them. As a result, it is possible that the usage of source specific terms may reduce when looking for results from specific source. Since, due the introduction of the aggregated search and it being around for almost five years, it is possible that users will expect results to be included from the sources without having them to mention explicitly.

Furthermore, users' preferences regarding source combinations, number of sources and the sequence in which they access these results may also result in similar or different observations when compared to the observed behavior from 2006 log data. However, the aim of this study was not just to obtain an insight into users search pattern, but mainly to highlight the possibility of source related users' interaction, and to suggest that such behaviors should be taken into account when designing an aggregated search interface. As this may lead to issues like – information overload and user dissatisfaction.

6.11 Summary

In this chapter, users' searching behavior with respect to different source-orientations were analyzed. For this purpose dynamics of search behavior with respect to three domain; image, video and map, and three genre; blog Wikipedia and news were studied. More specifically, six research questions were investigated to uncover the patterns in which users access documents of different types, i.e. the dynamics of source-orientations.

First, effect of rank order was analyzed, where it was hypothesized users clicks are biased towards the top positioned results, further reflecting on some of the past studies (e.g., Keane et al. [2008]), but with respect to different source-orientations. Results obtained supported the hypothesis thus suggesting that, indeed, position (ranking) of search results continue to influence users searching behavior. Furthermore, the results also backed up the results from user studies described in Chapter 4.

Second, the aim was to see if there exits any order in accessing results with different source-orientations. That is, if results from certain source relevant results are accessed earlier or later in the session. It was observed that, image being an exception, remaining source relevant results seemed to accessed when the search session begins. The pattern of these accessed intents were further analyzed by research question four.

With respect to third research question, the frequent combinations of source-orientations were looked at. That is, how different source-orientations co-occur within a session. It was observed that users do not often mix orientations, and if they do, they had at most two source-orientations in a session. Furthermore, when there were two source-orientations, these are often a web and any of the non-web one (e.g., image, news). These results supports the approach followed by all major search engines, where search results contain mostly web results, and when appropriate, either images, videos, blogs, etc, are provided.

Fourth, if these source-orientation combinations evolve according to some patterns was investigated. The results show that there is a common sequence in all except with Wikipedia. Users in general tend to seek results from the same source for a while and then switch to results from another source. For most cases, users do not switch back and forth between sources.

Fifth, the aim was to see if there were relations between query re-formulation and change of source-orientation. We observed that for video, news and map orientations, more users used completely different queries, whereas, for blog and Wikipedia ori-

entation, more users used the same query, when they change their source-orientation. Further, most users use source specific terms (photo, newspapers, video, etc) to modify their query when they change their source-orientation.

Finally, association between the category of the submitted query and the selected source-orientation were investigated by the last research question. It was observed that there is an association between a user query (category) and the source-orientation. In addition, for certain query category, there is a dominance of specific sources, which indicates that not all sources are "relevant" to all information needs (here approximated as the queries submitted within a session).

To conclude, results from Part II suggested that identifying the relevance of source (image, map, etc) for a given query is critical in aggregated search scenario. Today, when most search engines attempts to combine results from different sources on their result page, relevance of result extends beyond search topics. In order to satisfy users' information need, relevance of result to the search topic, as well as relevance of the source from which the results are fetched is required.

In this Part III, two main contributions were made; 1) a combination method of rule based and machine learning technique were used to identify the source-orientation behind users' information need (Chapter 5). 2) The results obtained were then further exploited to uncover the dynamics of users' searching behavior with respect to different source-orientations (Chapter 6). In the next Part IV , discussions and conclusions drawn from the research from this thesis and avenues for future work is presented.

Part IV

Discussion, Conclusion and Future Work

Part IV

In part III, through log-data analysis attempt to understand users' information need and their dynamics were made. In chapter 5 ways to identify source-orientation behind a users' query was provided. The method was used to classify the log data where, source-orientation for the given queries were identified. The classified log-data was further analyzed to gain insight on how users searching behavior evolve when they have information needs requiring results from various sources. In this part, the main findings of this work are discussed first (chapter 7) and then this thesis is concluded in chapter 8 where the avenues for future work is also presented. The conclusions and future directions are drawn from findings obtained in the user experiment and research described throughout this thesis.

Chapter 7

Discussion

This chapter draws together the findings of this thesis and discusses their implications. In particular, the outcomes of the five research questions stated in introduction chapter (Section 1.3) of this thesis will be discussed here. In the previous parts of the thesis, various aspects of aggregated search were presented and investigated. The problem definition was described in part I, while part II investigated the result presentation aspect of the aggregated interface. Finally, dynamics of various information need were investigated in part III.

Aggregated search is an emerging paradigm and its complexity makes it different from the conventional search paradigm. However, despite the complexity, there have been a limited number of studies looking at the various aspects of aggregated search such as result presentation and search intent, on users' search behavior.

Aggregated search can be seen as an extension of federated and meta-search. However, the differences in the "typeÕÕ of result obtained from different verticals are more distinct in case of aggregated search. That is, aggregated search deals with more heterogeneous results when compared to earlier approaches like; federated search, distributed search and metasearch. Therefore, aggregating results of different types (image, news, maps, videos, etc) makes aggregated search more complex. Furthermore, the diversity in media and document content makes result organization in aggregated search more challenging.

This chapter is structured as follows. The effectiveness of an aggregated interface in task completion is discussed in section 7.1. Section 7.3 discusses the findings of how different interface related factors affect users' click-behavior. The concept of relevance within the context of aggregated search is discussed in section 7.2. Section 7.4 sums up the findings of log analysis, where dynamics of information need with respect to

different sources are discussed. Finally, section 7.5 highlights the design issues existing with respect to an aggregated interface, and suggests preliminary design guidelines.

7.1 Effectiveness of an Aggregated Search Interface

Aim of the first research question of this thesis was to investigate the effectiveness of an aggregated interface in information access and task completion. As discussed in the introduction chapter (Section 1.1), there are evidences suggesting the "need" for an aggregated search, and that more and more users are using this new search approach. However, whether aggregating results from multiple sources facilitates in task completion was not known.

In order to determine the effectiveness of an aggregated search interface, a user study was carried out (Chapter 3) where an aggregated interface was compared to the conventional tabbed search interface. The overall hypothesis of this study was that an aggregated result presentation can facilitate informational search tasks by offering diversified search results.

Informational search tasks can also be described as information gathering task, where the goal of the information gathering tasks involve the collection of information for which there is no one specific answer Liu et al. [2010]. Therefore while investigating effectiveness of an aggregated search interface, it was hypothesized that, search result aggregation was most useful for supporting informational queries.

The findings from the study suggested that for informational queries, presenting results aggregated from various sources (image, video, news, etc) help in improving information access when compared to conventional search interface. That is, when information need behind a search requires information to be collected from more than one source, result aggregation approach can prove to be effective. Furthermore, it was observed that the participants were able to collect more diverse results when using an aggregated search interface, as compared to the conventional search interface. Therefore, aggregated search interface can be seen as an *"effective"* means to gather information.

Furthermore, most of the participants were able to complete the tasks in the specified time. Over the decades, conventional search engines have proved to be efficient in helping their users to complete their respective search tasks Ali et al. [2009]. During the study, the duration of the task and the resources (i.e., results and number of results) for both the interfaces (aggregated and conventional) was kept similar. Using the given resources, participants were able to complete the tasks efficiently, with conventional as

well as aggregated interface. Although with aggregated interface, gathering information was much easier and faster.

Finally, the study expected the aggregated interface to have a better assessment than the conventional interface. The results showed that the participants tended to find the aggregated interface easier to find relevant information to complete a task. That is, participants had a positive attitude towards the aggregated interface.

To conclude, overall these results suggest that an aggregated search interface is effective for completing informational search tasks. **These observations provided evidence to support the first research question**. Next section will discuss the dimensionality of relevance within the context of aggregated search.

7.2 Concept of Relevance in Aggregated Search

The aim of the second research question was to understand the concept of relevance within the context of aggregated search. The overall aim of this research question was two-folds: (1) To understand, which non-topical characteristics contributes to the relevance of results within the context of an aggregated search; and (2) whether the suggested characteristic of relevance (source-orientation in this work) can be of different degrees.

"Relevance" is considered a key element in the IR and is often an important criterion in measuring the effectiveness of an IR system. Therefore, estimating relevance is critical in IR systems. Before estimating the relevance of the fetched results, it is first important to understand the characteristics which contributes in estimating the relevance of the fetched results to the given search task.

It is known that different interpretations of relevance exist (as discussed in Section 2.1, Chapter 2), and that there is no universal definition or understanding of the concept relevance. However, the relevance is often broadly classified into two main categories: system-based relevance – which lies in its pragmatism for systems applications and evaluation; and user-based relevance – where relevance is seen as the manifestations of topical, cognitive and situational relevance. While, some use genre information to determine the relevance. Based on Mizzaro [1998]'s concepts of dimensions of relevance, estimating the relevance of the source to the search task was proposed in this thesis. For this purpose, the notion of source-orientation was introduced which adds another element (source) to the relevance space in the context of an aggregated search.

The research presented in this thesis investigated the relevance of the source to the search task, and its effect on users' searching behavior. To capture this aspect, this thesis introduced the notion of the ***"source orientation"*** of an information need, which, refers to the degree to which documents *from* a specific source would be relevant for the task completion.

Results from the study suggested that the effect of the source-orientation was a significant factor in an aggregated search scenario. Users' click-behavior was equally effected in both the tiled and the vertical design with respect to source-orientation of the task at hand. Furthermore, results also suggested that this trend is slightly more consistent in the tiled design than the vertical design. However, in terms of the frequency, the results were comparable in both form of aggregation. Taken together, the study suggests that the effect of the source-orientation is an important factor to be investigated in aggregated search research.

The second goal of the research question was to see if source-orientation can be of different degree. For this purpose, three levels of source-orientation: high, medium and low, depending on the degree of orientation were also investigated. Degree of relevance refers to the rating and indication of the value of relevance of a given assessed information object Borlund [2003]. For instance, one request could be *highly* video oriented (e.g., learning salsa dance), where some visual learning is required. While, some request could have *medium* orientation towards video results (e.g., baking a pancake), where video results may or may not be required. As some users may prefer to see a video showing how bake a pancake, while some may prefer to read an online recipe.

Results from the study presented in this thesis showed that the click-through frequency decreases as the level of source-orientation weakens. Thus suggesting that participants clicked the non-web results more frequently when an information need has a strong orientation towards a particular information source type. The effect of the levels of source-orientation was found to be equally significant in both the aggregation designs: tiled and vertical.

From system perspective, determining source-orientation would require estimating the relevance of the source to the given query (terms). Since aggregated search involves selection and aggregation of different sources, therefore, it is important to estimate the relevance of the source to the given information need. While from the user perspective, source-orientation can be considered to fall into category of situational relevance. Situational relevance expresses the relationship between the user's perception of usefulness of a retrieved information object, and a specific work task situation Borlund [2003]. Therefore, the degree of the source-orientation may depend on the task interpretation.

For instance, one request could have 80% image and 20% web, for example, in the case of two sources. Here, the orientation of the search request is stronger towards results from an image source.

Therefore on the design level, devising a means of capturing a searcher's intent about the source is an important problem to tackle. As this would help suitable selection of sources for aggregation. Furthermore, capturing the degree of source-orientation is also important, as this information can help the designers of an aggregated interface in deciding the position of the non-web results. For instance, when the task at hand is highly image-oriented, it might be useful to provide image results on the top, since the users would be looking for image results.

These findings from the study showed that that source-orientation is a key factor to be measured while estimating the relevance of the result. Thus **providing answer to the second research question addressed in this thesis**. Next, how different result presentation aspects affected users' click-behavior on an aggregated result page is discussed.

7.3 Interface Related Factors

The aim of the third research question addressed in this thesis was to understand how do the factors associated with an aggregated interface effect users' searching behavior. That is, how different aspects of an aggregated result page, for instance position of image results, relevance of image results for search task at hand, etc affect users' searching behavior. Therefore, in order to gain further understanding of the aggregated interface and its impact, this thesis aimed to investigate the effect of aggregated result presentation approach. For this purpose, factors affecting users click-through behavior on aggregated search interfaces (tiled and vertical) were studied (Chapter 4).

Due to the addition of non-web results (e.g., image, maps, video) on an aggregate result page, there arises two main questions: *(1) Where to place these additional results (position)?* and, *(2) When to provide these additional results (source-orientation)?* However, before answering these question, it important to first verify if these questions are important to be investigated. That is, whether these factors affect users' searching behavior and hence, finding an answer to these question is required. Therefore, in addition to estimating suitable position and relevance of these sources, it is important to investigate if these factors have any impact of users' searching behavior. In order to achieve this, factors as position and source-orientation on two different modes of aggregation were studies, tiled and vertical. These two modes of aggregation can be found in most existing aggregated search interfaces on the web.

