THE USE OF SOCIAL TAGS IN TEXT AND IMAGE SEARCHING ON THE WEB

by

Yong-Mi Kim

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in the University of Michigan
2011

Doctoral Committee:

Associate Professor Soo Young Rieh, Chair
Professor Karen Markey
Associate Professor Christopher L. Quintana
Assistant Professor Mark W. Newman
Associate Professor Diane Kelly, University of North Carolina at Chapel Hill

Dedication

To my father,
Heung Soo Kim

Acknowledgements

I would like to thank Soo Young Rieh for seeing me through the process of my doctoral studies with great patience and faith.

I would like to thank Yan Chen for the unique opportunities for collaboration and mentoring she provided.

The BlearyTheory lab group provided emotional support, advice, and valuable feedback in the early phases of this work. We've all managed to get out of here!

I would like to thank Matt "mathowie" Haughey, Paul "pb" Bausch for providing access to Ask MetaFilter, and Josh "cortex" Millard, for the MetaFilter infodumps. This research could not have been done without you guys.

I would like to thank my parents for supporting me through difficult times, including my decision to return to graduate school. I would like to thank my brothers for providing encouragement during my long doctoral journey.

I would like to thank the Rackham Graduate School for their generous financial support. This dissertation was partially funded by the Beta Phi Mu Eugene Garfield Doctoral Dissertation Fellowship.

Finally, I would like to thank my husband Art for generously building the system used in my experiments, all the last minute proofreading, and his patience and understanding as I completed the dissertation.

Table of Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

1.1 Background

Searching for information on the Web is now an everyday part of life, to the extent that "googling" has become a common verb. People search not just for text documents, but for music, videos, software, websites, and other types of resources as they emerge on the Web. People search the Web using general search engines such as Google, and also search within sites using site search engines. But despite such variety in online searching, the fundamental questions of information retrieval have not changed since the days of the Cranfield retrieval experiments in the 1960's. These two fundamental and interrelated questions are, how to find the information one wants, and how should the information retrieval system support this activity (Sparck Jones, 1981; Ingwersen & Järvelin, 2005; Ruthven, 2008; Xie, 2008). The focus of the first question is the searcher and their behavior, while the focus of the second question is the design and implementation of information retrieval systems, including algorithms and interfaces. The searcher's behavior or use of an information retrieval system will depend on features of the system, while ideally information retrieval systems should have designs and implementations informed by the search behavior of users.

Online searching, despite its outward simplicity – enter search terms in a search box, get search results back, then click on the items that look useful – is actually a complex, multi-stage process (Saracevic, 1997; Belkin, 2000). The appropriate search terms are not always obvious, especially for topic areas unfamiliar to the user. Using terms that are not specific enough can return an overwhelming number of search results, often several million on Google. The user must then judge which of these results are worth clicking on and reading. While reading the selected Web page the user may decide it does not contain useful information, despite the initial judgment based on the search results. The user may decide to redo their search, but is now faced with the problem of how to reformulate the search terms to get better results.

Social tagging is a relatively recent development on the Web, aimed at improving findability and discovery of information. Tags are descriptive terms people attach to Web content, either their own or other people's. In social tagging, both the tags and the objects tagged are publicly viewable online. Social tagging has been rapidly adopted across a variety of websites, ranging from popular commercial ones such as YouTube to academic ones such as that of the University of Michigan Library. The benefits cited for tagging are that tags make it easier to find or re-find information, and tags help organize information (Hammond, Hannay, Lund & Scott, 2005; Golder & Huberman, 2006; Guy & Tonkin, 2006; Macgregor & McCulloch, 2006).

Given the difficulties in current Web search for users, could tags help users in their searches? Tags are the words that people associate with the information object being tagged – in the absence of malicious intent (such as in tag spam) tags represent what people think the tagged object is about, or what it is, or what other things are similar to it.

Tags can provide additional information to help decide whether a particular search result is useful or not, or be a source of alternative search terms when redoing a search. The role and relative importance of tags may also differ depending on the type of content tagged, such as whether it is text or multimedia.

While there is a widespread perception that tags improve findability and discovery of information (Hammond, Hannay, Lund, & Scott, 2005; Guy & Tonkin, 2006; Macgregor & McCulloch, 2006), there is little empirical data on how tags do so, particularly from the perspective of the searcher. Very little is known about how people make use of tags during the course of finding information – do people click on them or simply look at them to get ideas for search terms? Do they rely on tags to help them decide whether a search result is worth clicking on or not? When do they even notice tags? This study examines these questions, focusing on how people use tags during the online search process.

1.2 Definition of Tagging

The following are some of the definitions of tags provided by popular tagging sites:

> Tags are one word descriptors that you can assign to your bookmarks (Delicious)
>
> Tags are like keywords or labels that you add to a photo to make it easier to find later (Flickr)
>
> A tag is a keyword or short phrase that writers assign to articles to describe or identify the content: the subject matter, the people involved, the type of article, themes addressed. This helps people searching for a particular type of content to find articles using those tags. (Technorati)

Researchers have defined tags as "free-form labels assigned by the user and not drawn from any controlled vocabulary" (Hammond, Hannay, Lund, & Scott, 2005) and tagging

as "marking content with descriptive terms" (Golder & Huberman, 2006). The basic idea of tagging, that of attaching descriptive labels to information resources, is not new. It is tagging in the context of the Web that is regarded as a novel phenomenon.

Due to this novelty, various terms have emerged for tagging on the Web: tagging, collaborative tagging, social tagging, and folksonomy. Tagging is the most general term, with the other terms emphasizing different aspects of tagging as practiced on the Web. *Collaborative tagging* and *social tagging* both emphasize the public or shared nature of tags – that is, tags being visible to people other than the tagger. Non-public tagging occurs when the tagger chooses to restrict visibility of the tags to only the tagger or a select set of users, or when the application in which tagging is occurring is not of a public nature, such as a user's email folders. The term collaborative tagging may be somewhat misleading in that it implies that taggers are somehow collaborating in their selection of tags. Studies indicate this is not usually the case (Sen et al, 2006; Wash & Rader, 2007). The term social tagging avoids this implication while specifying that the tagging of interest is of a social or public nature, open to social influences that are part of the tagger's context. The term *folksonomy* is used almost synonymously with social tagging (Hammond, Hannay, Lund, & Scott, 2005; Guy & Tonkin, 2006; Macgregor & McCulloch, 2006), although the connotations are different. The implication with folksonomy is that a controlled vocabulary, or list of categories, will be generated in bottom-up fashion through the individual contributions of numerous taggers on the Web.

1.3 Problem Statement

Social tags are now part of the landscape of the Web. They are ubiquitous especially on websites based on user-contributed content, such as Flickr, YouTube, and

blogs. Yet it is not clear what role they play or can play for those Web users who themselves are not taggers. Much of the research on tagging has focused on taggers and behaviors and motivations associated with tagging, or the tags themselves. Through this research, it has been found that taggers derive a number of benefits from tagging, such as self-expression (Cosley et al., 2009; Marlow, Naaman, Boyd & Davis, 2006), community building (Thom-Santelli, Muller & Millen, 2008), and signaling expertise to others (Thom-Santelli, Cosley & Gay, 2010). Yet taggers represent only a small proportion of the users of a website or service. For example, on MovieLens, 13.5% of active users had applied at least one tag (Sen, Harper, LaPitz & Riedl, 2007). On some systems, a small number of taggers produce the bulk of tags available on the system. Data from the Library of Congress Flickr project found that 40% of tags were added by a group of 10 taggers (Springer et al, 2008), while in the Australian Newspapers archive of the National Library of Australia, 57% of tags were created by the top 10 taggers (Holley, 2010). A large number of Web users are being exposed to tags created by others, and it is not clear what kind of benefit they are deriving from the tags.

There is a widespread perception that tags improve findability and discovery of information (Hammond, Hannay, Lund, & Scott, 2005; Guy & Tonkin, 2006; Macgregor & McCulloch, 2006). Tags have been integrated into search engines, enabling tag search on some websites, or as part of the "bag of words" for the ranking algorithms. In this sense, one might argue that tags have improved findability and discovery of information. However, little is known on how tags contribute to the search experiences of Web users when users interact with tags during search. Furthermore, no research has examined if and how searchers can indeed make use of tags created by others when trying to decide if

a particular search result is worth looking at or not, or when trying to come up with alternative search terms.

This research aims to obtain an understanding of the use of tags during the search process through an information retrieval experiment in which participants search for text documents and images, using information retrieval interfaces differing in their incorporation of tags. This study examines both user behavior and search interface design with respect to social tags. Questions explored include how people use tags during their online search process, and for what types of searches or stages in the search process people choose to use tags.

The significance of this study lies in its contributions to the fields of interactive information retrieval and social media, as well as implications for information retrieval system designers. By examining if and how tags are used during the interactive information retrieval process, the study contributes to the research on surrogates and relevance judgment. In addition, the study examines how the same surrogate element – in this case, tags – is used for different resource types. Research on social media has tended to focus on the content, and the practices around generating and sharing that content (e.g., Java, Song, Finin & Tseng, 2007; Lerman, 2007; Gilbert, Karahalios & Sandvig, 2008). Thus much of the research on social tagging has focused on the tags themselves, and the behavior and motivataion of taggers (Sen et al, 2006, 2007; Ames & Naaman, 2007; Heckner, Neubauer & Wolff, 2008; Lange, 2008; Nov, Naaman & Ye, 2008). This study brings together the two research areas of interactive information retrieval and social media, examining how social tags can be used to enhance interactie information retrieval. More broadly, it addressed the question of how social media can be used to support

people in their use of information retrieval systems during the course of information seeking.

Obtaining a clearer picture of how users interact with tags during the search process can help guide designers of search interfaces and information retrieval systems on using tags to improve the Web search experience. First, it provides guidance on how to incorporate tags into the interface, such as where to place tags on a Web page, or which types of pages should display tags. The results of this study can also provide guidance on designing the interactions with tags, such as what happens when a tag is clicked on.

1.4 Research Questions

The dissertation examines the effect of tags on the online search process by focusing on the following research questions:

1. What are the characteristics of tag and non-tag searches, for text searches and image searches?
    1.1. How does observed search behavior differ for tag and non-tag systems?
    1.2. How do perceptions of the search process differ for tag and non-tag systems?
2. How does the role of tags differ for text searches and image searches across stages of the search process?
    2.1. How are tags used for query reformulation in text searches and image searches?
    2.2. How are tags used for predictive judgments of relevance in text searches and image searches?

2.3. How are tags used for evaluative judgments of relevance in text searches and image searches?

3. How does prior knowledge of tags influence the use of tags during the search process?

3.1. What prior knowledge of tags do users have?

3.2. To what extent does prior knowledge of tags influence the use of tags in the experiment system?

3.3. To what extent does experience with other information retrieval systems influence the use of tags in the experiment system?

RQ1 focuses on identifying the respective characteristics of the search process in the presence and absence of tags, and comparing these search process, for resources of the same type. RQ2 examines how resource type affects the use of tags during the search process. RQ3 aims to shed light on reasons for tag use and non-use, by asking searchers about their perceptions of social tags on the Web.

1.5 Objectives and Scope of Study

This study examines people's use of tags during interaction with an information retrieval system. In contrast to previous studies on tagging, the focus is on tag use, not the act of tagging. The information retrieval systems of interest are websites where users can search or navigate the contents, such as Flickr. How taggers select the tags to attach to particular content, or methods to incorporate tags into searching or ranking algorithms, are outside the scope of this study. The study also restricts itself to tags that are publicly visible, and excludes tags that are only visible to the tagger.

The objectives of the study are:

1. To identify the effect of tags on the search process.
2. To understand how tags are used in image search and text search.
3. To identify factors that influence people's use of tags during the search process.

1.6 Research Design

To accomplish these objectives, the research is designed as a laboratory study in a controlled setting. There are two factors that are varied: the presence of tags in the search interface, and the resource type for the search. Search interfaces can either display tags or not display tags. The two resource types for this study are text and images. The design is within-subject, to control for individual characteristics when taking part in the search process.

1.7 Structure of the Dissertation

This dissertation is composed of seven chapters. Chapter 2 reviews related literature on social tagging, information retrieval interaction, interfaces for interactive information retrieval, and image retrieval on the Web. Chapter 3 presents the conceptual framework guiding this study. Chapter 4 discusses the research methodology for the study, including data collection, sample, and data analysis. Results are discussed in Chapter 5. Key findings of this study are discussed in Chapter 6. In the final chapter, implications for system design and theoretical implications are addressed along with future research.

Chapter 2

Literature Review

In this section four topic areas are presented: social tagging, information retrieval interaction, interfaces for interactive IR, and image retrieval on the Web. As this study examines the role of tags in IR interaction, background is provided on tags, and information retrieval interaction. Interfaces for interactive IR are discussed, as users interact with IR systems through their interfaces - if users encounter tags during IR interaction, it is likely through the interface. In addition to text, video, audio and images are some of the resource types currently available on the Web, all of which are searched for by Web users. Of the various types of non-text retrieval, image retrieval is examined in detail because of its longer history with end-user searching than other resource types.

The following questions are explored in this literature review:

1) What is the current state of understanding on tags and social tagging?
2) How have the findings on relevance criteria and document attributes been applied to the design of search result interfaces?
3) How have the findings on query reformulation been applied to the design of search result interfaces?
4) What is the current state of understanding of image retrieval on the Web?

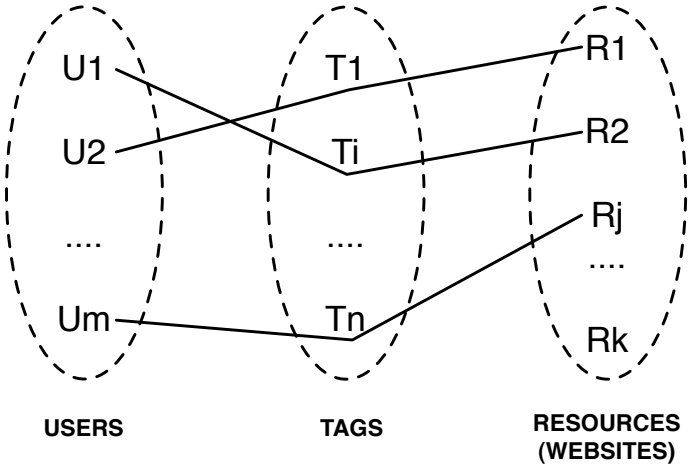
2.1 Tags

The general idea of attaching labels to describe information resources in digital form has been around since pre-Web days. Lotus Magellan was a desktop search package released in the 1980s that allowed users to add their own keywords to files in one's hard drive. Thus users could use these keywords to search for related files. Compuserve users in the 1990s could add keywords to documents they submitted to Compuserve forum libraries (Vander Wal, 2007). Such keywords are particularly valuable for files submitted in binary format or other formats for which full-text search is not possible.

Social tagging is a relatively new phenomenon, usually attributed to have originated with Delicious (http://www.delicious.com) in 2003. From there it spread to other Web sites centered around sharing user-contributed content, drawing the attention of researchers from a number of different fields. Early studies of tagging engaged in large-scale analysis of tags, focusing on identifying patterns and dynamics of tagging (Golder & Huberman, 2006; Marlow, Naaman, boyd & Davis, 2006). Other studies examined motivations (Hammond, Hannay, Lund & Scott, 2005; Ames & Naaman, 2007) and incentives (Sen et al, 2006; Sen, Harper, LaPitz & Riedl, 2007) for tagging. Other studies see tags as a data source for research on social networks, recommendation systems, and metadata generation and augmentation.

A generally accepted model of tagging systems (see Figure 2.1) has three entities: the tagger (or user), the item being tagged, and the tag (Marlow, Naaman, boyd & Davis, 2006; Halpin, Robu & Shepherd, 2007). The user is equated to tagger, or person doing the tagging. The model restricts tag use to tagging, and does not consider alternative uses of tags. Users who do not tag, but may still use tags are not included in the model.

Figure 2.1. Tripartite model of tagging (Halpin, Robu & Shepherd, 2007)



Much of the discussion on tags has taken place in blogs (Bray, 2005; Davis, 2005; Vander Wal, 2005). Tim Bray (2005) asked, "Are tags useful? Are there any questions you want to ask, or jobs you want to do, where tags are part of the solution, and clearly work better than old-fashioned search?" Ian Davis (2005) conjectured that tagging was expensive, in that "Tagging bulldozes the cost of classification and piles it onto the price of discovery." Thomas Vander Wal is credited with coining the term *folksonomy* and introducing the conceptual distinction of broad and narrow folksonomy. These two terms refer to two distinct models of tagging on the Web. In broad folksonomy, many users tag one document, while in narrow folksonomy, one or a few people, often the content creator(s), tag a document. Delicious exemplifies broad folksonomy, where many users attach tags to one Web page. Thus the collection of tags for a page represents the collective understanding of a group of people on what is noteworthy about the page. Flickr (http://www.flickr.com) is an example of narrow folksonomy – while users other than the photograph owner can tag photographs, this rarely occurs in practice (Marlow, Naaman, boyd & Davis, 2006). Table 1 summarizes some of these studies, focusing on

the larger, still-surviving Web sites implementing tagging. As with many Web 2.0 sites, several sites in early tagging studies were short-lived.

Table 2.1. Studies of social tagging

Application	Resource type	Studies
Delicious	Bookmarks	Golder & Huberman (2006); Guy & Tonkin (2006); Halpin, Robu & Shepherd (2007); Kipp (2007a); Wash & Rader (2007; 2008); Bischoff, Firan, Nejdl & Paiu (2008); Carman, Baillie & Crestani (2008); Heckner, Neubauer & Wolff (2008); Heymann, Koutrika & Garcia-Molina (2008)
Flickr	Photographs	Guy & Tonkin (2006); Marlow, Naaman, boyd & Davis, 2006; Ames & Naaman (2007); Bischoff, Firan, Nejdl & Paiu (2008); Nov, Naaman & Ye (2008); Heckner, Neubauer & Wolff (2008)
Connotea	Academic papers	Kipp (2007b); Heckner, Neubauer & Wolff (2008)
CiteULike	Academic papers	Kipp (2006); Kipp (2007b)
Last.fm	Music	Bischoff, Firan, Nejdl & Paiu (2008)
Technorati	Blog aggregator and search engine	Brooks & Montanez (2006)
YouTube	Video	Geisler & Burns (2007); Heckner, Neubauer & Wolff (2008)

Early empirical studies focused on characterizing and getting a snapshot of tags, and describing what tags are to a research audience. The former type of article came from researchers in computer science, while the latter came from researchers in library and information science. More recent studies have examined research questions such as the motivations of taggers and applications of tags to ranking algorithms and enhancing metadata. Tags in social tagging sites have also been used as dataset for exploring hypotheses and methods in social network analysis.

2.1.1 What are tags?

Three different characterizations of tags can be found in the literature: categories, keywords, and annotations. These characterizations are not mutually exclusive, but reflect the researcher's assumption of the nature of tags that then shape their research

questions and methods. Seeing tags as categories emphasizes the grouping together aspect, and relates to one of the key purposes of classification: "Classification brings like things together" (Svenonius, 2001, p. 10). Seeing tags as keywords emphasizes the extraction of the main idea(s) from the content. Annotations capture the idea that sometimes tags are about the tagger's reactions and uses for the item being tagged, and not about describing the item in the manner of a bibliographic catalog record.

Categories

In their early study of tags in Delicious, Golder and Huberman saw tags as being primarily about categorizing: "one makes sense of the things one encounters by categorizing them and ascribing meaning to them." (p. 201). Technorati, a blog search engine, instructs its users to “Think of a tag as a category name.” This perspective of creating categories is reflected in the term folksonomy. The idea is that a bottom-up accumulation of categories can eventually result in a taxonomy or classification system. Some of the literature on tagging recommends using tags for this purpose – instead of creating a taxonomy top-down by a small number of specialists, allow the users to tag and then use the collection of tags to generate the eventual number of categories used. Jacob (2004) makes an important distinction between classification and categorization. In the former, a classification system exists, and items are classified by being placed in the appropriate “bin” in the classification system. Categorization, on the other hand, involves perceiving similarity in a set of items and grouping them together in a category. Categories may be constructed on the fly. Tags as categories has a close connection to the literature on organizing information relating to grouping of information and folder structures.

Analysis of Delicious tags shows one of the functions of tags is task organizing, with bookmarks related to a task being assigned the same tag (Golder & Huberman, 2006). Thus tags perform a conceptual grouping function that is meaningful to the tagger. Most implementations of tagging on the Web allow tags to function as virtual folders – all items tagged with the same tag can be viewed as a group, analogous to selecting a labeled folder, whether physical or virtual. In the case of social tagging, a user can view the set of items that a large number of users have put in the same bin. It is interesting to see this in relation to Malone's (1983) study of filers and pilers regarding organization of work documents. One of the barriers to organization was deciding in which folder a document should be placed. In a social tagging system, a user can see how others have tagged an item, or examine the set of items for a particular tag, to help decide how to file an item. Since the items being tagged are digital, and not physical, multiple tags can be attached to an item, alleviating the difficulty in deciding which one particular category to assign to an item. Filing system complexity is not necessarily avoided, in that a user may have such a large set of tags they may not remember them all, or over time may have forgotten or misremembered the meanings of certain tags. Tags also allow grouping by factors external to the content of a document, such as those of topic and use identified by Kwasnik (1989; 1991) in her study of self-organization of work documents. A number of studies (Golder & Huberman, 2006; Kipp, 2007a) do indicate that users indeed do so, grouping information items by descriptions of the particular task they relate to.

Keywords

A textbook on cataloging and classification defines keyword as "a term that is chosen, either from actual text or from a searcher's head, that is considered to be a 'key'

to finding certain information." (Taylor, 2006) At least one definition of tags explicitly identifies them as being keywords. A distinguishing feature of keywords, making them different from categories, is that keywords are expected to be present in the document they are associated with. Thus sometimes keywords are equated with search terms, in that a document that contains the specified keywords is expected to be relevant to the user's information need. This characterization of tags motivates research in generating tags automatically from text present in the document or seeing them as a source of keywords (Al-Khalifa & Davis, 2006). If tags are terms already present in the information object, such as document keywords, it is not clear what added benefits they provide over keyword search using search engines.

But people are increasingly searching for non-text materials on the Web, such as images or video. In the case of image search queries or requests, images must have been indexed or have metadata associated with them. In the traditional model of information retrieval, items with indexing terms that match the user's query terms are returned as the search results. Unlike text retrieval, image retrieval offers the option of using text or image for indexing and querying. Prior to the Web, image searching was most likely to occur in the context of institutional collections such as libraries, museums, or newspaper photo archives. Such collections were professionally indexed and frequently offered rich search facilities, including reference librarians and extensive cataloging content. The Web now makes available a vast collection of images provided by Web users, but these images are not systematically indexed or described.

One early and extensive adopter of tagging was Flickr, an online photo-sharing site. While it provided a number of more traditional metadata fields, such as title and

description, it also allowed users to provide an unlimited number of tags to their own photographs. These tags often contain descriptive terms about the image, such as the location, the event, or people photographed. YouTube, a video-sharing site, also incorporates tags, which similarly to Flickr are often descriptive terms about the video being tagged. For non-text materials, descriptive terms provided in tags function in the manner of keywords for text search.

Annotations

Annotation in this section refers to "scribbles in the margins" (Abbas, 2007) made by users in the course of reading a document. There have been a number of studies specifically examining annotations made to documents in the course of reading (Marshall, 1997; 2004). Marshall initially examined annotations in used textbooks. More advanced students were more likely to seek annotated textbooks, and for specific types of annotations. These were annotations that provided guidance on how the textbook had been used in a course, with notes regarding what a professor had emphasized, or had indicated would appear in a test. These types of annotations, while initially made for personal use of the original textbook owner, provided valuable information to later users of the textbook. Other types of annotation, such as highlighting or underlining, were not perceived to be as useful to readers other than the original annotator. Marshall (2004) further explored the relationship between personal and public annotations by examining students' personal annotations on assigned papers and the annotations they shared using an online system. Only a small fraction of personal annotations were shared, and the annotations that were shared were substantially changed from their original form as personal annotations.

Studies of tags found that tags included affective terms, reflecting the tagger's emotional state or assessment of the information object to the tagger (Golder & Huberman, 2006; Kipp, 2007a). These types of tags cannot be seen as keywords, as they are not about the topical content of the document being tagged. Kipp explicitly examined two types of what she called non subject related tags, affective tags and time and task related tags. Affective tags were those tags consisting of words describing an emotional state (e.g., interesting, fun, cool). Time and task related tags were tags such as *toread*, *todo*, or *tobuy*. These types of tags constitute a minority of tags, with one study finding them to constitute about 16% of tags (Kipp, 2007a). Berendt and Hanser (2007) noted that for some users, tags are "just more content". This echoes Marshall's findings that annotations, when viewed at a later date by those other than the original annotator, become part of the content for these viewers.

2.1.2 Motivations for tagging

Given the rapid adoption of tags and tagging on the Web, one question that arises is, why tag? This question can be examined in two ways – what functions do tags serve, and the motivations of taggers. In an influential early examination of tags, Golder and Huberman (2006) identified the following functions for Delicious tags:

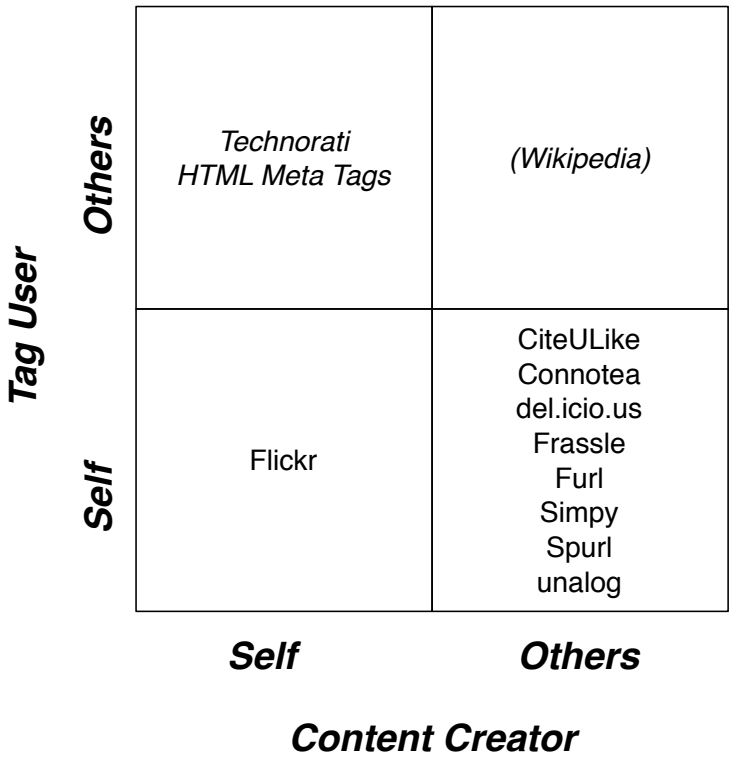
1) Identifying what (or who) it is about.
2) Identifying what it is (e.g., blog post)
3) Identifying who owns it (e.g., name of the author of the bookmarked content)
4) Refining categories. (e.g., tags that refine existing categories, often in the form of numbers)
5) Identifying qualities or characteristics (e.g., funny)

6) Self-reference (e.g., mystuff, mywork)

7) Task organizing (e.g. toread, jobsearch)

Several of these functions are about summarizing the content of the bookmarked pages. It should be noted that the primary purpose of Delicious is to bookmark Web pages for oneself. This stands in contrast to other sites such as Flickr or YouTube, where content contributors upload content to share with others.

Figure 2.2. Content creators and tag users (Hammond et al, 2005)

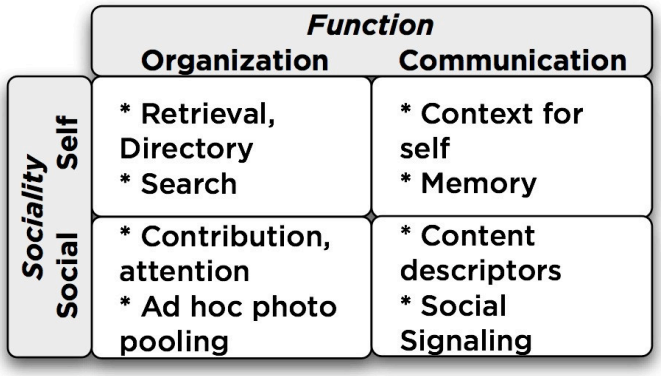


Hammond et al (2005) categorized the benefits of tagging against the content being tagged. The vertical axis of Figure 2.2 represents the tag consumer (tag user in the diagram) and the horizontal axis represents the content being tagged. Interestingly blogs or YouTube are not included in this chart. This characterization of tagging with respect to the content and tag consumption is useful for distinguishing between various types of

tagging sites on the Web. The characterization also illustrates possible reasons for variations in tagging patterns across different sites, as well as making clear problems in generalizing study results from one tagging site to another. Narrow folksonomy sites correspond to sites with self-created content, while broad folksonomy sites correspond to sites with content created by others.

Ames & Naaman (2007) examined motivations for tagging in two photo applications, ZoneTag and Flickr. Based on interviews of taggers in those systems, they derived a taxonomy of tagging motivations (Figure 2.3). Two dimensions are identified: sociality and function. Sociality relates to "whether the tag's intended usage is by the individual who took and uploaded the photo or by others, including friends/family and strangers." Function refers to the intended uses of a tag. The motivations in the social category of the sociality dimension were found to be the most common motivations for tagging.

Figure 2.3. A taxonomy of tagging motivations (Ames & Naaman, 2007)



Findability then appears to be the main function as well as a primary motivation for tagging. Findability can be broken down into findability by self and findability by others. Wash and Rader's (2007) study found that among Delicious users, findability of

one's own tagged items was poor. They also found that tagging was for oneself, and not for the benefit of others – that is, making it easier for others to find the material. In contrast, the Flickr users studied by Ames and Naaman do tag with findability by others in mind. But it is not findability aimed at the world at large, as in the case of indexing practiced in libraries. Flickr users want their photos found by specific audiences, and tag for findability by these audiences. Whether such focused findability is successful is an open question.

2.1.3 Tag production and consumption

In studies of social tagging, tag use and tagging are often used interchangeably. Only taggers or tag producers are seen as users, or consumers, of the tags. Tag producers have been extensively studied on a variety of systems such as Flickr (Marlow et al., 2006; Ames & Naaman, 2007), Delicious (Wash & Rader, 2007), enterprise tag applications (Thom-Santelli, Muller & Millen, 2008), and research-motivated systems such as MovieLens (Sen et al., 2006; Sen, Harper, LaPitz & Riedl, 2007) and MobiTag (Cosley et al., 2009; Thom-Santelli, Cosley & Gay, 2010).

Tag use or tag consumption behavior has not been studied as extensively, especially the use of tags created by others. In one of the few studies examining the value of tags to both taggers and non-taggers, Sen, Harper, LaPitz and Riedl (2007) found that in general taggers had a more favorable view regarding the usefulness of tags than the overall mix of users. A study of social tagging activity on the Australian Newspapers collection of the National Library of Australia found that two of the four "super-taggers" were tagging with the expectation that tags might help others, but were not using tags themselves, finding articles by keyword searching instead (Holley, 2010). These studies

suggest that tag production and consumption should be considered separately from each other. As yet there is a lack of studies examining tag consumption behavior of tag non-producers.

2.1.4 Tags and Information Retrieval

Information retrieval research on tags has focused on system IR issues such as their incorporation into relevance ranking algorithms (Hotho, Jäschke, Schmitz & Stumme, 2006; Aurnhammer, Hanappe & Steels, 2006; Freyne, Farzan, Brusilovsky, Smyth & Coyle, 2007; Yanbe, Jatowt, Nakamura & Tanaka, 2007), algorithms for personalizing retrieval (Carman, Baillie & Crestani, 2008), use in classification or clustering algorithms (Brooks & Montanez, 2006), and augmenting metadata (Hunter, Khan & Gerber, 2008).

Morrison (2008) compared the search retrieval performance of tags against search engines and subject directories. Study participants entered their queries into a search interface specifically constructed for the study, and this query was submitted to Google, Microsoft Live, AltaVista, Yahoo directory, the Open Directory Project, Delicious, Furl, and Reddit. Up to 20 results were retrieved from each of these eight sites, duplicates removed, and the rest presented in randomized order to the participant, who made a binary Yes/No judgment of relevance. Precision, relative recall, and retrieval rate (the number of documents returned compared to the maximum possible) were used to compare the eight sites. Six types of searches were examined: research, news, general, factual, entertainment and exact site. In general search engines outperformed folksonomies, or tagging systems, and subject directories across all search categories. He found that the folksonomies, or tagging systems, outperformed directories for news

searches in terms of precision and recall. Folksonomies fared particularly poorly for factual and exact site searches. Folksonomies and directories were particularly suited for searches for a set of items, compared to specific item searches. There was no statistical difference in performance for folksonomies and directories. This study suggests some of the types of searches for which tags may be suitable, and points out an intriguing similarity between folksonomies and subject directories in terms of IR performance. But it is difficult to isolate the effectiveness of tags in IR from this study. For example, a query entered into the Delicious search box returns search results produced by Delicious's search engine, which appear to combine tags and other information provided by Delicious taggers, such as title and description of the bookmark, in its relevance rankings. On the other hand, clicking on a tag in Delicious returns a list of the items that were tagged with that tag. So in effect Morrison's study is comparing search engine algorithms incorporating tags against those that don't, and in some ways the breadth of collections of the sites in the study. Search engine companies regularly crawl the Web, while folksonomy-based systems rely on what their users have found and entered into the system. Morrison acknowledges this limitation in his paper, but notes that different systems may then be better suited for different types of information needs. Heymann, Koutrika and Garcia-Molina's (2008) analysis of a large-scale crawl of Delicious appears to confirm this – Delicious users bookmarked new or frequently updated pages, with 25% of URLs bookmarked being new or unindexed pages. These results imply that tags may be particularly useful in searches for frequently updated information where people seek multiple documents.

2.2 Information retrieval interaction

Information retrieval (IR), as an academic field of study, examines "the processes involved in the representation, storage, searching and finding of information that is relevant to a requirement for information desired by a human user." (Ingwersen, 1992, p. 49) Alternately, IR is "finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)." (Manning, Raghavan & Schütze, 2009, p.1) The former definition reflects the state of IR prior to the advent of the Web, when IR systems were often costly to build and used mainly by academics and information professionals such as librarians. While the latter definition reflects most people's experience with information retrieval currently (i.e. using a Web search engine), it does not require human involvement in IR processes. Interactive information retrieval (IIR) denotes the subfield of IR that covers "research related to studying and assisting … diverse end users of information access and retrieval systems (Ruthven, 2008, p. 44). Within IIR, the study of IR interaction examines "the interactive communication processes that occur during the retrieval of information by involving *all* the major participants in IR, i.e. the user, the intermediary, and the IR system" (Ingwersen, 1992, p. viii), or the "dialogue between the participants – user and computer – through an interface, with the main purpose to affect the cognitive state of the user for effective use of information in connection with an application at hand." (Saracevic, 1997).

In this section two widely cited general models of IR interaction and one micro level model of IR interaction are discussed. In Belkin's model, the central process of IR is users' interactions with texts. Saracevic sees interactions as occurring at several levels

for both the system and the user. Neither of these general models explicitly account for the role of feedback in IR interaction, which is addressed by Spink's model of interactive feedback. An underlying concept in these models, and IR in general, is relevance. A user's goal in IR interaction is to find relevant information. Application of this concept in IR research has ranged from taking it at face value, as in many IR system evaluation studies, to deriving a theoretical understanding of the concept.

2.2.1 Models

Belkin's episode model of interaction with texts

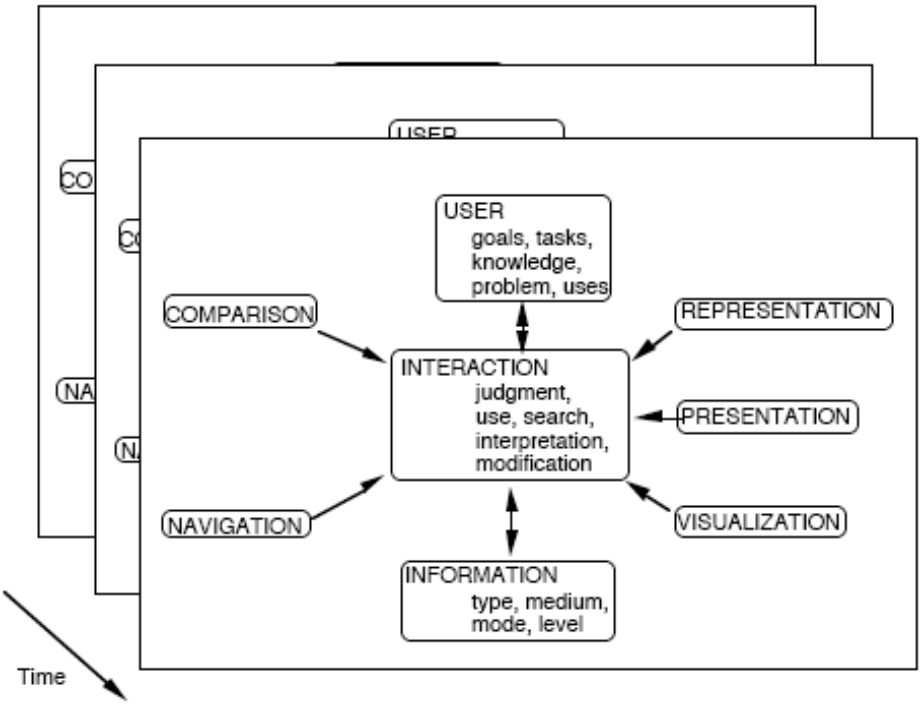
Belkin (1993, 1996) sees information retrieval as interaction with texts. Starting from this premise, he proposes the following questions need to be addressed in IR research (Belkin 1996):

- What are the kinds of interactions in which people engage?
- What situations or contexts or goals lead to specific kinds of interactions?
- How does the nature of the information objects interacted with affect the nature of the interaction itself?
- What leads to changes from one interaction to another?
- What are the different ways to support optimally different types of interactions?