Several interesting observations were made from the study, which provides directions for follow-up research for future work and can also have several design implications. The overall finding of this study was that different factor effects users' click-behavior differently, depending on the "type" of aggregation approach followed. For instance, it was found that participants' click- behavior was significantly affected by the position of augmented elements in the vertical design, but not in the tiled like design. Therefore, it might be possible to observe difference in the way users would interact with the search results, depending of the aggregation approach followed.

Another highlight of this study was that users' click-behavior on aggregated interface differs across "types" of results. At first glance, this finding may sound obvious and similar to some previous research findings (e.g., Halvey et al. [2009]; André et al. [2009]). Where, difference in users' searching behavior with respect to different types of search results were suggested. However, these studies focused primarily on **single** source search, for instance, video search, image search, web search and so on. The comparison was then made among these individual search conditions. Aggregated search presents results from different sources on one page, which creates a different search scenario. Hence, re-investigating users' search behavior with respect to an aggregated search interface is required. As this would ensure how consistent (or different) the findings of an aggregated search interface are to the previous search interfaces.

When put together, the findings from the study also suggested the difference in search behavior with respect to different types of results. However, similarities among some results were also observed during the study. For instance, although video results received different click-behavior when compared to image and news results, similar click-behavior pattern was observed between image and news results. Therefore, designers of aggregated interface should take into account the similarities and differences of type of result effect when presenting then on the aggregated interface. As this might help in better generalization of the retrieval models or interface designs for similar (in terms of effect) set of results.

One of the goals of the work presented in this thesis was to see if aggregated search was different (if at all) from the conventional search; and if there is a need to further explore this new paradigm. That is, if aggregated search is just a replication of conventional search and affects users' information seeking process similar to conventional search paradigm, and if the findings from conventional search also imply within aggregated search.

For instance, it was already known that in conventional search interface, users' clicks are biased towards top-ranking results (For example, Joachims et al. [2005]; Agichtein

and Zheng [2006]; Guan and Cutrell [2007]; Keane et al. [2008]). The results from the study showed that, users' clicks continued to be biased toward top ranking results in vertical aggregation, however, no such effect was observed in the case of tiled aggregation. Furthermore, in aggregated search source-orientation is another factor, which needs to be incorporated within the retrieval models (as already discussed in Section 7.2).

The findings from the studies suggest that, aggregated search, although an extension of conventional search engines, differs in many aspects. Therefore, further research is needed to investigate this new paradigm. Furthermore, factors like position and source-orientation effect users' searching behavior significantly, therefore estimating position of non-web results (in vertical design), and determining orientation of task at hand are important. **These observations provided answer to the third research question**. In the following section, users' searching behavior with respect different sources is discussed.

7.4 Dynamics and Patterns of Information Need

Understanding the users' search behavior is one of the important goals in web search, and information on how users interact and view search results helps improving the information seeking process (e.g., White and Morris [2007b]; White and Drucker [2007]). Furthermore, having an insight on searching behavior of users has proved to be useful in building better information retrieval systems (e.g., Halvey and Jose [2009]; Agichtein and Zheng [2006]; Agichtein et al. [2006]; Bilenko and White [2008]).

For aggregated search scenario, understanding users' searching behavior with respect to different sources is important; as it requires aggregation of results from multiple sources. In addition, as already discussed in previous section, results from the user studies (chapter 4) showed that source-orientation is an important factor in aggregated search. Therefore investigating users' information need oriented towards different sources is required. As this would help in the design of the aggregated search interfaces.

For instance, knowledge of how often users' have information need that requires results from different sources (and which sources) can help in selecting suitable sources for aggregation. Furthermore, information on how users' interact with search results when their information need is oriented towards different sources, or combination of sources, can help in designing better aggregated search interfaces. Therefore, the aim of the fourth research question of this thesis was to investigate the searching behavior when

the associated information need is oriented towards different sources. In addition, to see if there exists any pattern in searching behavior for such information needs.

In order to achieve this, search logs were analyzed to understand users' information need with respect to different sources (Chapter 6). In particular, the information needs behind usersÕ query oriented towards different source were investigated. The interest was to understand the dynamics of such searching behavior. For this purpose, information needs with respect to six sources were looked at, namely, image, video, map, news, blog and Wikipedia. All other categories of source were viewed as standard "web", i.e. the typical web search result.

The analyses performed during this work were different from some of the previous log analyses (For example, Joachims et al. [2005]; White and Morris [2007b]; Keane et al. [2008]) because, earlier analyses did not take into account of the underlying source-orientations. Furthermore, source-orientation is peculiar to aggregated search scenario. Hence investigating users' searching behavior with respect to source-orientation is important.

The goal of the research question to further look for patterns within searching behavior was inspired by the concept of **Pattern Language**, which was first introduced by Alexander et al. [1977]. He described a pattern language as a precise way of describing someone's experience. Using which he suggested new method of architecture and planning, by studying how users interacted with physical spaces. Later, the concept of the pattern language was used by many researchers in HCI for interaction design (e.g., Griffiths et al. [2000]; Lombardi [2000]; Wellhausen [2005]; Pauwels et al. [2010]). The term interaction design pattern was used to define design patterns in the HCI field because they state solutions in terms of perceived interaction behavior of an interface Pauwels et al. [2010]. Therefore, this thesis also aimed to investigate patterns in searching behavior so that the knowledge could be utilized in providing suitable design guidelines for an aggregated search interface.

Overall findings from the log analysis showed that there exists specific user searching behavior pattern when the information need is oriented towards results from different sources. For instance, there seem to be certain order in which results with different source-orientations are accessed. For example, a search sessions where the user has an information need oriented towards two sources, lets say A and B. Then the users tended to access results from source B first and then from source A. The order of sequences in which the results are accessed are also specific to the combination of source-orientations. That is if the sources were A and C, users might access C first, then results from A, and then results from C again (C $\rightarrow$ A $\rightarrow$ C). Whereas, in previous case the

pattern was B $\rightarrow$ A.

For example, sessions where the information need was oriented towards video and web results, or map and web results; the users accessed map and video results earlier while web results were accessed later in the search session. While, in case of image and web combination, users tended to access web results before the image results. Such searching behavior information can be useful in determining the ranks of selected sources, i.e., video results first, followed by web results in case of this example. Furthermore, results also showed that users do not move back and forth between sources. Therefore, if known through implicit feedback[7-1] (click-through, skips, page reads, etc.) that users have accessed video results (as in example case from above), adding more video results in real time can be avoided, as it is possible that users would access web results and exit the session.

Some findings also showed that users often use source specific keywords (e.g., image, video) to explicitly specify their orientation. But, this[7-2] was the case when results from different sources were not aggregated on one page (conventional search engines). With conventional search, users had to use specific keywords to indicate the need for results from a source (for example, photo of Madonna, etc). Else, they accessed results separately from dedicated verticals like, image search, video search, etc. With introduction of aggregated search, it is likely that the usage of such keywords may reduce, since in most cases image and video results will be added without having users to specify explicitly. This further indicates that aggregated search may change the way users seek information on the web. Therefore, further investigation of user behavior within the context of aggregated search is required.

So far, the results from log analysis provided an insight on how users' search for results when the underlying information need has a source-orientation. Similar analyses of user behavior with respect to source-orientation can help building better aggregated search systems. Above findings provide some evidence on how users perform search when their information need requires results from multiple sources, and that there exists pattern in searching behavior for such information needs. These patterns can be useful from the design perspective for an aggregated search interface. Thus **providing evidence in supporting answer to the fourth research question addressed in this thesis**. In the following section, some of the potential design issues, which might be crucial from the aggregated interface perspective, are discussed.

[7-1] Implicit Relevance Feedback aims to improve the precision and recall of information retrieval by utilizing user actions to infer the relevance or non-relevance of documents Cramer et al. [2009].

[7-2] It should be noted that the log-data was obtained from a conventional search engine and not an aggregated search engine.

7.5 Design Issues and Guidelines

The overall aim of this thesis work was to explore aggregated search, and to understand aspects of result presentation within this paradigm. The expectation was that during this process, some of the design issues may be identified, which might be useful from the interface perspective. Therefore, through the last research question, this thesis attempted to identify issues that might arise within an aggregated search interface.

Through the user studies reported in Chapters 3 and 4, some interface related issues were identified. Furthermore, searching behavior patterns were also observed when investigating search sessions in Chapter 6 (also discussed in previous section). Based on these search patterns, some design guidelines are also suggested. Overall, following six key design issues were identified during this work, for which, possible design guidelines are also provided.

7.5.1 Position

As already stated in Section 7.3, indeed the position of non-web results seems to be critical only in vertical aggregation. While tiled aggregation reduces the biases of users' click-behavior towards certain positions on the result page. Therefore, if the aim is to minimize the position effect, tiled like design might be a suitable option. In addition, situations when there is limited means to identify the correct position for the results, tiled designs might be safer approach again.

Furthermore, within vertical design approach, if there is a need to decide among middle and bottom position, either position might be suitable. Since it was observed that there were no significant difference in users' click behavior when image or video results were placed at the middle or at the bottom of the result page. Hence, position continues to be an issue from the design perspective in vertical design, while in tiled aggregation, it might be possible for the designers to ignore this issue safely.

7.5.2 Attraction Power

Search results, which are visually attractive, might attract users' attention more when compared to textual results. Sutcliffe [2009] suggested that dynamic media (video, audio, etc) and visual salience (pictures, etc) attract attention. For instance, pictures attract attention in preference to text, and any change in the image further stimulates attention.

This leads us to consider the attraction power of augmented elements present on the aggregated result page. The attraction power of an augmented element is the power of the element, which attracts users' clicks, irrespective of the relevance of the result to the query. For example, if participants performing a highly image oriented task and they clicked the image placed at the bottom of the result, then the image's attraction power is positively strong. On the other hand, when participants are performing a low video oriented task, and they click on video results at the bottom of the result page, then the video's attraction power is negatively strong. The latter suggesting the situation similar to that of click inversion. Clarke et al. [2007] described click inversion as situation when a lower ranked document receives more clicks than a higher ranked one.

During the study, it was found that with an interaction occurred between position and source-orientation effects, while video showed a different pattern from image and news. Users clicked the video more frequently than the other two sources at the middle level of source-orientation. A similar pattern was found at the middle position in the results.

These results suggest that the videos have a moderate level of negative attraction power. Therefore, users tended to click the videos more frequently than they did on the images and news. This tells us that, careful consideration is required when video results are added on the vertical aggregated search interface. Else, users might click the videos even when they should have been to complete a task. This effect was found to be stronger on videos than images or news.

Furthermore, results from an eye-tracking study[7-3] showed that, due to the addition of pictures on the result page, the point of focus of viewing results seem to change, leading to a distinctly different experience. The "F" pattern Nielsen [2009] for scanning results in conventional web interface seem to change to "E" pattern in an aggregated interface, as seen in Figure 2.9.

However, an eye-tracking study[7-4] performed by Google on their Universal interface (aggregated interface) observed that, visual results (image, video) within a result page did not affect the order of scanning the results and does not disrupt the information seeking process of users.

Therefore, aggregating image or video results within the search result can be beneficial, but considering the impact of attraction power of these results is also required. For the designers of an aggregated interface, estimating the correct position for placing these element might be a challenge. Furthermore, providing these visual results only when required (source-orientation) might help in minimizing the negative attraction power.

[7-3] http://searchengineland.com/eye-tracking-on-universal-and-personalized-search-12 233
[7-4] http://googleblog.blogspot.com/2009/02/eye-tracking-studies-more-than-meets.html

7.5.3 Layout

Another implication of the findings was that the layout of aggregation was likely to affect users' selection of information sources. That is, depending on the order of the sources on the result page, users' result selection process might be effected. It was found that the order of the web, image, and news was the most common selection in the conventional system (Web → Image → News). While the web, image, and Wikipedia (Web → Image → Wikipedia) results were the popular selection in the tiled aggregated system (Chapter 3, Section 3.3). These observed sequence of accessing results were exactly in the same order in which the sources were presented on the respective interfaces. The tab on the top of the controlled interface listed the sources in the order of web → image → news → Wikipedia (Figure 3.1). The top three panels of the aggregated interface were the web → image → Wikipedia, with news results at the bottom of the page (Figure 3.2).

This suggests that users' browsing of information sources can be sequential, and their attention moves horizontally rather than scrolling down the result page vertically. This is in consistent with the previous findings, which showed that users read results in "F" pattern Nielsen [2009]. Hence scan results from left to right.

The finding further suggests that an aggregated search interface might be able to offer an effective support by optimizing the order of information sources for different tasks or queries. Although, no position effect was found in case of tiled aggregation, and irrespective of the position of the source panels, the click-through rates were similar for all positions in a tiled design. However, it is possible that the sequence of accessing different sources may be influenced by the page layout, and hence may differ. In the tiled design, it is possible to observe similar click-though rates at different positions. Nevertheless, the order in which the results are accessed, might change due to the position. Therefore from the design perspective, impact of different layouts for aggregating sources should also be taken into account.

7.5.4 Textual Elements

While investigating the searching behavior of users with different source-orientations, it was found that, users do not seem to differentiate between textual search results from different sources (blog, Wikipedia, etc). The users' interaction did not change when they switched from a web result to a blog or Wikipedia result. Which otherwise was not the case when users switched from web page to an image or video or map results.

In other words, users do not visually differentiate between blog, Wikipedia, and web pages, in-spite of their difference in source and genre.

Studies like (e.g., Corritore et al. [2003]; Golbeck [2006]; Artz and Gil [2007]) have shown that, providing information about the source of the search results, helps users to trust the search result more. For instance, for news results, users might click more often on news result from BBC online news, when compared to news provided from other news sources. The brand[7-5] reputation of the source (Wikipedia or BBC in this case) are often considered by users when deciding to trust the content of the web page. The news result from BBC can be often be identified through the URL text[7-6] provided with the search results. This suggests that providing source knowledge helps users to identify their choice, and also provides confidence regarding the source of information.

In aggregated search, results from blog, Wikipedia, twitter are now often included within the search result page. Therefore, providing means to differentiate between these different sources might help users in differentiating and thus may help in making selections easier. That is, providing clear visual distinction between results from different sources. Using different font or color might be one possible way to provide this distinction, as this way users can easily distinguish a blog result from a Wikipedia results, and so on.

Although, interface design guidelines often stress on consistency of the interface, however studies reported by AlTaboli and Salvendy [2000], AlTaboli and Abou-Zeid [2007] and Finstad [2008] suggest that, a physically inconsistent web interface has no effect on users satisfactions. Furthermore, in a consistent interface, where many of the screen elements look very similar to each other, users may not be satisfied. However, in an inconsistent interface increased inconsistency may increase task variety, and thus users' satisfaction. Finstad [2008] stated that the performance for casual users is improved by superficially inconsistent interfaces.

Therefore, it is less likely that by changing font or textual colors of the blog, Wikipedia, twitter results on aggregated interface, will bring any major consistency issues on aggregated interface. Thus overall performance of users should not be affected by this. Hence from design perspective incorporating measures to provide distinction among textual results can help users. In addition, font or colors can be considered as one of the possible safer options in providing such distinction. Although, fonts or colors that are most appropriate should be further investigated in order to determine their right

[7-5] Brand: The reputation of a company independent of their website affects users' trust and thus willingness to do business with the website Golbeck [2006]

[7-6] as can be seen here, the name "bbc" occurs in the URL, http://www.bbc.co.uk/news/

selection.

7.5.5 Number of Sources

Another important design issue that might effect an aggregated interface is the number of sources that will be aggregated on it. Through log analysis, existences of frequent combinations of source-orientations in search sessions were investigated in Section 6.6 of Chapter 6. To this end, how often information need with respect to different sources co-occur within a same search session was computed.

The findings from the analysis suggest that users often have information need which requires results from two sources, and in few cases from three sources. Information needs looking for results from more than two or three sources are less likely. This is in accordance with the study reported in Arguello et al. [2009b] where, through manual assessments, only 30% of the queries were found to have more than one source-orientation.

Put together, these results suggest that, searchers might not prefer to see too many sources aggregated on a single search result page. Therefore for the designers of aggregated result page, selecting the appropriate number of sources is important. Providing results from many sources may cause information overload and might not be appreciated by the users. Information overload occurs when the information available exceeds the user's ability to process it. To manage information overload, a user is required to discriminate among useful, redundant, incorrect, and meaningless information Bergamaschi et al. [2010]. Therefore, for an aggregated result page, aggregating results from more than three sources should be avoided.

7.5.6 Combination of Sources

Another finding from the log analysis was that there exits co-occurrences of source-orientation within a search session. In other words, users' often have information need, which requires results from combination of sources. For instance, it is likely that users looking for image results may also be interested in video results.

Furthermore, it was also found that when there were two relevant sources, these were often web and any of the non-web one, e.g., image + web, news+web and so on. This confirms that aggregated search results page should continue to contain mostly web results, and when appropriate, images, videos, blogs, should be provided in addition. That is, web should continue to be the main source of information on an aggregated

interface. Designers of the aggregated interface devise means to estimate appropriate combination of sources to be displayed on the result page.

These are some of the key design issues that were identified during this work. Although, a formal evaluation would be required in order to investigate their actual effect. However, these results provide initial insight on some of the design challenges that exits within aggregated interface. Furthermore, from design perspective it might prove to be beneficial if these issues are taken into account. In addition, it should be noted that these issues and guidelines are specific to aggregated search interfaces; it is possible that many other very important design guidelines for other aspects of interface design may exist. **These design issues identified during the experiments provide evidence to support answer to the fifth research question addressed in this thesis**. In the next section, the overall discussions presented in this chapter are summarized.

7.6 Summary

To conclude, this chapter outlined the main findings and their implication of the work presented in this thesis. The chapter discussed the overall findings of the five research questions stated in the introduction chapter (Section 1.3). The answers obtained with regard to these research were described in this chapter. First, usefulness of an aggregated interface in informational search tasks was discussed. Next, how and which interface related factors affected users' click-behavior in an aggregated search interface were discussed. Furthermore, which non-topical characteristics of results that contributes to relevance within aggregated search was also described. Next, findings of searching behavior with respect to different information needs were also presented in this chapter. Finally, implications of design issues that may exists within aggregated search is discussed. The next chapter will present the conclusions drawn from the research presented in this thesis and will also discuss the avenues for future work.

Chapter 8

Conclusions and Future Directions

In order to facilitate information access, search engines are now providing access to diverse data in a unified manner, called **aggregated search**. An aggregated search interface is designed to aggregate retrieval results from different information sources (image, video, maps, etc) into a single result page. A traditional way of gathering relevant information from several *information sources* is to browse the search results of individual sources separately available in search engines; often referred as **vertical search**. In an aggregated search, users do not have to visit separate verticals to browse the search results to access a range of retrieved items. Therefore, aggregated search approach facilitates access to multiple information sources through a single interface.

This thesis is an exploration of this new area of aggregated search. It is a new field without much pre-existing work or knowledge available. In particular, this thesis topic is concerned with "understanding aggregated search", a very important topic for front-end technologies. To this end, a number of experiments to look into the effect of result presentation for aggregated search were performed. Furthermore, search logs were also looked at to understand user needs in the context of aggregated search.

Overall, this thesis is divided into four main parts. Part I introduced the concept of aggregated search, provided definition to terminologies associated with aggregated search, and outlined the main challenges and research questions. Part II investigated the aspects of result presentation for an aggregated interface. Understandings of users? information need with respect to different information sources were obtained through log analysis in part III. Finally, part IV discusses the main findings and their implications. This chapter of part IV, summarizes and concludes the main findings and contributions of this thesis.

8.1 Result Presentation in Aggregated Search

During part II, various aspects of aggregated interface were investigated. First, a comparison of an aggregated interface to a non-aggregated interface during information access for informational queries was preformed (Chapter 3). Next, comparison between different types of aggregated interfaces was performed during the second study in Chapter 4, where factors affecting click-through behavior are analyzed.

The first comparison provided empirical evidence to support that an aggregated presentation of information sources can increase the quantity and diversify of the retrieved items accessed to complete informational search tasks. Participants tended to find the aggregated presentation easier to access retrieved items and to find relevant information. Although these positive effects were not strong enough to increase the amount of relevant information collected, it can be speculated that an intelligent way of organizing information sources is a key to achieve such a goal.

Despite its limitations, the presented study was the first step into a new and unexplored domain and a contribution to the exploration of the understanding of aggregated interface. Another interesting observation, which came up during the analysis of this study, was that the layout of an aggregated page may have implications on users' click behavior.

During the second comparison, a task-based user study to investigate the position and orientation of task towards a source on users' click-through behavior were performed. In particular, the effect of position and task orientation for three different augmented elements; image, news and video were investigated. The analysis of participants' click-data suggested that the click pattern of users change with respect to the orientation and with respect to position of information source on the result page.

Results with stronger orientation and at higher position received more clicks as compared to results with weaker orientation and at lower position. Some indications regarding the usefulness of results were also obtained. For instance, it was found that it might be more useful to provide image and news results in an aggregated result page when the search task (topic) is strongly image- or news- oriented. Users seem to ignore image and news results once their relevance decreases. However, click-data for the video results showed different user behavior as compared to news and image. Users seem to click video results as long their relevance was strong or medium, and start ignoring once the need for video results became low.

Overall, there were three main findings of this study; first, it was found that the effect of

the position of augmented search results (image, video, news, etc) was only significant in the vertical and not in the tiled like design. Second highlight of the study was that the participants' click-through behavior on videos was different from other sources, and finally third finding suggested that, in aggregated search paradigm, capturing a task's orientation towards particular sources is an important factor for further investigation and research.

8.2 Source-Orientation in Aggregated Search

As the type of content available on the web is becoming increasingly diverse, a particular challenge is to properly determine the sources of documents sought by a user, namely, the domain orientation (e.g. image, video) and/or the genre orientation (e.g. blog, Wikipedia). In this thesis, the Microsoft 2006 RFP click dataset to obtain an understanding of source-orientation with respect to domain and genre were analyzed. There were two main goals of this part of work: first, to find an effective way to identify the source-orientation of the user, give the query. Second, to obtain an insight to the searching behavior when the users' information need is oriented toward results from multiple sources. Three domains, namely, image, video, and map, and three genres, namely, news, blog and Wikipedia were looked at during this analysis. All other source were viewed as standard "web" orientation, i.e., the typical web search result.

The first step was to find an effective way to identify the source-orientation of the given query. In order to achieve this, a combination of rule-based and machine learning technique was used to classify the log entries. Using which it was possible to classify approximately 8% of the total click dataset to have orientation towards one of the six domain or genre. The remaining (92%) were identified to have web orientation. To gain an understanding of users searching behavior with respect to different information needs, five users' searching patterns were uncovered using these classified log data.

First, effect of rank order was analyzed, where it was hypothesized users clicks are biased towards the top positioned results, further reflecting on some of the past studies (e.g., Keane et al. [2008]), but in the context of aggregated search. Results obtained supported the hypothesis thus suggesting that, indeed, position (ranking) of search results continue to influence users searching behavior. Furthermore, the results also backed up the results from user studies described in previous Chapter 4 where, a clear position effect was observed on users' click-through rates.

Second, the frequent combination source-orientations were looked at, that is, which source-orientations co-occur within a session. It was observed that users' information

need is usually limited to results from two sources, and rarely seeks results from three or more sources. Furthermore, users often seek results from 'web' and any of non-web related results. Thus suggesting that web results are prime source of information sought by users, in combination with some image, video, or blogs results. This result supports the approach followed by all major search engines today, where search results contain mostly web results, and when appropriate, either images, videos, blogs, etc, are provided.

Third, the aim was to investigated if these intent combinations evolve according to some patterns. The results showed that there is a common sequence in all intents except with Wikipedia. Users in general tend to follow the same intent for a while and then switch to another intent. For most intents, users do not switch back and forth between intents. In other words, users with information need requiring results from different sources, first access results from source (say A) and then from the second source (say B). It is unlikely that users would access results from A and then B and then again from A.

Fourth, the relations between query re-formulation and change of intent were analyzed. It was observed that for video, news and map intents, more users used completely different queries, whereas, for blog and Wikipedia intents, more users used the same query, when they change their information need required results from different source. Further, most users use source specific terms (photo, newspapers, video, etc) to modify their query when they change their orientation towards different source.

Fifth, association between the category of the submitted query and the selected intent were also investigated. It was observed that there is an association between a user query (category) and the source-orientation. In addition, for certain query category, there is a dominance of specific sources, which indicates that not all sources are "relevant" to all information needs (here approximated as the queries submitted within a session).

During the log analysis, a **random model** methodology was introduced to derive meaningful statistics from log-data analysis. The random model provides a novel evaluation methodology that allows the comparison of results obtained from log-data, in the absence of any previously computed statistics or baseline.

8.3 Overall Findings

This thesis aimed to explore the aggregated search paradigm from the users' perspective. To participate in building theories and in understanding constraints, as well as providing insights into the design space. This thesis primarily focused on understand-

ing and describing the phenomena related to the users' search process in the context of the aggregated search. In building this understanding, the thesis focuses on the click-behavior, information need, source relevance, and dynamics of search intents. The understanding came partly from conducting user studies (Part II) and, from analyzing search engine log data (Part III). Overall this thesis contributed in answering five main research questions, findings of which were discussed in detail in Chapter 7. The main highlights of this research were:

- For informational queries, presenting results aggregated from various sources (image, video, news, etc) help in improving information access when compared to conventional search interface. Therefore, when information need behind a search requires information to be collected from more than one source, result aggregation approach can prove to be a useful approach.