The model in Figure 2.4 is intended to take all the above questions into account (Belkin, 1996). In this model, an information seeking episode consists of a series of interactions or slices in time, in which these interactions occur to address a user's goal or problem. Interactions with information objects include judgments of relevance, decisions

to use the information, and interpreting the information encountered. Processes such as comparison, navigation, representation, presentation, and visualization support these interactions. An episode is terminated when the user achieves their goal or accomplishes their task. In this model what constitutes an episode is quite flexible, and so it is possible for an information need to evolve or remain unchanged within an episode.

Figure 2.4. Belkin's episode model (1996)



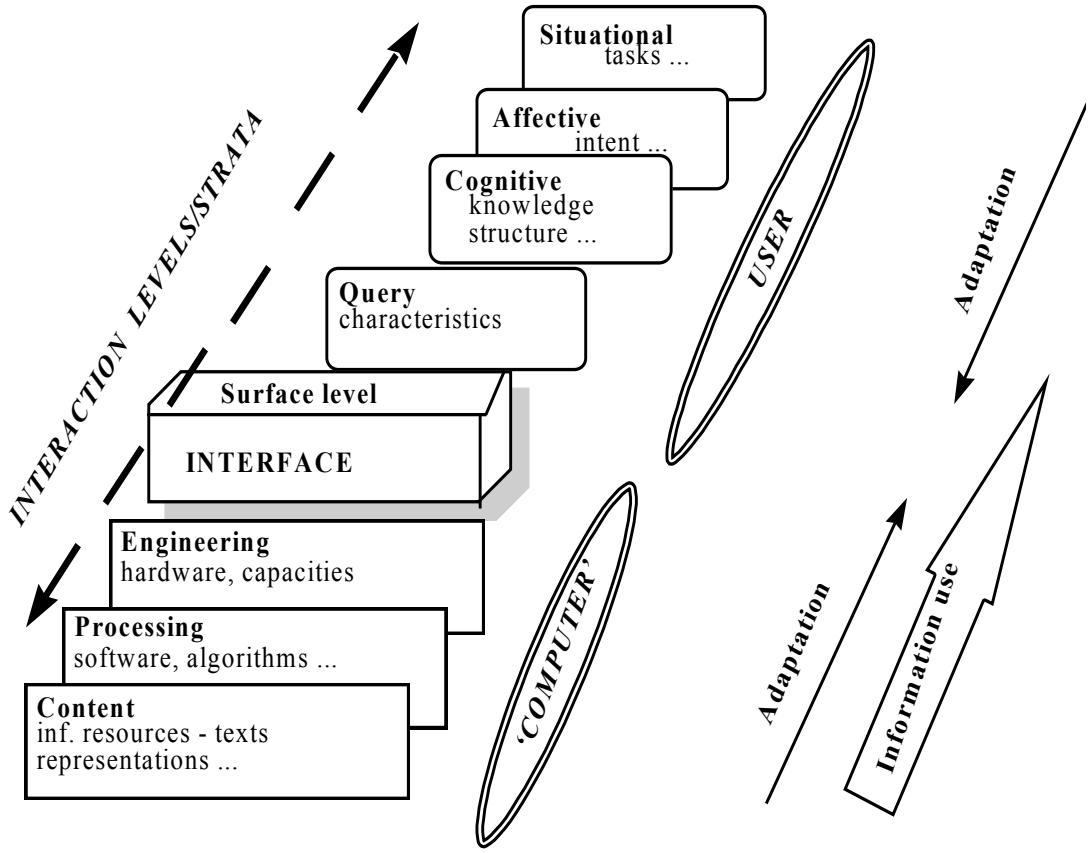
Despite the goal of addressing the five questions listed above, the model does not address the question of what leads to changes from one interaction to another. In this model it is difficult to represent how and when these shifts take place. It is also difficult to represent feedback occurring from one IR interaction to another.

Saracevic's stratified model of information retrieval interaction

Figure 2.5 illustrates Saracevic's stratified interaction model (1997). The model assumes that "(i) users interact with IR systems in order to use information, and (ii) that the use of information is connected with cognition and then situational application." (p. 315). The computer, or IR system, and the user communicate through the interface, which is the surface level of interaction. Both the system and the user have additional layers underneath that surface level of interaction. Users engage in three levels of interaction: cognitive, affective, and situational. At the cognitive level users are engaged in interpreting, judging, and assimilating the retrieved information. At the affective level users interact with their intentions. Investigations at this level involve users' intentions, beliefs, and motivations. The situational level involves the user's interaction with their environment or context, that is, the situation that brought about their information need and subsequent interaction with the IR system. User's assessment of the IR system will be affected by how useful they find the information retrieved for addressing their particular situation or problem. Investigations at this level will, for example, look at task - the interaction with the IR system occurs because of some task the user must perform. A significant contribution of this model is the recognition of different types of relevance. In this model each level has its own type of relevance. Relevance is discussed in more detail in a later section.

Saracevic (1997) acknowledged two limitations of this model: 1) as in other stratified models in linguistics and communication, decomposition and depiction of interplays between levels is difficult to specify; and 2) the model does not supply enough detail for experimentation and verification in larger interaction studies.

Figure 2.5. Stratified model of IR interaction (Saracevic, 1997)



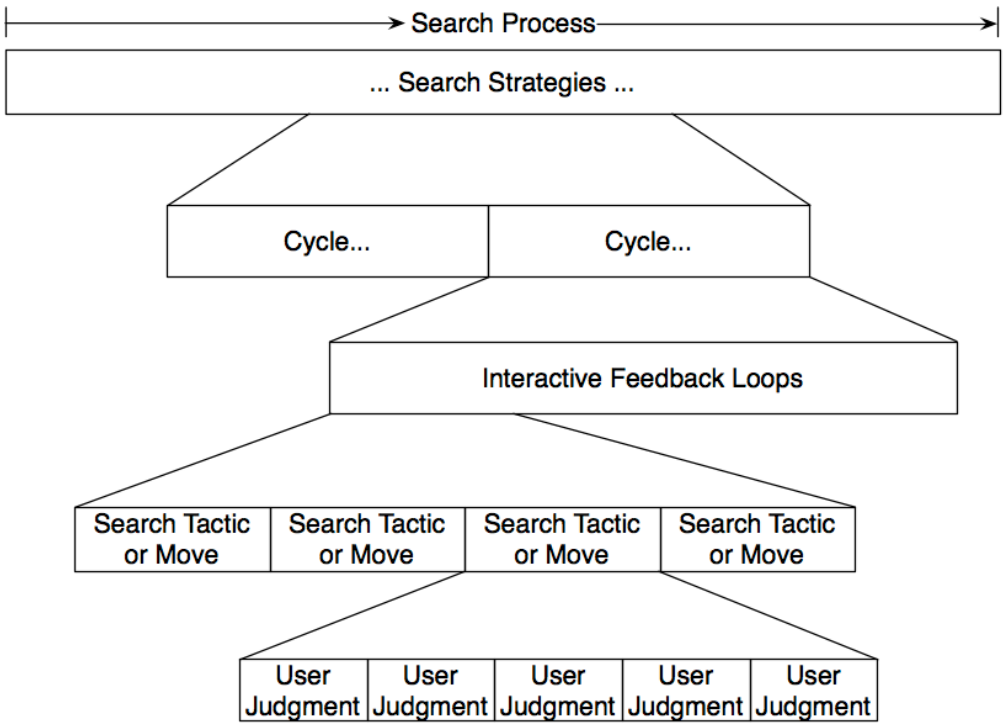
Spink's model of interactive feedback

Spink (1997) is considered a model of micro level IR interaction, in that it highlights one particular aspect of IR interaction, in this case the role of feedback in IR interaction. Existing models, such as Belkin's episode model or Saracevic's stratified model, do not account for the role of feedback in IR interaction. In addition, while previous research had only considered relevance feedback, Spink identified five types of interactive feedback (p. 387):

- Content relevance feedback: user query followed by an IR system output of retrieved items then judged by the user for relevance followed by a query or reformulation.

- Term relevance feedback: user query followed by an IR system output of retrieved items and user selection of a new search term(s) from the retrieved output used in a subsequent query.
- Magnitude feedback: user query followed by a judgment based on the size of the output from a query that effects the next query.
- Tactical review feedback: user input followed by a strategy related judgment to display the search strategy history influencing the subsequent query.
- Term review feedback: user input followed by a strategy related judgment to display terms in the inverted file influencing the subsequent query.

Figure 2.6. Spink's elements of the interactive search process (1997)



Spink also presents a search process model incorporating feedback (Figure 2.6). The interactive search process is constituted by a series of search strategies with one or more cycles, defined as "one or more search commands ending in the display of retrieved items". Each cycle consists of one or more interactive feedback loops ("user input, IR system output, user interpretation and judgment, user input") (p. 392). This model of the search process makes explicit how user interactions with an IR system are influenced by prior interactions with that system.

2.2.2 Relevance

In the models presented in the previous section, relevance is an underlying concept tied to user judgment. Several researchers consider relevance to be the central concept in information retrieval, if not information science (Saracevic, 1975; Saracevic, 2007a; Schamber, Eisenberg, & Nilan, 1990; Ruthven, 2005). Several comprehensive reviews have already been written on the topic (Mizzaro, 1997; Saracevic, 2007a; Saracevic, 2007b). This section summarizes the current consensus on the nature of relevance, and examines research on relevance criteria. Relevance criteria for non-text documents are briefly discussed.

The nature of relevance

Relevance is considered the central concept in information retrieval, if not information science, by several researchers (Saracevic, 1975; Saracevic, 2007a; Schamber, Eisenberg, & Nilan, 1990; Ruthven, 2005). Evaluation of IR systems still hinges largely on relevance, and measures derived from relevance (Voorhees & Harman, 2005). What has emerged from this research is that relevance is a multidimensional,

dynamic, situational, and subjective phenomenon (Schamber, Eisenberg & Nilan, 1990; Mizzaro, 1997; Borlund, 2003; Ruthven, 2005; Saracevic, 2007a, 2007b).

Saracevic summarizes previous research on relevance as taking five manifestations (see Table 2.2), system or algorithmic relevance, topical or subject relevance, cognitive relevance or pertinence, situational relevance or utility, and affective relevance.

Table 2.2. Manifestations of relevance (Saracevic, 2007a, p.1931)

System or algorithmic relevance	Relation between the query and information objects as determined by the IR algorithm
Topical or subject relevance	Relation between the topic of the query and the topic or subject of information objects; "aboutness".
Cognitive relevance or pertinence	Relation between the cognitive state of the user and information objects; informativeness, novelty, and information quality are some of the criteria for cognitive relevance.
Situational relevance or utility	Relation between the situation, task or problem and information objects; some criteria are usefulness in decision making and appropriateness of the information for resolving the problem at hand.
Affective relevance	Relation between intent, goals, emotions and motivations of a user and information objects; some criteria are satisfaction and success.

Borlund (2003) finds situational relevance to be particularly suitable for interactive information retrieval evaluation. Situational relevance is seen as "a relationship between the retrieved objects and the user's perception of a given work task situation" (p. 916). Situational relevance depends on the situation or task that generated the information need. Cosijn and Ingwersen (2000) expands upon Saracevic's manifestations of relevance by introducing socio-cognitive relevance, which is the

relation between a "situation, task, or problem at hand as perceived in socio-cultural context" and information objects (p. 547). Socio-cognitive relevance is distinct from situational relevance in that socio-cognitive relevance is "determined by the individual actor in *interaction* with other actors within a *community*." (p. 546) In this model affective relevance encompasses topical, cognitive, situational, and socio-cognitive relevances.

In practice a number of concepts are associated with relevance, which are used interchangeably with relevance, or are considered components of relevance. Usefulness or utility are often used interchangeably with relevance. Janes (1994) discusses relevance, topicality, and utility as separate concepts or measures. Topicality is "the relation of a document to the topic of a user's query" (p. 161) and is independent of the user. What this means is that as long as there is some match between query terms and the document, whether in the full text of the document or bibliographic record, which includes subject terms, the document is topical to the query. Utility, "the degree to which a document is useful to the user who requested it" (p. 161) corresponds to Saracevic's situational relevance. Janes argues that relevance and utility are not interchangeable, giving the example of a document that itself may not be relevant, but has utility by providing links to relevant documents. What this example points out is that while a document may not be topically relevant, it may be situationally relevant. Pertinence is another related concept, which Saracevic refers to as cognitive relevance. Kuhlthau differentiates relevance and pertinence and assigns them to different stages of her Information Search Process Model (2004). Relevance is a determination of topicality and utility, while pertinence is a determination that information has "a more decisive and significant relationship to a topic

than relevance and is related to personal information need." (p. 42) In this view, pertinent information is a subset of relevant information. In the initial stages of the information search process, the information seeker makes determinations of relevance, but once a focus is formed they shift to making determinations of pertinence. Some see usefulness as a broader concept that includes relevance (Cool et al. 1993).

Relevance judgment and relevance criteria

Another aspect of relevance is how we make judgments of relevance. That is, what are the criteria for relevance and how are these criteria applied? In conjunction with the relevance criteria applied by people, what about the information object itself leads people to judge it relevant or not relevant? There have been various studies examining relevance criteria. In general, by relevance criteria, what is meant are the criteria that a user applies to arrive at a judgment of relevance. But a number of other studies (Lan 2002; Saracevic, 2007b) examine the document attributes, or document clues, that people use to make judgments of relevance. Document attributes are aspects of the document used in making judgments of relevance, such as the title, summary, author, etc. Notably, Barry (1998) linked document attributes to relevance criteria categories. For example, one user criterion for relevance is recency, where newer documents are considered more relevant to the user's information need. Barry found that full text, abstract, title, and other document/source traits were used to make judgments of recency, while indexing terms were not applied for this purpose. It should be noted that relevance criteria, at least given the research methods used in the literature, is highly dependent on the nature of the documents presented and the information need. So while recency may be mentioned as a criterion, the user may have to rely on other criteria for relevance if the particular

document set has no attributes related to recency. The nature of the information need or task cannot be ignored - while recency may be important for finding news stories, it may not be as important a criterion for finding historical images of Shackleton's Antarctic expeditions. The two types of studies, those aiming to identify user relevance criteria, and those identifying document attribute and relevance judgment relationships, have different implications. The former help us elucidate the concept or nature of relevance. But, they do not yield direct implications for system design. The latter type of study, examining the document attributes people use to make judgments of relevance, has direct implications for system design across different types of information retrieval systems, ranging from what document attributes to make visible or salient to the user to weighting document attributes in relevance ranking algorithms.

Lan (2002) examined in detail document clues used for relevance judgment of surrogates and Web pages. When examining search engine results, the title, summary and URL were most frequently used. When examining Web pages, the text itself was most frequently used. Participants were allowed to use the search engine of their choice to address their own information needs, resulting in the use of 11 different search engines. Tombros, Ruthven and Jose (2005) examined how Web page features affected people's perceptions of utility, in an experimental setting with assigned search tasks. In contrast to the Lan study, only the features of Web pages were examined. The assigned search tasks were also of a general nature, such as shopping for hi-fi speakers or tourist information for Kyoto, while Lan's participants carried out searches related to their academic work. Tombros et al found that the content or text was the Web page feature category most frequently used, followed by structure and quality features. Text and structure categories

are based on document clues or attributes, but quality features are related to user relevance criteria, including items such as recency, general quality, and content novelty. Crystal and Greenberg (2006) examined what attributes in surrogates (search results) and documents (Web pages) were used for predictive and evaluative judgments of relevance, employing Rieh's (2002) model of two-stage judgment. Toms and Latter (2007) examined selection criteria for hyperlinks in search results and relevance criteria for Web pages.

Kelly, Murdock, Yuan, Croft and Belkin (2002) took a different approach, examining features of documents relevant for different types of information needs. They distinguished between task- and fact-oriented questions and found document features used to assess relevance differed for these two types of questions. Lists and FAQs occurred more often in documents relevant to task questions, while the number of links was greater for documents relevant to fact questions. There are two possible interpretations of these results: different types of documents are relevant for different tasks, or different relevance criteria are applied for different types of tasks. While not examined by the researchers, this suggests that items in a surrogate used to assess relevance may also differ by type of search task.

Earlier studies of user relevance criteria attempted to identify criteria, and did not distinguish between surrogate and full-text documents, or focused on only the surrogate or the full-text. Tang and Solomon (2001) identified document evaluation as occurring in two stages, the evaluation of document surrogates and evaluation of full-text documents. They had two criteria related to topicality, topical focus and topical relatedness. Topical focus was found to be more important and topical relatedness less important moving from

Stage 1 to Stage 2, indicating narrowing of the topical focus. Newness was more important in the later stage. The results indicate that the relative importance of relevance criteria changes across stages of the information retrieval process.

Rieh's (2002) model of two-stage judgment has been the basis for a number of subsequent studies of Web search behavior. In this model, users make a predictive judgment from the search results list, and make an evaluative judgment after viewing the Web page. Savolainen and Kari (2006) also used Rieh's model of predictive and evaluative judgment to examine the criteria used to judge relevance of hyperlinks and relevance of Web pages. In their study of the search process of consumers searching for health information, Toms and Latter (2007) distinguished between criteria employed when selecting links from the results page and when examining pages for pertinence to the task, but did not explicitly refer to Rieh's model.

Relevance and non-text documents

So far the discussion has focused on relevance regarding text documents. Relevance for non-text materials has been comparatively under-examined. Non-text materials include images, video, and music. Research in retrieval of non-text materials divides along two lines, content- and concept-based retrieval. Content-based retrieval is analogous to full-text search in text retrieval, in that the query is matched to the contents of the actual document. An example of such a query is to request all images that have the color red in over 50% of the image. Additional complexity lies in how the content is represented or stored, as images, video and music are frequently stored in compression formats that may involve some loss of information. In concept-based retrieval, the query involves an abstraction of the content and may involve matching to the metadata, and not

the content itself. An example of such a query is to request images of Renaissance Madonna's by Italian painters. In content-based image retrieval (CBIR), measurable attributes in an image, such as color or shapes, are used as the basis for search and retrieval, applying pattern recognition and machine learning techniques. Concept-based retrieval relies on text about the image, usually in the form of image metadata. It is not the image itself, but the text about it, that is the basis for search and retrieval.

Earlier research on video retrieval, when considering relevance, focused on system relevance (Wactlar, Christel, Gong & Hauptmann, 1999). More recently, Yang and Marchionini (2004) explored users' video relevance criteria. Cunningham and Nichols (2008) identified some of video and Web page criteria used in making relevance judgments when searching for videos online. Viewer ratings or view counts emerged as one of these criteria. Research on music retrieval relevance criteria has also focused on system relevance (Uitdenbogerd & Zobel, 1999). Some recent studies have identified a number of user relevance criteria, although the focus of the studies lay elsewhere. Cunningham, Reeves and Britland (2003) studied the behavior of people searching and shopping for music in libraries and music stores. Inskip, MacFarlane and Rafferty (2008) examined how creative professionals search for music to use in movies and commercials. Both studies identified some selection criteria employed by their respective user groups. The relevance criteria of creative professionals in Inskip et al's study ranged from budget or cost of licensing the music to "gut instinct". Genre and recency appeared to influence the consumers in Cunningham et al's study. There has not been as yet a study systematically examining user-defined relevance criteria for music information retrieval.

User relevance criteria for images have received relatively more researcher attention than music or video, and are discussed in more detail in a separate section.

2.3 Interfaces for interactive information retrieval

In the IR interaction models discussed previously, whether explicitly or implicitly stated, interactions between the user and the IR system occur through an interface. Thus the design of the interface can direct the IR interactions, supporting (or hindering) the user in their interactions with the IR system. While existing models of IR interaction acknowledge the role of the interface, they do not provide guidance for IR interface design. Shneiderman, Byrd and Croft (1997) present a four-phase framework for search intended to be used in the design of IR interfaces. In this model, IR interfaces need to support the user appropriately at each of these four phases:

- Formulation: what happens before the user starts a search
- Action: starting the search
- Review of results
- Refinement: what happens after review of results and before the user goes back to formulation with the same information need

The formulation and action phases are minimal or almost non-existent for typical Web searching. As the typical Web searcher uses a search engine for searching, and rarely uses advanced searching features, the decisions proposed for the formulation phase in the original model are not applicable. In a typical Web search session, the action phase consists of clicking the Search button to start the search and waiting for the results. The bulk of information retrieval interaction occurs in the review of results and refinement

phases. Thus this section focuses on topics pertinent to two phases, review of results and refinement. In particular, for review of results, research on presentation of search results and surrogates is examined. While review of results can be divided into two subphases, review of search results and review of documents, in current Web searching the user typically encounters different systems or interfaces for the two subphases. The user usually interacts with one IR system (e.g. Web search engine) for reviewing search results and refining their queries, while review of documents entails leaving the search engine for another Web site. Review of search results and refinement are the phases occurring within the context of one IR system.

2.3.1 Search results and surrogate presentation

This section examines the research literature regarding the presentation of search results and surrogates in online information retrieval systems. Surrogates are "representations of information or search objects presented to the searcher at the interface level." (Ruthven et al, 2008, p. 437) Surrogate and document attributes serve as input for making judgments with respect to the information need, such as relevance or usefulness. Especially in the case of Web search, users often only examine the surrogates in the search results display without further examination of documents. Eye-tracking studies found that users reformulated queries from scanning the surrogates alone, without examining any Web pages (Lorigo et al, 2008). During a Web search session, people spend at least as much time examining the contents of the search results page as they do examining web pages linked to from the list (Toms & Latter, 2007). The amount and quality of results descriptions in the search results page influences users' perceptions of the quality of a Web search engine (Lewandowski, 2008). At the same time, results

descriptions can misrepresent actual content, whether deliberately or by accident. Thus a user's assessment of satisfaction, ease, or success regarding the search rely a great deal on the contents of the surrogates, and how these surrogates are displayed. Some questions that arise are how much is the right amount of material to show, and what to show, depending on the device, the context, the task, and the user.

In this section we discuss surrogates from two perspectives, *what* should be in a surrogate, and *how* these components of the surrogate should be presented. Excluded from the discussion are novel search interfaces such as those providing query alternatives to the typical search box, or those providing novel interaction modalities. As stated, the focus is on surrogates and their presentation.

What to present

The creation of document surrogates has long been a research issue in library and information science (LIS). From the days of card catalogs to the web OPACs of today, library users interact extensively with document surrogates. Much of the research focus has been on the content of the bibliographic record – that is, what type of information about the document should be included in the surrogate, and specific rules on the creation of these document surrogates (e.g., AACR2). What should actually be displayed in the search results has not been extensively studied.

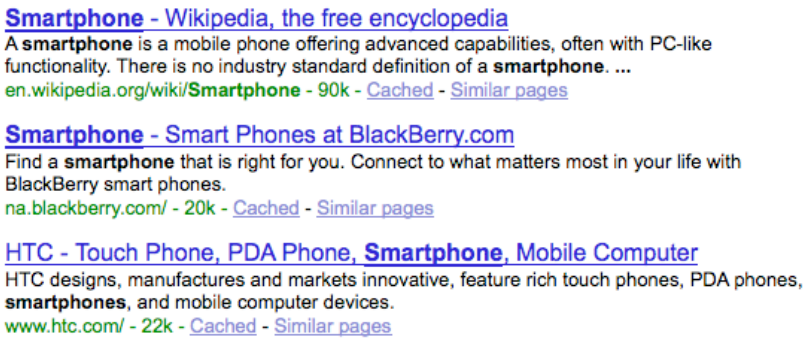
While ostensibly a report on intelligent interfaces and retrieval methods, Hildreth (1989) focused more on what would now be considered system aspects. The discussion on search interfaces was on browsing and navigation, and presentation of subject headings and thesauri for searching, without the search result display being discussed separately. Yee (1991) reviewed studies on OPAC user interfaces and identified a

number of search result display issues. In her list of 21 user problems with OPACs, she identified two that were directly related to search result displays: difficulty scanning through long displays, and difficulty understanding codes and abbreviations in displays. Another problem identified, difficulty due to brief displays, comes from a study comparing OPAC and card catalogue use, and so is more related to the display of bibliographic records rather than search results. In pre-Web OPAC studies, discussions of the search results interface do not differentiate between the display of search results and display of the bibliographic record. This is because in OPACs, the user is dealing with surrogates up to the point where they decide to end their OPAC session. A surrogate in the list of search results typically includes the title, the resource type, call number and/or location of the item. Selection of one of these surrogates results in the display of a more detailed surrogate that includes author and publisher names and sometimes a table of contents. This contrasts with Web search, where the search result display includes a snippet from the document, and selection of an item in the search results list leads to the actual document.

Web search has now shaped expectations regarding other types of information retrieval systems, including OPACs. A number of recent studies have shown that users now expect the library catalog to function similarly to Internet search engines (Fast & Campbell, 2004; Novotny, 2004). In addition to system expectations, such as relevance ranking of results, users now also expect summaries in the search results, as well as information related to the use of the resource, such as ratings and reviews, that are common in many Web sites (OCLC, 2009).

The contents of the surrogate, or what is displayed in the search engine results page (SERP), have become quite standardized for Web search results. As shown in Figure 2.8, the typical Web search engine's SERP includes the Web page title (more precisely, the text accompanying the HTML <title> in the header section), some summary text with the search term(s) in bold, the URL, and the size of the file. Depending on the Web document, its format, such as PDF or DOC, will also be included. Typically ten search results are displayed per page, and studies have shown that users rarely examine beyond these first ten results (Jansen, Spink & Saracevic, 2000; Jansen & Spink, 2006).

Figure 2.7. A typical Web search results page.



This type of search result display is a fairly recent development. Early search engines displayed a list of URLs without document summaries – gradually, additional components were added to arrive at the current *de facto* standard. Research has been ongoing on document surrogates for search results, such as the inclusion of document thumbnails or query-biased summaries. In actuality, in addition to these document summary or surrogate, users would see a number of other elements depending on the search engine they are using (Table 2.3).

What users see on the SERP undeniably influences relevance judgments, and perceptions regarding the search engine being used. Jansen, Zhang and Schultz (2009) found that search engine branding influenced perceptions regarding the performance of the search engine. In their study, participants were presented with SERPs that were identical in content except for branding elements (e.g. logo). Perceived performance as rated by participants differed for the four brands presented (Google, Yahoo!, MSN, and No Name), with the No Name or unbranded interface faring the worst, and Google and Yahoo! having the best perceived performance.

Table 2.3. Elements on SERPs (Höchstötter and Lewandowski, 2009)

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

Tombros and Sanderson (1998) compared query-biased summaries to typical output of IR systems of the time, composed of the title and the first few sentences of retrieved documents. Query-biased summaries are summaries customized to the user's

query, in this case sentences from the document selected based on the distribution of query terms in the sentences. Users performed better on recall and precision with the query-biased summaries, and also were more satisfied with the search. They also examined full-text documents less frequently than the control condition, leading to the conclusion that query-biased summaries provided enough information for relevance judgments without needing to refer to the full-text document.

Dumais, Cutrell and Chen (2001) developed and evaluated seven interfaces for integrating semantic category information with Web search results. Category labels derived from the text were automatically assigned to each search result using text classification techniques. In all cases category interfaces were faster than list interfaces.

Woodruff, Faulring, Rosenholtz, Morrison and Pirolli (2001) compared three different types of document summaries: text summaries, plain thumbnails, and enhanced thumbnails, for web search tasks. Participants used three different types of summaries: enhanced thumbnails, plain thumbnails, and text summaries. Plain thumbnails were simple reduced-size images, while enhanced thumbnails featured enhanced text and image contrast levels. Participants were given four types of search tasks, varying by the type of information they were expected to find: Picture, Homepage, E-commerce, and Side-effects. In terms of total search time, text summaries took the longest overall, but the relationship between summary type and total search time depended greatly on the question category. Plain thumbnails were the worst in terms of minimizing the number of visits to content pages. With the exception of the Picture task, participants spent more time on the summary page per visit with text summaries than with either type of

thumbnail. Across various measures the Picture task differed from the other 3 types of questions.

The use of thumbnails was further examined by Dziadosz and Chandrasekar (2002). They characterized the interaction between the end-user and a text-based IR system as having three steps: query formulation; relevance prediction ("the user inspects the search results page and guesses which items will lead to the desired information."); and relevance evaluation ("the user attempts to locate the desired information in the documents s/he predicted would be helpful.") (p. 365) Their study focused on relevance prediction, and participants were presented with 3 variations on summaries, text-only, thumbnails-only, and text plus thumbnails. Participants made more accurate decisions with combination of text and thumbnails than with text-only or thumbnails-only. They also tended to give the benefit of the doubt to many more results when there were thumbnails, compared to the text-only case.

White, Jose and Ruthven (2003) further examined query-based summarization using four different systems, two using query-biased summarization, and two using the standard ranked titles/abstracts approach - Google, Google with WebDocSum, AltaVista, and AltaVista with WebDocSum. Users preferred the simpler interface of WebDocSum systems. Users took longer on the ranked titles/abstracts systems and completed less tasks. While users liked both styles of interface, they disliked the interfaces for different reasons.

Clarke, Agichtein, Dumais, and White (2007) examined clickthrough data to study user behavior when interacting with search result captions. Clicking on a link was interpreted as an implicit judgments of relevance. Captions were defined as being

comprised of title, snippet and URL. They found that the presence of all query terms, readability of the snippet, and length of the URL shown in the caption significantly influence users' Web search behavior. Missing snippets, short snippets, missing query terms and complex URLs negatively impacted clickthroughs.

Ruthven et al (2008) focused on the contextual factors influencing the effectiveness of surrogates, or more specifically summaries of information sources. Personal contextual factors and the context in which summaries, or information, are displayed were explored. The study was done as part of a TREC[1] question-answering track with data provided by TREC assessors. Assessors were asked to rate the importance of four criteria solely based on the questions before seeing the answers. Answers in this case are essentially summaries of Web documents, in that they contain snippets answering the question from Web documents. While recency was not predicted to be important, after being presented with the answers, there was a significant preference for recent answers. Regarding good versus weak sources, there was a slight preference for information from good sources, and a tendency to rate as poor information from weak sources. Assessors were also more likely to accept answers to the questions if there was an assertion that supporting evidence existed from particular sites, even though information from those sites was not presented.

Joho and Jose (2008) examined the addition of top-ranked sentences and thumbnails to the baseline search result presentation typically found in SERPs. There were four layouts used in the experiment: baseline (identical to Google: title, snippet,

[1] The Text REtrieval Conference (TREC) is a workshop series co-sponsored by the National Institute of Standards and Technology (NIST) and U.S. Department of Defense, with the purpose of large-scale evaluation of text retrieval methodologies. http://trec.nist.gov/

URL, size, hyperlinks of cached page and similar page), baseline and top-ranked sentences, baseline and thumbnails, and baseline with top-ranked sentences and thumbnails. They were interested what constituted a good surrogate, the relationship between the task and surrogate effectiveness, and how current search engine surrogates could be improved. They did not find significant differences in task completion time between the different layouts. Participants submitted more queries for the baseline layout than for the other three types of layouts. Clickthrough data indicated that participants were more likely to make relevance judgments based on the surrogates for the augmented layouts than for the baseline one. The higher clickthrough rate for the baseline layout may not be an indication of a higher number of relevant documents being found, but the need to examine the document before a relevance judgment can be made.

How to present

This section examines issues of result presentation, such as grouping or clustering, layout, and relevance ranking. If the previous section was on surrogates and what should be in the surrogates, this section examines how to present surrogates and their components. These represent efforts at finding alternatives to the current default ranked list presentation.

Zamir and Etzioni (1999) compared post-retrieval clustering to ranked list display of search results. In Grouper, their clustering interface, users followed more documents than in a ranked list display, and were more likely to follow multiple documents. Finding the first few interesting documents required more effort in Grouper than in the ranked list presentation, but appeared to require less effort after those first few documents. The authors surmise that time and or effort is spent understanding the clusters, but after that

point clusters help users find the information faster. They found that the clustering interface was not suitable for all search tasks. Chen and Dumais (2000) evaluated an interface that organizes Web search results into hierarchical categories, again comparing it to a ranked list interface of search results. There were significantly more tasks on which the users gave up in the list interface than in the category interface. They also found that users took 50% longer to find answers using the list interface.

Capra, Marchionini, Oh, Stutzman, Zhang (2007) examined the effects of structure and interaction style on search tasks. Three user interfaces were tested in the study: standard web site, hierarchical text-based faceted interface, and a dynamic query faceted interface. In the first study, no significant differences were found for either of the two main effects- task type and interface - for the four measures of accuracy, confidence, satisfaction, and mental effort. In the second study, they found a general preference for the standard web site over the other interfaces. Participants liked some features of the facet interface, but preferred the familiarity afforded by the standard interface.

Clusty (http://clusty.com) and KarTOO (http://www.kartoo.com) are two search engines – or more precisely, metasearch engines collecting search results from several different search engines – that display search results in alternative ways. Clusty (previously known as Vivisimo) is a Web metasearch engine that dynamically clusters users' search results (Koshman, Spink & Jansen, 2006). In addition to the typical list of search results, the user is also shown the clusters identified along with the cluster labels and number of search results in each cluster. Clicking on a cluster label displays the search results in that cluster. KarTOO calls itself a visual metasearch engine. It displays search results retrieved from other Web search engines on a two-dimensional map,

organizing the search results by topic. Each search result is represented by a ball, with the size of the ball corresponding to the relevance of the result to the query. There have not been as yet user studies comparing these search result presentations to the typical list presentation.

Perhaps because of the dominance of the list presentation for SERPs, we have not seen as much variety in search results in presentation in recent years. Alternative displays, such as faceted or hierarchical ones, require significant redesign of data representations at the system level. In studies of IR systems implementing facets, categories, and hierarchies, these have not been tested on large heterogeneous collections such as the Web, but on small, focused document collections. It may not be practical to create metadata to the extent some of the novel interfaces require, for collections such as the Web. The dominant search engines all use list presentation (Höchstötter & Lewandowski, 2009). While novel ways of presenting search results can be expected for non-text materials, the list form is still prevalent for SERP on the Web. As Web searchers are now used to this format for SERP, the promise of being more effective searchers for future searches may not be enough of a lure to make people learn to interpret a novel SERP.

2.3.2 Query reformulation

Any Web search session starts with the user thinking of some query terms, and entering them into the search box. If the user selected their query terms well, then that particular information search episode can end without the user reformulating their initial query. But approximately half of all Web users find they have to reformulate their initial queries (Spink, Jansen, Wolfram, & Saracevic. 2002). In general, people search on the

Web because they have an information need. This information need is ultimately expressed as a query to a search engine. But, as Belkin (2000) notes, "it is difficult for people to specify what they don't know" (p. 59). Even if a person has some idea of what he or she is looking for, as in the case of a known-item search, the person's representation of their information need may not necessarily match up with the representation in the IR system. That is, the words the searcher uses to represent their information need may not be the ones used in the documents that can address that information need. This mismatch may arise from the person simply not knowing the appropriate terminology, or as a byproduct of reducing a complex information need to a few query terms. Query reformulation may then result once the searcher ascertains whether their original query formulation was "in the ballpark" based on the search results. Thus, query reformulation is unavoidable much of the time, and given how frequently it occurs, it seems natural interactive IR systems should provide some support for this activity.

Two approaches have been examined to support query reformulation: relevance feedback and term suggestion (Belkin, 2000). In relevance feedback, the system reformulates the query based on feedback provided by the user on the relevance or non-relevance of retrieved information objects. The user is not directly modifying the query, but provides input to the system in the form of relevance judgments, which are then used by the system to reformulate and rerun the query. Belkin considers relevance feedback to be a system-controlled type of interaction. In term suggestion, the system shows the user new terms that can be used to reformulate the original query. These new terms are derived from the original query and/or documents retrieved by the original query. The

user may choose to reformulate the query based on these term suggestions. In contrast to relevance feedback, term suggestion is a user-controlled type of interaction.

A number of studies on IR systems implementing relevance feedback indicate that relevance feedback features are not used in interactive searching (Beaulieu, 1997; Belkin et al., 2001; Ruthven, Tombros, & Jose, 2001). One factor in the lack of use may be the lack of control or visibility provided to the user during query reformulation. Koenemann and Belkin (1996) compared different levels of visibility and interactivity in systems implementing relevance feedback for automatic query reformulation. They found that users performed better and preferred the system that allowed them to manipulate the list of suggested terms. Another factor may be the "habit-driven behavior" (Anick & Kantamneni, 2008) of searchers. Relevance feedback is not a feature typically encountered in search engines, whether general ones such as Google or site-specific ones. In addition, explicit relevance feedback requires the user to provide relevance judgments to information objects, which may seem like an extraneous activity to the user. Implicit relevance feedback has been explored as an alternative to the traditional explicit relevance feedback model. One approach to implicit relevance feedback is to use measures based on interaction with documents (e.g. reading time, scrolling, mouse clicks) as the feedback for query reformulation (Kelly & Teevan, 2003).

In contrast to relevance feedback, term suggestion has been adopted by the major search engines. For example, entering the term "vancouver" in the Google search box results in a list of suggested query terms being displayed, with suggestions such as "vancouver bc" and "vancouver weather". Yahoo! allows the user to turn off the term suggestion feature as part of the term suggestion display, giving the user even more

control. One question that arises is the source of these term suggestions. Major search engines such as Yahoo! appear to be using two sources for these term suggestions: frequently occurring queries mined from search logs, and terms derived from top search results (Anick & Kantamneni, 2008). Researchers in term suggestion or query refinement have examined techniques deriving terms from subject thesauri (Schatz, Johnson, Cochrane, & Chen, 1996), documents in the search results (Vélez, Weiss, Sheldon, & Gifford, 1997), and search engine query logs (Huang, Chien, & Oyang, 2003). The use of query logs has been further expanded with incorporation of a user's query history and clickthrough data from the search session (Sriram, Shen, & Zhai, 2004). Belkin et al (2001) found a preference for term suggestions from the top-ranked documents retrieved by a query compared to term suggestions based on documents selected as relevant by the user. In the latter case term suggestions are derived from a smaller set of documents than the former.