- The factors that affect participants' click-through behavior differ between the vertical and tiled designs. Furthermore, the designers of aggregated search interfaces need to concentrate on different aspects over the aggregation styles. In a vertical aggregation, one needs an accurate estimation of the best position of augmented elements, whereas in a tiled aggregation approach, one needs an accurate selection of the type of augmented elements. The results also suggest that the relevance of source for the search task at hand is a key element for deciding on both the position and the type of the augmented elements.

- Findings from the studies showed that source-orientation is an important concept within the context of aggregated search. Therefore on the design level, devising a means of capturing a searcher's intent about the source is an important problem to tackle.

- Results from the log analysis helped to understand the information need behind a user query with respect to different source orientations. It was observed that users require results from at most two sources within a single session, and that there exists intent-related patterns. Furthermore, source specific terms (e.g. video, map) are added to or removed from the query whenever there is a switch of orientation. Finally, association between the category of the submitted query and the selected intent were also observed. Results showed that there is an association between a user query (category) and the orientation. In addition, for certain query category, there is a dominance of specific sources, which indicates that not all sources are "relevant" to all information needs.

- From the design perspective, it was observed that in certain situations selecting vertical over tiled design and vive-versa might help in restricting some of the factors which affect users' click-behavior. Findings from the studies also suggested that attraction power might be another important aspect that needs to be considered when designing an aggregated result page. Furthermore, the study also indicated that the layout of an aggregated result page may affect users' click behavior, therefore, depending on the position of the results, it may be possible to observe difference in users' click behavior at different positions. Finally, results suggested that, searchers might not prefer to see too many sources aggregated on a single search result page. Therefore for the designers of aggregated result page, selecting the appropriate number of sources is important.

To conclude, aim of this thesis was to explore the emerging aggregated search from users' perspective, since it is a very important for front-end technologies. An additional goal was to provide empirical evidence for influence of aggregated search on users searching behavior, and identify some of the key challenges of aggregated search. During this work several aspects of aggregated search were uncovered. For instance, it was found that aggregated search is useful for completion of informational tasks, and that source-orientation is one the key challenges in building an aggregated interface. Although, the studies reported in this thesis do not investigate every aspect of aggregated search, as exploring all possible factors and challenges is beyond the scope of this thesis. Further research is needed to gain a deeper understanding of the factors and their challenges. However, this thesis provides a foundations for future research in aggregated search and highlighted the potential research directions. Finally, this thesis contributes in rectifying the lack of scientific literature in aggregated search. In the following section, some of these opportunities for future works are outlined.

8.4 Future Directions

So far, this PhD research focused on "Aggregated Search: Result Presentation and User Behavior". During which, the aggregated search paradigm from the users' perspective was explored. The findings from the research so far lead us to formulate new research questions for future work and provide directions for research in future. Although aggregated search evolved from approached like federated search and metasearch, yet, there exits certain issues and challenges, which are peculiar to aggregated search paradigm only. Therefore, further investigations and research is required in order to overcome these difficulties and challenges Diaz et al. [2010].

8.4.1 Performance Measures

In this thesis various comparisons were made between: aggregated and non-aggregated interface, vertical and tiled like interface. The comparisons primarily aimed at result presentation aspects of an aggregated interface. For instance, usefulness in task completion and factors affecting click-through behavior, which were very important topic for front-end technologies. From the system point of view, measuring the performances of different aggregated interface would be useful. It would be interesting to further investigate how different types of aggregated perform during task completion. There could be several measure that can be investigated in order to compare the performances. For instance, measuring the diversity of the results provided by the system, comparing retrieval models, selection of appropriate sources, etc. Performance measurement of an aggregated system is a challenging work for future research.

8.4.2 Optimization

From future point of view, research on how to optimize an aggregated result page is required . That is given a query, what is the optimal number of sources to retrieve results from (two, three, four....or more)? What is the optimal number of results to be shown on the aggregated result page, 10 or more? From cognitive point of view, providing results from many sources may cause information load on users , or may distract users from their intended search task (Janssen and de Poot [2006]; Chen et al. [2009]). For instance, Chen et al. [2009] indicated that rich information leads to a perception of high information overload; and leads consumers to a worse subject state towards decision. Furthermore, Janssen and de Poot [2006] through a study showed that the extent to which people suffer from information overload is closely related to the strategies they use to deal with it. Therefore, achieving optimal aggregated result page without distracting or overloading user is challenging. Results from the log analysis of this thesis (chapter 6) and the study reported by Arguello et al. [2009b] suggested that users do not have information needs, which requires results from several different sources. From the log analysis it was observed that there are very few sessions with more than “two” different query intents. In other words, users seek information from at the most from two different sources. However, further testing of different combinations of sources is required in order to investigate how users interact given different combinations of sources; and what happens when the ‘number’ of sources are increased or decreased.

8.4.3 Source-Orientation

Findings from the user study described in chapter 4 suggested that, capturing a task's orientation towards particular sources is an important factor for further investigation and research. That is, in order to aggregate results from various sources, determining the relevance of the source is further required. Or, given an a query, whether to aggregate image or video or maps results with web results needs to be determined. Furthermore, it is also required to determine if a result from a specific source should be provided or not. For instance if for the query "nutrition", an image result would be required satisfy the information need or not?

In chapter 5, a combination of rule base and machine learning method was described. Using this method, it was possible to identify the orientation of the given query towards the result from a particular source. However, there were some limitations of this method. First, it made use of data generated by a search engine where general web results were predominant. This had consequences in the analyses, since this lowers the probability of observing more specific sources (e.g., image, video, etc.) and also forces users to use source specific terms in order to retrieve intent specific documents. Due to this low probability, more data (sessions) would have been necessary to obtain more reliable estimates of the different statistics. Therefore, further investigation is required on log data obtained from a current aggregated search engine to obtain more reliable statistics. Furthermore, only six different sources were tested during this research, there is a scope to extent this work to include more sources, as for example, now provided by search engine verticals.

8.4.4 Levels of Source-Orientation:

As already discussed above, the results from the user study reported in chapter 4 showed that source-orientation is a key factor in aggregated search. The study also tested different levels of source-orientations: high, medium and low. It was speculated that a highly source-oriented search task was one where results from the source would be highly desirable. A medium source-oriented search task was one where it was assumed that source results may or may not be required. Finally, a low source-orientated tasks were those where it was less likely that the results from the source would be needed.

The aim of manipulating levels of source-orientation was to see if different levels of effect users click behavior differently and, hence if determining levels of source-orientations is also important. The results showed that different levels of source-orientation

significantly affected users click behavior. Therefore, in aggregated search determining levels of source-orientations can also be an important factor. Determining levels of source-orientations can be useful in building better aggregated result pages. For instance, levels of source-orientation can be used for determining the position of augmented results. That is, when the user's information need has low orientation towards image result, then image results could be placed at the bottom. Furthermore, if between two sources one needs to be selected, in order to augment it on the aggregated result page, then selecting the one with higher orientation might be useful. For example, if for a given query, image an video results are found to be relevant, and the query has high image orientation and low medium orientation, then selecting image results might be beneficial.

8.4.5 Interface Design

The design of an aggregated search interface is more complex when compared to conventional search interfaces. Due to the addition of media data (e.g., image, video) and different document genre (e.g., blog, twitter) the design of an aggregated search interface becomes more challenging. There has been significant amount of research devoted to the effective design of conventional search interface, yet there is very limited understanding of the design for an aggregated interface. For instance, what is a suitable "surrogate" for representing different types of results (image, video, map, etc) on an aggregated interface is not known. What makes a well-designed, attractive, trustworthy and engaging aggregated search interface? In addition, how different design and visualization approach can be used to improve information access in an aggregated search interface is still unexplored. Furthermore, results from the studies indicated that layout of the aggregated result page might have an influence on users' searching behavior. Therefore testing suitable layout options for designing an aggregated interface is also required. The user studies carried during this work highlighted some of the design related issues that might be important from the interface design perspective. Some preliminary design guidelines are also provided in this thesis. However, these guidelines needs to be further investigated to see their effectiveness from an interface point of view. Furthermore, one of the findings of the study also indicated that the layout of an aggregated result page may affect users' click behavior. Therefore, there is potential to be exploited in investigating different layout options for an aggregated result page.

8.4.6 Evaluation

Due to complexity and heterogeneity of data involved, evaluation of different aspects of an aggregated search is difficult. Diversity due to the type of media and document data makes the comparison harder in aggregated search. For example, comparing a result page containing an image and web results, to a result page containing maps and web results is difficult due to the difference in the *type* of results involved; image and maps in this case. Designing user studies for an aggregated interface is hard because several factors need to be taken into account because of the diversity in the information types. With addition of each new source on the result page, a new variable measurement is required. Furthermore, incorporating standard recall/precision measures to evaluate an aggregated interface will also be difficult, as they need to be evaluated in two stages. First, the retrieval of relevant sources will be required, and then retrieving relevant results with the selected sources if further required. Furthermore, ranking of results on an aggregated result page also involves two stages. For instance, in an attempt to provide image and video results along with the web results; first it requires ranking of results within the source (for example, top 5 image results), and then ranking of sources on the result page is further required. That is, whether to provide image results before web and video results or to provide video results at the middle positions, etc. Therefore, measuring different aspects of aggregated search requires more complex evaluation strategies. Future research addressing evaluation for aggregated search should also aim to find ways to generalize the evaluation strategies. So that, the evaluation strategy measuring aggregation of image and web results could also be implied on aggregation of video and web results or map and news results, etc.

8.5 Summary

To conclude, this chapter summarized the overall work of this thesis and the major findings. This research has investigated and explored the aggregated search paradigm from the users' perspective. The ramifications of this work are notable and warrant further investigation. Several potential avenues for such investigation for future work were also outlined in this chapter.

Appendix A

Definitions

Since the aggregated search is a new paradigm, terminologies are still evolving. Here, the terminologies associated to the aggregated search is listed and defined.

- **Aggregated Search** – An aggregated search is a technique that aims to aggregate results from different physical sources on a single search result page. Aggregated search is also referred as *integrated search*[1-1][1-2], *universal search*[1-3] (name given by Google) and *horizontal search*[1-4]. In this thesis, this search approach is referred to as *"aggregated search"*.

- **Sources** – An source is a collection of a type of document or media, which is searched using a dedicated search engine (or a vertical or a server). In context of aggregated search, results are provided from different sources; for instance, from an image, video, maps, etc. In search engine market, these distinct sources (collection) of information are often referred as verticals. And, the dedicated vertical used to search within these verticals are referred as vertical search engines[1-5]. In this thesis, term *source or information source* is used to address these collections.

- **Source-Orientation** – Degree to which documents from a specific source would be relevant to complete the corresponding search task.

- **Vertical aggregation** – Is one where results from different sources are blended in together to form a single ranked list. Such type of aggregation is also referred to

[1-1] http://www.holisticsearch.co.uk/2010/09/28/integrated-search-the-future-of-seo/
[1-2] http://www.zelst.co.uk/
[1-3] http://searchengineland.com/google-20-google-universal-search-11232
[1-4] http://searchengineland.com/google-20-google-universal-search-11232#what
[1-5] http://en.wikipedia.org/wiki/Vertical_search

as merged or slotted aggregation, since results from image, video, etc are merged in with the web results or slotted between web results.

- **Tiled aggregation** – A tiled aggregation is one where results from different sources are provided in their dedicated panels. There is no blending or merging of results in such type of aggregation, and the distinction among sources is more evident.

Bibliography

The open directory project. http://www.dmoz.org/, 2010. 145

Eugene Agichtein and Zijian Zheng. Identifying "best bet" web search results by mining past user behavior. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 902–908, New York, NY, USA, 2006. ACM. ISBN 1-59593-339-5. doi: http://doi.acm.org/10.1145/1150402.1150526. 12, 44, 46, 48, 49, 76, 102, 130, 158, 159

Eugene Agichtein, Eric Brill, and Susan Dumais. Improving web search ranking by incorporating user behavior information. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 19–26, New York, NY, USA, 2006. ACM. ISBN 1-59593-369-7. doi: http://doi.acm.org/10.1145/1148170.1148177. 130, 159

Dan Albertson. Influences of users' familiarity with visual search topics on interactive video digital libraries. *J. Am. Soc. Inf. Sci. Technol.*, 61:2490–2502, December 2010. ISSN 1532-2882. doi: http://dx.doi.org/10.1002/asi.v61:12. URL http://dx.doi.org/10.1002/asi.v61:12. 131

Christopher Alexander, Sara Ishikawa, and Murray Silverstein. *A Pattern Language: Towns, Buildings, Construction*. Oxford University Press, New York, 1977. ISBN 0195019199. URL http://www.amazon.fr/exec/obidos/ASIN/0195019199/citeulike04-21. 160

Hilal Ali, Falk Scholer, James A. Thom, and Mingfang Wu. User interaction with novel web search interfaces. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, OZCHI '09, pages 301–304, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-854-4. doi: http://doi.acm.org/10.1145/1738826.1738880. URL http://doi.acm.org/10.1145/1738826.1738880. 11, 12, 154

N. Aloia, M. Matera, and F. Paternò. A semantics-based approach to designing presentations for multimedia database query results. In *AVI '96: Proceedings of the workshop on Advanced visual interfaces*, pages 91–100, New York, NY, USA, 1996. ACM. ISBN 0-89791-834-7. doi: http://doi.acm.org/10.1145/948449.948463. 36, 38

Ahamed AlTaboli and Mohammad Raafat Abou-Zeid. Effect of physical consistency of web interface design on users' performance and satisfaction. In *Proceedings of*

the 12th international conference on Human-computer interaction: applications and services, HCI'07, pages 849–858, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 978-3-540-73109-2. URL http://portal.acm.org/citation.cfm?id=1769451.1769550. 165

Ahamed AlTaboli and G Salvendy. Measuring consistency of web page design and its effects on performance and satisfaction. *Ergonomics*, 43:443–460, April 2000. 165

Ioannis Anagnostopoulos and Ilias Maglogiannis. Adapting user's browsing behavior and web evolution features for effective search in medical portals. In *SMAP '06: Proceedings of the First International Workshop on Semantic Media Adaptation and Personalization*, pages 37–42, Washington, DC, USA, 2006. IEEE Computer Society. ISBN 0-7695-2692-6. doi: http://dx.doi.org/10.1109/SMAP.2006.7. 130

Paul André, Edward Cutrell, Desney S. Tan, and Greg Smith. Designing novel image search interfaces by understanding unique characteristics and usage. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II*, INTERACT '09, pages 340–353, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 978-3-642-03657-6. doi: http://dx.doi.org/10.1007/978-3-642-03658-3_40. URL http://dx.doi.org/10.1007/978-3-642-03658-3_40. 158

Jaime Arguello, Jamie Callan, and Fernando Diaz. Classification-based resource selection. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, Hong Kong, China, November 2-6, 2009*, pages 1277–1286. ACM, 2009a. 75

Jaime Arguello, Fernando Diaz, Jamie Callan, and Jean-Francois Crespo. Sources of evidence for vertical selection. In *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 315–322, New York, NY, USA, 2009b. ACM. ISBN 978-1-60558-483-6. doi: http://doi.acm.org/10.1145/1571941.1571997. 75, 77, 122, 127, 138, 166, 174

Donovan Artz and Yolanda Gil. A survey of trust in computer science and the semantic web. *Web Semant.*, 5:58–71, June 2007. ISSN 1570-8268. doi: 10.1016/j.websem.2007.03.002. URL http://portal.acm.org/citation.cfm?id=1265608.1265746. 165

Azin Ashkan and Charles L.A. Clarke. Characterizing commercial intent. In *Proceedings of the 18th ACM conference on Information and knowledge management*, CIKM '09, pages 67–76, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-512-3. doi: http://doi.acm.org/10.1145/1645953.1645965. URL http://doi.acm.org/10.1145/1645953.1645965. 127

Javed A. Aslam and Mark Montague. Models for metasearch. In *SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 276–284, New York, NY, USA, 2001. ACM. ISBN 1-58113-331-6. doi: http://doi.acm.org/10.1145/383952.384007. 32

Javed A. Aslam, Virgiliu Pavlu, and Robert Savell. A unified model for metasearch and the efficient evaluation of retrieval systems via the hedge algorithm. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 393–394, New York, NY, USA, 2003a. ACM. ISBN 1-58113-646-3. doi: http://doi.acm.org/10.1145/860435.860517. 32

Javed A. Aslam, Virgiliu Pavlu, and Robert Savell. A unified model for metasearch and the efficient evaluation of retrieval systems via the hedge algorithm. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 393–394, New York, NY, USA, 2003b. ACM. ISBN 1-58113-646-3. doi: http://doi.acm.org/10.1145/860435.860517. 32

Thi Truong Avrahami, Lawrence Yau, Luo Si, and Jamie Callan. The fedlemur project: Federated search in the real world. *J. Am. Soc. Inf. Sci. Technol.*, 57(3):347–358, 2006. ISSN 1532-2882. doi: http://dx.doi.org/10.1002/asi.v57:3. 29, 30, 32

M. Beaulieu. Experiments with interfaces to support query expansion. *Journal of Documentation*, 53(1):8–19, 1997. URL http://www.emeraldinsight.com/Insight/viewPDF.jsp?Filename=html/Output/Published/EmeraldAbstractOnlyArticle/Pdf/2780530102.pdf. 69

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: Proceedings of the Fifth IEEE International Conference on Data Mining*, pages 42–49, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2278-5. doi: http://dx.doi.org/10.1109/ICDM.2005.80. 110

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. ISSN 1046-8188. doi: http://doi.acm.org/10.1145/1229179.1229183. URL http://doi.acm.org/10.1145/1229179.1229183. 127

Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida. Characterizing user behavior in online social networks. In *IMC '09: Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 49–62, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-771-4. doi: http://doi.acm.org/10.1145/1644893.1644900. 130

Sonia Bergamaschi, Francesco Guerra, and Barry Leiba. Guest editors' introduction: Information overload. *IEEE Internet Computing*, 14:10–13, November 2010. ISSN 1089-7801. doi: http://dx.doi.org/10.1109/MIC.2010.140. URL http://dx.doi.org/10.1109/MIC.2010.140. 166

Mikhail Bilenko and Ryen W. White. Mining the search trails of surfing crowds: identifying relevant websites from user activity. In *WWW '08: Proceeding of the 17th*

international conference on World Wide Web, pages 51–60, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-085-2. doi: http://doi.acm.org/10.1145/1367497.1367505. 12, 130, 159

Elizabeth Sugar Boese. Stereotyping the web: Genre classification of web documents. Master's thesis, Colorado State University, Fort Collins, Colorado, 2005. 11, 17, 21

Christine L. Borgman. End user behavior on an online information retrieval system: a computer monitoring study. In *SIGIR '83: Proceedings of the 6th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 162–176, New York, NY, USA, 1983. ACM. ISBN 0-89791-107-5. doi: http://doi.acm.org/10.1145/511793.511818. 43, 47

Christine L. Borgman. The study of user behavior on information retrieval systems. *SIGCUE Outlook*, 19(3-4):35–48, 1987. ISSN 0163-5735. doi: http://doi.acm.org/10.1145/964080.964086. 43

Pia Borlund. Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56:71–90, 2000. 62

Pia Borlund. The concept of relevance in ir. *J. Am. Soc. Inf. Sci. Technol.*, 54:913–925, August 2003. ISSN 1532-2882. doi: 10.1002/asi.10286. URL http://portal.acm.org/citation.cfm?id=941240.941242. 11, 17, 18, 156

Pavel Braslavski. Marrying relevance and genre rankings: An exploratory study. In Alexander Mehler, Serge Sharoff, Marina Santini, and Nancy Ide, editors, *Genres on the Web*, volume 42 of *Text, Speech and Language Technology*, pages 191–208. Springer Netherlands, 2011. ISBN 978-90-481-9178-9. 21

David J. Brenes, Daniel Gayo-Avello, and Kilian Pérez-González. Survey and evaluation of query intent detection methods. In *WSCD '09: Proceedings of the 2009 workshop on Web Search Click Data*, pages 1–7, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-434-8. doi: http://doi.acm.org/10.1145/1507509.1507510. 109

Andrei Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002a. ISSN 0163-5840. doi: http://doi.acm.org/10.1145/792550.792552. 58

Andrei Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002b. ISSN 0163-5840. doi: http://doi.acm.org/10.1145/792550.792552. 110

Andrei Z. Broder and Yoelle S. Maarek. Sigir'2006 workshop on faceted search. *SIGIR Forum*, 40(2), 2006. ISSN 0163-5840. doi: http://doi.acm.org/10.1145/1189702.1189705. 28

BŐnŐdicte Bucher, Paul Clough, Hideo Joho, Ross Purves, and Awase Khirni Syed. Geographic ir systems: requirements and evaluation. In *In Proceedings of the 22nd International Cartographic Conference*, 2005. 20

R. G. G. Cattell. An entity-based database user interface. In *SIGMOD '80: Proceedings of the 1980 ACM SIGMOD international conference on Management of data*, pages 144–150, New York, NY, USA, 1980. ACM. ISBN 0-89791-018-4. doi: http://doi.acm.org/10.1145/582250.582273. 38

Yu-Chen Chen, Rong-An Shang, and Chen-Yu Kao. The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment. *Electron. Commer. Rec. Appl.*, 8:48–58, January 2009. ISSN 1567-4223. doi: 10.1016/j.elerap.2008.09.001. URL http://portal.acm.org/citation.cfm?id=1508321.1508476. 174

Charles L. A. Clarke, Eugene Agichtein, Susan Dumais, and Ryen W. White. The influence of caption features on clickthrough patterns in web search. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pages 135–142, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-597-7. doi: http://doi.acm.org/10.1145/1277741.1277767. URL http://doi.acm.org/10.1145/1277741.1277767. 41, 163

Edward C. Clarkson, Shamkant B. Navathe, and James D. Foley. Generalized formal models for faceted user interfaces. In *JCDL '09: Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, pages 125–134, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-322-8. doi: http://doi.acm.org/10.1145/1555400.1555422. 28

C. W. Cleverdon, J. Mills, and M. Keen. Factors determining the performance of indexing systems. 1966. URL https://dspace.lib.cranfield.ac.uk/handle/1826/863. 18

Nick Conference Chair-Craswell, Rosie Conference Chair-Jones, Georges Conference Chair-Dupret, and Evelyne Conference Chair-Viegas. Wscd '09: Proceedings of the 2009 workshop on web search click data. New York, NY, USA, 2009. ACM. ISBN 978-1-60558-434-8. 110

William S. Cooper. A definition of relevance for information retrieval. *Information Storage and Retrieval*, 7(1):19–37, 1971. 17

Cynthia L. Corritore, Beverly Kracher, and Susan Wiedenbeck. On-line trust: concepts, evolving themes, a model. *Int. J. Hum.-Comput. Stud.*, 58:737–758, June 2003. ISSN 1071-5819. doi: 10.1016/S1071-5819(03)00041-7. URL http://portal.acm.org/citation.cfm?id=941177.941183. 165

Erica Cosijn and Peter Ingwersen. Dimensions of relevance. *Inf. Process. Manage.*, 36(4):533–550, 2000. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/S0306-4573(99)00072-2. 11, 17, 18, 19, 20

Mark Cramer, Mike Wertheim, and David Hardtke. Demonstration of improved search result relevancy using Real-Time implicit relevance feedback. In *Understanding the user - workshop in conjuction with SIGIR'09*, 2009. 161

Koby Crammer and Yoram Singer. On the learnability and design of output codes for multiclass problems. *Mach. Learn.*, 47(2-3):201–233, 2002. ISSN 0885-6125. doi: http://dx.doi.org/10.1023/A:1013637720281. 121

Nicholas Eric Craswell. *Methods for Distributed Information Retrieval*. PhD thesis, The Australian National University, May, 2000. vi, 29, 30, 31

W Bruce Croft. *Organising and Searching Large Files of Documents*. PhD thesis, University of Cambridge, October, 1978. 26

Kevin Crowston, Barbara H. Kwaśnik, and Joe Rubleske. *Problems in the use-centered development of a taxonomy of web genres*, volume 42 of *Series: Text, Speech and Language Technology (Series Editors: Ide, Nancy & Véronis, Jean)*, chapter 3. Springer, New York, 2010 2010. ISBN 978-90-481-9177-2. doi: 10.1007/978-90-481-9178-9. 22

Steven S. Curl, Lorne Olfman, and John W. Satzinger. An investigation of the roles of individual differences and user interface on database usability. *SIGMIS Database*, 29 (1):50–65, 1997. ISSN 0095-0033. doi: http://doi.acm.org/10.1145/506812.506821. 38

Douglass R. Cutting, David R. Karger, and Jan O. Pedersen. Constant interaction-time scatter/gather browsing of very large document collections. In *SIGIR '93: Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 126–134, New York, NY, USA, 1993. ACM. ISBN 0-89791-605-0. doi: http://doi.acm.org/10.1145/160688.160706. 26

Fernando A. Das Neves and Edward A. Fox. A study of user behavior in an immersive virtual environment for digital libraries. In *DL '00: Proceedings of the fifth ACM conference on Digital libraries*, pages 103–111, New York, NY, USA, 2000. ACM. ISBN 1-58113-231-X. doi: http://doi.acm.org/10.1145/336597.336648. 47

Lynne Davis. Designing a search user interface for a digital library. *J. Am. Soc. Inf. Sci. Technol.*, 57(6):788–791, 2006. ISSN 1532-2882. doi: http://dx.doi.org/10.1002/asi.v57:6. 39