As to whether term suggestion helps users, the results appear mixed. In laboratory experiments, users were positive regarding term suggestion features yet did not use them as often as researchers were expecting (Belkin et al, 2001). Users expressed preference for term suggestion implementations that did not intrude on their search (White & Ruthven, 2006). Bruza, McArthur, and Dennis (2000) found that term suggestion added to the cognitive load of searchers. The suggested terms were additional data for the searcher to evaluate and make a decision on whether to use or not.

In contrast to these laboratory experiments, Anick and colleagues examined use of term suggestion features in production systems in the wild. Anick (2003) examined use of term suggestion as implemented in AltaVista by examining logs covering a contiguous

five-day period. Visitors to the AltaVista site were assigned to feedback (term suggestion) or baseline (no term suggestion) conditions. While both groups showed similar rates of reformulation, only about 6% of initial queries in the feedback group were followed by refinement using feedback terms. Anick also examined reuse of the feedback feature over time by capturing data from the feedback group over a 2-week period. Nearly half the users used the feature more than once during the two-week window. When the data was analyzed as search sessions, where a search session may contain multiple queries, the percentage of refined sessions using feedback increased from 25% to 38% during the study period. Anick and Kantamneni (2008) found similar results in their study of the use of query refinement features in Yahoo!'s Search Assist interface. While overall use was low when compared to the total number of queries, use increased over the month-long study period. Increased uptake and the relatively large percentage of people reusing the feature, indicate that term suggestion provides benefits worth the additional cognitive load for some users.

2.4 Image retrieval on the Web

With the advent of the Web and increased access to fast network connections, users can directly access a multitude of images online. The Web has allowed ordinary users to share the images they created, such as photographs and drawings, through their own web pages or specialized sites. The Web has also created a demand for images that users can appropriate for their own use, such as for use in their own web sites or participation in online communities. The Web has made it possible for ordinary users to search for and retrieve images directly, without the use of intermediaries, and to do so for purposes such as entertainment and sharing experiences with others.

This situation presents a marked contrast to that faced by image retrieval researchers prior to the Web. Prior research focused on image retrieval occurring in work settings (Ørnager, 1997; Conniss, Ashford, & Graham, 2000), specialized collections (Keister, 1994; Armitage & Enser, 1997), and search through intermediaries (Enser, 1993; Keister, 1994; Fidel 1997). Rasmussen (1997) notes, "relatively little is known about information-seeking behavior as it relates to images." (p. 173).

Image retrieval research has proceeded along two parallel paths, content-based and concept-based image retrieval. In content-based retrieval, measurable attributes in an image, such as color or shapes, are used as the basis for search and retrieval, applying pattern recognition and machine learning techniques. Concept-based retrieval relies on text about the image, usually in the form of image metadata created by a specialist other than the image creator. It is not the image itself, but the text about it, that is the basis for search and retrieval. Different research communities have pursued these two approaches: mainly library and information science in the case of concept-based image retrieval, and computer science in the case of content-based image retrieval (CBIR). This divide persists, although there are continuing calls to integrate the two approaches and communities (Enser, 2005; Jaimes, 2006).

These two approaches should be considered in light of what CBIR researchers refer to as the semantic gap. In their survey of CBIR, Smeulders, Worring, Santini, Gupta and Jain (2000) define this as "the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation" (p. 1353). Analyses of user queries show it is rare for users to formulate queries in terms of primitive features of images (Hare, Lewis, Enser &

Sandom, 2006). Concept-based image retrieval attempts to address queries in terms of the meaning and subject of images. While it may appear that this approach is more suited to actual user queries, a number of studies have shown a gap between image indexing and user query terms (Markey, 1986; Enser, 1993).

Image retrieval on the Web has taken a direction different from these two approaches. The retrieval approach taken by a Web image search engine such as Google Images takes advantage of the fact that images on the Web have some text associated with them, with the file names at the most minimal level. Google image search examines the text around the image, the image file name, HTML ALT tag if specified and links to the image to match against the user query. Photos uploaded to Flickr, an online photo sharing site, have two types of metadata associated with them: author-created, which includes title, description, and tags, and are entirely optional, and automatic, which is embedded in the photo file itself if the photo was taken by a digital camera, and includes information about the type of camera, exposure, and date and time. The textual data supplied by the user is the basis for searching. Browsing options, such as "interesting" photos, use information such as "where the clickthroughs are coming from; who comments on it and when; who marks it as a favorite; its tags and many more things which are constantly changing" (Flickr, 2006). Flickr, then, implements concept-based retrieval with the image creator also supplying the metadata in their own words, not professional indexers using controlled vocabularies.

Social navigation must be considered as another access mode or mechanism for image retrieval – as the popularity of web sites such as Reddit, MetaFilter, or YouTube show, people want to look at what (many) other people have seen. Such sites are

examples of social navigation tools, "designed to enable users to be aware of, and be guided by, the activities of others during information seeking." (Foster, 2006, p.349) User behavior with respect to information objects becomes another mode for searching and browsing, not just the information itself. For example, in YouTube this can mean providing "Videos Being Watched Now" as a browsing option.

There have been a number of comprehensive reviews pertaining to image retrieval. Enser (1995; 2008) and Rasmussen (1997) focused on the concept-based image retrieval approach. Smeulders, Worring, Santini, Gupta and Jain (2000) and Datta, Joshi, Li, and Want (2008) focus on computational methods in CBIR, while Rui, Huang, and Chang (1999) provide a system-based review. Kherfi, Ziou and Bernardi (2004) survey existing experimental Web image retrieval systems. Lew, Sebe, Djeraba and Jain (2006) examine content-based multimedia information retrieval, covering image, video, and audio retrieval.

In this section, different approaches to categorizing the images used in image retrieval research are presented. Then user studies are discussed in terms of the categories of users examined. Lastly user interactions with image retrieval systems are examined. CBIR literatures on specific systems or algorithms without a user evaluation component are not covered.

2.4.1 Images

Intuitively it would appear that a collection of family photographs intended for personal use will be organized and searched for differently than a university art history department's collection of high resolution slides of paintings from 16th century Italy. A graphic design studio's interest in images will lie more with how they can be used as part

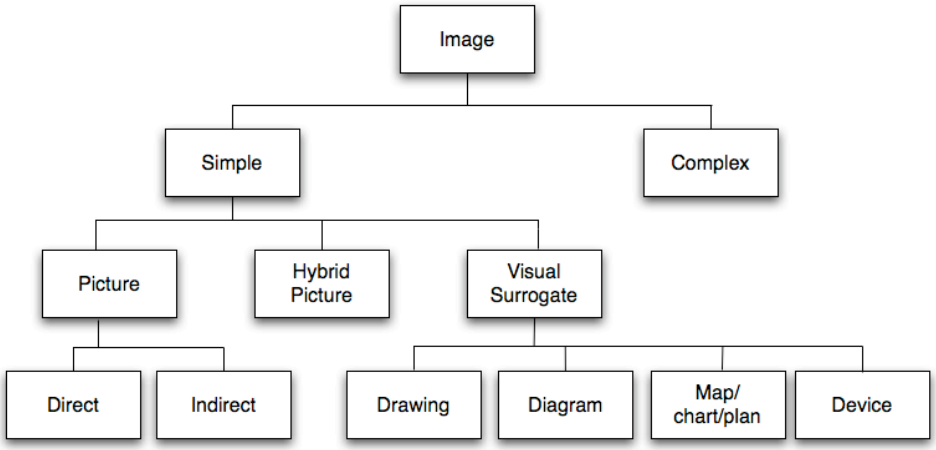
of brochures and web sites rather than as objects of inherent interest, as would be the case with the source materials for the art history slide collection. In addition to the context of use, images may be categorized by inherent properties – photographs are different from architectural drawings, maps are different from X-ray images, and so forth.

Whether implicitly or explicitly, such distinctions have been used by researchers to define more precisely the problem to be researched, or to make the problem more tractable. For example, one commonly used type of categorization is that of art and non-art images. Aside from stating which type of image will be used in the study, no further explanation of the distinction is usually provided, the assumption being that these are clear distinctions without need for further explanation. Even an art historian, Elkins (1999), regards this distinction as self-evident before embarking on his exploration of non-art images as objects of study for art historians. Some of the non-art images he examines are scientific diagrams and crystallography images.

Enser, Sandom, and Lewis (2005) present a taxonomy of images based on CBIR research, intending it to inform their research on the semantic gap in the context of still images. This taxonomy is based on a survey of image retrieval activity, with particular focus on types of image and user. Definitions are provided in the table below. The different types of images presented at the lowest level (Figure 8) are distinguished by inherent properties of images that can be recognized by computers. A drawing remains a drawing, whether it was originally intended as a scientific illustration and later came to be regarded as art. Ordinary photographs, regardless of subject matter or intention, are examples of direct pictures. X-ray images or radio telescope images are examples of indirect pictures. CBIR systems have been implemented that can distinguish drawings

from photos, or can recognize trademarks. The intended or potential use of an image is not used as a basis for categorization.

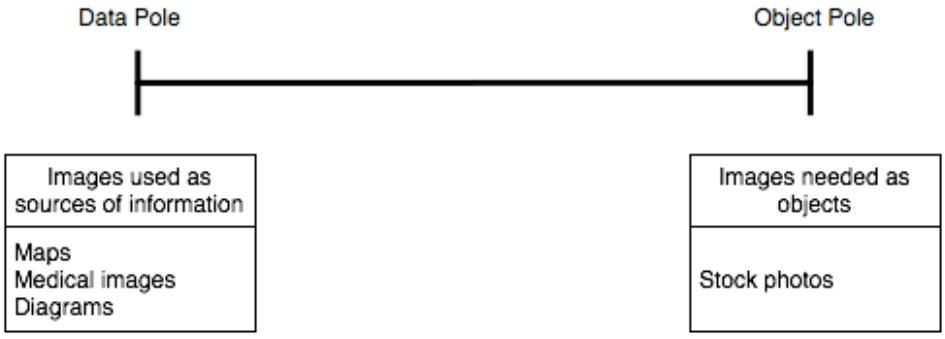
Figure 2.8. Taxonomy of images (Enser, Sandom & Lewis, 2005)



Fidel (1997) proposes a conceptual framework for image retrieval tasks, which provides a way to characterize images solely on intended use. Image retrieval tasks lie in a continuum, with the Data Pole at one end and the Objects Pole at the other end (Figure 9). For image retrieval tasks in the Data Pole, images are used as sources of information; types of images commonly associated with these tasks are maps and medical images. In the Object Pole, images are needed as objects, that is, as elements of potential products, in Turner's terminology. Stock photos are commonly associated with the Object Pole. Fidel notes most tasks and associated images will lie somewhere in-between the extremes – users may retrieve images both as information sources and as objects. An academic art historian searching for images to include in lecture slides is regarding these images as both objects (e.g. elements to be included in a final product, the lecture slides) and as information sources (e.g. information on artistic practices of a particular era or geographical region, common themes in an artist's body of work). The same image may

be associated with the Data Pole or the Object Pole, depending on the image retrieval task and user.

Figure 2.9. Continuum of image retrieval tasks (Fidel, 1997)



2.4.2 Image retrieval interaction

This section will focus on studies focusing on interaction with an image retrieval system, where the user interacts directly with an automatic system. Image retrieval system evaluation studies are also discussed in this section. Another perspective from which to study image retrieval is to examine who looks for what images and how. Some of the questions examined then are:

- What kinds of queries do different types of users submit?
- What is their information-seeking process with respect to images?
- What are their relevance criteria?

The population studied in such user studies has been users of particular image collections (Enser, 1993; Keister, 1994; Armitage & Enser, 1997; Fidel, 1997) or members of occupational groups, such as art historians (Hastings, 1995; Markey, 1986) or journalists (Ørnager, 1997; Markkula & Sormunen, 2000). Image retrieval is usually examined in a work context (Connis, Ashford, & Graham, 2000). More recent studies

have asked users to search for images on the Web in controlled experimental settings (Hollink et al, 2004).

CBIR studies (McDonald & Tait, 2003; Fukumoto, 2006) usually have two image search tasks, performed by the same subject:

- Known item search: subject is first shown an image, then asked to find the image in the image retrieval system; researcher may specify search strategy to be used (e.g., browsing, sketching an image) or leave it up to the subject.
- General search: subject is asked to find image(s) fulfilling a particular purpose, such as to accompany a newspaper article or to include in a web page on a particular topic.

More recently there have been a number of studies analyzing Web image searches, examining search logs (Goodrum & Spink, 2001; Jörgensen & Jörgensen, 2005) or examining digital reference queries (Cunningham, Bainbridge & Masoodian, 2004; Goodrum, 2005). In contrast to previous user query studies, these studies examine queries that come directly from the user, and not search forms or logs filled out by an intermediary. Thus the queries in previous studies may be the intermediary's interpretation of the user's search request.

Image searches on the Web

Goodrum and Spink (2001) examined 33,149 user image requests submitted to the Excite search engine. Users averaged 3.36 queries per session and 3.74 terms per query, with a high rate (59.6%) of search modification. Most terms appeared infrequently. Jörgensen and Jörgensen (2005) analyzed search logs from a commercial image provider. Users of this subscription service were image professionals involved in areas such as

advertising, marketing, and graphic design. Across the two sample sets, the mean number of search terms per query was close to 2, and 61.7% of total queries were modified. Defining a successful search as one that resulted in preview or download action, 26.3% of queries across the two samples were successful. They found this group of users used more descriptive and thematic queries than earlier research had shown. They also heavily employed Boolean searching, albeit ineffectively. Browsing was heavily used, occurring in 85.6% of sessions.

Cunningham, Bainbridge and Masoodian (2004) analyzed a set of 404 queries submitted to Google Answers (http://answers.google.com). This is Google's 'ask an expert' service, where a user submits a question along with how much they are willing to pay for an answer, and answers can be submitted by approved Google 'researchers'. Past queries and their answers are publicly available, arranged into a number of categories, the question category being selected by the person posing the question. Google answer queries averaged 62 words, in contrast to the 2 to 4 terms found in search engine query logs. Queries for analysis were selected from the Visual Arts category in Google Answers. The majority of queries involved two-dimensional media such as paintings and photographs. About 80% of the queries provided some bibliographic metadata, such as artist, date or title, providing justification for the use of these access points in image retrieval systems. Also mentioned in queries are colors used in a work (10.9%) and examples of the desired work (10.4%), indicating the utility of CBIR methods for some users. The contents of the picture are mentioned in 41.6% of the queries, in contrast to 6.2% for abstract concepts. This provides support to including descriptions of visual

elements in image records (Keister, 1994) and prior research indicating the low utility of subject headings for images (Markey, 1986; Enser, 1993).

Goodrum (2005) examined questions submitted to the Virtual Reference Desk (VRD) and AskERIC, both projects supported by Syracuse University and the U.S. Department of Education. Of 7,257 digital reference requests, 590 requests were identified as being requests for visual materials. Requests had 21 terms on average. The majority of requests (77.3%) were for images "meant to be faithful reproductions of reality," such as photographs, drawings, or paintings. Images categorized as "Models," which included maps, diagrams, timelines and architectural plans, accounted for 24.4% of requests. Goodrum acknowledges these queries may reflect the particular needs of the educational community who use AskERIC and VRD, and not that of the population in general.

These query analysis studies show that in general users search in terms of the literal contents of the image, not according to the subject or emotional impact. The type of query mechanism offered (natural language/asking a human, search box/search engine) significantly affects how the query is formulated in term of length. The one study with somewhat different results (Jörgensen & Jörgensen, 2005), such as in the use of thematic queries and Boolean search, involved one particular user group and a commercial system targeted to that user group. This hints that studies looking at large numbers of queries from a heterogeneous group of users may actually be misleading regarding user needs and behavior. For example, there may be significant differences in query type between art historians and graphic designers, which may not be apparent in studies that aggregate all

queries and analyze them as a group. This indicates a need to study particular types of users in depth.

Relevance criteria for images

User relevance criteria for images have received relatively more researcher attention than criteria for music or video. As part of their study of journalists' image searches, Markkula and Sormunen (1998) also examined their photo selection criteria. Selection or relevance criteria changed according to the stage in the search process. Journalists engaged in a two-stage search process, first selecting a set of candidate photos, and then selecting a final photo to be published from this candidate set. Topicality was the first criterion applied for selecting candidate photos, and the associated caption text was the most important source of information for judging topicality. Interestingly the cost of the photo was one of the criteria applied after topicality. But, final selection was based solely on the visual attributes of a photograph.

Hirsh (1999) specifically examined relevance criteria in her study of children's information seeking on electronic resources. This study is also notable for requiring text and image search for the same information need and directly comparing text and image relevance criteria. Hirsh applied Kuhlthau's ISP model and interviewed students at two different points in the search process. While topicality was the most frequently mentioned relevance criterion for text, how interesting a picture was the most frequently mentioned relevance criterion for images. Hirsh observed a decrease in the importance of topicality as a text criterion in the later stages of a search, while the importance of interestingness increased for images.

Choi and Rasmussen (2002) investigated relevance criteria for images related to American history, as applied by scholars of American history. Participants were asked to rate the relative importance of nine criteria at two stages of the information retrieval process: before search and after examination and identification of relevant documents. The researcher carried out the searches and presented participants with the result set. Her findings indicate that participants needed both image and textual description to make a relevance judgment. The relative importance of relevance criteria changed across the two stages, with topicality decreasing while novelty and accessibility increased in importance in the later stage. As in the Markkula and Sormunen study, topicality was not as important a criterion once a candidate set of images had been selected.

The above studies were carried out prior to the explosion of user-generated content on the Web, which has come to include images, video and music, in addition to blogs and social networking sites. Web users seek images not just for academic or work purposes, but for their own everyday uses. There is as yet no research on the image relevance criteria of Web users who are looking for images for reasons that are not academic or work-related, such as reappropriating images for use in their own blogs or social networking site profiles.

Chapter 3

Conceptual Framework

This chapter introduces the research framework that guides this study with respect to understanding how tags are used during the search process. Ultimately, the study seeks to understand the utility of tags for Web users other than the taggers, and one area in which tags are expected to be of use is in the online search process. First, the definition of tags and tag use that will be used in this study is presented, followed by the model of tag use in interactive information retrieval guiding this study.

3.1 Tags and their use

3.1.1 Tags

As discussed in Chapter 2, tags have been conceptualized in diverse ways that reflect the intended use, in the case of websites, or analysis focus, in the case of researchers. The general understanding of tags that has emerged is that, in the context of social tagging, tags are descriptive terms people attach to online content. We will use this as our working definition of tags in this study, adding that the tags and the content they describe should be publicly viewable. This definition excludes tags used in private collections such as email or personal photographs, but does not necessarily exclude tags used in sites open only to registered members. Lastly, we note that this definition

specifies that tags are generated by people, although they may make use of system features designed to aid tagging, such as recommended tag suggestions. This definition does not address how tags will be used, so their purpose remains open-ended.

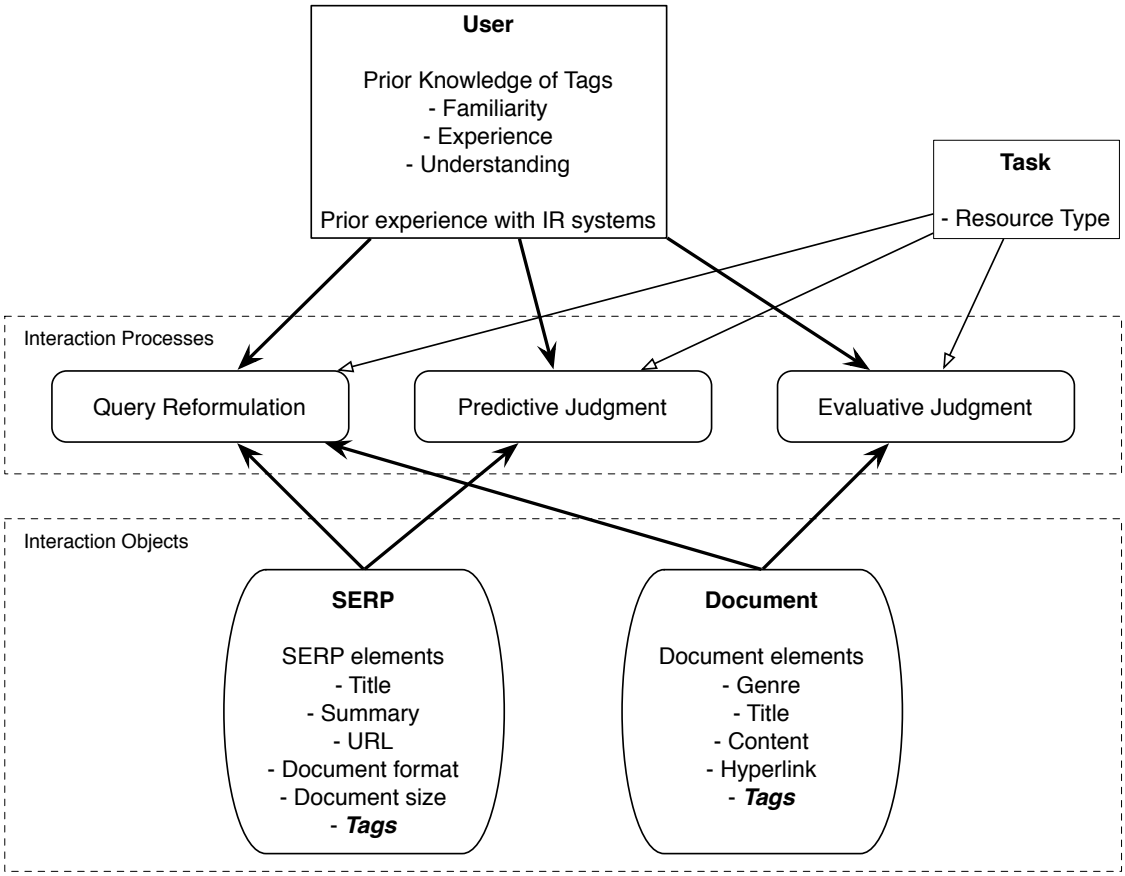
3.1.2 Use of tags

We distinguish tag production from tag consumption. Tag production is the act of attaching tags to online content. Tag consumption is the use of existing tags excluding tag production. *Tag use* will refer to tag consumption activities. In this study, tag use is restricted to the use of tags in the interaction processes described in Section 3.2.

3.2 Model of the role of tags in interactive information retrieval

The model presented in Figure 3.1 proposes that the Web search process using an interactive IR system has two components, interaction processes and interaction objects. The model is inspired by Belkin's (1996) model of interaction with texts. The user, or searcher, engages in the interaction processes of query reformulation, predictive judgment, or evaluative judgment during the course of the search. The interaction objects involved in an interaction process differ for the interaction processes. SERPs are involved for query reformulation and predictive judgment, while Web pages are involved for query reformulation and evaluative judgment. Tags, as one of the information elements present in an interaction object, are thus one of the elements with which a user can interact during an interaction process. Interaction objects are distinct from information objects, in that some Web pages are information objects, but SERPs are not necessarily information objects. The terms document and information object will be used interchangeably.

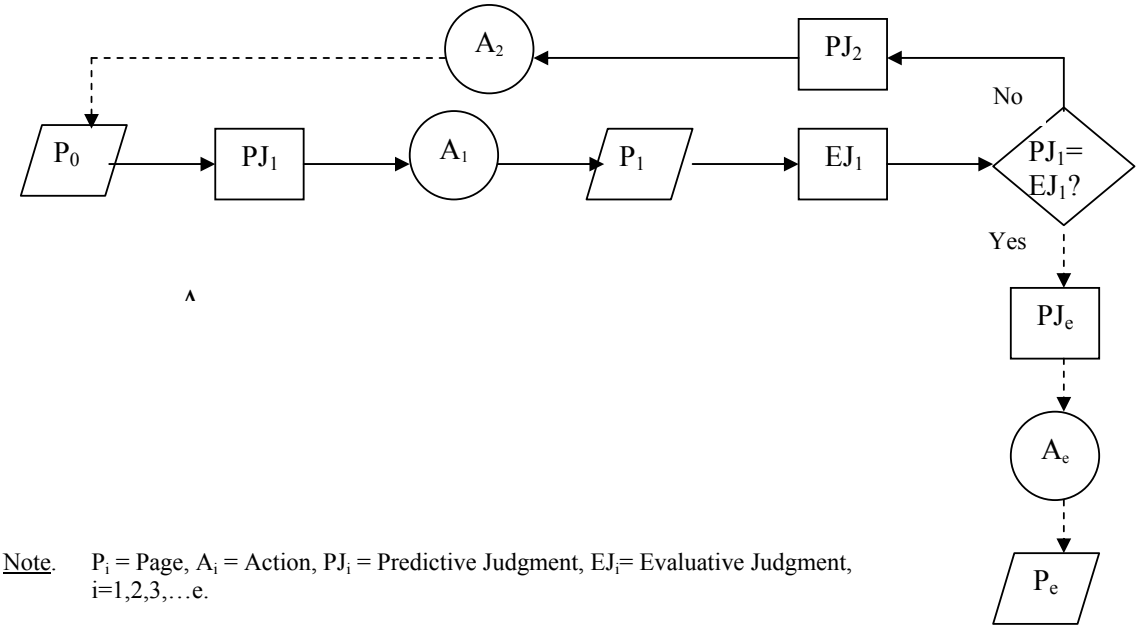
Figure 3.1. Use of tags in interactive information retrieval



The model presented in Figure 3.1 does not include initial query formulation, as the study is concerned with interaction processes involving the SERP and the Web pages that become available after the initial query has been submitted to the information retrieval system. In initial query formulation, the searcher is often starting from an "anomalous state of knowledge" (Belkin, 1982) on the search topic itself, and likely is not aware of what types of information objects exist that may be related to the search topic. By the time the searcher is expressing their reformulated query, they have engaged in several interaction processes with information objects through the IR system. The cognitive state of the searcher is very different for initial formulation and query

reformulation. While this model presents the components of the Web search process, it does not illustrate how a searcher shifts from one interaction process to another over the course of their search session.

Figure 3.2. Judgment processes on the Web (Rieh & Belkin, 2000)



Note. P_i = Page, A_i = Action, PJ_i = Predictive Judgment, EJ_i= Evaluative Judgment, i=1,2,3,…e.

A typical web search session involves two kinds of judgment, predictive judgment and evaluative judgment (Rieh, 2002). Predictive judgment takes place during examination of the search results, which present surrogates of the actual documents. Evaluative judgment takes place when examining the actual documents. When examining the search results, users are making a judgment regarding the usefulness, relevance, pertinence or interestingness of the search results presented. This judgment leads to action, that of clicking on a link to examine the document it links to, or reformulating the query.

Rieh and Belkin (2000) characterized the search process on the Web as consisting of a series of actions (A_i), which are associated with predictive judgments (PJ_i), and

evaluative judgments (EJ_i), which occur after viewing a Web page (P_i). This process model (Figure 3.2) illustrates the process of making several decisions and judgments during the interactive search process. The searcher, after viewing the initial page of search results (P_0), makes predictive judgments (PJ_1) on which Web page to view or not, which results in an action (A_1). If the action was to click on a search result, this leads to the display of a Web page (P_1), and the searcher makes an evaluative judgment (EJ_1) of relevance. In combination with the model of Figure 3.1, the search process is described both in terms of components (Figure 3.1) and iterative process (Figure 3.2). When the two models are considered together, one of the possible actions is query reformulation, and the Web pages viewed can be either the search results page or the actual document page.

Ruthven (2005) suggests viewing relevance as a "process of human decision making" (p. 61). Document and surrogate elements serve as input for making judgments of usefulness (or relevance), where the user applies relevance criteria based on task and resource type. What research on relevance has shown is that there are a limited number of relevance criteria employed across different types of information needs and contexts (Schamber, 1991; Barry, 1994; Choi & Rasmussen, 2002; Yang & Marchionini, 2004; Balatsoukas & Ruthven, 2010). For example, topicality emerged as the primary relevance criteria whether evaluating Google search results (Balatsoukas & Ruthven, 2010), video (Yang & Marchionini, 2004), or images (Markkula & Sormunen, 1998; Choi & Rasmussen, 2002).

During the Web search process a searcher engages in a combination of interaction processes. The model assumes the interaction processes are occurring within one Web IR

system. Depending on the Web IR system the searcher is using, interaction with the document can occur inside or outside of the IR system. For example, when a Flickr user clicks on a photograph in the Flickr SERP, the user lands on the Web page for that photograph, still within the Flickr system. On the other hand, when a searcher using Google clicks on a link, the user "leaves" Google and is no longer interacting with it. The model assumes interaction processes of the former kind.

3.2.1 Interaction processes

Query reformulation

Tags can be a source of query terms for refining the initial search query. In some social tagging site implementations, a tag can itself be the query. For example, clicking on a tag in YouTube results in that tag being entered into the query box by the system and a search is carried out. From the perspective of the user, what happened was that clicking on a tag resulted in a SERP being displayed, where the query term used was the tag, and which is now shown in the SERP query box. Tags can also be used indirectly in query reformulation. The user may simply type in a tag term into search box exactly as is, or may add a tag term to an existing query. Another way to use tags indirectly is to base the query terms on tag terms seen, for example, using "Brooklyn" as a query term after seeing the tag *manhattan*.

Tags are expected to play a larger role in query reformulation from the SERP than in query reformulation after examining a Web page or document. This is because in the latter case the searcher has had a chance to examine the actual document and encounter a larger set of words than provided in the SERP.

Predictive judgment

The primary value of tags in relevance judgment is expected to be for predictive judgment when evaluating search results in the SERP. One aspect in which tags can provide additional information is that tags have been applied after evaluation of the information object as a whole, while in the SERP the searcher is only provided a snippet of the information object. Thus tags can provide an overall view of the document not available from the SERP snippet alone. In addition, tags provide information not typically found in metadata or the content itself, such as use information (e.g. 'toread') or evaluative information (e.g. 'boring') (Kipp, 2007). Another aspect in which tags can provide information on the object tagged is that a person found it worthwhile to invest the effort to tag it. A tag is in this sense a vote of confidence that the information object is worth remembering or revisiting, in the case of Delicious, and a signal that the information object has been put out there to be found, as in the case of Flickr. For non-text materials, tags can provide information not evident from thumbnails or screenshots used as surrogates, such as the location in which events in a photograph took place, the photographer or director, or even the type of equipment used to produce the material.

Evaluative judgment

Tags likely will not play as large a role in evaluative judgment of a document, as there are other more prominent sources of data in the Web page itself, such as text or images, from which to make a judgment. In addition, many Web pages or documents do not display tags or are not designed to display them (e.g. an online PDF document). Even on sites implementing social tagging, such as Flickr or YouTube, tags are not prominently displayed on the actual photo or video page, making them difficult to be

noticed. In YouTube's case the default setting is to not display the tags, and a user who wants to see them has to click on the "More information" link. Studies have shown that for image searches, metadata elements used to make relevance decisions change depending on the stage of the search process (Markkula & Sormunen, 1998; Choi, 2002). Similarly, the role of tags in relevance judgment is expected to be different for predictive judgment and evaluative judgment.

3.2.2 Interaction objects

Interaction processes take place with interaction objects, in this case the SERP and the document. Each interaction object has elements or constituent pieces. The term document is used to refer to information objects, whether in text form or non-text form, such as multimedia and software. Web pages are a particular type of document, written in HTML and that can be displayed in a browser. A SERP can link to documents that are not Web pages, such as software or PDF files, in addition to Web pages.

In the model presented in Figure 3.1, tags can be part of both the SERP and the document, but are not required elements. Tags can be present in one and not the other, or not be present in either of the interaction objects. The SERP is used in the interaction processes of query reformulation and predictive judgment, and the document is used in the interaction process of evaluative judgment and query reformulation.

3.2.3 Prior Knowledge

Prior knowledge is a concept that has been used in fields such as education and consumer research to account for the fact that when people encounter new situations, concepts, or products, often they are not doing so as *tabula rasa*. Prior knowledge is

knowledge that is available prior to a certain learning task (Dochy, Segers & Buehl, 1999), and is sometimes equated with expertise (Wood & Lynch, 2002) or familiarity (Park & Lessig, 1981; Rao & Monroe, 1988). Prior knowledge can both help task performance, by allowing the person to process information faster, and hinder it, through overconfidence (Wood & Lynch, 2002). While prior experience is related to both expertise and familiarity, it appears to be different from these two, in that experience does not necessarily lead to expertise, and familiarity, or the extent of awareness of a product or technology is not the same as having experience with it. Kerstetter and Cho (2004) proposed that prior knowledge is a multidimensional construct, composed of three dimensions: familiarity ("how much an individual knows or perceives"), expertise ("the ability to apply a solution to task-related problems"), and past experience ("previous purchase or usage of the product").

In this study prior knowledge is conceptualized as a multidimensional construct with the following dimensions: familiarity, experience, and understanding. It differs from Kerstetter and Cho's construct in making it applicable to an information object, and not a consumer product, as is the case with their definition. Familiarity with tags is defined as how much an individual perceives they know about tags. Experience with tags is defined as past exposure or use of tags, and understanding is what the individual knows about tags. Prior knowledge of tags is expected to influence how tags are used in an unfamiliar or not previously encountered IR system with tags.

Users bring their experience of using other IR systems to searching on the Web (Kim, 2001; Slone, 2005). So in addition to prior knowledge of tags, we also consider prior experience with other IR systems, in particular library database systems and

OPACs, and its relationship to tag use. Kim (2001) found that previous experience with library database systems was the main predictor for Web search efficiency. Library database systems provide a wide range of advanced search features, including searching using controlled vocabulary and other metadata elements. Experience using such systems may provide a different type of prior knowledge than using Web search engines.

3.2.4 Task characteristics

Characteristics of the search task can also influence search interactions. The particular task characteristic of interest in this study is resource type, or whether the search is for text documents or for images. As described in the literature review in Chapter 2, concept-based image retrieval is the preferred mode of image retrieval for users in general. To support this kind of retrieval images must have text related to the image attached to them. Tags are one such type of text.

Tags are expected to contribute to relevance judgment, especially predictive, for images. Studies of image retrieval show that it is a 2-step process (Markkula & Sormunen, 1998), where the searcher first identifies a set of candidate images, then makes a final selection from these candidate set of images. The selection or relevance criteria are different for these two steps – while information about the image is valued in the first step, in the second step inherent qualities of the image are more important. As tags provide some information about the tagged image, they are expected to help in the selection of candidate images for further evaluation.

Tags are expected to contribute to image retrieval to a different extent compared to text search when present in the SERP. In the case of text search, a text snippet is shown in addition to the title and tags. In the case of image search, typically the image

title or file name is shown along with a thumbnail image. Tags are one of the few sources of textual information on an image search SERP, and so may play a larger role in predictive judgment and query reformulation in image searches than text searches.

3.3 Summary

The interactive information retrieval process is conceptualized as the user engaging in interaction processes with interaction objects. The interaction processes are query reformulation, predictive judgment, and evaluative judgment. A user makes predictive judgments of relevance when interacting with the SERP, and evaluative judgments of relevance when interacting with the document. In query reformulation a user can interact with the SERP or the document. During an interaction process, users interact with one or more elements of the interaction object. The outcome of the interaction is based on decisions made using these interaction object elements as input.

The role of tags in the interactive information retrieval process is as one of the elements of an interaction object, in this case the SERP and the document (Web page). When present in the information object, it is up to the user whether to include tags or not in the interaction process at hand. Influencing their use of tags is their prior knowledge of tags, and prior experience using other information retrieval systems. Task characteristics also affect both interaction processes and how interaction object elements are used. In this model tags have the potential to be used in all three of the interaction processes.

Chapter 4

Methodology

In this chapter, the initial section presents the rationale for the interactive information retrieval experiment methodology adopted for this study. Findings from preliminary studies are presented to provide additional background on the selection of the methodology. Then the research design is presented, followed by descriptions of the collections, tasks, experiment system to be used in the study, and the study subjects. The next sections address the data collection procedure, data sources, and data analysis.

4.1 Rationale

The intent of this study is to understand how people use tags during the search process. In addition to gaining an understanding of when and how tags are used during the course of searching, the study also seeks to understand how people's searching behavior changes in the presence of tags compared to when tags are not present in the search interface. The study is not intended to be a feature evaluation, but rather, aims to obtain typical search behaviors for different types of searches and search interfaces. The study methodology thus demands both the controlled conditions and comparisons possible through a controlled laboratory experiment, but also realism in the search tasks assigned. The methodology of interactive information retrieval experiments addresses these issues and thus was selected as the methodology for this study.

According to Tague-Sutcliffe (1992), "A laboratory test is one in which the sources of variability stemming from users, databases, searchers, and search constraints are under the control of the experimenter" (p. 469). While a naturalistic study can capture behaviors as they occur in real life, it provides no control over the sources of variability identified by Tague-Sutcliffe. Thus comparisons, whether across users or across systems, are difficult. This study seeks to identify patterns of tag use that apply to different searchers, as well as factors relating to tag use in searches. Establishing relationships between factors through replication in a controlled setting can only be achieved through a laboratory experiment. By assigning search tasks and providing the experiment system to be used, the experimenter can control search constraints as well as the databases (system) used. Therefore an interactive information retrieval experiment is the most appropriate method to address the study aims.

4.2 Research Design

This study has a 2×2 factorial design. The 2 two-level factors are interface type (Tags, No Tags) and resource type (Text, Image) (Figure 4.3). A within-subjects design is used, where each subject carries out one search task for each interface and resource type combination, for a total of four search tasks. In order to avoid fatigue as well as learning effect from doing four searches in a row, counterbalancing was used. Counterbalancing was done through factorial rotation (Kelly, 2009). As there are four interfaces or systems – Text/No Tags, Text/Tags, Image/No Tags, Image/Tags – there are 4! or 24 possible system presentation orders. Given a system presentation order, there are 2 possible orders for the text tasks and 2 possible orders for the image tasks. Thus the total number of possible combinations is 96. For this study 48 of the possible system-task combinations

were used. Although tasks were a consideration for counterbalancing, the tasks themselves are not considered factors in this study, as text tasks are considered equivalent to each other and image tasks are considered equivalent to each other.

Figure 4.1. Research design

		System		
		No Tags	Tags	Total
Resource	Text	12	12	24
	Image	12	12	24
Total		24	24	48

4.3 The Collections

In general, studies evaluating the effectiveness of an IR system have used test collections, while studies examining certain aspects of the search process of Web searchers have used the Web at large or specific Web sites. This study examines the effect of social tags on the online search process, which places specific requirements on the collection to be searched. The most important one is that the majority of the content in the collection must have been tagged, preferably by multiple users. This presents a problem for using the Web at large as the collection, as only a fraction of Web content has been tagged. Where tagged content is likely to be found is in Web sites implementing social tagging, such as Flickr, or Delicious. As pointed out in the tagging section of the literature review, websites differ in terms of what information objects or resources are tagged, how, and by whom, as well as user motivations for tagging. For example, in Delicious users tag bookmarks, and several users can bookmark and tag the same Web page. Thus the set of tags for a given bookmark in Delicious is the aggregate of several users' tags. On Flickr, by contrast, in general users tag their own photographs. Although Flickr users who have been given permission to tag a photo by the photo's owner can do

so, in practice people rarely tag other people's photographs. In addition, Delicious users tag "selfishly", while Flickr users tag with findability by others in mind.

Some additional desired characteristics of collections for use in this study are that the search scope of the text and image collections is comparable, and the site content must be of general interest. What is meant by search scope is whether clicking on a search result has the effect of taking one away from the search site, or whether one remains on the site. A Flickr user who clicks on a link in the Flickr search results page remains within Flickr, and will do so for the entire search session unless they explicitly choose to go to another website. On the other hand, clicking on a link on a Delicious search page results in being taken to that Web page, away from Delicious. One of the goals of this study is to compare tag use in image and text search, and so differences in the collections other than the resource type should be minimized if possible.

Flickr was initially selected for use as the image collection, as Flickr is a widely used and well-established photo site with tagging functionality. Since Flickr already had many of the desired characteristics of a collection for image search, a comparable collection for text searching had to be found. After examination of several websites implementing social tagging, Ask MetaFilter (http://ask.metafilter.com) was selected as the site for text-oriented search tasks. The remainder of this section describes the characteristics of these two sites, Flickr and Ask MetaFilter.

4.3.1 Flickr

Flickr is perhaps the most well known photo-sharing site on the Web. In addition to photos, short video clips can also be uploaded. Flickr is also used to share images other than photographs, such as illustrations and scans of book pages. The large user base and

the large number of tagged or otherwise annotated images has led to Flickr being used as a source of collections for use in TREC-style evaluations of multimedia information retrieval (Huiskes & Lew, 2008) and interactive cross-language information retrieval (interactive track for the Cross Language Evaluation Forum, iCLEF).

Flickr itself is composed of collections of images. At the most basic level, each individual user's images form individual collections. There are also collections of images that result from Flickr users forming "groups" within Flickr. These collections, or "group pools", are composed of photographs group members share with the group. The associated tags also become part of the group photo pool. Another type of collection on Flickr are institutional collections, such as the Getty Image Collection, a collection of Flickr photographs selected by Getty Images to be part of a stock photo pool, or The Commons[2], referred to by Flickr as "the world's public photo collections". The Commons is a collection consisting of images posted to Flickr by 27 cultural heritage institutions worldwide, including the Library of Congress.

The Commons collection was selected for use in the study, as it was a public photo pool with the intention of wide dissemination and annotation by Flickr users. There were two reasons for not using the entirety of Flickr: (1) one user's set of photos can dominate search results if he or she uploads a set of photographs with similar text descriptions and tags – because group photo pools typically restrict the number of photos that can be added by one user, there is greater variety in the search results of a group photo pool; (2) as found by Ames and Naaman (2007), many Flickr users upload photos with findability by selected users in mind, and not that of the public at large, and thus text

[2] http://www.flickr.com/commons/

descriptions and/or tags may not be aimed at the general public. The Commons has a total of 26,493 images as of July 29, 2009, with the majority being historical images, and a substantial number are tagged or annotated in a foreign language or not following U.S. English usage. This limits the type of search tasks possible for the collection. The Getty Image Collection, with 31,697 images as of August 1, 2009, was also considered for use but rejected. Flickr images that are part of the collection are only identifiable as such by Flickr – there is no publicly viewable image metadata identifying the Flickr photographs as being part of the Getty Image Collection. While searches can be restricted to this collection, no other information can be collected, such as the tags for that collection.

Figure 4.2. Tag cloud for The Commons collection on Flickr

The tag cloud for The Commons is shown in Figure 4.4. Tags describe the originating institution for the photograph (e.g. *libraryofcongress*, *smithsonianinstitution*), the contents of the photograph (e.g. *horse*, *trees*), and the topic of the photograph (e.g.

baseball, *worldsfair*). It can be seen a number of tags are not in English, and despite Flickr allowing multi-word tags, phrases have been turned into one term by eliminating spaces between words.

4.3.2 Ask MetaFilter

Ask MetaFilter[3]is an offshoot of the community blogging site MetaFilter[4]. MetaFilter was started in 1999, and Ask MetaFilter followed in 2003. While relatively small in terms of registered users, MetaFilter is widely read and influential on the Web. For example, in 2009, The New York Times withdrew a photo essay published in its Sunday Magazine after a MetaFilter user posted evidence of digital alterations in the photographs in a MetaFilter thread discussing the photo essay (Collins, July 8, 2009). Ask MetaFilter (AskMe for short) is considered one of the premier question-answering sites on the Web (Agger, February 23, 2009). Primarily moderated by a rural librarian, Jessamyn West, all questions are archived, searchable, linkable, tagged, and categorized.

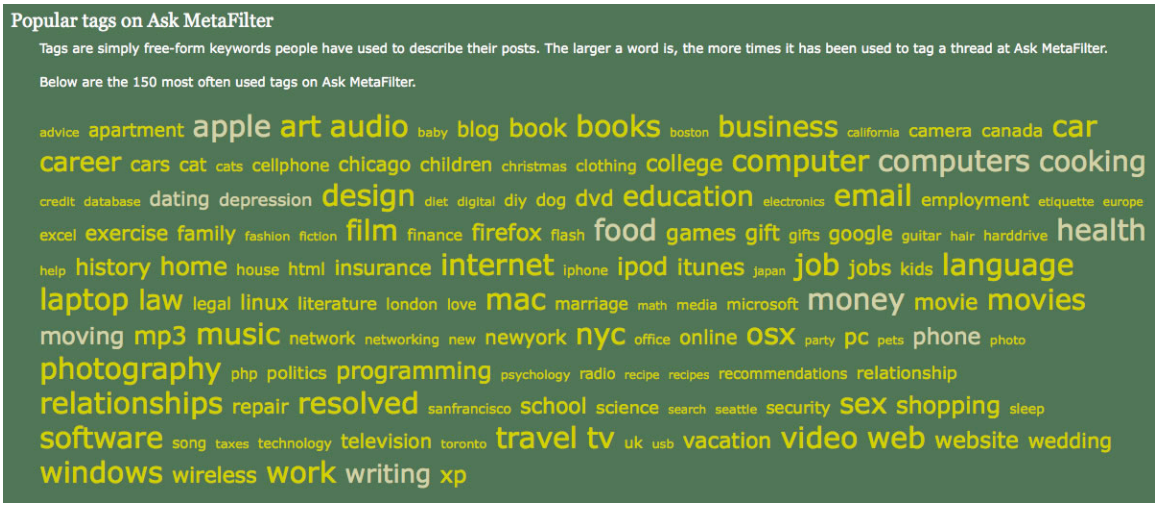
Tagging was implemented across MetaFilter in January 2005. A sitewide backtagging project started in May 2007 and was completed in March 2008, during which 43,000 old posts on the main MetaFilter site and AskMe were tagged by hundreds of volunteers. During the backtagging project, volunteers were presented with ten randomly selected untagged posts. Once they had tagged all ten posts and submitted them, they were presented with another set of ten posts. Once submitted, volunteers could not go back and change the tags on those posts. Currently, the author of a post (referred to as OP for original poster) and users on the OP's contact list can edit the tags on a post

[3] http://ask.metafilter.com
[4] http://www.metafilter.com

at any time. Moderators can also edit tags, and have been known to delete frivolous tags or add tags to posts. There is great emphasis that AskMe be a useful, reusable resource for all, both inside and outside the MetaFilter community.

As can be seen from the tag cloud for AskMe (Figure 4.5), questions cover a wide range of topics such as personal finance, health, and a variety of hobbies. The tags also show that computer and location-specific questions are quite popular. As of July 30, 2009, there were 123,652 posts or questions on AskMe, with 1,701,313 associated comments, and 496,488 unique tags. A post has on average about 14 comments and 4 tags attached to it.

Figure 4.3. Tag cloud from Ask MetaFilter



4.4 Tasks

In this experiment subjects were asked to perform two different types of search tasks, a text search task and an image search task. The tasks are intended to allow for comparison as asked in RQ2. Both of these tasks are each carried out in the two systems, Tags and No Tags. These tasks need to be plausible to the sample population while also

allowing for the collection of data addressing the research questions. Borlund (2000) compared "simulated work tasks" with real information needs of her test subjects, and discovered no significant difference in the test subjects' treatment of the information needs. Three characteristics of "good" simulated work task situations were identified: "(1) the situation has to be one to which the test persons can relate and with which they can identify; (2) the topic of the situation has to be of interest to the group of test persons; and (3) the situation has to provide enough imaginative context in order for the test persons to be able to apply the situation." (p. 86)

The text tasks and the image tasks were designed using the following criteria:

- The task is realistic for the sample population (e.g. University of Michigan undergraduates)
- The task is appropriate for the collection (e.g. information being sought should be well represented in the collection being searched)
- The task description is easily understood

4.4.1 Text task

The text tasks are shown below:

```
1. You want to buy a new laptop computer and need to
decide what kind to get. To help you make this decision,
you would like to know what other people recommend, as
well as their own experiences using different models of
laptop computers. Save 3 pages you found useful.

2. You are interested in visiting Chicago for a weekend
trip, and would like to find information about hotels,
restaurants, and interesting things to do in the city.
Save 3 pages you found useful.
```

4.4.2 Image task

The image tasks are shown below:

> 1. You are preparing flyers advertising events for Women's History Month. You want to find 5 photographs to use in the flyers, showing images of American women at work through the years. Save the 5 photographs you intend to use.
>
> 2. You are taking a class on the history of cities. For your next homework assignment you have to present to the class historical images of New York City. You need to find 5 photographs for your presentation. Save the 5 photographs you intend to use.

4.5 Experiment System

In the search experiments subjects were presented with the interfaces of the experiment system, and did not interact directly with any existing social tagging system or search engine. The experiment system is in effect a wrapper, presenting a generic interface to the user, but which does not itself include a search engine component.

From a subject's perspective, there are four interfaces in this study, Text/No Tags, Text/Tags, Image/No Tags, and Image/Tags. The No Tags interfaces for both text and images do not display tags at all (Figure 4.4 and Figure 4.5). The Tags interfaces display tags as part of the search results and for the text or image page (Figure 4.6 and Figure 4.7). The interfaces were designed to look as generic as possible, and drew from existing search interfaces on various websites. The Tags interfaces display tags in two locations in the SERP: as a list of related tags on top of the search results, and accompanying each search result. The related tags are ten of the most popular tags of the set of tags for that particular page of search results. So even for the same query, the list of related tags for

the first page of the search results will be different from the second page, and so forth. For the tags accompanying each search result, up to five tags were randomly selected and displayed from the full set of tags for each item. If there were less then five tags all of the tags were displayed. The full set of tags was displayed in the document page.

One reason for using an experiment system without identifying characteristics is to reduce subject bias in the evaluation of search results. Jansen, Zhang, and Zhang (2007) found brand awareness could affect by as much as 25% the evaluation of search results. Another reason for using an experiment system, and not an existing tagging system, for the experiment is that companies make changes to their interfaces, including the SERP, without announcement, leading to the risk that experiment subjects will be presented with different interfaces depending upon their time of participation in the study. For example, several changes in the presentation of tags in YouTube's SERP were observed between the time of the first pilot study and start of the main study data collection.

Figure 4.4. No Tags interface for text search

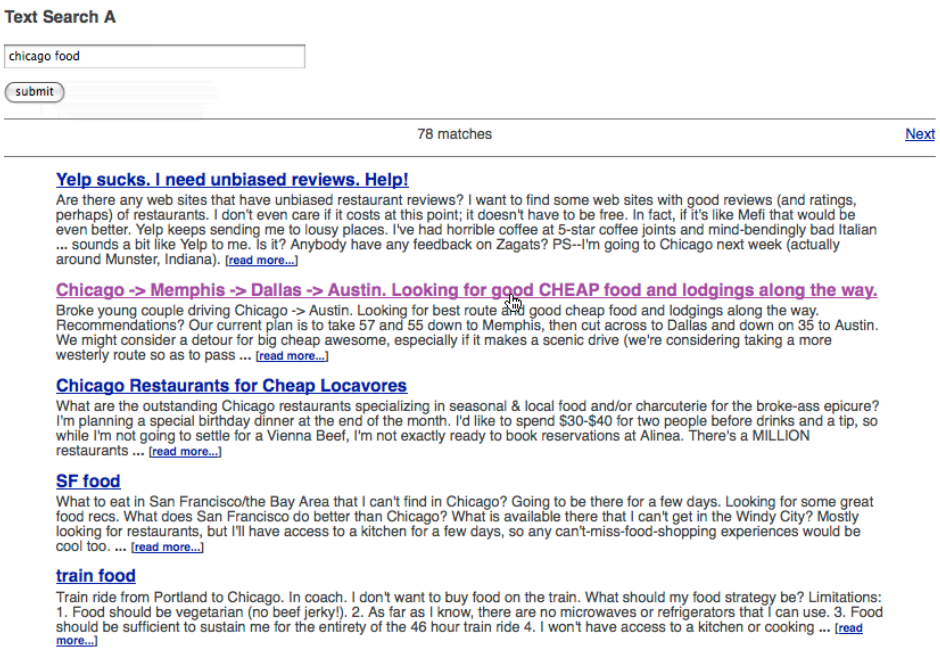


Figure 4.5. No Tags interface for image search

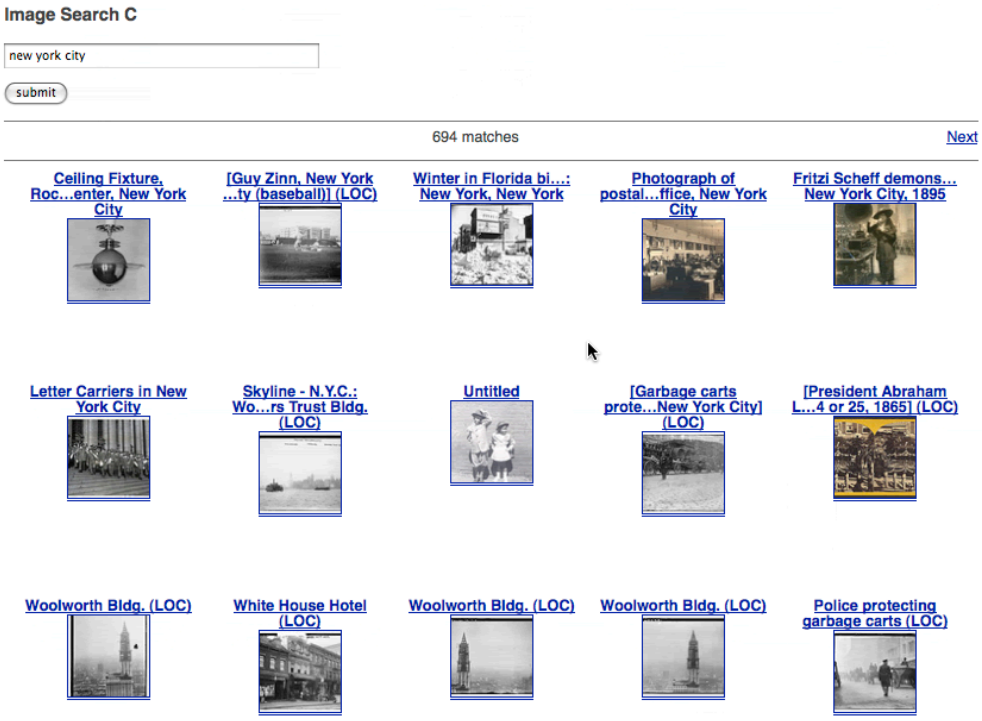


Figure 4.6. Tags interface for text search

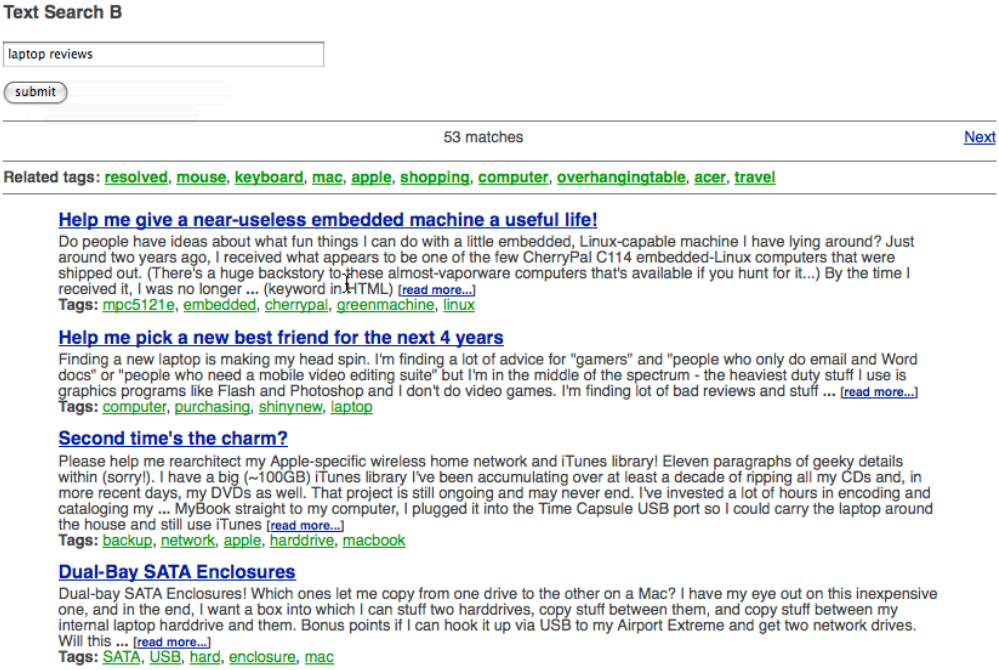
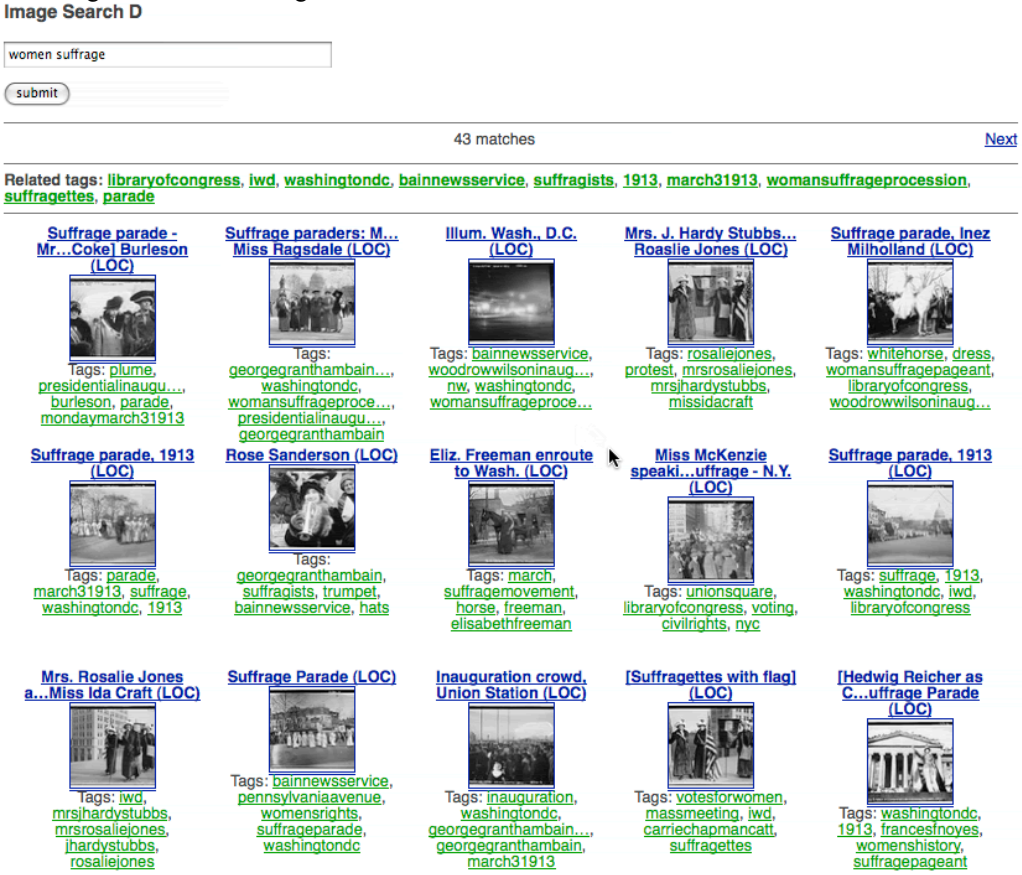
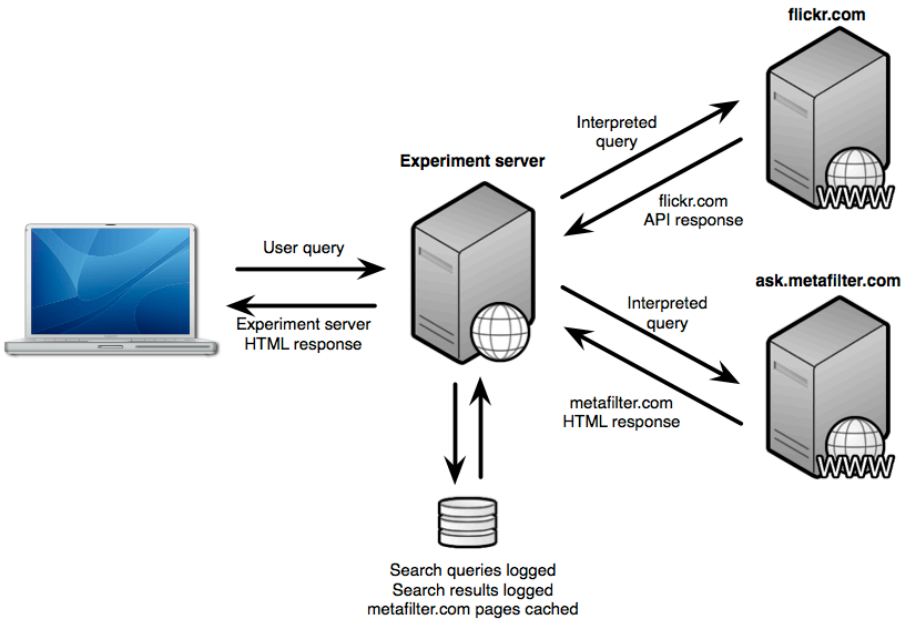


Figure 4.7. Tags interface for image search



The experiment system (Figure 4.8) includes the interface, a scraper component, and a component for database reading and writing. When a user enters a query, the query is sent to either Flickr or AskMe, depending on whether it is an image search or a text search. The query is processed by the site's search engine, and the search results generated are received and processed by the experiment system. The cleaned up search results are then displayed to the user by the experiment system.

Figure 4.8. Experiment system architecture



The experiment system is implemented using PHP and a MySQL database. PHP was selected because as a widely used scripting language for Web applications, code and libraries already exist for handling many of the required experiment system functionalities. Flickr searches are handled using the Flickr API, while AskMe searches required sending search requests and extracting content from the SERP's HTML. As the

default AskMe search results do not display the tags for the search results, for each of the search results the corresponding Web page has to be retrieved and the tags scraped from that page. This results in a long delay between the time a query is submitted and search results are displayed for the Text/Tags system. After the first day of pilot testing this delay was judged to be excessive and a tag index was constructed to speed up performance. Subjects judged system speed acceptable after this modification to the experiment system.

4.6 Subjects

A total of 48 (30 females and 18 males) undergraduate students from the University of Michigan in Ann Arbor participated in the study. The subjects were recruited through on-campus flyers and Facebook advertisements. Subjects were self-identified native English speakers who also self-identified as regularly looking for information on the Web. Subjects were not required to be tag users or have tagging experience, as we wanted to avoid priming them regarding tags in the search experiment. By requiring subjects to be regular Web users, it was predicted that subjects who had used tags or had tagging experience would be included in the subject sample. Subjects were required to be native English speakers so as to reduce the language comprehension variability. No restrictions were placed on their fields of study. Requiring subjects to be undergraduates was a way to restrict variability in the population in terms of age and also computer and Internet experience. Participation in the study was voluntary. The subjects were given $20 as compensation for their time spent in the study.

4.7 Data Collection Procedure

The procedure for the experiment is as follows:

1. Consent form and objective of experiment explained to subject (5 minutes)
2. Search task 1 (15 minutes)
3. Search task 2 (15 minutes)
4. Search task 3 (15 minutes)
5. Search task 4 (15 minutes)
6. Post-search interview (40 minutes)
7. Background questionnaire (5 minutes)

In the 15 minutes allocated for each search task, 10 minutes were allocated to actual searching, while 5 minutes were allocated for completing pre- and post-search questionnaires. The searches with the tagged systems were reviewed, with 20 minutes being allocated for each post-search interview. An experiment session was thus estimated to take an hour and 50 minutes. Subjects typically completed all four search tasks in 40 minutes or less, including the pre- and post-task questionnaires and the searches. The duration of post-search interviews depended on the duration of the search tasks, as lengthier searches resulted in lengthier search reviews. Typically experiment sessions lasted between 1.5 and 2 hours.

The experiment setup was as follows:

- MacBook Pro laptop with 15” screen running OS X 10.5.8
- Firefox browser version 3.5.2
- Pearl Crescent Page Saver Basic 2.1 (Firefox add-on)

- Silverback[5] usability testing software

Silverback can capture both screen activity as well as record the subject's voice. When the experimenter is ready to capture a session, the screen goes dark and recording is started by pressing the space bar on the keyboard. This allowed the experiment subjects to start the recordings at their convenience. Once a search task was completed and the subject was completing the post-task questionnaire, at this time the search session data captured by Silverback was exported to a QuickTime movie. This allowed for screen capture and voice recording of the search review using Silverback while viewing the recorded searches. All questionnaires were administered on paper, allowing the experimenter to export a search session screen recording and set up the laptop for the next search task while the subject was completing the questionnaires. Pearl Crescent Page Saver is a Firefox add-on for taking screenshots. When installed, a camera icon is visible in the toolbar area of the browser. Clicking on the icon results in a screenshot of the entire page being saved to the desktop.

4.8 Data Sources

Data was collected from multiple sources, including screen recordings, transaction logs, interviews after the searches, questionnaires, and observer notes.

Screen recordings

Each subject generated 5 screen recording files: one screen recording for each of the four searches, and one screen recording of the post-search interview, in which two search recordings were reviewed and general questions on tag usage and experience

[5] http://silverbackapp.com

asked. Silverback usability testing software was used to obtain both screen and audio recordings of the searches and the post-search interview. Silverback records screen activity and records audio, but does not capture keypress or mouse click events in a separate file. This captured data is exported into a QuickTime format file, which can be played like a movie.

Transaction logs

Transactions were logged through server-side logging, capturing communication between the subject's web browser and the server. The subject's queries, results shown, and the search result selected by the subject were stored in a MySQL database. Client-side events such as the user using the back button on the server, or the subject taking a screenshot were not captured in the transaction logs.

Post-Search Interview

The post-search interview consisted of two parts: in the first part the screen recordings of the two searches on the tagged systems were reviewed with the subject, during which the subject was asked questions on their tag usage; and in the second part, subjects were asked questions about their experience of tags in general, such as tagging sites they had seen or if they had tagged anything themselves. Both parts of the post-search interview were recorded on the laptop using Silverback software. Care was taken in the retrospective interview to not bias the subject about tags. At least one subject expressed surprise about the topic of the second part of the interview: "This whole thing was about tags?" (S22) The interview questions are included in Appendix B.7.

Questionnaires

There were three sets of questionnaires: pre- and post-search questionnaires for each search task, and a background questionnaire administered at the completion of the post-search interview. Questionnaires were administered on paper. The questionnaires are included in Appendix B.

4.9 Data Analysis

Out of 48 subjects, the data of one subject (S16) was dropped from analyses of text searches, because this subject did not complete the text search task on the Text/Tags system. Interviews of all 48 subjects were transcribed, and the transcripts were marked with timestamps. The timestamps allow direct linkage of interview questions and answers to particular events in the screen recordings.

In order to analyze the searches in detail, the transaction log data from the server was enhanced with client-side events to create detailed search logs for each search. Search recordings were reviewed to obtain client-side events. These detailed search logs were the main data source for RQ1, characterizing tag and non-tag searches for text and image searches. To address RQ2, comparing differences in tag use for text and image searches, tag use had to be identified as well as categorized. The basic unit of analysis was the query interval, " a segment of a search session that starts with a search query formulation and ends when the search session is concluded or a new query is formulated" (Kim, 2010). These are similar to the query reformulation intervals proposed by Liu, Gwizdka, and Liu (2010). Figure 4.9 shows the query log of one of the study subjects (S05) for the laptop information search task. This particular example has 9 query intervals. The query 'buying laptop' shows up twice, as the first and sixth queries of the

search session. According to our definition they are separate query intervals as there were several new queries that were formulated between these two queries.

Figure 4.9. Sample queries from S05, for laptop search

userid	queryterms
5	buying laptop
5	Mac
5	mac options
5	laptop options
5	hard drive laptop
5	buying laptop
5	bestbuy
5	electronic store
5	laptop comments

Tag use for query reformulation was identified from query logs and the post-search interviews reviewing searches on tagged systems. Whether a query came from a tag click or from the search box was logged as part of the query log. During the post-search interview, subjects were asked how they came up with search terms. If the mouse pointer was hovering over tags in the search recording, subjects were asked what was going on, to determine if subjects were using tags indirectly.

Tag use for predictive judgment was identified from the post-search reviews of screen recordings. Subjects were asked what items in the SERP made them select or not select a particular search result while reviewing the searches. Tag use for evaluative judgment was similarly identified from the search recording reviews. Subjects were asked what items they were paying attention to on the document page, as well as the reasons for saving or not saving a page. The list of questions can be found in Appendix B.

Prior knowledge of tags, consisting of familiarity, experience, and understanding, was identified from the post-search interview, primarily from the part of the interview focused on their general knowledge and usage of tags. Subjects were asked about their

familiarity with tags (familiarity), whether they had noticed or used tags, or tagged on the Web (experience), and their definition of tags, as well as who or what created the tags and the purpose of tags (understanding). Subjects also provided information about their tag knowledge and experience during the search recording reviews, and these were incorporated into the answers for the general part of the post-search interview.

4.10 Summary

The methodology adopted for the study was an information retrieval experiment under controlled laboratory conditions. This approach was adopted in order to be able to compare searches in the presence and absence of tags, and searches for different resource types (text and image). An experimental system was developed which presented different interfaces with respect to tags and resource types. The underlying collections for the searches were an online question-answering community for text searches and The Commons collection on Flickr for image searches. The collections were selected on the basis of the extent of tagging and their orientation of tagging for others. Data was collected from multiple sources, including screen recordings, transaction logs, interviews after the searches, questionnaires, and observer notes.

Tag use for query reformulation was identified from query logs and the post-search interviews reviewing searches on tagged systems. Tag use for predictive judgment and for evaluative judgment were identified from the post-search interviews. Prior knowledge of tags, consisting of familiarity, experience, and understanding, was identified from the post-search interview, primarily from the part of the interview focused on their general knowledge and usage of tags.

Chapter 5

Results

This chapter presents the results of the research on how people use tags during the search process. The analyses were based on 192 searches by 48 subjects recruited from the University of Michigan. The first section describes the demographic characteristics of the subjects and their experience with tags. Section 5.2 presents general characteristics of the searches, and provides definitions of the measures used. The next three sections summarize the results and discuss the findings for each of the three research questions. Section 5.3 compares the searches for the two interface types, Non Tags and Tags, for text and image searches, respectively. Section 5.4 compares the use of tags in text searches and image searches. Section 5.5 describes the prior knowledge of tags of the subjects and its relationship to tag use in the experiment.

5.1 Subject Profiles

The subject group included 48 undergraduate students, who were self-identified as native English speakers. The 48 participants (30 women and 18 men) ranged in age from 18 to 23 years old, with a mean age of 20 years. There were 25 different majors represented, from fields such as engineering, social and natural sciences, business, and languages.

Self-rated ability to find information online was relatively high for both information in text form (M=5.56, SD=0.74) and images (M=5.38, SD=1.23), measured on an ordinal scale (from 1 is "Poor" to 7 is "Excellent"). The most popular site for text searches was Google, used by 47 (98%) subjects, followed by Yahoo!, used by 3 (6%) subjects. No other major search engines were mentioned by the subjects. The most popular site for image searches was Google/Google Images, used by 47 (98%) subjects, followed by Yahoo!, used by 2 (4%) subjects. All participants reported spending time online at both school and at home on an average day. Thirty-one (65%) subjects spent at least an hour online at school, and 46 (96%) subjects spent at least an hour online at home, with over half (N=25) spending at least 3 hours online at home.

The most commonly used sites were Facebook, Wikipedia and YouTube. Facebook was used by all subjects, and Wikipedia and YouTube by 47 subjects each. Of the well-known tagging sites, Flickr was used by 21 (44%) subjects, while Delicious was used by 2 subjects, who said they had used it because it was required for a class but had not continued using it afterward. Last.fm was used by 10 (21%) subjects, LiveJournal by 9 (19%) subjects, and Twitter by 7 (15%) subjects. CiteULike, Connotea, LibraryThing, and Technorati were not used by the subjects.

In terms of tagging experience, 21 (44%) of the 48 subjects said they had never tagged online. Of the remaining 27 subjects, 15 had tagged on Facebook, 5 had tagged on YouTube, 4 had tagged on blogs, 3 had tagged on Flickr, and 2 had tagged on Delicious. Three subjects had tagged on multiple sites. One subject reported using Twitter hashtags and another subject reported using MTagger, a tagging feature available on the University of Michigan library web pages. The sample of this study can be characterized

as undergraduate students who are online regularly and are familiar with social media and user-generated content.

5.2 Characteristics of the Searches

Each of the subjects carried out two text searches and two image searches, for a total of four searches per subject. For each resource type (text and image), one search was on a system with no tags displayed (No Tags) and the other on a system that displayed tags (Tags). All 48 subjects carried out four searches, for a total of 192 search sessions. Searches were compared using a number of measures, which are described in Section 5.2.1. Section 5.2.2 discusses the comparability of the tasks for the two resource types. Some overall search characteristics are presented in section 5.2.3. Lastly, 5.2.4 presents direct and indirect tag use, a distinction used in some of the results for the research questions.

5.2.1 Measures of search characteristics

Several measures were used to characterize the searches and allow comparison of search characteristics. The measures fall into three classes: search session measures, search perception measures, and use category measures.

Search session measures

Search session measures directly measure characteristics of the search, such as search session duration or number of unique queries issued during a search session. Values for these measures are obtained from the transaction logs. Some of these measures were not obtained directly from the transaction logs, but required processing of

transaction log data. These included SERP dwell time and document dwell time, as well as the number of unique queries.

Search session duration was measured as the number of seconds from search start to the termination point. The termination point of a session was defined as the time of the third save for text searches and the time of the fifth save for the image searches, as subjects were required to save 3 Web pages for text search and 5 photographs for image search.

Number of unique queries is the number of unique queries from a search session. If a subject does not repeat a query during the search, the number of unique queries is the same as the number of query intervals in a session (in Chapter 4 query intervals were defined as "a segment of a search session that starts with a search query formulation and ends when the search session is concluded or a new query is formulated"). If a subject repeats one or more queries during a search, then the number of unique queries will be smaller than the number of query intervals.

Number of results is the sum of the number of results that were returned for all of the queries in the session. This is not the number of unique results, but simply sums the number of results returned for each query; if a result is returned more than once in response to different queries, it will be counted multiple times.

Number of unique documents counts how many unique documents were viewed during a search session. Subjects sometimes may revisit a document during a search session, hence the need to specify unique documents.

First query time is the number of seconds from search start to the time the first query is submitted. At the start of a search session, all subjects are shown an empty

search box, and have no information regarding the search system or the collection being searched. This measure can give an indication of how difficult it was to get started searching on a topic, as the only information subjects have at this point is the search task description.

SERP dwell time and *document dwell time* are respectively number of seconds spent on the SERP and number of seconds spent on the document.

Search perception measures

Search perception measures are intended to measure perceptions of the search process as well as expectations regarding the search process, and were measured on 5-point ordinal scales. They are used in the search questionnaires (Appendix X).

Search questionnaires were administered prior to a search and after a search. Prior to the search, subjects were asked for their *previous search experience* on the topic or images on the topic, how clear an idea or *plan for search* they had, and how clear an idea they had of the type of information or images they were looking for (*expected information*). After completing a search, subjects were asked to rate the (1) easiness of the search (*easiness*); (2) whether the search results were as expected (*expected information found*); (3) whether they felt lost at some point during the search (*lost during search*); (4) whether they felt they had used all available search features (*utilization of search features*); (5) perception of search duration (*search duration perception*); and (6) their satisfaction with the results of the search (*satisfaction*).