Ronak Desai, Qi Yang, Zonghuan Wu, Weiyi Meng, and Clement Yu. Identifying redundant search engines in a very large scale metasearch engine context. In *WIDM '06: Proceedings of the 8th annual ACM international workshop on Web information and data management*, pages 51–58, New York, NY, USA, 2006. ACM. ISBN 1-59593-525-8. doi: http://doi.acm.org/10.1145/1183550.1183562. 32

Fernando Diaz. Integration of news content into web results. In *WSDM '09: Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 182–191, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-390-7. doi: http://doi.acm.org/10.1145/1498759.1498825. 75, 110

Fernando Diaz, Mounia Lalmas, and Milad Shokouhi. From federated to aggregated search. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 910–910, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0153-4. doi: http://doi.acm.org/10.1145/1835449.1835682. URL http://doi.acm.org/10.1145/1835449.1835682. 173

Daniel Dreilinger and Adele E. Howe. Experiences with selecting search engines using metasearch. *ACM Trans. Inf. Syst.*, 15(3):195–222, 1997. ISSN 1046-8188. doi: http://doi.acm.org/10.1145/256163.256164. 32

Sergio Duarte Torres, Djoerd Hiemstra, and Pavel Serdyukov. Query log analysis in the context of information retrieval for children. In *SIGIR '10: Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 847–848, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0153-4. doi: http://doi.acm.org/10.1145/1835449.1835646. 46

Yael Dubinsky, Tiziana Catarci, and Stephen Kimani. Using catalogue browsing for speech-based interface to a digital library. In *IASTED-HCI '07: Proceedings of the Second IASTED International Conference on Human Computer Interaction*, pages 130–135, Anaheim, CA, USA, 2007. ACTA Press. ISBN 978-0-88986-655-3. 39

Georges Dupret and Benjamin Piwowarski. A user behavior model for average precision and its generalization to graded judgments. In *SIGIR '10: Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 531–538, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0153-4. doi: http://doi.acm.org/10.1145/1835449.1835538. 47

Susan Dziadosz and Raman Chandrasekar. Do thumbnail previews help users make better relevance decisions about web search results? In *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 365–366, New York, NY, USA, 2002. ACM. ISBN 1-58113-561-0. doi: http://doi.acm.org/10.1145/564376.564446. 40

Andy Field and Graham J Hole. *How to Design and Report Experiments*. Sage Publications Ltd, 2003. 45, 59, 86

Kraig Finstad. Analogical problem solving in casual and experienced users: When interface consistency leads to inappropriate transfer. *Human-Computer Interaction*, 23:381–405, October 2008. doi: 10.1080/07370020802532734. 165

Richard H. Fowler, Wendy A. L. Fowler, and Bradley A. Wilson. Integrating query thesaurus, and documents through a common visual representation. In *SIGIR '91: Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 142–151, New York, NY, USA, 1991. ACM. ISBN 0-89791-448-1. doi: http://doi.acm.org/10.1145/122860.122874. 26

Luanne Freund. *The Janus Faced Scholar: A Festschrift in Honour of Peter Ingwersen*, chapter Genre Searching: A Pragmatic Approach to Information Retrieval. Det Informationsvidenskabelige Akademi (Royal School of Library and Information Science, Copenhagen); ISSI, 2010. 11, 17, 21

Luanne Freund and Justyna Berzowska. The goldilocks effect: task-centred assessments of e-government information. In *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47*, ASIS&T '10, pages 58:1–58:10, Silver Springs, MD, USA, 2010. American Society for Information Science. URL http://dl.acm.org/citation.cfm?id=1920331.1920415. 11, 17, 21

Luanne Freund, Charles L. A. Clarke, and Elaine G. Toms. Towards genre classification for ir in the workplace. In *Proceedings of the 1st international conference on Information interaction in context*, IIiX, pages 30–36, New York, NY, USA, 2006. ACM. ISBN 1-59593-482-0. doi: http://doi.acm.org/10.1145/1164820.1164829. URL http://doi.acm.org/10.1145/1164820.1164829. 21

Eric J. Glover, Steve Lawrence, William P. Birmingham, and C. Lee Giles. Architecture of a metasearch engine that supports user information needs. In *CIKM '99: Proceedings of the eighth international conference on Information and knowledge management*, pages 210–216, New York, NY, USA, 1999. ACM. ISBN 1-58113-146-1. doi: http://doi.acm.org/10.1145/319950.319980. 32

Jennifer Golbeck. Trust on the world wide web: a survey. *Found. Trends Web Sci.*, 1: 131–197, January 2006. ISSN 1555-077X. doi: 10.1561/1800000006. URL http://portal.acm.org/citation.cfm?id=1373448.1373449. 165

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. ISBN 1-58113-881-4. doi: http://doi.acm.org/10.1145/1008992.1009079. 131

Richard Griffiths, Lyn Pemberton, Jan Borchers, and Adam Stork. Pattern languages for interaction design: building momentum. In *CHI '00 extended abstracts on Human factors in computing systems*, CHI EA '00, pages 363–363, New York, NY, USA, 2000. ACM. ISBN 1-58113-248-4. doi: http://doi.acm.org/10.1145/633292.633510. URL http://doi.acm.org/10.1145/633292.633510. 160

Zhiwei Guan and Edward Cutrell. An eye tracking study of the effect of target rank on web search. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 417–420, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: http://doi.acm.org/10.1145/1240624.1240691. 76, 102, 133, 159

Qi Guo and Eugene Agichtein. Exploring mouse movements for inferring query intent. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference*

on Research and development in information retrieval, pages 707–708, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-164-4. doi: http://doi.acm.org/10.1145/1390334.1390462. 109

Martin Halvey and Joemon M. Jose. The role of expertise in aiding video search. In *CIVR '09: Proceeding of the ACM International Conference on Image and Video Retrieval*, pages 1–8, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-480-5. doi: http://doi.acm.org/10.1145/1646396.1646407. 12, 130, 159

Martin Halvey, David Vallet, David Hannah, and Joemon M. Jose. Vigor: a grouping oriented interface for search and retrieval in video libraries. In *JCDL '09: Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, pages 87–96, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-322-8. doi: http://doi.acm.org/10.1145/1555400.1555415. 103, 158

Kuan-Yu He, Yao-Sheng Chang, and Wen-Hsiang Lu. Improving identification of latent user goals through search-result snippet classification. In *WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 683–686, Washington, DC, USA, 2007. IEEE Computer Society. ISBN 0-7695-3026-5. doi: http://dx.doi.org/10.1109/WI.2007.137. 115

Marti A. Hearst. Uis for faceted navigation: Recent advances and remaining open problems. In *iIn the Workshop on Computer Interaction and Information Retrieval, HCIR 2008*, 2008. 29

Marti A. Hearst. *Search User Interfaces*. Cambridge University Press, New York, NY, USA, 2009. ISBN 0521113792, 9780521113793. 9, 36, 41

Marti A. Hearst and Jan O. Pedersen. Reexamining the cluster hypothesis: scatter/gather on retrieval results. In *SIGIR '96: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 76–84, New York, NY, USA, 1996. ACM. ISBN 0-89791-792-8. doi: http://doi.acm.org/10.1145/243199.243216. 26

Charles R. Hildreth. Accounting for users' inflated assessments of on-line catalogue search performance and usefulness: an experimental study. *Inf. Res.*, pages –1–1, 2001. 18

Frank Hopfgartner, Thierry Urruty, Pablo Bermejo Lopez, Robert Villa, and Joemon M. Jose. Simulated evaluation of faceted browsing based on feature selection. *Multimedia Tools Appl.*, 47(3):631–662, 2010. ISSN 1380-7501. doi: http://dx.doi.org/10.1007/s11042-009-0340-6. 29, 45

Jian Hu, Gang Wang, Fred Lochovsky, Jian-tao Sun, and Zheng Chen. Understanding user's query intent with wikipedia. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 471–480, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-487-4. doi: http://doi.acm.org/10.1145/1526709.1526773. 109

Meishan Hu, Aixin Sun, and Ee-Peng Lim. Event detection with common user interests. In *Proceeding of the 10th ACM workshop on Web information and data management*, WIDM '08, pages 1–8, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-260-3. doi: http://doi.acm.org/10.1145/1458502.1458504. URL http://doi.acm.org/10.1145/1458502.1458504. 131

Hilary Browne Hutchinson. Children's interface design for hierarchical search and browse. *SIGCAPH Comput. Phys. Handicap.*, (75):11–12, 2003. ISSN 0163-5727. doi: http://doi.acm.org/10.1145/976261.976265. 40

Peter Ingwersen. Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory. *Journal of Documentation*, 52:3–50, 1996. 18

Bernard J. Jansen. Search log analysis: What is it; what's been done; how to do it. *Library and Information Science Research*, 28(3):407–432, 2006. 46, 132

Bernard J Jansen and Amanda Spink. *An analysis of document viewing patterns of web search engine users*. Idea Publishing Group, Hershey PA, 2005. 9, 26, 46, 48, 131

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. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/j.ipm.2004.10.007. 26

Bernard J. Jansen, Amanda Spink, and Tefko Saracevic. Real life, real users, and real needs: a study and analysis of user queries on the web. *Inf. Process. Manage.*, 36(2): 207–227, 2000. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/S0306-4573(99)00056-4. 26, 143

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. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/j.ipm.2007.07.015. 7, 11, 59

Bernard J. Jansen, Gerry Campbell, and Matthew Gregg. Real time search user behavior. In *CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, pages 3961–3966, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-930-5. doi: http://doi.acm.org.libproxy.usc.edu/10.1145/1753846.1754086. 49

Ruud Janssen and Henk de Poot. Information overload: why some people seem to suffer more than others. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, NordiCHI '06, pages 397–400, New York, NY, USA, 2006. ACM. ISBN 1-59593-325-5. doi: http://doi.acm.org/10.1145/1182475.1182521. URL http://doi.acm.org/10.1145/1182475.1182521. 174

Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Syst.*, 20:422–446, October 2002. ISSN 1046-8188. doi: http://doi.acm.org/10.1145/582415.582418. URL http://doi.acm.org/10.1145/582415.582418.

Myoung Hoon Jeon, Dae Yol Na, Jung Hee Ahn, and Ji Young Hong. User segmentation & ui optimization through mobile phone log analysis. In *MobileHCI '08: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, pages 495–496, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-952-4. doi: http://doi.acm.org/10.1145/1409240.1409326. 44

Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. Accurately interpreting clickthrough data as implicit feedback. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 154–161, New York, NY, USA, 2005. ACM. ISBN 1-59593-034-5. doi: http://doi.acm.org/10.1145/1076034.1076063. 76, 102, 158, 160

Hideo Joho and Joemon M. Jose. Effectiveness of additional representations for the search result presentation on the web. *Inf. Process. Manage.*, 44(1):226–241, 2008. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/j.ipm.2007.02.004. 36, 41, 45

Yvonne Kammerer and Peter Gerjets. How the interface design influences users' spontaneous trustworthiness evaluations of web search results: comparing a list and a grid interface. In *ETRA '10: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 299–306, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-994-7. doi: http://doi.acm.org/10.1145/1743666.1743736. 40

Hyunjeong Kang. *The effect of interface design on e-commerce web sites for product information learning and search performance*. PhD thesis, Buffalo, NY, USA, 2005. Chair-Sanders, Lawrence G. 36, 40

Alexandros Karatzoglou, David Meyer, and Kurt Hornik. Support vector machines in r. *Journal of Statistical Software*, 15(9):1–28, 4 2006. ISSN 1548-7660. URL http://www.jstatsoft.org/v15/i09. 121

Mark T. Keane, Maeve O'Brien, and Barry Smyth. Are people biased in their use of search engines? *Commun. ACM*, 51(2):49–52, 2008. ISSN 0001-0782. doi: http://doi.acm.org/10.1145/1314215.1314224. 76, 102, 132, 133, 149, 159, 160, 170

Diane Kelly. Methods for evaluating interactive information retrieval systems with users. *Found. Trends Inf. Retr.*, 3(1—2):1–224, 2009. ISSN 1554-0669. doi: http://dx.doi.org/10.1561/1500000012. 45

Rinat Khoussainov and Nicholas Kushmerick. Specialisation dynamics in federated web search. In *WIDM '04: Proceedings of the 6th annual ACM international workshop on Web information and data management*, pages 112–119, New York, NY, USA, 2004. ACM. ISBN 1-58113-978-0. doi: http://doi.acm.org/10.1145/1031453.1031474. vii, 33

Mieczyslaw A. Klopotek, Slawomir T. Wierzchon, and Krzysztof Trojanowski. *Intelligent Information Processing and Web Mining: Proceedings of the International IIS: IIPWM'06 Conference held in Ustron, Poland, June 19-22, 2006 (Advances in*

Soft Computing). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 354033520X. 122

Harshit Kumar, Sungjoon Park, and Sanggil Kang. A personalized url re-ranking methodology using user's browsing behavior. In *KES-AMSTA'08: Proceedings of the 2nd KES International conference on Agent and multi-agent systems*, pages 212–221, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 3-540-78581-7, 978-3-540-78581-1. 130