Interaction process measures

Interaction process measures count how many times tags were used during the searches in the study, for each of the interaction processes identified in the conceptual framework. The counts were obtained from the post-search interviews reviewing searches on the Text/Tags and Image/Tags systems. For each query, SERP, and Web page viewed, subjects were asked about their use of tags. Then these answers were coded based on the following operationalization of interaction processes:

- query reformulation: clicking on a tag, including a tag term in search query, or using terms in search query that were suggested by tags; query reformulation can happen from the SERP or a document page
- predictive judgment: using tags to make a predictive judgment of relevance during a query interval
- evaluative judgment: using tags to make an evaluative judgment of relevance during a query interval

The predictive and evaluative judgments of relevance draw from Rieh's model of two-stage judgment (Rieh, 2002). In this model, users make predictive judgments from the SERP, and make an evaluative judgment after viewing a Web page. This model is adopted as it maps well to user search behavior, where a user selects a subset of documents to examine from a usually much larger list of search results.

Query reformulation took the following forms: typing a query into a search box, or clicking on a tag. Typing into a search box could take place on the SERP or the document page. Clicking on a tag could take place in two places on the SERP – the list of related tags, and the tags accompanying each search result – and on the document page.

Subjects sometimes typed in a tag term in a search box, or typed in a term based on the tags they saw:

[S03]: “Some of the related tags were a good catalyst for thinking of new ideas.”

[S39]: “Yeah, I saw the *laptop*, *purchasing*, and decided to try that out.”

Predictive judgment takes place when a subject is examining the SERP to determine whether to click on a search result link or not. Tags are used in predictive judgment when a subject explicitly references it as a factor in his or her decision to click on a link or not. A predictive judgment is positive when tags contribute to the decision to click on a search result, and is negative when tags contribute to the decision to ignore a search result. Some examples of positive and negative predictive judgments with respect to tags are shown below:

[S05]: "I looked at the thumbnail and I had to like decide, like a lot of these things I didn't know exactly what they were, so I looked at the tags, like *manhattan*, I'm like okay, that actually is in New York, and like... And it said *newyork* in its tags so then I'm like, okay it actually is in New York."

[S46]: "... it was just that the tag *tenements* drew me to it."

[S14]: "I looked at the tags and it said *linux* and *cheap* and that's, I don't want, so I guess I thought that would have been a good page to look at but then I saw the tags and I didn't."

[S27]: "It was these two [tags, *tennessee* and *indiana*] that made me not want to click on this."

Evaluative judgment was coded when the subject mentioned such judgments while he/she was looking at the page during the search review. This is related to his or her decision to save or not to save the page they were examining at the moment. Here are some examples of evaluative judgment with respect to tags:

[S37]: “When I got to about here [document page], I did notice *laptop* and *notebook* and *budget*. And I was like, oh, this must be, this is the exact thing or pretty close to what I want.”

[S44]: “If it wasn't for the tags, I wouldn't really know what this picture was about.”

5.2.2 Comparability of tasks

For each resource type, text and images, there were two search tasks, so that subjects could carry out searches on the Tags and No Tags interfaces without repeating the same search task. By having different tasks for each interface, experimental confounds from learning or fatigue can be avoided. But this requires the tasks that will be interchanged for a particular system to be comparable in a number of ways, especially task difficulty. In pilot tests the tasks for each resource type were found to be comparable. But the pilot tests were on small numbers of subjects, with 7 subjects for the first pilot and 6 subjects for the second pilot. After the conclusion of data collection for the dissertation study, search session and pre- and post-search questionnaire data from the No Tags systems were analyzed to determine if the two tasks for each resource type were significantly different.

For the text searches, the Chicago and Laptop tasks carried out on the Text/No Tags interface were compared. Both searches asked subjects to find and save the three Web pages they found most helpful. In the Chicago task the searcher had to find information on places to eat, hotels, and things to do in Chicago. For the Laptop task, the subject had to find information on purchasing a laptop. For the image searches, the NYC and Women searches carried out on the Image/No Tags interface were compared. For the image search tasks subjects had to find and save five photographs, of historical images of New York City for the NYC task and images of American women at work suitable for use in a Women's History Month flyer for the Women task. The tasks were compared on

the No Tags interfaces as these were the baseline systems without the additional feature of tags.

The Mann-Whitney test was carried out to compare the tasks for each resource type. This test was selected as it does not require the dependent variables to be normally distributed interval variables. Tests for normality showed the interval variables did not have normal distributions across the variables of interest, while the variables from the pre- and post-search questionnaires were ordinal variables. As subjects used each system (Text/No Tags, Image/No Tags) only once, comparison of two tasks on the same system was between subjects. Each text task was used 24 times on the Text/No Tags system, as the 48 subjects had to do 2 text searches, one each on the Text/No Tags and Text/Tags systems. So half of the subjects used the Text/No Tags system for the Chicago task and half for the Laptop task. Similarly, the NYC and Women tasks were used 24 times each on the Image/No Tags system.

The following variables were compared for the task types: duration of search in seconds, number of unique queries, number of unique documents viewed, first query time, and the search questionnaire items. These variables can provide an indication whether the tasks are comparable in terms of difficulty as well as the subject's previous knowledge or experience on the topic. The first query time, for example, can indicate how difficult it was to get started searching on a topic – it would be difficult to argue that two search tasks with significantly different first query times were similar in terms of difficulty. The number of unique queries, the number of results, and the number of unique documents viewed can indicate the similarity of the tasks for the particular information retrieval system and collection in question – while two tasks may be similar

in difficulty, a collection may have significantly more relevant documents for one task than the other.

Table 5.1. Search session measures - text searches

	Chicago		Laptop		
Variable	M	SD	M	SD	*z*
Search session duration	313.46	133.45	358.25	126.80	-1.464
Number of unique queries	5.71	4.10	4.33	2.37	1.050
Number of results	72.04	40.90	63.83	30.96	0.465
Number of unique documents	4.67	1.79	4.67	1.63	-0.127
First query time	9.96	4.62	9.79	4.83	0.259

Table 5.2. Search session measures - image searches

	NYC		Women		
Variable	M	SD	M	SD	*z*
Search session duration	333.83	160.44	323.33	143.01	-0.072
Number of unique queries	7.33	5.87	7.21	5.38	-0.083
Number of results	145.25	109.44	120.21	103.28	0.773
Number of unique documents	7.75	2.71	8.21	4.04	0.126
First query time	14.38	7.29	11.04	4.20	1.427

Suppose a collection has a substantial number of photographs of birds, but very few of lizards. Then an image search task for images of birds and an image search task for images of lizards are not comparable in that collection. The number of unique queries might be substantially larger for the lizards task, while the number of unique documents viewed would be smaller. If the searches are being carried out to compare different interfaces for an information retrieval system, then the measures would not provide information about differences in the interfaces. The search questionnaires measure perceptions of the search task as well as the search process, providing additional data on the comparability of search tasks.

No significant differences were found between the two text search tasks and between the two image search tasks for all the variables compared. Tables 5.1 to 5.4

show the results of the statistical tests, as well as the mean and standard deviation of the variables. For example, while the laptop search task took almost 45 seconds longer on average than the Chicago search task, the difference was not statistically significant. The standard deviations for the search session durations indicate there was considerable individual variation on search session duration, and search session duration was not consistently longer for the laptop task compared to the Chicago task.

Table 5.3. Search questionnaires - text searches

	Chicago		Laptop		
Variable	M	SD	M	SD	*z*
Previous search experience	2.71	1.23	2.83	1.13	-0.270
Plan for search	4.00	.72	3.79	.72	1.252
Expected information	3.92	.72	4.13	.34	-0.956
Easiness of search	3.46	.93	3.50	.93	-0.286
Expected information found	2.79	.93	3.29	1.20	-1.637
Lost during search	2.42	1.02	2.17	.87	0.852
Utilization of search features	3.29	1.12	3.33	.70	-0.011
Search duration perception	3.38	1.38	3.04	1.12	0.978
Satisfaction	3.38	1.13	3.79	.98	-1.375

Table 5.4. Search questionnaires - image searches

	NYC		Women		
Variable	M	SD	M	SD	*z*
Previous search experience	1.71	.95	1.83	1.01	-0.597
Plan for search	3.71	.55	3.46	.83	1.120
Expected information	3.67	.96	3.67	.96	-0.087
Easiness of search	3.54	.83	3.75	.99	-0.601
Expected information found	2.71	.91	3.00	1.18	-0.889
Lost during search	2.46	.88	2.50	1.14	0.140
Utilization of search features	3.38	.88	3.17	.70	1.018
Search duration perception	3.38	1.13	2.88	1.15	1.499
Satisfaction	3.50	.88	3.63	1.06	-0.667

Similarly for the two image tasks, while there was considerable individual variation for the variables of interest, there was no consistent difference indicating one task took significantly longer than the other task, or was significantly easier than the

other task. As the tasks for each resource type were not found to be significantly different across a number of variables for the baseline system (No Tags), in subsequent analyses the tasks from each resource type will not be analyzed separately and instead regarded as either Text or Image tasks.

5.2.3 Search characteristics

All 48 subjects carried out four searches, for a total of 192 search sessions. One of the subjects did not complete his search task on the Text/Tags system, so is excluded from the analyses for the Text/Tags system. Table 5.5 summarizes characteristics of the text searches and Table 5.6 summarizes characteristics of the image searches.

While image and text searches both lasted about 6 minutes and 28 seconds in the No Tags interface, image searches on the Tags interface lasted about 21 seconds less than text searches on the Tags interface. Image searches on the Tags system took about 8 seconds less than on the No Tags system, while text searches on the Tags system took over 13 seconds longer than on the No Tags system. Image searches had a larger number of unique queries than text searches for both No Tags (7.98 vs. 5.04) and Tags (7.00 vs. 5.72) interfaces. The larger number of unique queries, results, and unique documents in image searches can be attributed to subjects being required to save 5 images for the image tasks, compared to 3 pages for the text tasks.

Subjects tended to examine 1 to 2 documents more than the number they were required to save, although for image searches some subjects examined a considerably larger number of documents. The first query times were larger for image searches than for text searches, regardless of system. This suggests that initial query formulation for image searches may be more difficult than for text searches. Maximum SERP depth

measures how far into the search results a subject looked, that is, the largest page number the subject examined across all queries in a search session. In the case of text searches, the maximum SERP depths of 1.36 (No Tags) and 1.26 (Tags) indicate subjects rarely looked beyond the first page of search results, which is consistent with previous research on Web searching behavior. But for image searches, subjects on average went beyond the second page of results, with maximum SERP depths of 2.5 (No Tags) and 2.42 (Tags). Some subjects examined up to the 9th or 10th page of results.

Table 5.5. Characteristics of text searches

	Text/No Tags M (SD)	Text/No Tags Min / Max	Text/Tags* M (SD)	Text/Tags* Min / Max
Search session duration	328.11 (120.51)	76 / 663	341.43 (152.41)	68 / 715
Number of unique queries	5.04 (3.42)	1 / 20	5.72 (3.88)	1 / 21
Number of results	68.28 (36.44)	12 / 155	82.53 (50.30)	13 / 208
Number of unique documents	4.68 (1.71)	3 / 9	4.60 (1.48)	3 / 9
Maximum SERP depth	1.36 (0.64)	1 / 3	1.26 (0.49)	1 / 3
First query time	9.81 (4.73)	3 / 23	10.36 (6.75)	4 / 42
SERP dwell time	130.21 (69.82)	20 / 265	152.23 (105.66)	20 / 516
Document dwell time	186.32 (94.96)	36 / 495	176.75 (87.49)	38 / 450

Note: Computed for 47 sessions

Table 5.6. Characteristics of image searches

	Image/No Tags M (SD)	Image/No Tags Min / Max	Image/Tags M (SD)	Image/Tags Min / Max
Search session duration	328.58 (150.44)	117 / 706	320.75 (134.30)	80 / 651
Number of unique queries	7.27 (5.57)	1 / 22	7.00 (4.98)	1 / 27
Number of results	132.73 (106.03)	20 / 472	124.10 (78.47)	17 / 359
Number of unique documents	7.98 (3.41)	5 / 21	7.88 (2.41)	5 / 16
Maximum SERP depth	2.5 (1.95)	1 / 9	2.42 (1.84)	1 / 10
First query time	12.71 (6.07)	4 / 31	12.15 (6.11)	4 / 35
SERP dwell time	226.40 (138.10)	38 / 619	216.85 (108.51)	37 / 494
Document dwell time	89.81 (43.86)	31 / 207	91.27 (43.33)	20 / 249

SERP dwell time and document dwell time measure the amount of time spent on the SERP and document pages, respectively. In the case of text search, document dwell time is larger than SERP dwell time for both Tags and No Tags systems. In the Text/No Tags system, subjects spent on average 56 seconds longer on document pages than on SERP pages, while on the Text/Tags system, subjects spent 24 seconds longer on document pages than on SERP pages. This reflects the fact that subjects spent time reading the documents they selected from the SERP. For Text/No Tags searches, document dwell time was significantly larger than SERP dwell time, $t(46) = -3.341$, $p < 0.001$ (one-sided t-test). Document dwell time was not significantly larger than SERP dwell time for the Text/Tags searches, $t(46) = -1.360$, $p = 0.0902$ (one-sided t-test).

In the case of image searches, document dwell time was significantly smaller than SERP dwell time for both Image/No Tags ($t(47) = 6.747$, $p<0.0001$, one-sided t-test) and Image/Tags searches ($t(47) = 8.837$, $p<0.0001$, one-sided t-test). Subjects spent over twice the amount of time on the SERP as on document pages in both the Image/No Tags system (226.4 seconds on the SERP vs. 89.8 seconds on document pages) and the Image/Tags system (216.8 seconds on the SERP vs. 91.3 seconds on document pages). When searching for images, subjects spent considerable time on the SERP examining the thumbnails, frequently going beyond the first page of search results for a query. Once on the actual image page, subjects did not stay long on that page, deciding very quickly whether to save the image or not.

5.2.4 Direct and indirect tag use

An alternative characterization of tag use is based on the modality of tag use, or whether use happens by clicking on a tag or without clicking on it. We refer to the former

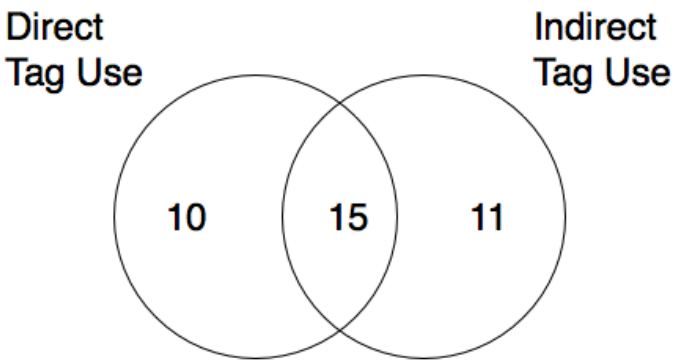
as direct tag use, and to the latter as indirect tag use. In query reformulation tags can be used directly or indirectly – directly by clicking on a tag, or indirectly by including in the search query terms seen in tags (e.g., adding "recommendation" to a search query on laptops after seeing "recommendation" in the list of related tags) or based on tags (e.g., using "brooklyn" as a search term after seeing the tag "manhattan"). Tags are used indirectly when they are part of the information used to decide whether to click on a search result item, or when used to decide whether a document should be saved. That is, tags are used indirectly when making predictive or evaluative judgments of relevance, in addition to query reformulation. Direct tag use can be identified from query logs, while indirect tag use is identified through communication from the user, such as think-aloud or post-search interviews.

In the experiment system used in this study, direct tag use can take place in three ways: clicking on a tag that came with the search result, clicking on a tag that is in the list of related tags on the SERP, or clicking on a tag on the document page. When reviewing the searches on the tagged systems during the post-search interviews, subjects were asked which items on a Web page they used to help decide whether to click on a search result or to save a document page. When entering query terms, subjects were asked how they came up with the search terms. If subjects mentioned using tags in their answers to these questions, they were counted as instances of indirect tag usage.

Of the 36 subjects who used tags in their searches, 10 subjects only used them directly, 11 subjects only used them indirectly, and 15 subjects used them both directly and indirectly (Figure 5.1). Thirty-one percent, or nearly a third of tag users in the search sessions, used tags only indirectly, which would not have been detected from only the

transaction logs. Table 5.7 shows the frequency of direct and indirect tag use for query reformulation, predictive judgment, and evaluative judgment. For text searches nearly a third of tag use in query reformulation is of the indirect type. The proportion of indirect tag use for image query reformulation is much smaller, being a ninth of direct tag use.

Figure 5.1. Direct and indirect tag users



Distinguishing between direct and indirect tag use allows us to see that a relatively large proportion of tag use can be missed if only tag click behavior is observed. The post-search interviews indicate that some indirect tag users were perceiving the tags to be additional text on the page, and not necessarily live links, while other indirect tag users were reluctant to click on links because they were unsure what would happen if they did so. Observing the modality of tag use provides additional insight into differences in tag usage for text and image searches.

Table 5.7. Frequency of direct and indirect tag use in interaction processes

	Text		Image		Total
Interaction process	Direct	Indirect	Direct	Indirect	
Query reformulation	21	6	45	5	77
Predictive judgment		18		15	33
Evaluative judgment		3		1	4
Total	21	27	45	21	114

5.3 Research Question 1: Characteristics of Tag and Non-Tag Searches

This section presents the results for Research Question 1: What are the characteristics of tag and non-tag searches? This research question examines differences in search behavior between tag users and tag non-users across text search and image search contexts. The question focuses on understanding how searches when tags are present are different from searches when tags are not present. For text search and image search each, the following questions are examined:

RQ1.1 How does observed search behavior differ for tag and non-tag systems?

RQ1.2 How do perceptions of the search process differ for tag and non-tag systems?

RQ1.1 draws from the screen recordings and transaction log data, and reviews of screen recordings from the post-search interviews. RQ1.2 draws from questionnaires, and the reviews of screen recording. In the next sections, the results for questions RQ1.1 and RQ1.2 for text searches, then RQ1.1 and RQ1.2 for image searches are presented.

Searches on Text/No Tags and Text/Tags systems were compared on a number of variables drawn from the search questionnaire items and search session data. The intent was to determine if people searched differently on the two systems. In addition to the measures defined in section 5.2.1, four derived measures were used for search efficiency.

Query interval duration (QI duration) measures the average duration of a query interval, and is computed by dividing the search session duration by the number of queries during the search session.

Query efficiency is computed by dividing the number of pages viewed by the unique number of queries. The value will be close to zero when only a small number of

pages is viewed despite a large number of queries, while the number will be large if a large number of pages are viewed from a small number of queries.

Save efficiency is computed by dividing the number of pages saved by the unique number of queries. This is another form of query efficiency, measuring how many documents were found useful enough to save per query.

Predictive efficiency is the number of pages saved divided by the number of pages viewed and denotes how efficient the subject was in predicting whether a page was useful enough to save.

5.3.1 Text search

RQ1.1 How does observed search behavior differ for tag and non-tag searches?

RQ1.1 examines characteristics of the search processes for Text/Tags and Text/No Tags systems as measured through behaviors and actions during the search process. As each subject in the study used both the Text/No Tags and Text/Tags systems, search processes for the two systems are compared within subjects. Tests for normality indicated that the distribution was not normal for the variables of interest except for search session duration. Because of the non-normal distributions, the Wilcoxon signed-rank test was used to compare the variables. Comparisons were done for 47 subjects[6].

Table 5.8 summarizes text search measures for Text/No Tags and Text/Tags systems. On average, search sessions on the Text/Tags system took 13 seconds longer than search sessions on the Text/No Tags system, with more unique queries issued (5.72 vs. 5.04), but the differences were not statistically significant. The number of results

[6] One subject ended his search session after saving only 2 documents, past the allotted 10-minute search time

viewed on the Text/Tags system was larger than on the Text/No Tags system, with 14 more results viewed on average, or about one more page of search results. These larger number of search results seen may be due to the larger number of unique queries on the Text/Tags system, about one more than on the Text/No Tags system. As each tag click counts as a query, and tags were clicked on 21 times in 12 text search sessions, it is possible that direct tag use contributed to the larger number of queries and results viewed on the Text/Tags system. The first query time is slightly larger on the Text/Tags system than the Text/No Tags system, but this cannot be attributed to tags, as subjects had not yet been exposed to the SERP or documents at this time. This difference was not statistically significant.

Table 5.8. Text search session measures across systems

Variable	Text/No Tags M (SD)	Text/Tags M (SD)	*z*
Search session duration	328.11 (120.51)	341.43 (152.41)	-0.159
Number of unique queries	5.04 (3.42)	5.72 (3.88)	-1.211
Number of results	68.28 (36.44)	82.53 (50.30)	-1.730
Number of unique documents	4.68 (1.71)	4.60 (1.48)	-0.005
Maximum SERP depth	1.36 (0.64)	1.25 (0.49)	0.631
First query time	9.81 (4.73)	10.36 (6.75)	-0.048
SERP dwell time	130.64 (69.86)	152.68 (105.68)	-0.783
Document dwell time	186.32 (94.96)	176.74 (87.49)	0.884

Differences in SERP dwell time and document dwell time can be seen for the Text/Tags and Text/No Tags systems. While document dwell time was longer than SERP dwell time for both systems, the difference between the two is larger for the Text/No Tags system (56 seconds) than for the Text/Tags system (24 seconds). This difference arises from the larger SERP dwell time on the Text/Tags system (22 seconds more than Text/No Tags) and smaller document dwell time on the Text/Tags system (10 seconds

less than Text/No Tags). That is, for Text/Tags searches, 45% of the search time is spent on the SERP and 52% on document pages, while for Text/No Tags searches, 40% of the search time is spent on the SERP and 57% on document pages. There was a slightly larger number of unique queries in the Text/Tags system, but almost no difference in the number of unique documents viewed in both systems (4.68 for Text/No Tags vs. 4.60 for Text/Tags). This suggests that when using the Text/Tags system, subjects were able to make judgments of relevance from only the surrogate, without having to click through to the document. The maximum SERP depth indicates subjects on both systems tended to only look at the first page of search results. The differences in number of unique documents viewed, SERP dwell time, and document dwell time were not statistically significant.

We next examine some derived measures (Table 5.9). The average query interval (QI) duration is computed by dividing the search session duration by the number of queries in the session. Query intervals were 15 seconds shorter on average on the Text/Tags system (M=71.1 sec) compared to the Text/No Tags system (M=86.9), indicating that a larger number of shorter query intervals took place on the Text/Tags system. This difference was not statistically significant from a Wilcoxon signed-rank test. The Text/No Tags system appeared to have somewhat higher query efficiency and save efficiency than the Text/Tags system, where query efficiency is defined as the number of unique documents viewed divided by the number of unique queries, and save efficiency is the number of saved documents (in this case 3) divided by the number of unique queries. The larger number of unique queries in Text/Tags search sessions likely contributed to the lower query efficiency and save efficiency values, as the number of

unique documents viewed were very similar (4.60 for Text/Tags, 4.68 for Text/No Tags) and the number of saved documents was fixed by the experimental design. Given the similarity of the number of viewed documents, it is not surprising that there was no difference in predictive efficiency – the number of saved documents divided by the number of unique documents viewed – between the systems.

Table 5.9. Text search derived measures across systems

Variable	No Tags (Text/No Tags) M (SD)	Tags (Text/Tags) M (SD)	*z*
Average QI duration	86.90 (62.68)	71.10 (49.20)	1.439
Query efficiency	1.40 (1.25)	1.09 (0.78)	1.021
Save efficiency	0.94 (0.79)	0.77 (0.52)	0.948
Predictive efficiency	0.72 (0.23)	0.71 (0.20)	0.404

While all 47 subjects were exposed to both the Text/No Tags and Text/Tags systems, not all of the subjects used tags on the Text/Tags system, either directly or indirectly. Of the 47 subjects, 25 subjects (52%) used tags at least once in their text searches, with 13 subjects (27%) using them more than once. The 22 subjects who did not use tags in their searches in effect might have been using the Text/Tags system similarly to the Text/No Tags system. In the post-search interviews, these subjects indicated they had not really noticed the tags and thus their search behavior on either system would be expected to be similar. Therefore the search process variables were re-examined for the 25 subjects who had used tags at least once on the Text/Tags system (Table 5.10).

Table 5.10. Text search measures for 25 subjects who used tags at least once

Variable	No Tags (Text/No Tags) M (SD)	Tags (Text/Tags) M (SD)	z
Search session duration	365.24 (96.44)	399.68 (140.69)	-0.834
Number of unique queries	5.28 (2.84)	7.44 (4.40)	-2.608**
Number of results	72.76 (33.51)	106.80 (52.75)	-2.692**
Number of unique documents	5.00 (1.58)	5.00 (1.66)	0.055
Maximum SERP depth	1.36 (0.64)	1.40 (0.58)	-0.624
First query time	8.36 (3.57)	10.56 (7.74)	-0.950
SERP dwell time	128.68 (55.76)	190.76 (114.33)	-2.220*
Document dwell time	225.68 (94.56)	195.28 (86.73)	1.709

Note: * $p < 0.05$; ** $p<0.01$

Table 5.11. Text search derived measures for 25 subjects who used tags at least once

Variable	No Tags (Text/No Tags) M (SD)	Tags (Text/Tags) M (SD)	*z*
Average QI duration	93.05 (71.01)	59.77 (22.60)	2.462*
Query efficiency	1.44 (1.51)	0.83 (0.48)	1.669
Save efficiency	0.82 (0.70)	0.55 (0.34)	2.210*
Predictive efficiency	0.66 (0.20)	0.66 (0.19)	0.177

Note: * $p < 0.05$

For these 25 subjects, a number of search characteristics differed significantly in their Text/No Tags and Text/Tags searches. On average, subjects issued 2 more unique queries when using the Text/Tags system compared to the Text/No Tags system, and viewed significantly more search results. This is probably why subjects spent more time on the SERP when using the Text/Tags system, as there were more SERPs to view and more results to view on a SERP. On the Text/Tags system, subjects issued a significantly larger number of unique queries ($p = 0.0091$) with significantly shorter query interval durations ($p = 0.0138$) compared to when using the Text/No Tags system. One interpretation is that subjects had a larger number of shorter query intervals during a search session on the Text/Tags system than on the Text/No Tags system. A possible reason is that subjects were able to tell quickly on the Text/Tags system whether a

particular query was fruitful or not, based on what they saw on the SERP. Another possible reason is that on the Text/Tags system subjects were able to reformulate their queries faster, due to the information present on the SERP. On the Text/Tags system, subjects used tags for query reformulation, predictive judgment and evaluative judgment. This usage is addressed in more detail in RQ2.

Subjects allocated SERP dwell time and document dwell time differently for the Text/No Tags and Text/Tags systems. SERP dwell time was significantly larger for the Text/Tags system, with subjects spending over a minute more on the SERP for Text/Tags searches than for Text/No Tags searches. In contrast, subjects spent 30 seconds less on document pages on Text/Tags searches than on Text/No Tags searches. In the Text/Tags system, 48% of search session time was spent on the SERP and 49% on document pages. In the Text/No Tags system, 35% of search session time was spent on the SERP and 62% on document pages. The number of viewed documents was the same for both Text/No Tags and Text/Tags systems, so more time was spent reading documents when searching on the Text/No Tags system. Subjects explored the search space more on the Text/Tags system while spending proportionately less time reading documents. This suggests that tags on the SERP helped with query reformulation as well as predictive judgment, contributing to the larger number of queries and search results seen.

Query efficiency was higher for Text/No Tags searches than for Text/Tags searches, which can be attributed to the larger number of unique queries in Text/Tags searches (Table 5.11). Save efficiency differed significantly (p=0.0271) for the two systems. Since the number of documents saved is constant, this measure depends on the number of unique queries, of which there are significantly more in the Text/Tags

searches. Predictive efficiency, or the ratio of pages saved to pages viewed was similar in both systems, which is not surprising given the constant number of documents saved and the similarity in the number of documents viewed during a search.

Table 5.12. Text search measures for 13 subjects who used tags more than once

Variable	Text/No Tags M (SD)	Text/Tags M (SD)	z
Search session duration	381.92 (114.02)	425.15 (155.19)	-0.734
Number of unique queries	5.08 (3.57)	8.62 (5.22)	-2.809**
Number of results	71.85 (36.03)	110.15 (54.10)	-2.450*
Number of unique documents	5.31 (1.80)	5.08 (1.93)	0.387
Maximum SERP depth	1.31 (0.48)	1.46 (0.66)	-1.00
First query time	8.85 (4.34)	9.46 (4.65)	0.036
SERP dwell time	128.0 (68.09)	216.54 (115.22)	-2.271*
Document dwell time	244.92 (97.63)	199.15 (102.24)	1.503

Note: * $p < 0.05$; ** $p<0.01$

Table 5.13. Text search derived measures for 13 subjects who used tags more than once

Variable	Text/No Tags M (SD)	Text/Tags M (SD)	z
Average QI duration	113.11 (90.34)	51.39 (12.84)	3.040**
Query efficiency	1.88 (1.98)	0.68 (0.22)	2.308*
Save efficiency	1.02 (0.92)	0.45 (0.21)	2.875**
Predictive efficiency	0.64 (0.23)	0.66 (0.21)	-0.210

Note: * $p < 0.05$; ** $p<0.01$

As shown in Tables 5.12 and 5.13, these measures are examined for the 13 subjects who used tags more than once, to determine more clearly if the differences arise from tag usage. Subjects who used tags only once cited curiosity as to what would happen as one of the reasons they used tags. Those who used tags more than once would have done so because they had seen some value from using tags in their searches. The results are similar to those of the 25 subjects who used tags at least once. When searching using the Text/Tags system, subjects searched for a longer time, issuing more queries, and viewing a larger number of search results, but viewed about the same number of

documents. Each query interval was shorter in the Text/Tags system. Using the Text/Tags system, subjects were able to explore a larger search space while taking less time for each query instance and making many more predictive judgments.

Next, we examine the 22 subjects[7] who did not use tags in their searches, to follow up on the intuition that these subjects in effect were using the Text/Tags system similarly to the Text/No Tags system. A number of differences in search characteristics observed when tags were used were reversed for the subjects who did not use tags in the Text/Tags system. For example, search sessions were shorter by 10 seconds on the Text/Tags system, query intervals were longer, and less time was spent on the SERP. But none of these differences in search characteristics between searches on the Text/No Tags and Text/Tags systems (Table 5.14 and Table 5.15) were statistically significant. In effect, subjects who did not use tags in the Text/Tags system did not search differently on the two systems.

Table 5.14. Text search measures for 22 tag non-users

Variable	TN M (SD)	Text/Tags M (SD)	z
Search session duration	285.91 (132.95)	275.23 (140.12)	0.584
Number of unique queries	4.77 (4.03)	3.77 (1.85)	1.142
Number of results	63.18 (39.67)	54.95 (29.38)	0.503
Number of unique documents	4.32 (1.31)	4.14 (1.13)	-0.184
Maximum SERP depth	1.36 (0.66)	1.09 (0.29)	1.484
First query time	11.45 (5.40)	10.14 (5.58)	0.781
SERP dwell time	132.86 (84.42)	109.41 (76.40)	1.347
Document dwell time	141.59 (74.79)	155.68 (85.41)	-0.796

[7] While there were 23 subjects who did not use tags in their searches on the Text/Tags system, one of these subjects did not complete his search on the Text/Tags system and is not included in the analysis in this section.

Table 5.15. Text search derived measures for 22 tag non-users

Variable	TN M (SD)	Text/Tags M (SD)	z
Average QI duration	79.92 (52.42)	83.99 (66.27)	-0.308
Query efficiency	1.35 (0.90)	1.38 (0.95)	-0.292
Save efficiency	1.08 (0.88)	1.02 (0.58)	-0.293
Predictive efficiency	0.79 (0.25)	0.77 (0.19)	0.434

When tags are used, there are significant differences in the number of unique queries issued and the number of search results seen. When using tags, subjects spend more time on the SERP, reformulate their queries quickly and view relatively few pages compared to how many unique queries they issue during a search. Tables 5.16 and 5.17 show differences in search characteristics when using the Text/Tags system when tags are not used (non-users, N=22), tags are used only once (low users, N=12), and tags are used more than once (high users, N=13). Search session duration increases with increasing tag use, as do the number of unique queries and number of unique documents viewed. Search sessions of subjects using tags at two or more times were 53 seconds longer than those of subjects who used tags only once, whose search sessions were in turn 97 seconds longer than those of subjects who did not use tags. High users had over twice the number of unique queries and number of results as non-users. High users also saw one more unique document on average than non-users. Patterns of SERP dwell time and document dwell time also change with increasing tag use: 40% and 57% of search session time for non-users, 44% and 51% of search session time for low users, and 51% and 47% of search session time for high users. Increased tag use is associated with increasing proportion of time on the SERP and diminishing proportion of time viewing document pages. Tags can be used for query reformulation and predictive judgment on the SERP, and for query

reformulation and evaluative judgment on the document page. This usage is addressed in more detail in RQ2.

Table 5.16. Text/Tags search measures for tag non-users (N=22), low users (N=12), and high users (N=13)

Variable	Non-users M (SD)	Low users M (SD)	High users M (SD)
Search session duration	275.23 (140.12)	372.08 (123.75)	425.15 (155.19)
Number of unique queries	3.77 (1.85)	6.17 (3.01)	8.62 (5.22)
Number of results	54.95 (29.38)	103.17 (53.39)	110.15 (54.10)
Number of unique documents	4.14 (1.13)	4.92 (1.38)	5.08 (1.93)
Maximum SERP depth	1.09 (0.29)	1.33 (0.49)	1.46 (0.66)
First query time	10.14 (5.58)	11.75 (10.20)	9.46 (4.65)
SERP dwell time	109.41 (76.40)	162.83 (111.35)	216.54 (115.22)
Document dwell time	155.68 (85.41)	191.08 (70.51)	199.15 (102.24)

Table 5.17. Text/Tags search derived measures for tag non-users (N=22), low users (N=12), and high users (N=13)

Variable	Non-users M (SD)	Low users M (SD)	High users M(SD)
Average QI duration	83.99 (66.27)	68.85 (27.59)	51.39 (12.84)
Query efficiency	1.38 (0.95)	1.00 (0.62)	0.68 (0.22)
Save efficiency	1.02 (0.58)	0.65 (0.42)	0.45 (0.21)
Predictive efficiency	0.77 (0.19)	0.65 (0.17)	0.66 (0.21)

RQ 1.2 How do perceptions of the search process differ for tag and non-tag systems?

RQ1.2 examines perceptions of the search processes for Text/Tags and Text/No Tags systems as measured through the search questionnaires. The pre-search questionnaire asked about the subject's previous experience with the topic, if they had a clear plan for their search, and had an idea of what kind of information they expected (Appendix B.4). The post-search questionnaire asked about the easiness of the search, if the expected information had been found, if the subject had felt lost during the search, if the subject had fully utilized the search features, the subject's perception of search duration, and their satisfaction with the search (Appendix B.5).

As each subject in the study used both the Text/No Tags and Text/Tags systems, search processes for the two systems are compared within subjects. The Wilcoxon signed-rank test was used to compare the search questionnaire variables. Initially the completed searches from 47 subjects were compared (Table 5.18). There were no significant differences in the perceptions of the search tasks for each of the systems prior to starting the search, in terms of previous experience with the topic, the plan for search, or knowing clearly what information they were looking for. No significant differences were found in perceptions of the search, such as easiness, expected information being found, utilizing all search features, perception of search duration, or satisfaction.

Table 5.18. Perceptions of search process (N=47)

Variable	Text/No Tags M (SD)	Text/Tags M (SD)	z
Previous experience with topic	2.81 (1.15)	2.70 (1.35)	0.339
Plan for search	3.94 (0.67)	3.85 (0.55)	1.013
Expected information	4.06 (0.48)	4.06 (0.70)	-0.385
Easiness of search	3.51 (0.91)	3.51 (0.98)	0.000
Expected information found	3.02(1.09)	3.26 (1.11)	-0.996
Lost during search	2.30 (0.95)	2.72 (0.97)	-2.209*
Utilization of search features	3.34 (0.92)	3.30 (0.88)	0.255
Search duration perception	3.19 (1.26)	3.09 (1.16)	0.427
Satisfaction	3.60 (1.08)	3.62 (0.90)	-0.305

Note: * $p < 0.05$

Subjects felt more lost using the Text/Tags system than the Text/No Tags system, $z = -2.209$, $p = 0.0272$. This finding was investigated in more detail taking into consideration that 22 subjects did not use tags either directly or indirectly on the Text/Tags system. Pairwise comparisons were conducted for the 25 subjects who had used tags at least once during their text searches (Table 5.19). Again, subjects felt significantly more lost using the Text/Tags system, $z = -1.986$, $p = 0.0471$. But, for the 13

subjects who had used tags more than once (Table 5.20), there was no significant difference in how lost they felt while searching using Text/No Tags or Text/Tags systems, $z = -0.963$, $p = 0.3354$.

Table 5.19. Perceptions of subjects who used tags at least once (N=25)

Variable	Text/No Tags M (SD)	Text/Tags M (SD)	z
Previous experience with topic	2.76 (1.20)	2.64 (1.50)	0.247
Plan for search	4.04 (0.73)	3.80 (0.65)	1.611
Expected information	4.20 (0.41)	4.00 (0.71)	1.200
Easiness of search	3.44 (1.04)	3.24 (1.01)	0.826
Expected information found	3.00 (1.08)	2.96 (1.14)	0.224
Lost during search	2.36 (0.86)	2.80 (1.04)	-1.986*
Utilization of search features	3.36 (1.04)	3.28 (0.89)	0.466
Search duration perception	3.56 (1.16)	3.20 (1.19)	0.933
Satisfaction	3.48 (1.08)	3.44 (0.96)	0.098

Note: * $p < 0.05$

Table 5.20. Perceptions of subjects who used tags at least twice (N=13)

Variable	Text/No Tags M (SD)	Text/Tags M (SD)	z
Previous experience with topic	2.46 (1.33)	2.46 (1.20)	0.000
Plan for search	3.77 (0.83)	3.69 (0.75)	0.404
Expected information	4.08 (0.28)	3.92 (0.95)	0.173
Easiness of search	3.08 (1.12)	3.23 (0.83)	-0.357
Expected information found	3.08 (0.95)	3.00 (1.00)	0.182
Lost during search	2.54 (0.88)	2.85 (0.90)	-0.963
Utilization of search features	3.46 (0.88)	3.38 (0.87)	0.290
Search duration perception	3.69 (1.11)	3.31 (1.25)	0.603
Satisfaction	3.38 (1.04)	3.46 (0.88)	-0.251

Examining tag non-users, low users, and high users on the Text/Tags system (Table 5.21), we see that the perception of the easiness of the search decreases with increasing tag use, while the perception of feeling lost and search duration perception increases with tag use. Satisfaction is lower for tag users than non-users.

Table 5.21. Text/Tags perceptions of tag non-users (N=22), low users (N=12), and high users (N=13)

Variable	Non-users M (SD)	Low users M (SD)	High users M (SD)
Previous experience with topic	2.77 (1.19)	2.83 (1.80)	2.46 (1.20)
Plan for search	3.91 (0.43)	3.92 (0.51)	3.69 (0.75)
Expected information	4.14 (0.71)	4.08 (0.29)	3.92 (0.95)
Easiness of search	3.82 (0.85)	3.25 (1.22)	3.23 (0.83)
Expected information found	3.59 (1.01)	2.92 (1.31)	3.00 (1.00)
Lost during search	2.64 (0.90)	2.75 (1.22)	2.85 (0.90)
Utilization of search features	3.32 (0.89)	3.17 (0.94)	3.38 (0.87)
Search duration perception	2.95 (1.13)	3.08 (1.16)	3.31 (1.25)
Satisfaction	3.82 (0.80)	3.42 (1.08)	3.46 (0.88)

It is possible that a subject using the Text/Tags system notices and tries using tags when he or she is feeling lost during a search. If the result of using the tag was not as expected, the subject would continue feeling lost and also would not be inclined to click on tags any more. We found some evidence for this from the search interviews of subjects who had clicked on tags once and reported feeling quite lost during a search:

> [S05]: [When I clicked on the tag *Mac*] I thought it would tell me about Macs. ... I was not happy with what it told me. ... I wanted things that talked about Mac, and like the different memories, and the different pros and cons of each one.

> [S23]: [When I clicked on the tag *restaurants*] I thought it would be Chicago restaurants. And then all of a sudden it brought it out to different cities in the country. And so I was like, oh that's not helpful.

> [S24]: [When I clicked on the tag *laptop*] And you know, that's when I started running out of ideas. ... I thought that there'd be some more results that I want because I figured someone had done it before. Like those are kind of like most popular or like... To me, related tags are kind of like "If you're looking for this you'll also be looking for this." So I thought maybe they have some good results. ... it was like the same thing I already had seen.

> [S44]: So then, I just clicked on it [the tag *Chicago*] to see. ... I just thought that it would give me like, a list of attractions. ... It wasn't nearly what I was expecting.

For these subjects, tags did not behave as expected when clicked on. Subjects had noticed tags when their searches were not going as expected or were "running out of ideas" (S24) on how to reformulate their queries. So then using tags, and finding it "not helpful" (S23) would likely not help subjects feel less lost during their search. It is even possible that subjects exposed to a feature that does not work as expected will feel more lost than if they had not been exposed to the feature at all. Subjects who are feeling lost during a search will likely also feel that the search is taking longer than expected and will experience less satisfaction with the search.

5.3.2 Image search

RQ1.1 How does observed search behavior differ for tag and non-tag searches?

RQ1.1 examines characteristics of the image search processes for Image/Tags and Image/No Tags systems as measured through behaviors during the search process. As each subject in the study used both the Image/No Tags and Image/Tags systems, search processes for the two systems are compared within subjects. Tests for normality indicated that the distribution was not normal for the variables of interest. Because of the non-normal distributions, the Wilcoxon signed-rank test was used to compare the variables.

On average, subjects in the Image/No Tags system took 8 seconds longer in their searches than on the Image/Tags system. They also issued more queries and saw more results on the Image/No Tags system than the Image/Tags system. However, the differences were not statistically significant. Average query interval length was similar for both Image/No Tags and Image/Tags systems. No statistically significant differences were observed for the measures in Table 5.22 and Table 5.23.

Table 5.22. Image search characteristics - search session measures (N=48)

Variable	Image/No Tags M (SD)	Image/Tags M (SD)	*z*
Search session duration	328.58 (150.44)	320.75 (134.30)	0.062
Number of unique queries	7.27 (5.57)	7.00 (4.98)	-0.036
Number of results	132.73 (106.03)	124.10 (78.47)	0.477
Number of unique documents	7.98 (3.41)	7.88 (2.41)	-0.124
Maximum SERP depth	2.50 (1.95)	2.42 (1.84)	0.262
First query time	12.71 (6.07)	12.15 (6.11)	0.678
SERP dwell time	226.69 (138.08)	217.33 (108.54)	0.133
Document dwell time	89.81 (43.86)	91.27 (45.33)	0.041

Table 5.23. Image search characteristics - derived measures (N=48)

Variable	Image/No Tags M (SD)	Image/Tags M (SD)	*z*
Average QI duration	60.37 (39.41)	61.23 (43.76)	0.503
Query efficiency	1.91 (1.91)	1.76 (1.46)	0.108
Save efficiency	1.25 (1.07)	1.25 (1.24)	0.113
Predictive efficiency	0.71 (0.22)	0.68 (0.17)	0.817

While all 48 subjects were exposed to both the Image/No Tags and Image/Tags systems, not all of them used tags, either directly or indirectly, on the Image/Tags system. Of the 48 subjects, 25 subjects (52%) used tags at least once in their image searches, with 16 subjects (33%) using them more than once. The 23 subjects who did not use tags in their searches in effect might have been using the Image/Tags system similarly to the Image/No Tags system. In the post-search interviews, these subjects indicated they had not really noticed the tags and thus their search behavior on either system would be expected to be similar. Therefore the search process variables were re-examined for the 25 subjects who had used tags at least once on the Image/Tags system (Tables 5.24 and 5.25).

Table 5.24. Image search measures for 25 subjects who had used tags at least once

Variable	Image/No Tags M (SD)	Image/Tags M (SD)	z
Search session duration	360.20 (162.65)	371.04 (131.79)	-0.377
Number of unique queries	8.00 (5.76)	8.04 (4.34)	-0.391
Number of results	137.36 (114.68)	147.44 (83.86)	-0.229
Number of unique documents	8.52 (4.04)	8.76 (2.83)	-0.624
Maximum SERP depth	2.24 (1.71)	2.56 (2.18)	-0.502
First query time	12.16 (5.60)	10.96 (6.39)	1.293
SERP dwell time	242.92 (148.73)	249.72 (108.29)	-0.471
Document dwell time	106.32 (49.93)	110.36 (48.20)	-0.283

Table 5.25. Image search derived measures for 25 subjects who had used tags at least once

Variable	Image/No Tags M (SD)	Image/Tags M (SD)	z
Average QI duration	56.57 (35.86)	59.91 (46.66)	0.578
Query efficiency	1.68 (1.64)	1.54 (1.13)	0.242
Save efficiency	1.03 (0.75)	0.97 (0.98)	0.458
Predictive efficiency	0.69 (0.24)	0.62 (0.17)	1.292

A number of differences can be observed for these 25 subjects compared to measures of the 48 subjects overall. Tag users had longer search sessions on the Image/Tags system and also saw a larger number of search results, while having the same number of unique queries. This indicates that tag users went beyond the first page or search results for a given query. This is supported by the maximum SERP depth, as the value is larger for searches on the Text/Tags system. As presented in Table 5.24 and Table 5.25, no statistically significant differences were found for these 25 subjects with respect to their searches on the Image/No Tags and Image/Tags systems.

In Table 5.26 and Table 5.27, we examine the sixteen subjects who used tags at least twice when using the Image/Tags system. As in text search, the intent is to examine the impact of repeated tag use during a search session. Subjects issued on average two more unique queries on the Image/Tags than the Image/No Tags system. They also saw

more search results on the Image/Tags system than on the Image/No Tags system, but the number of search results pages they saw per query were similar for both systems. Search sessions were over 50 seconds longer on the tagged system, which is probably related to the longer times spent on the SERP and the document pages when using the tagged system. Nevertheless the proportions of SERP dwell time and document dwell time were similar for Text/No Tags (66% and 31%) and Text/Tags systems (67% and 30%). When searching for images subjects on average looked beyond the first two pages of search results for a given query.

Table 5.26. Image search measures for16 subjects who had used tags at least twice

Variable	Image/No Tags M (SD)	Image/Tags M (SD)	z
Search session duration	352.00 (150.96)	405.25 (137.65)	-1.474
Number of unique queries	6.81 (5.22)	8.81 (3.97)	-1.842
Number of results	138.69 (116.20)	166.38 (89.36)	-0.827
Number of unique documents	8.75 (4.60)	9.25 (2.44)	-0.758
Maximum SERP depth	2.69 (1.96)	2.63 (2.47)	0.474
First query time	12.50 (6.37)	9.88 (4.00)	1.583
SERP dwell time	233.0 (129.86)	272.38 (118.60)	-1.241
Document dwell time	108.38 (54.39)	123.0 (50.46)	-0.983

Query interval duration was significantly shorter for the Image/Tags system (p=0.0437), indicating subjects on the Image/Tags system had many more shorter queries than when searching on the Image/No Tags system. When searching on the Image/Tags system subjects had significantly lower save efficiency (p=0.0361), which is explained by the larger number of unique queries on the Image/Tags system. Query efficiency and predictive efficiency were higher on the Image/No Tags system. So it appears that when searching on the Image/Tags system, subjects view more pages that they end up not saving, and issue many queries that do not result in a page view.

Table 5.27. Image search derived measures for16 subjects who had used tags at least twice

Variable	Image/No Tags M (SD)	Image/Tags M (SD)	z
Average QI duration	64.92 (41.11)	50.89 (30.99)	2.017*
Query efficiency	2.01 (1.93)	1.27 (0.72)	1.914
Save efficiency	1.21 (0.84)	0.68 (0.29)	2.096*
Predictive efficiency	0.69 (0.26)	0.57 (0.12)	1.915

Note: * $p < 0.05$

As in text search, subjects were categorized as non-users, low users, and high users based on their use of tags when searching on the Image/Tags system. Low users used tags only once during their searches, while high users used tags at least twice. Compared to text search, in which the numbers of low and high users were similar, in image search the number of high users is almost twice that of low users. Search session duration, number of unique queries, number of results seen, number of unique documents seen, maximum SERP depth, SERP dwell time and document dwell time all increase with tag use (Table 5.28). SERP dwell time and document dwell time as proportions of search session duration remain relatively constant for different levels of tag use: 68% and 27% for non-users, 68% and 28% for low users, and 67% and 30% for high users. This contrasts with text search, where the proportion of time spent on the SERP increased and the proportion of time spent on documents decreased with tag use. Additionally, first query time decreases with tag use, raising the possibility that subjects who had less trouble getting started with their image search were also more likely to use tags.

When examining the derived measures for image search (Table 5.29), it is interesting that average query interval duration is highest for low tag users. Otherwise non-users and low users are quite similar to each other, but different from high users. High users viewed more documents and issued more queries than non-users or low users,

resulting in a query efficiency value close to 1, and smaller save efficiency and predictive efficiency values than non-users or low users.

Table 5.28. Image/Tags search measures for tag non-users (N=23), low users (N=9), and high users (N=16)

Variable	Non-users M (SD)	Low users M (SD)	High users M (SD)
Search session duration	266.09 (116.57)	310.22 (100.38)	405.25 (137.65)
Number of unique queries	5.87 (5.46)	6.67 (4.87)	8.81 (3.97)
Number of results	98.74 (64.75)	113.78 (64.27)	166.38 (89.36)
Number of unique documents	6.91 (1.35)	7.89 (3.41)	9.25 (2.44)
Maximum SERP depth	2.26 (1.42)	2.44 (1.67)	2.63 (2.47)
First query time	13.43 (5.64)	12.89 (9.27)	9.88 (4.00)
SERP dwell time	182.13 (99.43)	209.44 (77.17)	272.38 (118.60)
Document dwell time	70.52 (31.48)	87.89 (36.15)	123.0 (50.46)

Table 5.29. Image/Tags search derived measures for tag non-users (N=23), low users (N=9), and high users (N=16)

Variable	Non-users M (SD)	Low users M (SD)	High users M(SD)
Average QI duration	62.67 (41.38)	75.94 (65.42)	50.89 (30.99)
Query efficiency	2.01 (1.75)	2.02 (1.57)	1.27 (0.72)
Save efficiency	1.56 (1.43)	1.48 (1.49)	0.68 (0.29)
Predictive efficiency	0.75 (0.15)	0.71 (0.22)	0.57 (0.12)

RQ 1.2 How do perceptions of the search process differ for tag and non-tag systems?

RQ1.2 examines perceptions of the search processes for tag (Image/Tags) and non-tag (Image/No Tags) systems. The analysis procedure followed was similar to the one for text searches, using the Wilcoxon signed-rank test for within-subjects comparisons. Initially all 48 subjects were analyzed as a group, followed by analysis of 25 subjects who had used tags at least once, and lastly the 16 subjects who had used tags at least twice in their searches.

In the initial analysis of all 48 subjects, subjects on the Image/Tags system felt they had utilized all the search features more than on the Image/No Tags system. Search

duration was perceived to be lower on Image/Tags searches than on Image/No Tags searches, while satisfaction was higher. Nevertheless, no significant differences were found in the perceptions of the search process for Image/No Tags or Image/Tags systems (Table 5.30).

Table 5.30. Perceptions of search process (N=48)

Variable	Image/No Tags M (SD)	Image/Tags M (SD)	*z*
Previous experience with topic	1.77 (0.97)	1.69 (1.03)	0.661
Plan for search	3.58 (0.71)	3.63 (0.82)	-0.250
Expected information	3.67 (0.95)	3.63 (0.91)	0.315
Easiness of search	3.65 (0.91)	3.65 (0.91)	0.111
Expected information found	2.85 (1.05)	3.04 (1.15)	-0.887
Lost during search	2.48 (1.01)	2.67 (1.04)	-0.888
Utilization of search features	3.27 (0.79)	3.50 (0.83)	-1.834
Search duration perception	3.13 (1.16)	2.81 (1.12)	1.585
Satisfaction	3.56 (0.97)	3.77 (0.97)	-1.341

Similarly to text search, subjects who had used tags at least once for image search were investigated separately. Some changes that can be seen from the measures for the entire set of 48 subjects are that tag users on the Image/Tags system were less satisfied with their searches than Image/Tags users overall, and tag users felt more lost during search than overall users of the Image/Tags system. Searches on the Image/Tags system were also perceived to be longer than searches on the Image/No Tags system. No significant differences were found for any of the measures for searches on Image/No Tags and Image/Tags systems (Table 5.31).

Table 5.31. Search perceptions of subjects who used tags at least once (N=25)

Variable	Image/No Tags M (SD)	Image/Tags M (SD)	z
Previous experience with topic	1.76 (0.93)	1.68 (1.07)	0.577
Plan for search	3.60 (0.65)	3.68 (0.80)	-0.501
Expected information	3.64 (0.99)	3.64 (0.81)	0.083
Easiness of search	3.64 (0.95)	3.60 (0.87)	0.307
Expected information found	3.00 (1.08)	3.00 (1.08)	0.169
Lost during search	2.52 (1.08)	3.00 (1.04)	-1.484
Utilization of search features	3.16 (0.69)	3.44 (0.71)	-1.464
Search duration perception	2.92 (1.15)	3.08 (1.04)	-0.631
Satisfaction	3.64 (0.95)	3.56 (1.00)	-0.165

As some differences had been observed between tag users who used tags at least once and those who used tags at least twice in their text searches, tag users who had used tags at least twice for image search were examined separately. There were 16 subjects who used tags two or more times in their image searches (Table 5.32). These subjects felt more strongly that searches on the Image/Tags system took longer than expected than on the Image/No Tags system. Search sessions on the Image/Tags system took about 50 seconds longer than on the Image/No Tags system for these 16 subjects, which may have contributed to this perception. Interestingly, when searching on the Image/Tags system, subjects started out with a significantly clearer idea on how they would search than when searching on the Image/No Tags system ($p=0.0190$). This may be attributed to task order effect, as the Image/Tags search was the fourth search for 7 of the 16 subjects (44%), by which point subjects would have been exposed to the Text/Tags system as well as already completed one image search task. Yet after completing their searches, these 16 subjects felt their searches took longer, felt more lost during their search, and were not as satisfied with their searches on the Image/Tags system. Perception of search length could be related to subject fatigue, which may also influence satisfaction. But it is interesting that

subjects felt more lost even though 10 of these 16 subjects (62.5%) were on their third or fourth search in the experiment. The 16 subjects also felt they had utilized search features more fully when searching on the Image/Tags system. This suggests as with text search that subjects noticed and started using tags when the search was not going well. Subjects were also likelier to use tags repeatedly in image searches compared to text searches. This may be due to tags being more salient text elements on the SERP for image searches than on the SERP for text searches.

Table 5.32. Perceptions of repeat tag users (N=16)

Variable	Image/No Tags M (SD)	Image/Tags M (SD)	z
Previous experience with topic	1.69 (0.60)	1.50 (0.89)	1.134
Plan for search	3.38 (0.62)	3.88 (0.72)	-2.345*
Expected information	3.56 (1.09)	3.88 (0.72)	-0.996
Easiness of search	3.69 (0.95)	3.63 (0.88)	0.320
Expected information found	3.06 (1.12)	3.06 (1.18)	0.164
Lost during search	2.50 (1.10)	2.88 (1.15)	-0.864
Utilization of search features	3.06 (0.77)	3.50 (0.82)	-1.585
Search duration perception	2.69 (1.01)	3.19 (1.11)	-1.933
Satisfaction	3.88 (0.89)	3.69 (1.14)	0.150

Note: * $p < 0.05$

Perceptions of searches on the Image/Tags system were examined for tag non-users, low users (only used tags once), and high users (used tags at least twice) (Table 5.33). Subjects who had only used tags once during their searches were the least satisfied of the three groups, while perceiving the searches to be more difficult, feeling more lost during search, and not finding the information they had expected. Interestingly subjects who did not use tags were the most satisfied with their searches. They also rated their utilization of search features higher than both types of tag users. Tag non-users found the searches to be easiest, and felt the least lost during search. This further supports the idea

that subjects turn to tags when they are feeling lost during a search or feel a search is difficult. If the search is going well, subjects will not use tags and report high satisfaction with their search. When a search is not going well, a subject may try using tags. If the tags were found to be helpful, they will be used again, but if they were not helpful, the subject will not use them again and their difficulty and dissatisfaction with the search will not be alleviated.

Table 5.33. Image/Tags perceptions for tag non-users (N=23), low users (N=9), and high users (N=16)

Variable	Non-users M (SD)	Low users M (SD)	High users M (SD)
Previous experience with topic	1.70 (1.02)	2.00 (1.32)	1.50 (0.89)
Plan for search	3.57 (0.84)	3.33 (0.87)	3.88 (0.72)
Expected information	3.61 (1.03)	3.22 (0.83)	3.88 (0.72)
Easiness of search	3.70 (0.97)	3.56 (0.88)	3.63 (0.88)
Expected information found	3.09 (1.24)	2.89 (0.93)	3.06 (1.18)
Lost during search	2.30 (0.93)	3.22 (0.83)	2.88 (1.15)
Utilization of search features	3.57 (0.95)	3.33 (0.50)	3.50 (0.82)
Search duration perception	2.52 (1.16)	2.89 (0.93)	3.19 (1.11)
Satisfaction	4.00 (0.90)	3.33 (0.71)	3.69 (1.14)

5.3.3 Summary of RQ1

The use of tags during the text search process was associated with more unique queries being issued, more search results being seen, more time being spent on the SERP, and shorter query intervals than when searching on a system without tags. On the Text/Tags system, subjects who used tags were able to explore a larger search space while taking less time for each query instance and making many more predictive judgments. Despite the larger number of unique queries and number of results seen, the number of documents seen was the same as for searches on the No Tags system. Comparing perceptions of searches, subjects felt more lost searching on the Text/Tags system than the Text/No Tags system. This effect was not present for subjects who had

used tags two or more times in their text searches. Examining subjects who reported a high level of feeling lost and had used tags in their searches, it appears subjects who were lost or frustrated about their search attempted to use tags, and tags did not behave as expected for them.

While a number of significant differences were seen in the text searches using tags compared to searches without tags, differences were not as evident in image searches. Similar patterns as in text search were observed, with more search results being seen, more time being spent on the SERP, and shorter query intervals when tags were used in search. There was no significant difference in the number of unique queries issued, although the number was larger when tags were used two or more times during a search. When searching for images, on average subjects looked beyond the second page of search results, compared to text searches, in which subjects rarely went beyond the first page. This increases the number of results viewed without increasing the number of unique queries.

For both text and image searches, it appears subjects use tags when the search is not going well in some way, such as feeling lost during the search or finding the search difficult. If tags are found to be helpful, tags will be used more than once, but if tags are not helpful because they do not behave as expected by the searcher, tags will not be used again. Meanwhile, subjects who are experiencing little or no difficulties with their searches may feel no need to use tags. Overall, perceptions regarding searches are more positive for those not using tags and those using tags more than once, and least positive for those using tags only once.

5.4 Research Question 2: Use of Tags in Text and Image Searches

This section presents the results and discussion for Research Question 2: How does the role of tags differ for text searches and image searches across the search process? Tag use for image and text searches is compared through the following sub-questions:

RQ2.1 How are tags used for query reformulation in text searches and image searches?

RQ2.2 How are tags used for predictive judgments of relevance in text searches and image searches?

RQ2.3 How are tags used for evaluative judgments of relevance in text searches and image searches?

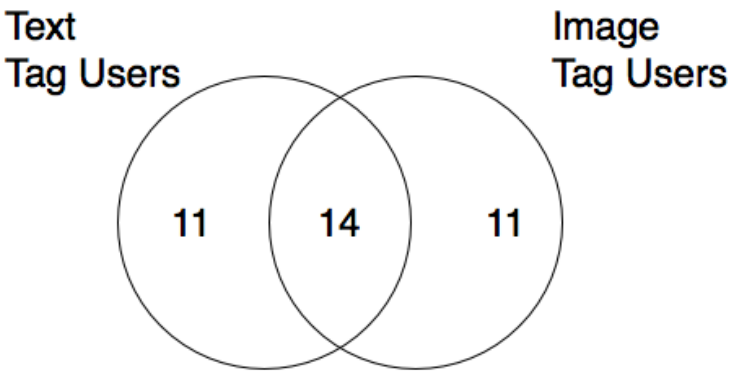
RQ2 extends the findings from RQ1 by examining in detail how tags are used for each different type of resource, across different parts of the search process. Thus the analysis focuses on the 36 subjects who used tags in their text or image searches on the two tagged interfaces, Text/Tags and Image/Tags.

5.4.1 Characteristics of image and text searches

Thirty-six out of 48 subjects (75%) used tags during one or more of their searches using the tagged systems. Fourteen subjects used tags for both image and text searches, 11 subjects used them only for text searches, and 11 subjects only for image searches (Figure 5.2). Thus equal numbers of subjects used tags for text searches (N=25) and image searches (N=25). Tags were used more frequently for image searches (66 times) than for text searches (48 times). Tags were used an average of 1.92 times (SD=1.22) by

subjects who used them in text searches, and 2.64 times (SD=2.46) by subjects who used them in image searches. Text searches on average took 6 minutes and 40 seconds (N=25), and image searches took 6 minutes and 11 seconds (N=25). Some of the searches took longer than 10 minutes as subjects were allocated extra time upon request, with the longest session, a text search, taking nearly 12 minutes.

Figure 5.2. Number of tag uses for text searches and image searches



Tags were used for query reformulation, predictive judgment, and evaluative judgment in similar proportions of sessions for both text and image searches (Table 5.34). Tags were used most frequently for query reformulation, somewhat less so for predictive judgment, and rarely used for evaluative judgment. The picture is similar when tag use frequency per session is counted (Table 5.35). Again, tag use in evaluative judgment is rare, while tags are used for predictive judgments somewhat more frequently in text searches than in image searches. But there is a considerable difference in the number of times tags were used for query reformulation in text searches compared to image searches. Tags were more likely to be used for query reformulation in image searches than text searches. These differences are explored in more detail in the following sections.

Table 5.34. Use of tags per search session

Resource Type	Query reformulation	Predictive judgment	Evaluative judgment
Text	18 (72.0%)	14 (56.0%)	3 (12.0%)
Image	17 (68.0%)	13 (52.0%)	1 (4.0%)

Note: percentages are computed from N=25

Table 5.35. Tag use frequency during search process

Resource Type	Query reformulation	Predictive judgment	Evaluative judgment	Total
Text	27 (56.3%)	18 (37.5%)	3 (6.3%)	48 (100%)
Image	50 (75.8%)	15 (22.7%)	1 (1.5%)	66 (100%)

Table 5.36 shows the number of sessions in which search interaction processes were used singly and in combination. There was only one session for which tags were used for query reformulation, predictive judgment and evaluative judgment. For both text and image searches, nearly half of the sessions using tags only used them for query reformulation. Comparable numbers of sessions used tags for predictive judgment only, and for both query reformulation and predictive judgment. Of the 50 sessions using tags, 35 sessions used tags in only one stage of the search process.

Table 5.36. Tag use in search sessions across search process

Type of tag use	Text (N=25)	Image (N=25)
Query reformulation only	11	12
Predictive judgment only	5	7
Evaluative judgment only	0	0
Query reformulation and predictive judgment only	6	5
Query reformulation and evaluative judgment only	0	0
Predictive judgment and evaluative judgment only	2	1
Query reformulation, predictive judgment, evaluative judgment	1	0

5.4.2 Tag use in query reformulation

Tags can be used for query reformulation directly, by clicking on a tag, or indirectly, by using a tag term in a search query. Indirect use occurs when entering the tag term as is, as an addition to other query terms, or using a query term based on a tag term. For example, some of the participants entered a tag term (e.g. *newyorkcity*) into the search box instead of clicking on the tag itself, or added a term they saw in the list of tags into their query (e.g. composing the query "laptop purchase" after seeing the tag *purchase*). While direct tag use for query reformulation can be extracted from query logs, obtaining indirect tag use information is not as straightforward.

In this study, direct tag use in query reformulation was extracted from the query log for the experiment system. Indirect tag use was identified through review of the searches during the post-search interviews. In addition, the type of direct tag use was identified through the search recording review in the post-search interviews, as this information was not recorded in the query logs. There are three types of direct tag use in this study, which are based on the location of the tags in the experiment system: (1) list of related tags on top of list of search results; (2) tags displayed with each search result; and (3) tags displayed in the document page.

Tags were used for text query reformulation 27 times in 18 search sessions, and used for image query reformulation 50 times in 17 search sessions. Tags were used nearly twice as many times in image query reformulations (2.9 times per session) compared to text query reformulations (1.5 times per session). Broken down into direct and indirect use, tags were used directly in 12 text search sessions and indirectly in 6 text search sessions. No text search sessions had both direct and indirect use of tags in query

reformulation. In contrast, 4 subjects used tags both directly and indirectly for query reformulation in their image search sessions. Tags were used only directly for query reformulation in 12 image search sessions, and used only indirectly for image query reformulation in only one session. So tags are much more likely to be used directly in image searches compared to text searches.

Table 5.37 summarizes the frequency of tag use for query reformulation according to the location of the tag used, per search session. In the case of text searches, tags in the list of related tags are more likely to be used than tags in the individual search results or the document page. For image searches, tags in the document page were used less frequently than tags in the search results or list of related tags, which are both in the search results page. Tags were used from the document page in similar proportions for text and image searches: 18.5% of the time for text searches, and 20% of the time for image searches.

Table 5.37. Location of tags used in query reformulation

Resource - Tag use type	Search results	Related tags	Document	Total
Text - direct	2 (9.5%)	16 (76.2%)	3 (14.3%)	21
Text - indirect	3 (50.00%)	1 (16.7%)	2 (33.3%)	6
Image - direct	15 (33.3%)	21 (46.7%)	9 (20.0%)	45
Image - indirect	2 (40.0%)	2 (40.0%)	1 (20.0%)	5

We next examined at which point during the search tags were used directly for query reformulation. First, for each instance of tag use for query reformulation, the time at which it occurred relative to the beginning of the search was identified. For direct tag use, this was the time at which the tag was clicked. This tag use time, which was the number of seconds since the beginning of the search session, was divided by the overall duration of the search session, also in seconds, to obtain the relative time into search.

Thus, a value close to 0 indicates that tag use occurred near the beginning of the search, while a value close to 1 indicates that tag use occurred near the end of the search. The reason for using this relative measure instead of actual time is that search session lengths differ, so that using a tag 45 seconds into a search may place it near the beginning of the search for long searches, but it may actually be close to the end of the search for short searches.

Figure 5.3. Relative tag use times for direct use of tags in query reformulation

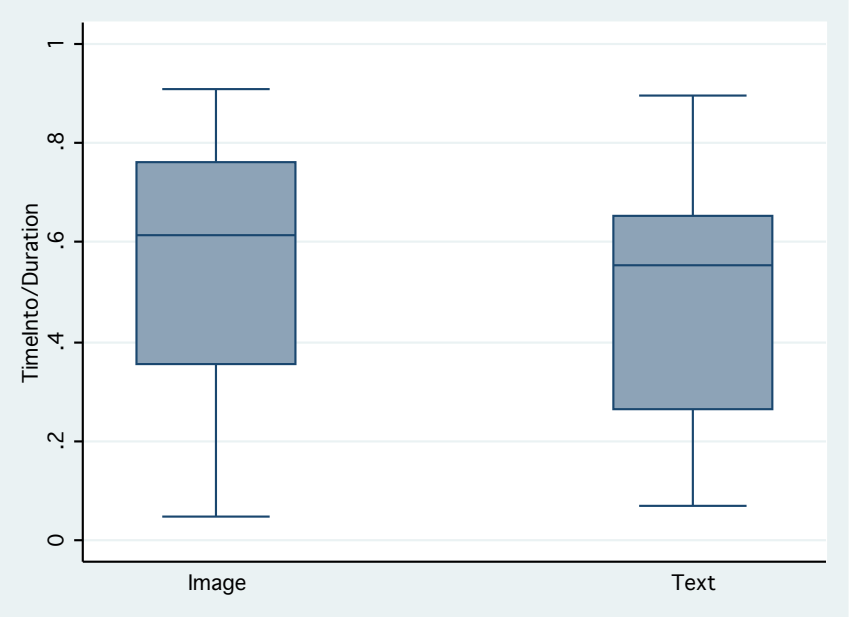


Figure 5.3 shows the distribution of relative tag use times for direct use of tags. For both image search (Median=0.61) and text search (Median=0.55), the median relative tag use time for query reformulation is after the search halfway point. A value of 0.5 for relative time into search represents the halfway point in a search session. The distribution indicates that direct tag use in text search tends to occur earlier in the search than in image search. More direct tag use may be occurring later in an image search because

subjects may not have noticed tags until then, or may choose to use them as a way to change or refine their queries:

> [S02]: "But I noticed the tags more and more as I kept going. ... I'm way better off doing these tags because you're just wasting your time with trying to find keywords and trying to do flips to find the best keyword."
>
> [S03]: "Nothing came up for 'women's rights'. So then I decided to click on *suffrage* just to see what would show up."
>
> [S22]: "And now I noticed the tags. I was like, oh, tags, green. Because I was getting frustrated. I was like, what do I do? Oh, there's tags. And this is when I noticed there's tags. I was like, that would be really helpful. *womensday* sounded like something women's, we're looking for Women's History Month, *womensday*, I thought that would be interesting. So I clicked on that."

In these examples, subjects were in the final quarter of their image search sessions, and clicked on tags as a response to frustration with their search queries. Frustration can lead to noticing tags, but tags are also noticed from the continuing exposure to them as the search session progresses. Subject who noticed tags in a text search session tended to notice tags fairly early in the search session. This seems to be because in text search, tags are also text and get noticed in the course of reading the other text on the screen, while in image search, subjects tend to remain focused on the thumbnail or image.

5.4.2.1 Predictive judgment

The use of tags for predictive judgment is exclusively an indirect use of tags, where tags are used as a way to help decide whether to click on a search result or not. Thus, the use of tags for predictive judgment was identified through the post-search interviews. In the interviews, participants mentioned using tags differently for image and text searches when evaluating search results.

For text searches, tags were seen as a quicker and more convenient way of getting a snapshot of the content than reading the accompanying text snippet. Tags were also used in conjunction with the title for judgments of relevance. That is, the combination of title and tags was used as an alternative to reading the text snippet in the search results. Thus, tags allowed subjects to make quick predictive judgments:

> [S14]: "Because there's just so many laptops out there and different operating systems and preferences so if I saw like tags like I mentioned before, *linux*, or *cheap* or something like that, I would just ignore it. I guess it just helps me filter through quicker and I automatically assume that it's kind of like the topic. Instead of reading this whole chunk of text I think it helps me decide faster if I want to go that page."

> [S40]: "I don't wanna read like all of the fine print. So I'll look at the bold [title] and then look at the tags and if the tags are there and the tags correspond to what the bold blue says then I'll click on it."

In particular, tag use was frequently mentioned in relation to negative predictive judgments:

> [S14]: "I looked at the tags and it said *linux* and *cheap* and that's, I don't want, so I guess I thought that would have been a good page to look at but then I saw the tags and I didn't."

> [S23]: "The green [tags] in terms of like, oh *Cincinnati*, I don't want to go there [look at that search result]. Just looking at *sanfrancisco*, I don't want to go there [look at that search result].

> [S27]: "It was these two [tags *tennessee* and *indiana*] that made me not want to click on this."

In contrast to text searches, in image searches tags were used to gain additional information about the image depicted in the thumbnail, and were not used as an alternative to the thumbnails. In image searches the image thumbnail remained the main SERP item relied on for predictive judgments, supplemented by the title and sometimes the tags. The information provided by tags could be informative, but also cause confusion:

[S04]: "Then I noticed like *penn*, *pennstation* then I wasn't sure if it was in Pennsylvania, even though it says like New York. I just wasn't... It's probably why I didn't pick it."

[S05]: "I looked at the thumbnail and I had to like decide, like a lot of these things I didn't know exactly what they were, so I looked at the tags, like *manhattan*, I'm like okay, that actually is in New York, and like... And it said *newyork* in its tags so then I'm like, okay it actually is in New York."

Tags could also be misleading for participants, especially for image searches. Images in The Commons include in their tags the name of the originating institution (e.g. *smithsonianinstitution* for images from The Smithsonian Institution). Participants interpreted this provenance tag as having something to do with the image itself, such as the image being of a Smithsonian Institution building. Participants would then conclude that a photograph was not of the correct geographic location for the assigned search task (Washington, D.C. instead of New York City) and thus not relevant for the search task.

The appearance of the tags sometimes contributed to tags being used for negative predictive judgments. S45, when trying to decide which of two search results to click on, chose one over the other because "I saw the [tag] *laptoprecommendation*, but I didn't like the fact that it wasn't like, two separate words."

5.4.2.2 Evaluative judgment

As with predictive judgment, the use of tags for evaluative judgment was identified through the post-search interviews. Use of tags for evaluative judgment was rare. Once on the document page, participants focused on the content – question and answer text for text searches, the image in image searches – and tended to ignore metadata, such as the date when a question was asked, who posted an image, or the tags.

Tags contributed to evaluative judgment in three text search instances and one image search instance.

In the three text search instances, tags played a confirmatory role, assuring subjects that the document they had selected to view was indeed relevant to their search task:

> [S19]: "Well, those are like the tags for the main points of this article. So it just gave me a clue that that's around about what I was going for."
>
> [S37]: "When I got to about here, I did notice *laptop* and *notebook* and *budget*. And I was like, oh, this must be, this is exactly, or pretty close, to what I want.
>
> [S48]: "I saw *gourmet*, and I was like, that's always a good way to go. ... it gave me an idea of what kind of food. It wasn't just a pub on the street, you know, gourmet food is more higher-end."

S19 and S48 were both looking at the same document, titled "Chicago Restaurants for Cheap Locavores." The title may have contributed to their looking at the tags for confirmation – they were not familiar with the term "locavore" but had selected the document based on the rest of the title. The tags for the document, *food*, *gourmet*, *slowfood*, *seasonal*, and *local*, provided an indication of the types of restaurants being discussed. S19 was puzzled by the tag *slowfood* ("It's a little strange. I don't know what that means.") but found the other tags indicative of the topic of the document. S19 later saved the document, but only after seeing the document turn up in 2 subsequent query intervals. So the use of tags for evaluative judgment did not lead to saving the document during the query interval in which tags were used. S48 did save the document at the conclusion of the query interval in which tags were used for evaluative judgment, as did S37. S37 had selected the document from the SERP based on the title, "One cheap laptop with everything, please" and the tags provided further confirmation the document was

relevant to his search. S19 used tags for evaluative judgment only 27 seconds into a search lasting 5 minutes and 43 seconds. S37 and S48 used tags for evaluative judgment in the latter half of their search sessions.

In the case of image search, tags contributed to the decision of whether or not to save the image by providing additional information about the photograph:

> [S44]: " On this one, I read the tags and that's how I knew that it was about strikes and stuff. And then like what the picture was about... So, then I clicked [took a screenshot of] that picture.

S44 selected this photograph, titled "Union Square, New York. J.J. Ettor speaking to striking barbers (LOC)," from the search results because she thought it had to do with Union Station, based on the title. Once on the document page, she turned to the tags instead of the title or descriptive text to determine what the photograph was about. She pointed out that in the tags for the photograph, "it says 'strike' like five times," making clear the subject matter, which she considered part of the history of New York. The photograph had 25 tags, 6 of which included the term strike: *strike*, *strikes*, *laborstrikes*, *barberstrikes*, *barbersstrikes*. Other tags described the location as well as components of the picture, such as *unions* and *crowds*. The photograph was the 4th of 12 photographs S44 viewed during her search, and it was in the second page of search results she saw for her first search query. She did not use tags for evaluative judgment in subsequent query instances or in text search on the Text/Tags system later.