Martin Kurth. The limits and limitations of transaction log analysis. *Library Hi Tech*, 11(2):98–104, 1993. 46

Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *WWW '10: Proceedings of the 19th international conference on World wide web*, pages 591–600, New York, NY, USA, 2010. ACM. ISBN 978-1-60558-799-8. doi: http://doi.acm.org/10.1145/1772690.1772751. 130

Barbara H. Kwaśnik and Kevin Crowston. A framework for creating a facetted classification for genres: Addressing issues of multidimensionality. In *Proceedings of the 37th Hawai'i International Conference on System Sciences (HICSS)*, Big Island, Hawai'i, January, 2004. doi: 10.1109/HICSS.2004.1265268. 20, 21

Barbara H. Kwaśnik and Kevin Crowston. Introduction to the special issue: Genres of digital documents. *Information, Technology & People*, 18(2):76–88, 2005. doi: 10.1108/09593840510601487. 22

Barbara H. Kwaśnik, You-Lee Chun, Kevin Crowston, J. D'Ignazio, and Joe Rubleske. Challenges in creating a taxonomy for genres of digital documents. In *Proceedings of the ISKO Conference*, Vienna, Austria, 2006. 22

Krista Lagus, Samuel Kaski, and Teuvo Kohonen. Mining massive document collections by the websom method. *Inf. Sci.*, 163(1-3):135–156, 2004. ISSN 0020-0255. doi: http://dx.doi.org/10.1016/j.ins.2003.03.017. 26

Mounia Lalmas and Anastasios Tombros. Evaluating xml retrieval effectiveness at inex. *SIGIR Forum*, 41(1):40–57, 2007. ISSN 0163-5840. doi: http://doi.acm.org/10.1145/1273221.1273225. 76, 103

Uichin Lee, Zhenyu Liu, and Junghoo Cho. Automatic identification of user goals in web search. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 391–400, New York, NY, USA, 2005. ACM. ISBN 1-59593-046-9. doi: http://doi.acm.org/10.1145/1060745.1060804. 109, 111

Anton Leouski and James Allan. Visual interactions with a multidimensional ranked list. In *SIGIR '98: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 353–354, New York, NY, USA, 1998. ACM. ISBN 1-58113-015-5. doi: http://doi.acm.org/10.1145/290941.291043. 26

Xiao Li, Ye-Yi Wang, and Alex Acero. Learning query intent from regularized click graphs. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, pages 339–346, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-164-4. doi: http://doi.acm.org/10.1145/1390334.1390393. URL http://doi.acm.org/10.1145/1390334.1390393. 127

Hsin-Hui Lin and Yi-Shun Wang. Predicting consumer intention to use mobile commerce in taiwan. In *ICMB '05: Proceedings of the International Conference on Mobile Business*, pages 406–412, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2367-6. doi: http://dx.doi.org/10.1109/ICMB.2005.80. 130

Jingjing Liu, Jacek Gwizdka, Chang Liu, and Nicholas J. Belkin. Predicting task difficulty for different task types. In *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47*, ASIS&T '10, pages 16:1–16:10, Silver Springs, MD, USA, 2010. American Society for Information Science. URL http://portal.acm.org/citation.cfm?id=1920331.1920355. 154

Yiqun Liu, Min Zhang, Shaoping Ma, and Liyun Ru. User behavior oriented web spam detection. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 1039–1040, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-085-2. doi: http://doi.acm.org/10.1145/1367497.1367645. 49

Victor Lombardi. Pattern languages for interaction design. Technical report, Razorfish Reports, 2000. 160

Malika Mahoui, Josette F. Jones, Derek Zollinger, and Kanitha Andersen. Leveraging user search behavior to design personalized browsing interfaces for healthcare web sites. In *HCD 09: Proceedings of the 1st International Conference on Human Centered Design*, pages 511–520, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 978-3-642-02805-2. doi: http://dx.doi.org/10.1007/978-3-642-02806-9_59. 130

Marcelo Maia, Jussara Almeida, and Virgílio Almeida. Identifying user behavior in online social networks. In *SocialNets '08: Proceedings of the 1st Workshop on Social Network Systems*, pages 1–6, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-124-8. doi: http://doi.acm.org/10.1145/1435497.1435498. 130

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008. ISBN 0521865719, 9780521865715. 121

M. M. Mantei and R. G. G. Cattell. A study of an entity-based database user interface. *SIGCHI Bull.*, 14(1):5–16, 1982. ISSN 0736-6906. doi: http://doi.acm.org/10.1145/1044183.1044184. 38

Gary Marchionini and Ryen. White. Effectiveness of additional representations for the search result presentation on the web. *International Journal of Human-Computer Interaction*, 23(3):205–237, 2007. 41

Karen Markey. Twenty-five years of end-user searching, part 2: Future research directions. *J. Am. Soc. Inf. Sci. Technol.*, 58(8):1123–1130, 2007a. ISSN 1532-2882. doi: http://dx.doi.org/10.1002/asi.v58:8. 45, 47, 48, 49

Karen Markey. Twenty-five years of end-user searching, part 1: Research findings. *J. Am. Soc. Inf. Sci. Technol.*, 58(8):1071–1081, 2007b. ISSN 1532-2882. doi: http://dx.doi.org/10.1002/asi.v58:8. 47

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. ISSN 1532-2882. doi: http://dx.doi.org/10.1002/asi.v56:9. 47

Richard Mateosian. Designing for discovery. *IEEE Micro*, 30(2):90–92, 2010. ISSN 0272-1732. doi: http://dx.doi.org/10.1109/MM.2010.33. 9, 12, 36, 40

Massimo Melucci and Ryen W. White. Utilizing a geometry of context for enhanced implicit feedback. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, CIKM '07, pages 273–282, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-803-9. doi: http://doi.acm.org/10.1145/1321440.1321480. URL http://doi.acm.org/10.1145/1321440.1321480. 12

Weiyi Meng, Zonghuan Wu, Clement Yu, and Zhuogang Li. A highly scalable and effective method for metasearch. *ACM Trans. Inf. Syst.*, 19(3):310–335, 2001. ISSN 1046-8188. doi: http://doi.acm.org/10.1145/502115.502120. 32

Weiyi Meng, Clement Yu, and King-Lup Liu. Building efficient and effective metasearch engines. *ACM Comput. Surv.*, 34(1):48–89, 2002. ISSN 0360-0300. doi: http://doi.acm.org/10.1145/505282.505284. vii, 31, 32, 34

Kenneth J. Mitchell, Jessie B. Kennedy, and Peter J. Barclay. A framework for user-interfaces to databases. In *AVI '96: Proceedings of the workshop on Advanced visual interfaces*, pages 81–90, New York, NY, USA, 1996. ACM. ISBN 0-89791-834-7. doi: http://doi.acm.org/10.1145/948449.948462. 38

Stefano Mizzaro. How many relevances in information retrieval? *Interacting with Computers*, 10(3):303–320, 1998. vi, 11, 17, 18, 19, 20, 23, 24, 155

Sylvain Mongy, Fatma Bouali, and Chabane Djeraba. Analyzing user's behavior on a video database. In *MDM '05: Proceedings of the 6th international workshop on Multimedia data mining*, pages 95–100, New York, NY, USA, 2005. ACM. ISBN 1-59593-216-X. doi: http://doi.acm.org/10.1145/1133890.1133902. 47

G. Craig Murray and Jaime Teevan. Query log analysis: social and technological challenges. *SIGIR Forum*, 41(2):112–120, 2007. ISSN 0163-5840. doi: http://doi.acm.org/10.1145/1328964.1328985. 46

Ivana Nakarada-Kordic and Brenda Lobb. Effect of perceived attractiveness of web interface design on visual search of web sites. In *CHINZ '05: Proceedings of the 6th*

ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction, pages 25–27, New York, NY, USA, 2005. ACM. ISBN 1-59593-036-1. doi: http://doi.acm.org/10.1145/1073943.1073949. 40

I. S. Nekrestyanov and N. V. Panteleeva. Text retrieval systems for the web. *Program. Comput. Softw.*, 28:207–225, July 2002. ISSN 0361-7688. doi: 10.1023/A:1016371117213. URL http://portal.acm.org/citation.cfm?id=597426.597506. 11

Cam-Tu Nguyen, Xuan-Hieu Phan, Susumu Horiguchi, Thu-Trang Nguyen, and Quang-Thuy Ha. Web search clustering and labeling with hidden topics. *ACM Transactions on Asian Language Information Processing (TALIP)*, 8(3):1–40, 2009. ISSN 1530-0226. doi: http://doi.acm.org/10.1145/1568292.1568295. 27

Jakob Nielsen. Usability engineering. In *The Computer Science and Engineering Handbook*, pages 1440–1460. Morgan Kaufmann, San Francisco, 1993. 9, 36, 41

Jakob Nielsen. *Eyetracking Web Usability*. Kara Pernice, 2009. 50, 163, 164

Stanislaw Osinski, Jerzy Stefanowski, and Dawid Weiss. Lingo: Search results clustering algorithm based on singular value decomposition. In *Intelligent Information Systems*, pages 359–368, 2004. 26

Marilyn Ostergren, Seung-yon Yu, and Efthimis N. Efthimiadis. The value of visual elements in web search. In *SIGIR '10: Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 867–868, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0153-4. doi: http://doi.acm.org/10.1145/1835449.1835656. 50

Hongseok Park. Relevance of science information: origins and dimensions of relevance and their implications to information retrieval. *Inf. Process. Manage.*, 33:339–352, May 1997. ISSN 0306-4573. doi: 10.1016/S0306-4573(96)00072-6. URL http://portal.acm.org/citation.cfm?id=254925.254934. 11, 17

Soyeon Park. Usability, user preferences, effectiveness, and user behaviors when searching individual and integrated full-text databases: implications for digital libraries. *J. Am. Soc. Inf. Sci.*, 51(5):456–468, 2000. ISSN 0002-8231. doi: http://dx.doi.org/10.1002/(SICI)1097-4571(2000)51:5<456::AID-ASI6>3.3.CO;2-F. 47

Stefan L. Pauwels, Christian Hübscher, Javier A. Bargas-Avila, and Klaus Opwis. Building an interaction design pattern language: A case study. *Comput. Hum. Behav.*, 26:452–463, May 2010. ISSN 0747-5632. doi: http://dx.doi.org/10.1016/j.chb.2009.12.004. URL http://dx.doi.org/10.1016/j.chb.2009.12.004. 160

Jan Polowinski. Widgets for faceted browsing. In *Proceedings of the Symposium on Human Interface 2009 on ConferenceUniversal Access in Human-Computer Interaction. Part I*, pages 601–610, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 978-3-642-02555-6. doi: http://dx.doi.org/10.1007/978-3-642-02556-3_68. 28

Jack Prins, Don McCormack, Di Michelson, and Karen Horrell. Product and process comparisons. http://www.itl.nist.gov/div898/handbook/, 2003. NIST/SEMATECH e-Handbook of Statistical Methods. 87

Ross Purves and Chris Jones, 2004. Workshop on geographic information retrieval, SIGIR. 103

FĞbio JosŐ Rauen. *Genre in a Changing World*, chapter Relevance and Genre: Theoretical and Conceptual Interfaces. The WAC Clearinghouse and Parlor Press, 2009. 21

C.J. Van Rijsbergen. *Information Retrieval, 2nd edition*. Butterworths, London, 1979. 17, 26

David Robins. Interactive information retrieval: Context and basic notions. *Informing Science Journal*, 3:57–62, 2000. 18

Daniel E. Rose. Reconciling information-seeking behavior with search user interfaces for the web. *J. Am. Soc. Inf. Sci. Technol.*, 57(6):797–799, 2006. ISSN 1532-2882. doi: http://dx.doi.org/10.1002/asi.v57:6. 131

Mark A. Rosso. User-based identification of web genres. *J. Am. Soc. Inf. Sci. Technol.*, 59:1053–1072, May 2008. ISSN 1532-2882. doi: 10.1002/asi.v59:7. URL http://dl.acm.org/citation.cfm?id=1364846.1364850. 20

D. Roussinov, K. Crowston, M. Nilan, B. Kwasnik, J. Cai, and X. Liu. Genre based navigation on the web. In *HICSS '01: Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 4*, page 4013, Washington, DC, USA, 2001. IEEE Computer Society. ISBN 0-7695-0981-9. 11, 17, 21

Tefko Saracevic. Saracevic, t. (1996). relevance reconsidered. information science: Integration in perspectives. *Proceedings of the Second Conference on Conceptions of Library and Information Science, Copenhagen, Denmark*, pages 201–218, 1996a.