Interestingly in all four cases none of the subjects noticed the tags for the search result they selected on the SERP prior to viewing the document page. All four subjects did save the pages they viewed. None of the subjects used tags for evaluative judgment for other documents in their respective searches, or other searches they did on tagged systems. So although for the three text searches tags had been used for predictive

judgment, this use was not connected to the use of tags in evaluative judgment. The findings indicate that tags are used for evaluative judgment sporadically, but when it does happen, it is due to lingering uncertainty after a document has been selected for viewing.

5.4.3 Summary of RQ2

The use of tags for query reformulation, predictive judgment, and evaluative judgment was examined. Tags were rarely used for evaluative judgment of relevance, whether for text searches or image searches. When tags were used for evaluative judgment, this use was not connected to use in predictive judgment. Tags were most commonly used for query reformulation for both text and image searches, but were more frequently used in query reformulation in image search. Query reformulation dominated the use of tags in image search. The majority of this use was of the direct type, or clicking on the tag. Tag use in text search was more evenly distributed; tags were used most frequently for query reformulation but this use of tags did not dominate to the same extent as in image searches. There was almost no difference in the number of text search sessions and image search sessions using tags.

Tags were used differently for predictive judgment in text searches than in image searches. For text searches, tags were seen as a quicker and more convenient way of getting a snapshot of the content than reading the accompanying text snippet. The combination of title and tags was used as an alternative to reading the text snippet in the search results. In image searches, tags were used to gain additional information about the image depicted in the thumbnail, and were not used as an alternative to the thumbnails.

Differences in patterns of tag use for text and image searches were observed when tag use in query intervals was counted. For example, for query reformulation, tags were

more likely to be used more than once for image searches than for text searches. Tag use in both text and image searches occurred most frequently from the search results page, the tags in the search page being used for both query reformulation and predictive judgment. When tags in the document page were examined, the purpose was to obtain more search terms or ideas for searching, and rarely for evaluating the relevance of the document.

5.5 Research Question 3: Factors Influencing Tag Use in Search

This section presents the results and discussion for Research Question 3: How does prior knowledge of tags influence the use of tags during the search process? The findings are discussed in association with the following related questions:

RQ3.1 What prior knowledge of tags do users have?

RQ3.2 To what extent does prior knowledge of tags influence the use of tags in the experiment system?

RQ3.3 To what extent does experience with other information retrieval systems influence the use of tags in the experiment system?

Findings are primarily based on the post-search interview conducted at the conclusion of all four searches. The post-search interview consisted of two parts: (1) retrospective interview conducted while reviewing the screen recordings of the two searches on the tagged systems; and (2) interview on the subject's experience of tags on the Web in general, and not restricted to the search systems used in the experiment. Section 5.1.1 draws primarily from the second part of the post-search interview, while section 5.1.2 draws from both parts of the post-search interview and search logs.

5.5.1 Prior knowledge of tags

Prior knowledge of tags is conceptualized as having the following components: familiarity, experience, and understanding. Familiarity is the subject's perception of their knowledge of tags. Experience is based on their actual behaviors relating to tags. Understanding is what the subject knows about tags. Thus familiarity ratings were obtained directly from the subject, while experience and understanding were derived based on the subject's interview answers.

For familiarity, subjects were asked to rate their familiarity with tags on a scale of 1 ("Not familiar") to 5 ("Very familiar"). Experience with tags was ascertained through the following interview questions: "Have you noticed tags on any of these sites?", "Have you used tags on any of these sites or other sites you can remember?", and "Have you tagged anything yourself?" These questions were asked after the subject had been asked to list websites they visited frequently, and the frequency of their visits. Subjects also provided information about their experience with tags during the course of answering other interview questions. Experience was coded by the researcher based on whether a subject had (1) noticed tags on websites, (2) used tags on websites, and/or (3) had tagged items. Understanding of tags was determined from answers to the following questions: "What do you think tags are?" for concept, "Who or what is creating or putting the tags there?" for origin, and "What is the purpose of tags?" for purpose. The analysis of understanding focused on deriving categories from the interview data.

5.5.1.1 Familiarity

Subjects expressed moderate familiarity with tags (M=2.67, SD=.91 on a 5-point scale where 1 is "Not familiar", 3 is "Moderately familiar" and 5 is "Very familiar"). Fourteen subjects (29%) rated themselves 2 and twenty-two subjects (46%) rated themselves 3. Subject responses for familiarity with tags are summarized in Table 5.38.

Table 5.38. Familiarity with tags

	N	%
1 ("Not familiar")	5	10.4
2	14	29.2
3 ("Moderately familiar")	22	45.8
4	6	12.5
5 ("Very familiar")	1	2.1
Total	48	100.0

Subjects rating themselves 2 or 3 in familiarity generally mentioned they did not usually use tags, although they knew what they were. Some examples of subjects rating themselves 2 are shown below:

> [S20]: I'm familiar with them and know about them, I just don't really use them.
>
> [S30]: I don't use them enough to be familiar with them. So I'm just slightly not familiar with them yet.
>
> [S32]: I know what they are and I've probably used them a couple of times. ... I usually skip over them just because if I'm gonna search for something I generally, like, search myself.
>
> [S41]: I know what they are and I know why they're there, but I don't choose to use them.

Subjects rating themselves 3 similarly mentioned lack of regular use as to why they considered themselves moderately familiar with tags:

> [S03]: I know what they do, but I don't really use them. So, I know how to use them, if I need to. I'm just not very experienced doing so.

[S16]: It's pretty much I understand how to use them, I don't really use them a whole lot, like in other searches.

[S17]: I am familiar but I hardly use them.

[S29]: I know basically what they're meant to do and stuff but I never use them. So I'm not familiar with using them in that case.

Three of the six subjects rating themselves 4 could not recall any sites in which they'd seen, used, or attached tags. The one subject who rated themselves 5 ("Very familiar") and another subject who rated themselves 4 had actually taken classes which explained tags to them as well as requiring uploading photos on Flickr and tagging bookmarks on Delicious, respectively. So although all seven subjects rated themselves highly on their familiarity with tags, their familiarity with tags was actually quite varied. This was in contrast with subjects rating themselves 1 ("Not familiar"), who indeed were quite unfamiliar with tags:

[S01]: I don't really know what tags are.

[S07]: Not very familiar at all. ... tags I don't use hardly at all, so I guess I wouldn't be familiar with those.

[S31]: I'm not really sure what tags are.

Subjects' comments in relation to familiarity indicate that familiarity is linked to knowing what tags are or knowing how to use them, but not necessarily to experience with tags from current or past usage. As we discuss later, subjects can rate their familiarity with tags relatively low while displaying a sophisticated understanding of tags, while others rate their familiarity high even while displaying an inaccurate understanding of tags.

5.5.1.2 Experience

Experience was initially derived as one of five categories based on whether a subject had noticed or seen tags on a website, used tags on a website, and/or tagged, (see Table 5.39). It was decided to restrict tagging experience to websites other than Facebook, for reasons described later in the section. There is a kind of hierarchy to prior experience with tags – that is, noticing tags but not using them implies less experience than using them, while to use tags one must have first noticed them. On the other hand, it is not clear whether somebody with tagging experience has necessarily used tags as described in this study, raising the possibility of having tagging experience without having used tags. Tag use includes clicking on a tag, as well as examining tags to additional information or help with query terms. Two subjects in the study had tagging experience but did not report using tags in the sense described. For both subjects their tagging was for "search purposes" (S12), so others could search for and find the items they had uploaded. Thirteen subjects (27.1%) had experience using tags and tagging on sites other than Facebook.

Table 5.39. Initial tag experience categories

	N	%
Have not seen or noticed tags	9	18.8
Have seen or noticed tags	8	16.7
Have used tags	16	33.3
Have tagged	2	4.2
Have both used tags and tagged	13	27.1
Total	48	100.0

Note: Facebook is excluded from the counts

Fifteen subjects, or 31% of the subjects, in this study had tagged content online, such as photos, videos, or blog posts. This is similar to the Pew Research Center's finding that 28% of internet users had tagged content online (Rainie, 2007). In addition, there

were 13 subjects in this study who had Facebook tagging experience. But subjects recognized Facebook tagging as being different from the type of social tagging seen on photo sites or blogs:

> [S30]: Because like, I think that for like Flickr or journals, you're like trying to categorize the pictures or information into certain categories. But then like for, Facebook it's just like saying that this person is in this picture. So, it's not really like categorizing it. Like on Flickr, you can say like this picture has water in it so you'll tag it as water. But then, you're not really tagging water on Facebook images.
>
> [S46]: Well, on Facebook and Twitter, tags are people. Tags are the people that you're mentioning. And so if you click on it, a tag on Twitter, it will take you to someone else's profile. And the same with Facebook. And they're noticeable on Facebook, because most of the time they're people. And having that format already in place, when somebody for instance tags a duck, it's silly.
>
> [S47]: It [Facebook tagging] is actually tagging people rather than tagging information.

At the time of data collection, tagging on Facebook was restricted to linking the profiles of Facebook users to photographs. Therefore, tagging experience was restricted to systems other than Facebook.

In examining the websites subjects mentioned in relation to their tagging experience, it emerged that on some blogs (e.g., PassiveAggressiveNotes.com) what subjects described as tags were actually labeled as categories. A number of blogging platforms allow the blogger to define categories and then attach one or more categories to a blog post. Functionally these categories are indistinguishable from tags, as new categories can be created at any time, and clicking on a category will return all blog posts associated with that category. In the case of WordPress, a popular blogging platform which has implemented both tags and categories, WordPress users continue expressing confusion regarding the difference between the two, or if both are necessary. So for the

typical blog reader, tags and categories can be very difficult to differentiate. For these reasons, if a subject identifies a blog site using categories as one using tags, it is still counted as a site implementing tags, and counted as having noticed or used tags on that site. Another issue that emerged with the websites mentioned by subjects was that some subjects reported having seen or used tags on websites that do not have tags, such as Google. If a subject's experience with tags was restricted to such sites, the corresponding tag experience was not counted.

Table 5.40. Web sites and tagging experience.

Site	Seen/noticed		Used		Tagged	
	N	%	N	%	N	%
YouTube	24	61.5	12	41.4	5	33.3
Blogs[a]	11	28.2	8	27.6	5	33.3
Flickr	6	15.4	6	20.7	3	20.0
News and media	4	10.3	2	6.9	0	0.0
Delicious	2	5.1	2	6.9	2	13.3
MTagger	1	2.6	1	3.5	1	6.7
Forums	1	2.6	1	3.5	0	0
Unspecified	3	7.7	2	6.9	0	0
Number of subjects[b]	39	100.0	29	100.0	15	100.0

Note: [a] Sites self-identifying as blogs are counted as such – for example, the TrueHoop blog from ESPN is considered a blog, and not a news or media site such as the Women's Health Magazine site.
[b] The number of subjects is the number of subjects with the specific type of tag experience. As a subject can interact with more than one site, this number is smaller than the column sum.

Table 5.40 summarizes which websites subjects mentioned in relation to their experience with tags. YouTube, blogs, and Flickr are the top three sites for all types of tag experience. While tags were seen and/or used on news and media sites such as ESPN or SOAPnet, there was no tagging activity since these sites do not allow user tagging. There were two subjects (S14, S42) who were required to use Delicious for a class, and so they saw tags, used them, and also tagged. Both personal and other types of blog sites

are included in the blogs category, but tagging only occurred in blogs the subjects contributed to as authors.

YouTube, Flickr, and LiveJournal were among the most frequently used sites according to the background questionnaire (Appendix B.6). Overall Facebook was the most frequently used site, being used by all subjects and nearly 90% using it everyday, and is included for comparison with other sites implementing tags. The most frequently used sites with tags were in descending order, YouTube, Flickr, Last.fm, and LiveJournal. Except for Last.fm, these were also the sites with the highest frequencies for tag experience, where the number of LiveJournal users is a subset of blog users. The frequency of use of these sites in summarized in Table 5.41.

Table 5.41. Frequency of use of selected sites

	Facebook	YouTube	Flickr	Last.fm	LiveJournal
Less than once	1 (2.1%)	7 (14.9%)	12 (57.1%)	2 (20%)	5 (55.6%)
1-2 times	1 (2.1%)	10 (21.3%)	5 (23.8%)	4 (40%)	0 (0.0%)
3-4 times	4 (8.3%)	15 (31.9%)	3 (14.3%)	3 (30%)	2 (22.2%)
Everyday	42 (87.5%)	15 (31.9%)	1 (4.8%)	1 (10%)	2 (22.2%)
Total users	48 (100%)	47 (100%)	21 (100%)	10 (100%)	9 (100%)

Frequently used sites, not surprisingly, appeared to be where subjects were getting their tag experience. But Last.fm was not mentioned by subjects as a site in which they had tag experience. This may be due to the nature of interaction with the site, which is frequently used to listen to music in the background while engaged in other tasks on the computer. YouTube, Flickr, and LiveJournal all require much more foreground interaction with the sites, as the primary purpose for using these sites is to view the content.

5.5.1.3 Understanding

Having obtained from subjects self-assessments of their familiarity with tags, and measured the extent of their experience with tags, this third component of prior knowledge investigates what subjects actually know about tags. User conceptualization of tags is operationalized as having the following three dimensions: definition ("What are tags?"), origin ("Who or what creates tags?"), and purpose ("What is the purpose of tags?").

Definition of tags

Subjects were asked to describe tags in their own words, through the questions "What do you think tags are?" and "How would you describe tags?" The following definitions of tags emerged from their responses: (1) categories (N =15, 31.3%); (2) keywords (N = 13, 27.1%); (3) related terms or topics (N = 8, 16.7%); (4) links (N = 4, 8.3%); and (5) indexing terms (N = 4, 8.3%). Four subjects responded they did not know what tags were. Definitions and examples of how subjects defined tags are presented below:

(1) Categories

Tags are a way to group related materials together. Here are some examples:

> [S16]: Just labels attached to pages or search results that put them in the broader categories even if the content of the results' text doesn't include that term.
>
> [S30]: I think tags are ways to categorize your information, or information on the Web. Because, on your computer you can make folders and then put your documents into your folder, but on Websites you can't just make a folder and then put it into the folder. So, tags kind of work as folders would work on a regular computer.
>
> [S48]: Broader categories in which certain things are under. So a tag of, if I say "buildings," it'll have like every "building" kind of thing.

(2) Keywords

Tags describe key aspects of the document. Some examples are:

> [S02]: Tags are just the words describing whatever subject that you're on, in different ways.
>
> [S27]: They're keywords or topic words that are pulled out of whatever it is so that you can easily find it.
>
> [S39]: I would say that they are crucial words that describe a page. If you were to summarize the page, in maybe five words or less, those would be the words that you would want to use to describe the content.

(3) Related terms or topics

Tags are terms or topics related to the current search – that is, tags are viewed as query suggestions. Examples include:

> [S34]: For me it's just like something I use to give me an idea or something. Or something like where I can't really think of something yet so I'll use it to give me other ideas. And obviously it's related to the term that I typed in the search engine.
>
> [S41]: ... related terms that you might wish to check out as well.
>
> [S44]: I guess they're just like something that's related to whatever you're looking at or like different parts of it. So, if you're talking about New York like different parts of New York would be tags. Things that would make up New York like if people think of New York they might think like baseball, so that might be a tag.

(4) Links

Tags are links to other pages or sites. Some examples follow:

> [S08]: It's just a connection between the current page to something that's slightly different, off topic or related.
>
> [S29]: They're links that basically take you to articles or sites that have information that is like the name of the tag.
>
> [S74]: I think they are recommended links for the reader that pertain to the subject area that you're looking at.

(5) Indexing terms

Tags are essentially indexing terms, where they may not necessarily be terms describing what is on the page to which the tags are attached, but provide a pointer to the object. The intent is simply to have the page with that tag turn up in the search results when somebody uses the tag term when searching. Some examples are shown below:

> [S22]: If you enter text somewhere, if you're searching for something, the tags are related to what you're entering.
>
> [S36]: You tag a photo with people or things in it so that people can find it. Like, so that when you search it in Google, the tag words lead you to that photo.
>
> [S43]: So a tag is something like whoever is posting it or posting whatever content is it, puts on there. ... If you wanted to find this post, what would you type in to find it?

Tag origin

In the dissertation study, subjects were asked who or what creates tags. In the pilot studies, the question was originally asked as "Who is putting the tags there?" A number of subjects answered that the search engine or "the system" was attaching tags to documents, so the question was modified for the main study to allow non-human sources of tags. Answers to the question of who or what creates tags were surprisingly diverse. In contrast to their more broadly applicable descriptions of tags, several participants identified different tag creators depending on the tagging system. For example, S38 thought content contributors tagged on YouTube and Flickr, but "on Google ... it's automated." Fifteen subjects (31.3%) provided multiple tag creators in their answers. At the same time, a larger proportion of subjects answered they didn't know for this question compared to the definition of tags, with 12 subjects (25.0%) saying they didn't know or were not sure who or what tagged. Some subjects were surprised by the question: "I never really thought about that" (S08).

Four types of tag sources were identified from responses from 36 subjects: (1) content creator (N = 26, 72.2%), (2) site owner or website builder (N = 11, 30.6%), (3) system (N = 11, 30.6%), and (4) general public (N = 3, 8.3%). Definitions and examples of the tag sources are as follows:

(1) Content contributor

The category of content contributor includes authors of the content as well as people who upload content they have not authored themselves to sites such as YouTube. Content contributors were mentioned in connection with personal blogs and sites such as Flickr or YouTube:

[S12]: ... the person who posted it saying what it's about.

[S13]: The person who uploads the picture on the site.

[S19]: Whoever who puts up the video.

(2) Site owner or website builder

Website owners or builders are distinct from content contributors – for example, while users upload videos to YouTube, the site owner is Google. Website owners or builders were usually mentioned in addition to content contributors, and were rarely mentioned on their own. Sites that do not rely on user-contributed content, such as SOAPNet, or sites with a restricted number of content contributors, such as Engadget, were associated with this answer. Some examples are shown below:

[S02]: ... just the website builders themselves.

[S11]: The people that make the website.

[S25]: The person who is running the website or the person who designed it

(3) System

Tags were perceived to have been automatically generated in some way by "the system," and not by people. Subjects usually mentioned this option in addition to content creator, for systems for which they had no first-hand experience tagging:

> [S06]: There couldn't be people putting them in. I mean, it has to be automatically generated. There's too many topics out there.
>
> [S10]: Sometimes on some sites you can tag it yourself. But on others, like, I think some search engines, just do it themselves.
>
> [S37]: I just always assumed whatever program kind of manages how specific my, all the keywords I put in the search bar is [generating the tags].

(4) General public

Tagging is perceived to be carried out by "the general public" or "other people just like me," and is not restricted to content contributors or website owners. Both of the subjects with experience using Delicious (S14, S42) had this perception:

> [S14]: Probably just other people like me. Other people just I guess searching for maybe similar topics.
>
> [S42]: Just the general public.
>
> [S45]: The people who enjoy the website.

Purpose of tags

Participants were asked what they considered to be the purpose of tags. Answers from the four subjects who had previously responded they did not know what tags were excluded from these counts. The purposes that emerged from responses from 44 subjects were findability (N = 29, 65.9%), organization (N = 10, 20.8%), leading people to content (N = 7, 14.6%), and description (N = 6, 12.5%). Eight participants identified multiple purposes for tags.

(1) Findability

Findability refers to making items easier to find, more specifically about making tagged items easier to find when using a search engine. Tags are thought of as links between the tagged item and the search engine:

> [S09]: So that their image or text will come up in your search
>
> [S17]: Search engines use tags as a way to reference searches. And people who create Websites use tags to identify to the search engines
>
> [S41]: So people can find it. Because there's no use of putting on a video if no one can find it.

(2) Organization

Organization is closely related to categorization, or grouping related items together. Tags are not for finding the tagged item itself, but for finding things in the same category as the tagged item. Some examples are shown below:

> [S10]: [Describing tagging their photographs on Flickr] I'm going to tag them like, so you could sort your own stuff, like you can tag them like family, dogs, friends, you can tag years. It was kind of like help sort it.
>
> [S14]: They're probably trying to simplify their own like clutter on the internet, like there's so much information overload these days and they're probably just trying to help themselves, maybe they're also trying to give other people too like with similar interests to put things in categories.
>
> [S46]: I tagged them because they were recipes … it was so I could click on something and make all of my recipes come up.

(3) Leading people to content

Leading people to content is different from findability. The intent is not to help people find what they are looking for, but to manipulate or sometimes mislead people to certain websites or web pages by use of tags. These sites or pages can be advertisements or simply misrepresented content. Users regarded such use of tags negatively, and did not see it as helping them:

[S05]: ... to string you along

[S27]: ... it's like they're trying to lure people in

[S36]: I'd say now, it's commercial, in the sense that even like photographers want people to see their photos, so they put like the list of tag words to help people, like lead people to the photo, people like, oh, like, so it's, there's basically like ulterior motives behind tags that I don't necessarily like want to be associated with.

(4) Description

Tags are intended to provide a concise description or additional information to the content they are attached to. Some examples are shown below:

[S15]: They are just supposed to like, they are just supposed to be like a very concise description, of what is in the article or website and things that are related to it.

[S16]: To identify things beyond what's in the text, because the text doesn't always completely satisfy that.

[S44]: ... it's just to give people different options and more information.

5.5.2 Prior knowledge of tags and use of tags in the search process

This section examines the question: To what extent does prior knowledge of tags influence the use of tags in the experiment system? In section 5.1.1, the extent of prior knowledge of tags of experiment subjects was assessed. This was the prior knowledge of tags subjects had gained from their exposure and experience of tags on the Web in general. This section investigates the relationship between this prior knowledge and the use of tags in the search experiment, looking separately at familiarity, experience, and understanding. Tag use is divided into direct and indirect tag use, where the former refers to clicking on a tag, and the latter refers to using a tag without clicking on it, such as for predictive or evaluative judgments of relevance.

5.5.2.1 Familiarity and use of tags

Whether familiarity has an effect on tag use is examined, where tag use includes both direct and indirect tag use, for both text and image searches. Figure 5.4 shows the proportions of tag users and non-users for each familiarity category, which appears to indicate the proportion of tag users is increasing with familiarity. There were originally five categories for familiarity, but only one subject rated herself as 5 or "Very familiar" in the original scale. Familiarity categories 4 and 5 were collapsed, resulting in a scale with 4 categories, "Not Familiar", "Somewhat Familiar", "Moderately Familiar," and "Very Familiar."

Figure 5.4. Distribution of tag users by familiarity

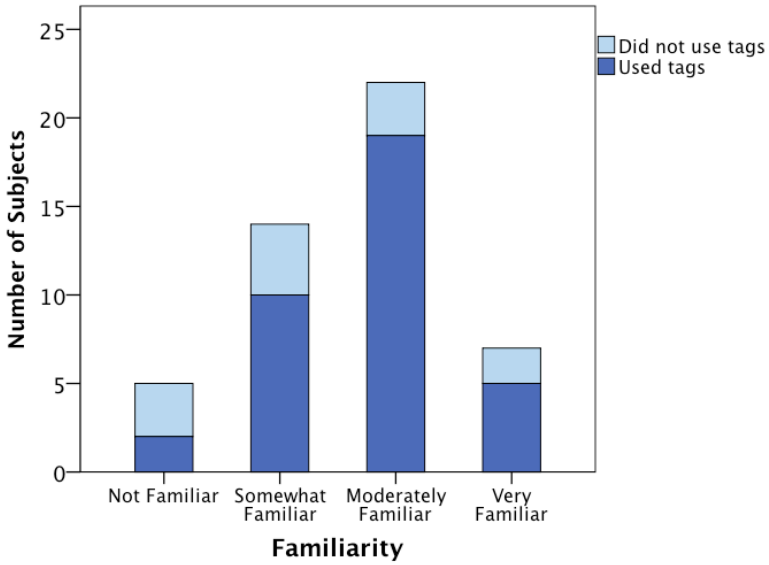


Table 5.42 summarizes how many subjects used tags for different types of tag use. There seems to be a trend for tag use percentage to increase with familiarity level, except for those very familiar with tags. For example, in the case of image search, 28.6% of subjects very familiar with tags used tags in their searches compared to 40% of those who were not familiar. The proportion of indirect tag users for all searches among those

very familiar with tags also relatively low (28.6%) compared to those in the somewhat familiar (50.0%) and moderately familiar categories (72.7%).

Table 5.42. Tag use according to familiarity level

	Familiarity			
Type of tag use	Not Familiar (N=5)	Somewhat Familiar (N=14)	Moderately Familiar (N=22)	Very Familiar (N=7)
All searches	2 (40.0%)	10 (71.4%)	19 (86.4%)	5 (71.4%)
Text search	1 (20.0%)	8 (57.1%)	13 (59.1%)	3 (42.9%)
Image search	2 (40.0%)	6 (42.9%)	15 (68.2%)	2 (28.6%)
Direct tag use	2 (40.0%)	6 (42.9%)	13 (59.1%)	4 (57.1%)
Indirect tag use	1 (20.0%)	7 (50.0%)	16 (72.7%)	2 (28.6%)
Text search - direct tag use	1 (20.0%)	4 (28.6%)	5 (22.7%)	2 (28.6%)
Text search - indirect tag use	1 (20.0%)	5 (35.7%)	10 (45.5%)	2 (28.6%)
Image search - direct tag use	1 (20.0%)	4 (28.6%)	9 (40.9%)	2 (28.6%)
Image search - indirect tag use	1 (20.0%)	4 (28.6%)	12 (54.6%)	0 (0.0%)

The relationship between familiarity and tag use was examined using Fisher's exact test. The analysis was carried out for all nine conditions on Table 5.42. The level of familiarity significantly affected indirect tag use in image searches ($p < 0.05$). Interestingly, none of the subjects who were very familiar with tags used them indirectly in their image searches. There were no common characteristics that could account for all seven subjects who were very familiar with tags not using tags indirectly in image search. In general, those who were moderately familiar with tags showed the highest percentage of tag use. For certain types of searches, familiarity level has an effect on tag use, but we cannot conclude that the more familiar a subject is with tags, the more likely they are to use them.

5.5.2.2 Experience and use of tags

The relationship of tag experience to tag use was examined. The intuition was that subjects who are more experienced with tags would be more likely to use them in the

search experiment. Based on the results from section 5.1.1.2, tag experience was classified as low, medium, or high. Subjects who had not seen tags, or only noticed them, were classified as having low tag experience, while subjects who had experience using them were classified as medium, and those with tagging experience were classified as high. Figure 5.5 shows the proportions of tag users and non-users for each experience category, which suggests a lower rate of tag usage for those in the medium category.

Figure 5.5. Experience and tag use

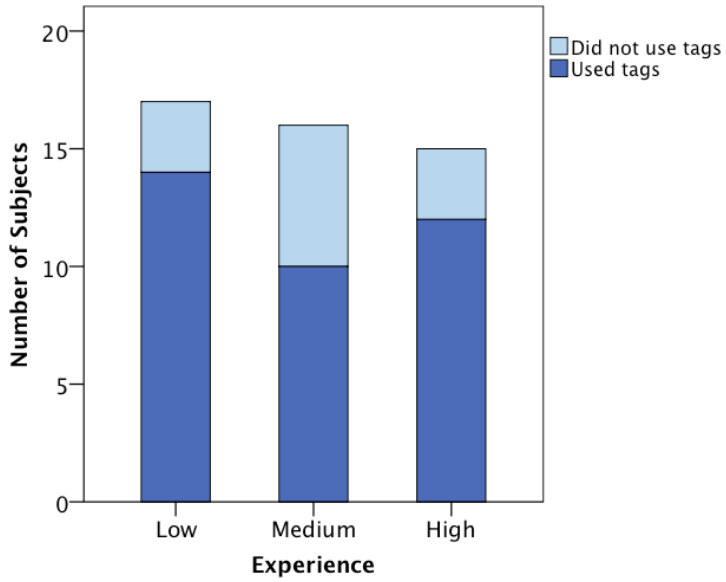


Table 5.43 summarizes how many subjects used tags for different types of tag use, for each experience category. The relationship between experience and tag use was examined using Fisher's exact test. The analysis was carried out for all nine conditions on Table 5.43. Although none of the relationships were statistically significant, there are some interesting patterns in the frequency of tag use. For example, subjects with a medium level of experience, that is, subjects who had used tags on the Web but had no tagging experience, overall used tags in the experiment less than those with tagging experience or those with low tag experience. Those with a medium level of experience

were also less likely to use tags indirectly than the other groups. Relatively high proportions of inexperienced tag users used tags indirectly for both text and image searches. This indicates that although inexperienced users are not clicking on the tags, they are looking at them and using them to get ideas for query reformulation as well as making predictive judgments. On the other hand, some experience with tags appears to make subjects more willing to click on tags. This may be due to their having more of an awareness of what tags are, and not simply seeing them as additional text on a Web page. Subjects with low tag experience had no dominant definition of tags, with nearly a quarter of them not knowing what they were. About 44% of subjects with medium experience thought tags were keywords, while 53% of those with high tagging experience thought they were categories. Nevertheless, previous experience with tags does not appear to be a predictor for tag use in a system not encountered previously.

Table 5.43. Tag use according to experience

	Experience		
	Low	Medium	High
Type of tag use	N=17	N=16	N=15
All searches	14 (82.4%)	10 (62.5%)	12 (80.0%)
Text search	9 (52.9%)	9 (56.3%)	7 (46.7%)
Image search	11 (64.7%)	7 (43.8%)	7 (46.7%)
Direct tag use	8 (47.1%)	8 (50.0%)	9 (60.0%)
Indirect tag use	11 (64.7%)	7 (43.8%)	8 (53.3%)
Text search - direct tag use	4 (23.5%)	5 (31.3%)	3 (20.0%)
Text search - indirect tag use	8 (47.1%)	5 (31.3%)	5 (33.3%)
Image search - direct tag use	5 (29.4%)	4 (25.0%)	7 (46.7%)
Image search - indirect tag use	4 (41.2%)	6 (37.5%)	4 (26.7%)

5.5.2.3 Understanding and use of tags

In this section we examine a number of hypotheses relating use of tags in the search experiment with findings from section 5.1.1.3 on what subjects know of tags.

Definition of tags

Whether subjects' prior conception of tags are related to their use of tags during the experiment is examined. Table 5.44 summarizes how many subjects used tags based on different categories. Fisher's exact test was used to examine the relationship between use of tags and subjects' idea of tags, for all nine conditions in Table 5.44. Although some interesting patterns can be observed, none of the relationships were statistically significant.

Table 5.44. Tag definitions and tag use

	Definition of tags					
Type of tag use	Category N=15	Keyword N=13	Related term (N=8)	Link N=4	Indexing N=4	Don't know N=4
All searches	12 (80.0%)	9 (69.2%)	7 (87.5%)	4 (100.0%)	2 (50.0%)	2 (50.0%)
Text	7 (46.7%)	8 (61.5%)	6 (75.0%)	3 (75.0%)	0 (0.0%)	1 (25.0%)
Image	7 (46.7%)	8 (61.5%)	3 (37.5%)	3 (75.0%)	2 (50.0%)	2 (50.0%)
Direct	8 (53.3%)	6 (46.2%)	4 (50.0%)	4 (100.0%)	1 (25.0%)	2 (50.0%)
Indirect	9 (60.0%)	7 (53.9%)	6 (75.0%)	2 (50.0%)	1 (25.0%)	1 (25.0%)
Text direct	3 (20.0%)	2 (15.4%)	3 (37.5%)	3 (75.0%)	0 (0.0%)	1 (25.0%)
Text indirect	6 (40.0%)	6 (46.2%)	3 (37.5%)	2 (50.0%)	0 (0.0%)	1 (25.0%)
Image direct	5 (33.3%)	6 (46.2%)	1 (12.5%)	2 (50.0%)	1 (25.0%)	1 (25.0%)
Image indirect	5 (33.3%)	5 (38.5%)	3 (37.5%)	2 (50.0%)	1 (25.0%)	1 (25.0%)

While it is hard to draw general conclusions from the smaller categories, it is interesting that all four subjects who considered tags to be links to other pages clicked on tags during the experiment. Indirect use of tags, that is, using tags for query reformulation or to make predictive judgments of relevance, is prevalent among those who think tags are categories (60.0%), keywords (53.9%), or related terms (75.0%), with indirect use being highest for those who thought tags were related terms. In the experiment text searches, subjects used tags to get a quick idea of the contents of a particular search result item, which corresponds with thinking of tags as keywords and their indirect use. Subjects also added to their queries terms they saw in the tags, again an indirect use of

tags. So there is some indication that subjects' understanding of tags was being reflected in their use of tags during the search experiment.

Tag origin

Whether subjects' understanding of tag origin is related to the use of tags in searching is examined in this section. Knowing how tags are created or where they come from could influence their use, either negatively or positively. For example, S42's awareness that Delicious tags could be created by anybody led to a distrust of tags: “I don't really have a trust for tags yet, because I feel anyone could tag it and it's just your opinion whether it's relevant to a topic or not." So it is reasonable to posit such users would be less inclined to use tags on a system they had not encountered before. On the other hand, a user who thinks tags are produced by "the site," to "help make it easier for people to search" (S48), could be more inclined to use tags in their searches.

Table 5.45. Tag origin and tag use

	Tag origin				
Type of tag use	Content contributor N=26	Website owner N=11	System N=11	General public N=3	Don't know N=12
All searches	19 (73.1%)	9 (81.8%)	6 (54.5%)	3 (100.0%)	9 (75.0%)
Text	15 (57.7%)	7 (63.6%)	5 (45.5%)	3 (100.0%)	4 (33.3%)
Image	13 (50.0%)	7 (63.6%)	2 (18.2%)	1 (33.3%)	7 (58.3%)
Direct	11 (42.3%)	6 (54.5%)	3 (27.3%)	3 (100.0%)	7 (58.3%)
Indirect	16 (61.5%)	7 (63.6%)	4 (36.4%)	3 (100.0%)	5 (41.7%)
Text direct	4 (15.4%)	4 (36.4%)	2 (18.2%)	2 (66.6%)	3 (25.0%)
Text indirect	12 (46.2%)	4 (36.4%)	3 (27.3%)	3 (100.0%)	3 (25.0%)
Image direct	8 (30.8%)	3 (27.3%)	1 (9.1%)	1 (33.3%)	5 (41.7%)
Image indirect	10 (38.5%)	5 (45.5%)	1 (9.1%)	1 (33.3%)	4 (33.3%)

Table 5.45 summarizes use of tags according to tag origin. Subjects provided multiple answers, attributing different tag creators for different sites. What jumps out is that a quarter of the subjects did not know or were not sure who or what created the tags

they saw on websites, yet 75% of them used tags in their searches. On the other hand, subjects who thought tags were automatically generated by the system in general used tags less frequently in their searches. Subjects who thought tags came from content contributors seemed to favor indirect use of tags over direct use in text searches. Subjects who thought tags were generated by website owners showed a relatively high rate of tag use in the experiment, with higher rates of use for both text and image searches as well as direct and indirect use of tags than subjects who attributed tags to content creators or automatic generation by the system.

One possible reason for differences in tag use depending on tag origin is the credibility of the tags. For example, S42, who had expressed a distrust of Delicious tags, did use tags in the experiment search sessions. She also expressed a willingness to use tags on MTagger, a tagging tool available for use on the University of Michigan Library Web pages, including its online catalog, as she thought the tags came from other students and librarians and would be "more academic." Subjects in the study were informed they would be using an experimental system developed by the researcher, and were not informed of the origin of the tags they were shown. Thus it is possible subjects trusted the tags on the experiment system more than they would tags they encountered on the Web at large, perhaps attributing the tags to the researcher, who could be considered the website owner from the subjects' perspective.

Purpose of tags

As with tag sources, subjects mentioned multiple purposes for tags, which sometimes were associated with specific websites. Some subjects saw tags as objects having multiple purposes. For example, S35 saw tags as being for "classification of

information," but also as providing some information or description in the case of photographs, of "what is this, the background of the photo." Tags can make an article more findable in searches, but also provide information on the "ideas and stuff that you think it covers" (S29). The perceived purpose could influence the use of tags in search – subjects who see tags as sources of additional information on the tagged item might be likely to use it for predictive judgment, while those who think the purpose of tags is to lead (or mislead) to content might avoid using tags.

Table 5.46. Purpose of tags and tag use

	Purpose of tags				
Type of tag use	Organization N=10	Findability N=29	Lead to content N=7	Description N=6	Don't know N=4
All searches	8 (80.0%)	20 (69.0%)	6 (85.7%)	5 (71.4%)	2 (50.0%)
Text	6 (60.0%)	15 (51.7%)	4 (57.1%)	5 (71.4%)	1 (25.0%)
Image	3 (30.0%)	14 (48.3%)	6 (85.7%)	4 (57.1%)	2 (50.0%)
Direct	6 (60.0%)	13 (44.8%)	5 (71.4%)	4 (57.1%)	2 (50.0%)
Indirect	5 (50.0%)	13 (44.8%)	6 (85.7%)	4 (57.1%)	1 (25.0%)
Text direct	4 (40.0%)	5 (17.2%)	2 (28.6%)	3 (42.9%)	1 (25.0%)
Text indirect	3 (30.0%)	12 (41.4%)	3 (42.9%)	2 (28.6%)	1 (25.0%)
Image direct	3 (30.0%)	10 (34.5%)	4 (57.1%)	1 (14.3%)	1 (25.0%)
Image indirect	2 (20.0%)	7 (24.1%)	5 (71.4%)	4 (57.1%)	1 (25.0%)

Table 5.46 summarizes use of tags according to tag purpose. Interestingly, despite the negative connotation of the purpose of leading to content, tag usage was relatively high among subjects who identified this as the purpose of tags. This could be a side effect of the experiment, where subjects were informed they would be using an experimental system developed by the researcher. So subjects may have considered the tags in the experiment to be different from the misleading tags encountered on the Web. Usage of tags was relatively low for subjects who thought the purpose of tags was findability. This purpose is associated with making a page or item findable to an automated system like a search engine, and does not entail the searcher interacting with the tag, which may

explain this relatively low usage. Among subjects who saw the purpose of tags as description, indirect use of tags was relatively low for text searches. Perceptions of tags on the Web did not appear to necessarily affect tag use during the experiment. More generally, the particular type of website or system a site is perceived to be may influence the use of tags. For example, if a site is perceived to be one likely to have tags that lead to content, then tag usage could be lower than seen in the study for those who attribute this purpose to tags.