Tefko Saracevic. Saracevic, t. (1996). relevance reconsidered. information science: Integration in perspectives. *Proceedings of the Second Conference on Conceptions of Library and Information Science, Copenhagen, Denmark*, pages 201–218, 1996b. 18, 19

Linda Schamber, Michael Eisenberg, and Michael S. Nilan. A re-examination of relevance: toward a dynamic, situational definition. *Inf. Process. Manage.*, 26:755–776, November 1990. ISSN 0306-4573. doi: 10.1016/0306-4573(90)90050-C. URL http://dl.acm.org/citation.cfm?id=95158.95168. 11, 17

S. Sendhilkumar and T. V. Geetha. User's search behavior graph for aiding personalized web search. In *PReMI'07: Proceedings of the 2nd international conference on Pattern recognition and machine intelligence*, pages 357–364, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 3-540-77045-3, 978-3-540-77045-9. 130

Ali Shiri. Metadata-enhanced visual interfaces to digital libraries. *J. Inf. Sci.*, 34(6):763–775, 2008. ISSN 0165-5515. doi: http://dx.doi.org/10.1177/0165551507087711. 39

Milad Shokouhi. Central-rank-based collection selection in uncooperative distributed information retrieval. In *ECIR'07: Proceedings of the 29th European conference on IR research*, pages 160–172, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 978-3-540-71494-1. 30

Milad Shokouhi, Mark Baillie, and Leif Azzopardi. Updating collection representations for federated search. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 511–518, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-597-7. doi: http://doi.acm.org/10.1145/1277741.1277829. 29, 30

Luo Si and Jamie Callan. Modeling search engine effectiveness for federated search. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 83–90, New York, NY, USA, 2005. ACM. ISBN 1-59593-034-5. doi: http://doi.acm.org/10.1145/1076034.1076051. 29, 30

Ahu Sieg, Bamshad Mobasher, and Robin Burke. Web search personalization with ontological user profiles. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 525–534, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-803-9. doi: http://doi.acm.org/10.1145/1321440.1321515. 130, 131

Yaxiao Song and Gary Marchionini. Effects of audio and visual surrogates for making sense of digital video. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 867–876, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9. doi: http://doi.acm.org/10.1145/1240624.1240755. 103

Amanda Spink, Bernard J. Jansen, Dietmar Wolfram, and Tefko Saracevic. From e-sex to e-commerce: Web search changes. *Computer*, 35(3):107–109, 2002. ISSN 0018-9162. doi: http://dx.doi.org/10.1109/2.989940. 9, 26, 46, 48, 131

Sofia Stamou and Lefteris Kozanidis. Impact of search results on user queries. In *WIDM '09: Proceeding of the eleventh international workshop on Web information and data management*, pages 7–10, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-808-7. doi: http://doi.acm.org/10.1145/1651587.1651591. 109

Nenad Stojanovic. Information-need driven query refinement. *Web Intelli. and Agent Sys.*, 3(3):155–169, 2005. ISSN 1570-1263. 109

Jonathan Stuart-Moore, Monte Evans, and Patricia Jacobs. Interface design for browsing faceted metadata. In *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, pages 349–349, New York, NY, USA, 2006. ACM. ISBN 1-59593-354-9. doi: http://doi.acm.org/10.1145/1141753.1141844. 28

Shanu Sushmita, Hideo Joho, and Mounia Lalmas. A task-based evaluation of an aggregated search interface. In *SPIRE '09: Proceedings of the 16th International Symposium on String Processing and Information Retrieval*, pages 322–333, Berlin, Heidelberg, 2009a. Springer-Verlag. ISBN 978-3-642-03783-2. doi: http://dx.doi.org/10.1007/978-3-642-03784-9_32. 57

Shanu Sushmita, Hideo Joho, Mounia Lalmas, and Joemon M. Jose. Understanding domain "relevance" in web search. In *WWW 2009 Workshop on Web Search Result Summarization and Presentation*, 2009b. 58, 75, 111

Shanu Sushmita, Hideo Joho, Mounia Lalmas, and Robert Villa. Factors affecting click through behavior in aggregated interface. In *The Conference on Information and Knowledge Management (CIKM)*, pages 322–333, New York, NY, USA, 2010a. ACM. ISBN 978-3-642-03783-2. doi: http://dx.doi.org/10.1007/978-3-642-03784-9_32. 72

Shanu Sushmita, Benjamin Piwowarski, and Mounia Lalmas. Dynamics of domains and genre intent. In *AIRS10; proceedings of The Sixth Asia Information Retrieval Societies Conference*, Secaucus, NJ, USA, 2010b. Springer-Verlag New York, Inc. doi: http://doi.acm.org/10.1145/1526709.1526773. 77, 114

Alistair Sutcliffe. *Designing for User Engagment: Aesthetic and Attractive User Interfaces*. Morgan & Claypool Publishers, 2009. ISBN 1608450236, 9781608450237. 162

Russell C. Swan and James Allan. Aspect windows, 3-d visualizations, and indirect comparisons of information retrieval systems. In *SIGIR '98: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 173–181, New York, NY, USA, 1998. ACM. ISBN 1-58113-015-5. doi: http://doi.acm.org/10.1145/290941.290987. 26

Saravadee Sae Tan, Gan Keng Hoon, Tang Enya Kong, Cheong Sook Lin, Chan Siew Lin, and Foo Wen Ying. Mice: Aggregating and classifying meta search results into self-customized categories. In *WI '04: Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 771–772, Washington, DC, USA, 2004. IEEE Computer Society. ISBN 0-7695-2100-2. doi: http://dx.doi.org/10.1109/WI.2004.96. 26

Muh-Chyun Tang. Browsing and searching in a faceted information space: A naturalistic study of pubmed users' interaction with a display tool. *J. Am. Soc. Inf. Sci. Technol.*, 58(13):1998–2006, 2007. ISSN 1532-2882. doi: http://dx.doi.org/10.1002/asi.v58:13. 29

Paul Thomas and David Hawking. Server selection methods in personal metasearch: a comparative empirical study. *Inf. Retr.*, 12(5):581–604, 2009. ISSN 1386-4564. doi: http://dx.doi.org/10.1007/s10791-009-9094-z. 32

Paul Thomas, Katherine Noack, and Cecile Paris. Evaluating interfaces for government metasearch. In *IIiX '10: Proceeding of the third symposium on Information*

interaction in context, pages 65–74, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0247-0. doi: http://doi.acm.org/10.1145/1840784.1840796. 32, 45

Misha W. Vaughan, Helmut Degen, Marc Resnick, and Peter Gremett. Best practices and future visions for search user interfaces: a workshop. In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*, pages 1052–1053, New York, NY, USA, 2003. ACM. ISBN 1-58113-637-4. doi: http://doi.acm.org/10.1145/765891.766144. 40

Olga Vechtomova and Hao Zhang. Articulating complex information needs using query templates. *J. Inf. Sci.*, 35(4):439–452, 2009. ISSN 0165-5515. doi: http://dx.doi.org/10.1177/0165551508101863. 109

Robert Villa, Iván Cantador, Hideo Joho, and Joemon M. Jose. An aspectual interface for supporting complex search tasks. In *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 379–386, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-483-6. doi: http://doi.acm.org/10.1145/1571941.1572007. 40, 45

Robert Villa, Martin Halvey, Hideo Joho, David Hannah, and Joemon M. Jose. Can an intermediary collection help users search image databases without annotations? In *Proceedings of the 10th annual joint conference on Digital libraries*, JCDL '10, pages 303–312, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0085-8. doi: http://doi.acm.org/10.1145/1816123.1816169. URL http://doi.acm.org/10.1145/1816123.1816169. 131

Agnès Voisard. Designing and integrating user interfaces of geographic database applications. In *AVI '94: Proceedings of the workshop on Advanced visual interfaces*, pages 133–142, New York, NY, USA, 1994. ACM. ISBN 0-89791-733-2. doi: http://doi.acm.org/10.1145/192309.192339. 38

Stefanos Vrochidis, Ioannis Kompatsiaris, and Ioannis Patras. Optimizing visual search with implicit user feedback in interactive video retrieval. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, CIVR '10, pages 274–281, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0117-6. doi: http://doi.acm.org/10.1145/1816041.1816082. URL http://doi.acm.org/10.1145/1816041.1816082. 131

Yitong Wang and Masaru Kitsuregawa. Evaluating contents-link coupled web page clustering for web search results. In *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*, pages 499–506, New York, NY, USA, 2002. ACM. ISBN 1-58113-492-4. doi: http://doi.acm.org/10.1145/584792.584875. 26

Michelle Q. Wang Baldonado. A user-centered interface for information exploration in a hetergeneous digital library. *J. Am. Soc. Inf. Sci.*, 51(3):297–310, 2000. ISSN 0002-8231. doi: http://dx.doi.org/10.1002/(SICI)1097-4571(2000)51:3<297::AID-ASI8>3.0.CO;2-N. 39

Tim Wellhausen. User interfaces for searching - a pattern language. In *EuroPLoP*, pages 219–236, 2005. 160

Ryen W. White and Steven M. Drucker. Investigating behavioral variability in web search. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 21–30, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-654-7. doi: http://doi.acm.org/10.1145/1242572.1242576. 12, 130, 159

Ryen W. White and Dan Morris. Investigating the querying and browsing behavior of advanced search engine users. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 255–262, New York, NY, USA, 2007a. ACM. ISBN 978-1-59593-597-7. doi: http://doi.acm.org/10.1145/1277741.1277787. 46, 48

Ryen W. White and Dan Morris. Investigating the querying and browsing behavior of advanced search engine users. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 255–262, New York, NY, USA, 2007b. ACM. ISBN 978-1-59593-597-7. doi: http://doi.acm.org/10.1145/1277741.1277787. 12, 130, 159, 160

Barbara M. Wildemuth. Evidence-based practice in search interface design. *J. Am. Soc. Inf. Sci. Technol.*, 57(6):825–828, 2006. ISSN 1532-2882. doi: http://dx.doi.org/10.1002/asi.v57:6. 40

Max L. Wilson, Bill Kules, m. c. schraefel, and Ben Shneiderman. From keyword search to exploration: Designing future search interfaces for the web. *Found. Trends Web Sci.*, 2(1):1–97, 2010. ISSN 1555-077X. doi: http://dx.doi.org/10.1561/1800000003. 41

Ian H. Witten. Customizing digital library interfaces with greenstone, 2003. 39

Zonghuan Wu, Weiyi Meng, Clement Yu, and Zhuogang Li. Towards a highly-scalable and effective metasearch engine. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 386–395, New York, NY, USA, 2001. ACM. ISBN 1-58113-348-0. doi: http://doi.acm.org/10.1145/371920.372093. 32

Ya Xu and David Mease. Evaluating web search using task completion time. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 676–677, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-483-6. doi: http://doi.acm.org/10.1145/1571941.1572073. URL http://doi.acm.org/10.1145/1571941.1572073. 11

Kwan Yi, Jamshid Beheshti, Charles Cole, John E. Leide, and Andrew Large. User search behavior of domain-specific information retrieval systems: An analysis of the query logs from psycinfo and abc-clio's historical abstracts-america: History and life: Research articles. *J. Am. Soc. Inf. Sci. Technol.*, 57(9):1208–1220, 2006. ISSN 1532-2882. doi: http://dx.doi.org/10.1002/asi.v57:9. 47

J. Yoon and S. H. Kim. A three-level user interface to multimedia digital libraries with relaxation and restriction. In *ADL '98: Proceedings of the Advances in Digital Libraries Conference*, page 206, Washington, DC, USA, 1998. IEEE Computer Society. ISBN 0-8186-8464-X. 39

Xiaojie Yuan, Zhicheng Dou, Lu Zhang, and Fang Liu. Automatic user goals identification based on anchor text and click-through data. *Wuhan University Journal of Natural Sciences*, 13(4):495–500, 2008. ISSN 1007-1202. 111

Oren Zamir. Visualization of search results in document retrieval systems. Technical report, General Examination Report, 1998. 26

Oren Zamir and Oren Etzioni. Grouper: a dynamic clustering interface to web search results. In *WWW '99: Proceedings of the eighth international conference on World Wide Web*, pages 1361–1374, New York, NY, USA, 1999. Elsevier North-Holland, Inc. doi: http://dx.doi.org/10.1016/S1389-1286(99)00054-7. 26

Ángel F. Zazo, Carlos G. Figuerola, José L. Alonso Berrocal, and Emilio Rodriguez. Reformulation of queries using similarity thesauri. *Inf. Process. Manage.*, 41(5): 1163–1173, 2005. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/j.ipm.2004.05.006. 109

Hua-Jun Zeng, Qi-Cai He, Zheng Chen, Wei-Ying Ma, and Jinwen Ma. Learning to cluster web search results. In *SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 210–217, New York, NY, USA, 2004. ACM. ISBN 1-58113-881-4. doi: http://doi.acm.org/10.1145/1008992.1009030. 26

Ying Zhang, Bernard J. Jansen, and Amanda Spink. Time series analysis of a web search engine transaction log. *Inf. Process. Manage.*, 45(2):230–245, 2009. ISSN 0306-4573. doi: http://dx.doi.org/10.1016/j.ipm.2008.07.003. 45