5.5.3 Relationship with use of other information retrieval systems

In this section we examine the question of whether experience using other information retrieval systems influences the use of tags in the experiment system. The alternative information retrieval systems considered are OPACs and library databases. As these systems provide subject headings or descriptors, which are analogous to tags for search purposes, we were curious whether experience using such systems was associated with tag use in our experiment. One subject drew an explicit connection between tags and what she called "embedded words" in library systems:

> [S23]: I think of them [tags] similar to, and I probably shouldn't, but I think of them as similar to, embedded like, look in the library system they have the embedded words. So if I type in women, I'm gonna get different search results than if I use their designed categories of women.

The question of library database and OPAC use was added to the post-search interview after main data collection had started, and so answers were obtained for 43 of the 48 subjects. Thirty-one subjects from this set of 43 subjects (72%) had used tags in the experiment search sessions, which was not very different from the proportion of

subjects in the full set of 48 subjects who used tags (75%). Of these 43 subjects, 22 subjects (51.2%) had experience using one or more library databases, while 31 subjects (72.1%) had experience using OPACs. In total, 38 subjects (88.4%) had experience using OPACs or library databases. Seventeen different library databases were mentioned by subjects as ones they had used, with the most popular ones being ProQuest, PubMed, ISI Web of Science, and PsycINFO. Also mentioned were more specialized databases such as Naxos Music Library and Beilstein, a database of organic chemistry molecules. Most subjects with OPAC experience had used Mirlyn, the University of Michigan library catalog, with some mentioning public libraries such as the Ann Arbor District Library and the Toledo-Lucas County Public Library.

The chi-square test of independence was used to investigate the relationship of library database experience with various types of tag use. Indirect tag use was dependent on library database use ($\chi^2(1,N=43) = 3.91$, $p<.05$, $phi = .30$), but library database use was not significantly associated with other types of tag use. Of the subjects with library database experience, 68.2% used tags indirectly, compared to 38.1% for those with no library database experience. No significant relationships were found when indirect tag use for image searches and indirect tag use for text searches were examined separately. As subjects' database experience came from a diverse set of databases, covering a wide variety of topic areas, tag use across both text and image searches may reflect this mix. No significant associations were found between previous OPAC use and tag use.

5.5.4 Summary of RQ3

Research Question 3 examines factors influencing tag use in search. It was found that prior knowledge of tags, composed of familiarity with tags, experience with tags, and understanding of tags, had no significant relationship to tag use in the experiment. Self-ratings of familiarity with tags were not found to be reliable indicators of prior knowledge of tags at the higher familiarity levels. Subjects developed their understanding of tags through their experience using them and tagging online content.

Subjects held diverse ideas as to what tags were and their origins. Tags were not always seen as originating from content creators, with some subjects attributing them to automatic generation by "the system" and others attributing them to site owners or operators. A number of subjects seemed to consider tags to be just another type of link. Several subjects recognized that tags were a way to make content findable to search engines, and described both positive and negative aspects of this. Tags make objects more findable by having tagged objects turn up in searches, but tags can also be used to turn up in unrelated searches or lead people to undesired content.

In general, prior knowledge of tags appeared to be system-specific, with knowledge of tags gained from one system not necessarily transferring over to other systems. Use of tags on the Web appeared to be both random and opportunistic for most of the subjects in the study. Tags were used if they caught the subject's eye in some way, but were not necessarily sought on their own. Interestingly, prior experience using a library database system was significantly related to indirect use of tags during the experiment, where indirect use is the use of tags for predictive and evaluative judgments, as well as query reformulation activities that do not involve clicking on the tag.

Chapter 6

Discussion

This chapter discusses the findings presented in the previous chapter, drawing connections to previous research. The first section relates the understanding of tags from the subjects of our study, in particular their characterization of tags and the purposes for tags they identified, with findings from previous studies on taggers. Section 6.2 discusses the use of tags for query reformulation as a type of term suggestion. Section 6.3 examines the use of tags in two types of relevance judgments, predictive and evaluative, in relation to a number of studies on augmented search result displays and document clues used when making judgments of relevance.

6.1 Understanding of tags

6.1.1 Characterization of tags

Tags were not an obvious or intuitive concept for several of the subjects in our study, with some identifying as tags objects that were not tags, despite previous exposure to and familiarity with tags on the Web. There are a number of reasons for these differences – the subjects of this study were not recruited because of their use of particular tagging sites or required to have previous tagging experience, nor were they provided instruction on tags or asked to apply tags themselves as part of the study. The

understanding of tags exhibited by the study subjects was derived from their own experiences on the Web, whether simply encountering them or tagging their own content.

In Chapter 2, three characterizations of tags were identified from the research literature on tags: categories, keywords, and annotations. In Chapter 5, study subjects characterized tags as *categories*, *keywords*, *related terms or topics*, *links*, and *indexing terms*. Categories and keywords were characterizations of tags shared by the study subjects and the research literature. Fifty-eight percent of subjects described tags as being categories or keywords, making these the dominant characterizations. When tags are characterized as categories, they are seen as a way to group related materials together. The characterization of tags as keywords emphasizes their role describing the key ideas of the document. Both research on tagging and websites implementing tagging tend to describe tags in terms of categories or keywords. These types of tags are likely to be encountered in both the sites study subjects were familiar with, such as YouTube or Flickr, as well as sites that have drawn researcher attention but were not familiar to the study subjects (e.g., Delicious, LibraryThing).

A third characterization of tags in the literature is as annotations, tags that function more as notes to the tagger. Studies of tags have found that tags include terms that are "non subject related" (Kipp, 2007a), reflecting the tagger's reaction to or assessment of an information object, such as "funny" or "toread" (Golder & Huberman, 2006; Kipp, 2007a). This type of characterization of tags was not identified by our study subjects, perhaps due to the types of tagging sites familiar to them and their own experiences with tagging. Annotations are found on social bookmarking sites such as Delicious, as well as sites devoted to academic articles such as CiteULike or Connotea.

These are sites on which tags are attached for the benefit of the tagger, who is bookmarking objects created by others. The subjects of this study were not familiar with social bookmarking sites. The sites in which they encountered social tags were sites for sharing content, often user-generated, such as YouTube, Flickr, or blogs. As the intent when uploading material is to share the content, attaching tags that are personal annotations naturally makes less sense to these content contributors. The concept of tags as personal annotations would also be unfamiliar to those with no tagging experience, as well as those who thought tags were system-generated or attached by website owners hoping to draw traffic to their sites. Tags as personal annotations appears to be a characterization of tags restricted to broad folksonomy sites such as Delicious.

The remaining characterizations of tags from our subjects might be considered "side effects" of tags. That is, some characteristic of the tag other than its describing the information object is more salient, such as it being a clickable link, or it being related to other terms in a collection in some way. Tags are then viewed as query suggestions, or as links to other pages or sites. Lastly, when tags are seen as indexing terms to make the content findable through a search engine (e.g., S36: "the tag words lead you to that photo"), tags do not even have to be related to the information object or its contents. This perception of tags is perhaps the most distanced from what proponents of social tagging had in mind, in that the focus is on having a search engine point to the tagged content. In this case what is important is that a tag is a metadata element used by a search engine to index and rank search results, and not that the tag describes the content or helps to group similar content together.

Seeing tags as related terms or topics (query suggestions) or links to other pages is not a conceptualization of tags found in studies of taggers. In our study, none of the study subjects with tagging experience considered tags to be links, and only two of the eight subjects who thought tags were query suggestions had tagging experience.

Subjects with tagging experience, on the other hand, are more likely to see tags as indexing terms. Findability of their content by others was important to them, as with the Flickr taggers studied by Ames and Naaman (2007). In Ames and Naaman's study some users described using tags to "game" the system, attracting more views to their photographs. Subjects in our study did not describe engaging in explicit "gaming," but described tagging their own material with terms they thought people would use to search for such material. Subjects without tagging experience were also aware of tags being used to have the tagged materials show up in certain searches. Overall, the perception of tags as indexing terms appears to be restricted to users who have experience with tags for user-generated content.

We defined tags in Chapter 3 as being "descriptive terms that people attach to online content." This reflects the general understanding of social tagging from the research literature and websites implementing tagging that *people* generate the tags. Yet 25% of our study subjects said they did not know who or what generated the tags, with some even being surprised by the question. A nearly equal number of subjects thought tags were generated automatically in some way, and not by people. Curiously this answer came from subjects who thought Google generated tags, including a number of subjects who had tagging experience on systems such as YouTube. If one thinks there are tags on Google, it is reasonable to conjecture that the same system that is generating the search

results is also generating the tags. It appears that given the variety of query reformulation support features that Google provides, such as lists of related queries or displaying query suggestions, some of the subjects may have been perceiving tags to be one of these features.

Content contributors and website owners were the most frequently mentioned sources of tags by our study subjects. This is a reflection of people's experiences with tags on the Web, as there are many sites (e.g., blogs, news media sites) that implement tags but restrict tagging to the site owner or people authorized by the site owner. On sites relying on user-generated content such as YouTube or Flickr, it is the person who uploads content who also attaches the tags. Subjects in our study had little experience with broad folksonomy sites such as Delicious or LibraryThing, where many people attach tags to content they have not created or contributed themselves. Thus, only three of the 48 subjects mentioned the general public as originators of tags. This was in contrast to XX of 48 who mentioned the content creator, or YYY of 48 who mentioned the website owner, as the creator of tags. Subjects' understanding of tag origin reflected their own experience with tags on the Web.

6.1.2 Purpose of tags

This study focused on the use of tags in interactive information retrieval, and in particular on the use of tags created by other users. Subjects identified the following purposes for tags: making items easier to find using a search engine (*findability*), grouping related items together making it easy to find them together (*organization*), leading or manipulating people to view content (*leading people to content*), and description of the tagged item (*description*). The purposes of findability and organization

both relate query terms to tags. In the former, the expectation is that when a searcher uses query terms that match with certain tags, then the information with those tags will be returned by the information retrieval system. In the latter, tags provide a mechanism for browsing and navigating a collection. We have already seen that tags describing the subject or content of the tagged document are used for relevance judgment. In a way, leading people to content is the flip side of findability – while the intent of findability is to make the tagged content findable by people who are interested in such content, the intent of leading people to content is to get people to the content regardless of interest or relevance, in effect describing tag spam (Heymann, Koutrika & Garcia-Molina, 2007). The purposes of tags identified by our study subjects are all related to information retrieval.

Previous studies have identified a variety of purposes for tagging, not all of them related to information retrieval: organization (Marlow et al., 2006; Ames & Naaman, 2007; Strohmaier, Körner, and Kern, 2010), finding/search (Marlow et al., 2006; Ames & Naaman, 2007), description (Strohmaier, Körner, and Kern, 2010), self-expression (Marlow et al., 2006; Cosley et al., 2009), reflection or thinking about the tagged items (Cosley et al., 2009), signaling of involvement and expertise (Thom-Santelli, Muller & Millen, 2008; Thom-Santelli, Cosley & Gay, 2010), and adding context or annotation (Ames & Naaman, 2007). Taggers also use tags to control or limit access to their content on sites such as YouTube (Lange, 2008) or Flickr (Ames & Naaman, 2007). Such content is "publicly private" (Lange, 2008), in that only people who know the particular tags can use them to find the tagged content. These studies either examined taggers (Ames & Naaman, 2007; Lange, 2008; Thom-Santelli, Muller & Millen, 2008) or provided

instruction on tags to the study subjects, effectively encouraging them to tag (Cosley et al., 2009; Thom-Santelli, Cosley & Gay, 2010). Thus these studies provide the taggers' perspective on the purpose of tags, as opposed to the tag consumer's perspective on the purpose of tags provided by our study.

What we see is that while there are some overlaps, taggers and tag consumers have somewhat different perceptions of the purposes of tags. Taggers perceived purposes for tags in addition to information retrieval, such as self-expression, signaling expertise to others, or reflection. Perhaps because of the search tasks they were assigned, our study subjects mainly saw tags as helping with information retrieval, by improving search and browsing. One subject (S27) mentioned her boyfriend using tags for self-presentation, showing how he was interested in different topics, but did not consider this to be an appropriate purpose for tags. On the other hand, a subject (S46) who was an avid blog reader enjoyed how one blogger used tags for self-expression: “They make me smile because they’re just PZ’s [blog author’s] way of viewing the things that he’s responding to.” A tagger's intended purpose for tags may not necessarily be apprehended by tag consumers contributing to the mismatch of perceptions. Tag consumers may also disagree with a tagger on the purpose of tags, leading to avoidance of tags. A particularly blatant example is tag spam, where tag consumers and taggers are at odds regarding the use of tags. Taggers' and tag consumers' perceptions of the purpose of tags differ, and these differing perceptions can in turn lead to divergent expectations of what constitute good or useful tags.

6.2 Tags and Query Reformulation

Researchers have explored two approaches to help with query reformulation, relevance feedback and query or term suggestion (Belkin, 2000). In relevance feedback, the searcher signals relevance of a search result to the system, and the information retrieval system makes use of this information to refine the search. Query reformulation is actually done by the system, not the searcher. In contrast, in term suggestion the system suggests query terms, and the searcher can choose to use them or not.

In our study, subjects used tags for query reformulation by treating the displayed tags as sources of query terms. In effect, the displayed tags served as query suggestions. This is further supported by the perception of some subjects that tags were related terms or topics that served as query suggestions. Nevertheless subjects did not consider query suggestion to be one of the purposes of tags. So while tags can be used as sources of query terms, this was not considered to be the intent of tagging.

Tag use was observed for all three interaction processes – query reformulation, predictive judgment, and evaluative judgment – in both text and image searches. In image searches 75.8% of tag use occurrences were for query reformulation, and in text searches, query reformulation accounted for 56.3% of tag use occurrences. A large part of such use occurred from the search results page: 81% for text searches, and 80% for image searches. Overall, the majority (83%) of tag use, whether direct or indirect, took place on the search results page for both text and image searches. When tags in the document page were examined, the purpose was to obtain more search terms or ideas for searching, and rarely for evaluating the relevance of the document.

Tags could be clicked on from two places on the search results page: the list of related tags, and the tags included with each search result. In the case of text searches, tag clicks skewed heavily towards the list of related tags, with 16 out of the 18 (89%) tag clicks on the search results page being for related tags. Tag clicks on the search results page for image searches was not as skewed, with 15 (42%) being for search results tags and 21 (58%) being for related tags. In the case of text searches, there was repeated use of related tags, being used more than once in five of the eight searches in which related tags were clicked. This contrasts with tags in the search results being clicked on once each by two subjects. For image searches similar numbers of searchers used tags repeatedly for either of the two tag locations. Interestingly when searching for images, subjects clicked on both types of tags on the search results page, while when searching for text, subjects clicked on either search results tags or related tags, but not both.

Anick and Kantamneni (2008) experimented with a combination of query extension and related concepts in the Yahoo! search interface. While overall use of query extension was higher than related concepts, of those using the features there was much more repeated use of related concepts than of query extension. The list of related tags in our study is similar to their related concepts, in our case obtaining the related tags from the most frequent tags for that page of search results, compared to the top search results for Anick and Kantamneni. Our study subjects also repeatedly used related tags. Both Anick and Kantamneni's and our results may derive from users' genuine preference for a list of related concepts or terms when in the process of reformulating a query. On the other hand, it is possible that for our study the placement of the related tags, under the

search box and above the search results, may have made them more visible than the tags accompanying individual search results.

If the placement of related tags played a role in their use, then a similar proportion of tag clicks per location should have been observed for both text and image searches. This was not the case, as the skew towards using related tags was not as prominent in image searches compared to text searches. The pattern of use and interview data indicates that when reformulating image search queries, subjects clicked on any text link that caught their attention. This suggests that on an image search results page, subjects distinguished between links and images, but did not distinguish between types of links. On a text search results page, subjects tended to fixate on particular text elements and ignore others – for example, a subject who mainly looked at the titles might not notice any of the tags. Tags displayed with images appear to be perceived differently at a visual level than tags displayed with other text items.

6.3 Tags and Relevance Judgment

A number of studies have found that augmenting Web search surrogates with category or subject metadata were helpful to the searchers (Drori; 2000; Chen & Dumais, 2000; Dumais, Cutrell & Chen, 2001). Drori found that adding "key words" extracted from the document improved users' ease of use and satisfaction when using the search results. Drori intended these key words to convey subject information. Category metadata was also found to be helpful by Chen and Dumais (2000) and Dumais, Cutrell and Chen (2001). Participants reported greater satisfaction with the interface that included category information than the one without. Participants were also able to find answers to questions faster on the enhanced interface. Drori's "key words" and the

category metadata used in the Chen and Dumais (2000) and Dumais et al (2001) studies are very similar to tags. Subjects in our study reported greater satisfaction and ease in their searches with the Text/Tags interface, when using tags more than once, although the differences were not statistically significant. This indicates tags or subject information are helpful for both simple search tasks (Drori; 2000; Chen & Dumais, 2000; Dumais, Cutrell & Chen, 2001) and the more complicated search tasks as used in our study.

In our study we found that tags were used for both predictive and evaluative judgments of relevance. As with previous research on document clues used for relevance judgment of surrogates and Web pages, the title was among the most frequently used relevance clues in the SERP (Lan, 2002), while content was the Web page feature most frequently used (Tombros, Ruthven & Jose, 2005). Unlike the subjects in Lan's study, our subjects did not consistently use the text snippet in the text search results for predictive judgments of relevance. Subjects mainly relied on the title, with some using a combination of title and tags. In image searches, the thumbnail was the primary surrogate element used by our searchers, and the title was not as important as in text searches. Searchers appeared to be relying on the most visually prominent surrogate elements for their predictive relevance judgments.

The lack of consistent use of text summaries may reflect the relevance criteria subjects used when making predictive judgments. In an eye tracking study, Balatsoukas and Ruthven (2010) found a relationship between relevance criteria for predictive judgments and surrogate components used. Subjects relied on the title when applying the relevance criterion of topicality, where topicality was the dominant criterion for deciding to click on a surrogate deemed at least partially relevant. Subjects in our study were

mainly concerned with topicality, and did not appear to be applying the relevance criteria associated with summary use in Balatsoukas and Ruthven's study. This does suggest an interesting avenue for future study identifying the relevance criteria associated with tag use.

The possibility exists that the quality of the text snippet was a factor in its usage by our study subjects. Tombros and Sanderson (1998) found that query-biased summaries were superior to summaries consisting of the first few lines of text, allowing users to identify relevant documents more accurately, as well as identifying more of them. For the text collection used in our study, Ask MetaFilter, each document page consists of a question and answers to the question. In particular, a question has two parts, a concise statement of the question and an optional extended explanation. The experiment system displays this first part of the question as the text snippet. So the snippet gives a reasonable indication of what the document will be about. One difference from current search engine text summaries is that, in our system, the query terms were not made to stand out visually, for example with a bold font. This was because the query terms did not necessarily appear in the concise question text. In contrast, tags, by their placement and bright green font color, were visually prominent. Searchers used to modern SERP displays may have been drawn to tags rather than a chunk of text, leading to their use in predictive judgments of relevance in lieu of the text snippet.

In the case of image searches, an image thumbnail was displayed instead of a text snippet. The image surrogate consisted of the title, thumbnail, and tags. Despite the search tasks being designed to encourage subjects to interact with the text accompanying the image, subjects focused on the images and tended not to pay attention to the text. For

text in the SERP, there was also the tendency to conflate the image title and the tags together, indicating that subjects grouped such text together and did not consider them as being distinct or different in some way. Once on the actual photo page itself, again subject attention was focused on the image. Although the image search tasks specified criteria that could not always be satisfied from the images alone (e.g. images of *American* women, images of *New York City*), subjects frequently selected images for use based on image content. Thus among the images saved by subjects were of a building in Buffalo, New York, or an Australian actress identified as such in the image description. Subjects were more concerned with images that "looked right" than with verifying the images depicted topics that were actually right for the task description..

Not surprisingly, tags were used less frequently in image searches compared to text searches for both predictive and evaluative judgments of relevance. Tags were used for predictive judgment in 37.5% of tag use occurrences in text searches and 22.7% of tag use occurrences in image searches. Tags were also used differently for predictive judgment in text searches from image searches. For text searches, tags were seen as a quicker and more convenient way of getting a snapshot of the content than reading the accompanying text snippet. The combination of title and tags was used as an alternative to reading the text snippet in the search results. In image searches, tags were used to gain additional information about the image depicted in the thumbnail, and were not used as an alternative to the thumbnails.

How the subjects of this study used tags and other textual material accompanying images for judgments of relevance differed from previous studies of image retrieval by journalists (Markkula & Sormunen, 1998) and historians (Choi & Rasmussen, 2002).

Both groups found textual data to be important in making judgments of relevance, as their relevance criteria required information that was not present in the image, such as background, provenance, and date. Both groups were also aware that the initial impression of what the image was about did not necessarily reflect its actual topic or content, and relied on textual information to determine topicality and subject matter. There are a number of differences between our study and these previous studies: our study subjects were undergraduate students who did not have specialized domain knowledge in the image search topics, they were assigned search tasks, and the search tasks were in Fidel's (1997) Object Pole. Historians engaged in image search tasks in the Data Pole and journalists in a mix of both. In image retrieval tasks in the Data Pole images are used as sources of information, while in the Object Pole images are used as objects, such as illustrating a poster.

There was some similarity in the predictive and evaluative judgment process of our study subjects and the image search process of journalists. Markkula and Sormunen found that journalists engaged in a two-stage search process, first selecting a set of candidate photos, and then selecting a final photo for use from this candidate set. The first stage involves making predictive judgments of relevance, while the second stage involves making evaluative judgments of relevance. Final selection was based solely on the visual attributes of a photograph. Similarly, for our subjects, evaluative judgment was based on the image and very rarely took into account tags or other text accompanying the image. Journalists made more extensive use of text information than our study subjects when making predictive judgments of relevance. This suggests that in more specialized

settings tags may be used much more extensively for predictive judgments in image searching than found in our study.

6.4 Summary

In assessing the understanding of tags held by our study subjects, who had a diverse range of tagging experience, we found that the understanding of tags held by taggers diverged in a number of ways from those of tag consumers. Tag consumers tended to perceive tags in terms of their role and value in information retrieval. Taggers saw purposes and benefits of tags arising from the act of tagging itself, in addition to those related to information retrieval. Taggers saw communicative purposes to tagging, such as self-presentation. Such aspects of tags are not likely to be perceived, or perceived favorably, by non-taggers, partly because they do not always assume that tags are generated by other people.

Tag use was observed for all three interaction processes – query reformulation, predictive judgment, and evaluative judgment – for both text and image searches. There were differences in the distribution of tag use occurrences for text and image searches. Tags were mainly used for query reformulation in image searches, while tag use was more evenly distributed among predictive judgment and query reformulation for text searches. Some of our subjects treated tags as query suggestions, although this was not seen as a primary purpose of tags. Tags supported query reformulation in image searches. Findings from our study and previous studies on augmented SERP displays indicate that the type of information found in tags are helpful for both simple search tasks (Drori; 2000; Chen & Dumais, 2000; Dumais, Cutrell & Chen, 2001) and more complicated search tasks. Subjects in our study relied less on textual data for image searches than did

journalists (Markkula & Sormunen, 1998) or historians (Choi & Rasmussen, 2002). This raises the possibility that tags may be used much more extensively in more specialized contexts such as work settings or by populations such as scholars.

Chapter 7

Conclusion and Implications

This chapter presents theoretical and practical implications of the study, suggesting several implications for system design. Limitations of the study are discussed, followed by future research directions.

7.1 Theoretical implications

The study makes a number of theoretical and methodological contributions to research in interactive information retrieval. Tags, a relatively novel search interface feature, increased interactions with the information retrieval system, as subjects issued more queries and saw more search results when using the tagged interfaces. These increased interactions were a result of tags being used for query reformulation and judgments of relevance. Tag use across the search process was not restricted to a particular resource type, as tags were used for query reformulation and relevance judgments in both text and image searches. Different patterns of use were found for the different resource types, indicating resource type is factor in tag use. The use of tags was also characterized by the modality of use: whether tags were used directly by clicking on them or indirectly without clicking on them. A substantial portion of tag use occurred indirectly, which has methodological implications for researchers as indirect tag use is

not readily captured in clickstream data. That is, relying on only direct use of a feature could result in substantially undercounting the actual frequency of use.

The study found that tags were used for all three interaction processes: query reformulation, predictive judgment, and evaluative judgment. While the majority of tag use was for query reformulation, supporting the view that tags help with discovery and findability of information, tags were also used for predictive and evaluative judgments of relevance. Tags were most frequently used for query reformulation, especially in image searches. The next most frequent use was for predictive judgment, while tag use for evaluative judgment was rare. When used for evaluative judgment, tags were used to resolve lingering uncertainty regarding the usefulness of a document. Interestingly, tags were used either for predictive judgment or evaluative judgment during a query instance but not both. Previous research on tags has tended to emphasize their role in organizing information and improving findability. The present research indicates tags help people to make predictive judgments of relevance quickly, in addition to their value for query reformulation.

While tags were used for all three interaction processes, tags serve different functions in an interaction process depending on the resource type. Seventy-six percent of the tag use occurrences for image searches were for query reformulation, while for text searches query reformulation accounted for 56% of tag use. Tags were more likely to be used for predictive judgment in text searches (37.5% of tag use occurrences) than in image searches (22.7% of tag use occurrences). Additionally, tags were used differently for predictive judgment in text and image searches – in text searches, tags were used as an alternative to the text snippet in the SERP, while in image searches, tags were used as

a source of additional information regarding the thumbnail image in the SERP. In a text search SERP, the text snippet, whether simply the first few lines of the document or a query-biased summary, does not provide a summary of the entire document. On the other hand, the set of tags for a document present the key aspects of the document as a whole. In such a situation the title and tags can provide a more complete picture of the document than the title and text snippet. In the case of images, the thumbnail presents the image as a whole, so unless the thumbnail size is very small or the resolution poor, a searcher can get the entire gist of the actual document from the thumbnail, so tags cannot substitute for the thumbnail.

In examining people's understanding of tags in this study, we found that subjects' understanding of tags was site-specific. Subjects ascribed different tag creators and purposes of tags to specific websites. Tags were not uniformly perceived as being user-generated; study participants also mentioned site owners and automatic generation as sources of tags. This is a consequence of people's experiences with tags on the Web having become much more diverse since the introduction of Delicious in 2003. Early studies of tagging occurred when there was a much more limited number of tagging systems and implementations of tags. Delicious represents social bookmarking and broad folksonomy, while Flickr, launched in 2004, represents narrow folksonomy. Tagging was open to all, whether bookmarking and tagging Web pages or uploading and tagging one's pictures or videos online. Currently, on a number of sites featuring tags tagging is not open to all users. In blogs and newspapers sites, where tags are displayed and can be used by all site visitors, tagging can only be done by content contributors or specialized "tag managers." Researchers should be cautious about assuming that Web users have a

consistent understanding of tags, or that Web users' understanding of tags is congruent with the researchers' own understanding of tags.

In addition to implications for research on social tagging, this study contributes to the methodology of interactive information retrieval research by distinguishing between direct and indirect use of a feature during the search process. In this study, both direct and indirect use of tags in search interaction processes were examined. Direct use involved clicking on a tag, while indirect use involved using tags without clicking on them, such as using them as sources of query terms or information to help make judgments of relevance. 42% of tag usage in this study came from indirect tag use, where indirect tag use rates differed depending on the interaction process and resource type. For example, 76% of the tag use occurrences in image searches were for query reformulation, and 90% of these occurrences involved clicking on a tag. This contrasts with text search, where tags were used for query reformulation in 56% of the tag use occurrences, with 78% of these occurrences involving clicking on a tag. To examine interface features in some systems, researchers may need to pay attention to both direct and indirect use of the feature. A large proportion of use may be happening indirectly, which may not be measured by the researchers' instruments if not designed to capture such use ahead of time.

7.2 Implications for system design

The finding of this study that tags were used for query reformulation, predictive judgment, and evaluative judgment has implications for the design of search interfaces as well as the presentation of documents or Web pages.

Display tags in SERP

The bulk of tag use occurs in the search results page, for both query reformulation and predictive judgment. The related tags are mostly used for query reformulation, while the tags accompanying each search result are used for predictive judgment. Thus displaying tags on the SERP can help users with query reformulation and predictive judgment. While a number of sites currently prominently show a list of related tags in the SERP, it is rare for sites to display tags with each search result. Displaying related tags supports users who think of tags as related terms that can be used to aid in their search. Tags are usually shown in the document page, but our study found that tag usage from the document page was one fifth of the usage from the SERP.

De-emphasize live link aspect of tags

The findings from our study suggest that de-emphasizing the live link or clickability aspect of tags could lead to increased use of tags. In our study tag use was analyzed in terms of direct (clicking on a tag) and indirect (use not involving clicking on tag) use of tags. 58% of tag use occurrences in the study were of the direct type, while 42% were of the indirect type. That is, over 40% of tag usage did not involve clicking on a tag, instead being used for predictive judgments, evaluative judgments, and assist in query reformulation. Given the study subjects' concerns that tags were a way to lead people to unwanted content, as well as a reluctance to click on tags because of not knowing what would happen, displaying tags in a way that emphasize their role as additional sources of information about the tagged content and possible query terms could be of benefit to users. Showing lists of tags that are not clickable links may be more suited for text resources than non-text ones. In our study, 90% of tag use for image query

reformulation was of the direct kind. So whether to make tags clickable or not should take the resource type into account.

Generate keywords automatically

In our study, when searching for text, subjects used tags in the SERP to get a quick idea of what a search result was about, using them as an alternative to the text snippet. As tags are expected to provide a summary of the document, tags that are keywords from the document can achieve this purpose. This type of usage of tags indicates that automatic generation of keywords from text content may be a viable approach to generating useful tags. Automatic generation of tags is also desirable as currently a large proportion of Web content is not tagged.

Contain tag spam

Tag spam is the phenomenon of applying a large number of often unrelated tags to content in order to turn up in more searches, or get more views on sites such as YouTube or Flickr (Heymann, Koutrika & Garcia-Molina, 2007). When asked to describe tags, several subjects in our study effectively described tag spam, seeing tags as a way to manipulate or mislead people to content. This indicates that for some users, any utility from tags has already been overshadowed by its misuse. Algorithmic spam detection is already an active area of research – for example, one of the tasks in the 2008 ECML PKDD Discovery Challenge was spam detection in social bookmarking systems. Limiting the number of tags that can be attached to an object is another way to control tag spam. Limiting the number of tags can also cut down on the number of tags displayed, which could be helpful for users.

7.3 Limitations of the study

There are a number of limitations to this study. First, this study had a relatively small number of participants, 48 subjects. While this is not considered a small number of subjects for an interactive information retrieval study, when analyzing for tag use in effect we had a smaller sample. Of the 48 subjects, 25 used tags in their image searches and 25 used them in their text searches, with 36 subjects using tags in at least one of their searches. The number of subjects was effectively 25, or possibly smaller if selecting for high tag users, subjects who had used tags at least twice in their searches. Subjects were also restricted to undergraduate students who self-identified as heavy Web users. While subjects may not be representative of all Web users, it can be argued that they are representative of Web users likely to have encountered and noticed tags in their Web use.

The second limitation of this study is related to the experimental setting used for data collection. Subjects were assigned tasks, and asked to carry out searches on an unfamiliar system, which searched pre-selected sites instead of the Web at large. Some subjects were frustrated the experiment system did not behave like Google. A number of subjects commented that searching on the assigned sites was not the way they would have typically done the search tasks if given a choice. In fact some of the subjects tried to use the experiment search system to point them to the sites they would typically go to for travel information, or reviews of consumer electronics.

Another limitation of this study is due to the search tasks, in particular the image search tasks. In general subjects had no problems with the text search tasks, and did not find them to be unusual, with subjects commenting that they frequently searched for travel information or had recently looked for information on purchasing laptops. The

image tasks were not as familiar to the subjects, although every effort had been made to create image search tasks that fit the criteria for Borlund's (1997) simulated work tasks. For example, the search task for historical images of New York City was based on an actual homework assignment from a course on the history of cities. Additionally, the image tasks may not have adequately captured the benefits to be gained from using tags in interactive information retrieval. Some of the subjects who had experience using Flickr commented they used tags on Flickr to explore and discover photographs, and not to find images satisfying certain criteria as required in the study. The nature of the image search tasks did not encourage exploration or serendipitous discoveries, which might have excluded certain types of tag use from the experiment.

7.4 Future directions

This study examined the use of tags when searching for documents in text form and in image form. The findings of this study imply that the type of tags that are found useful in searches vary depending on the interaction process and resource type. For example, the type of tags that are useful in predictive judgment in text searches are different from the type of tags useful in image searches. Currently there are other resource types being tagged on the Web, such as video or music. It is not clear if the findings of this study apply to these other resource types. In the case of music, "mood" or affective response appears to be one of the relevance criteria (Inskip, Butterworth & MacFarlane, 2008; Inskip, MacFarlane & Rafferty, 2010), and tags indicative of the mood of the music may be useful for predictive judgment when searching for music. Future research is needed on the use of tags when searching for resource types not

examined in this study, as well as the relevance criteria associated with tag use for the difference resource types.

Another direction for future research is to examine the usefulness of tags for tasks other than search. Sen et al (2006) found that different types of tags (factual, subjective, and personal) varied in their usefulness for different types of user tasks (self-expression, organizing, learning, finding, and decision support) in the MovieLens system. In our study all the user tasks were of the "Finding" type, as the focus of the study was on the use of tags during the interactive information retrieval process. Some subjects in our study appeared to be using tags to serendipitously encounter images they had not had in mind originally. By examining the use of tags for other types of tasks with goals such as exploration and inspiration, the concept of use and usefulness for tags can be extended.

In our study it appeared that knowledge of tags from one system did not translate to use in another system, but prior library database experience was related to the use of tags in the experiment. This study did not investigate why this is the case, only identifying the relationship. Previously, Kim (2001) found library database search experience had a strong impact on Web search performance, and it is intriguing that library database search experience continues to be a factor in studies on Web searching spaced ten years apart. Further study is needed on the specific aspects of library database use experience that influence people's use of different search features, such as tags, when searching on the Web. Such research also has implications for library instruction as well as in the training of librarians.

A final area for future study is development of methods to automatically (or implicitly) detect indirect use of tags. In this study we distinguished between direct and

indirect use of tags, depending on the modality of use, that is, whether tags were clicked on or not. 42% of tag usage came from indirect use. This indicates that relying on clickstream data for usage information may ignore a sizable portion of use of a feature, such as tags. But, clickstream data is easy to collect. Indirect use data in this study was collected through reviewing search recordings with subjects, and the researcher identifying episodes of tag use from these interviews. This is labor- and time-intensive, and while applicable for a research study with 48 subjects, obviously it cannot be used for larger numbers of subjects, such as all the visitors to a website. Findings in this study on the indirect use of tags for query reformulation suggest that some indirect use may be detectable implicitly. The use of tag terms in the query terms can be easily detected. It may also be possible to detect use of query terms that are semantically related to tag terms, such as synonyms. Such techniques could be extended to detecting implicit or indirect use of features other than tags.

7.5 Summary

This study investigated the use of social tags during the interactive information retrieval process. One of the motivations of this study was to examine an underlying assumption in much of the research on social tags – tags were useful – leaving unanswered in what way they are useful, or to whom. but it was not clear in what way, or to whom. Previous research on tagging focused on taggers, with little being known of the usefulness of tags to non-taggers. While much was known about the perceptions of tags held by taggers, it was not clear if these perceptions were also held by tag consumers.

This study found that people used tags during the search process, regardless of their level of tagging experience. The study also identified patterns of tag use in text and

image search with respect to the interaction processes of query reformulation, predictive judgment, and evaluative judgment. Tag use depended on resource type, in image searches being used mostly for query reformulation, and more likely to be used for predictive judgment in text searches than image searches. Subjects' understanding of tags depended on the websites in which they had encountered tags, and they did not generalize from one site to another.

In conclusion, this research has contributed to a better understanding of how social tags, a particular type of metadata element, is used during the search process for different types of resources. The research findings have practical implications for the design of search interfaces, especially in integrating or presenting tags to effectively support users in specific search interaction processes, and they suggest directions for future research that can further address these issues.

Appendix A

Recruitment Materials

Do you regularly look for information on the Web?

Are you a U-M undergraduate?

Are you a native English speaker?

If you answered yes to all three questions, you are invited to participate in a study conducted by a researcher from the University of Michigan School of Information.

You will conduct searches on different systems, and will talk about your search experience on these systems. Everything will be anonymous and confidential. The experiment will take an hour and a half and you will be paid $20 for your participation.

Interested? Please email kimym@umich.edu to schedule a session ASAP.
Participants must be at least 18 years old.

Appendix B

Data Collection Instruments

Appendix B.1

Informed Consent Form

Consent to Participate in a Research Study

ONLINE SEARCHING – EXPERIMENT

You are invited to be a part of a research study examining how people search the Web using different types of search systems. This study is being conducted by Yong-Mi Kim of the University of Michigan School of Information. The purpose of the research is to identify the features of online search systems that are used during different stages of the online search process. The results may be used to inform the design of future information retrieval systems.

If you agree to be part of the research study, you will be asked to take part in an experiment that will take approximately two hours. You will be given $20 to thank you for your participation.

Participating in this study is completely voluntary. Even if you decide to participate now, you may change your mind and stop at any time. You may choose not to answer questions for any reason.

If you have questions about this research study, you can contact the researcher, Yong-Mi Kim, University of Michigan, School of Information, 1075 Beal Avenue, Ann Arbor, MI 48109-2112, (734) 276-9260, kimym@umich.edu.

The University of Michigan Institutional Review Board Health Sciences and Behavioral Sciences has determined that this study is exempt from IRB oversight.

Appendix B.2

Participant Instructions

Participant Instructions

Thank you for agreeing to participate in this study.

In the first part of this study, you will be asked to carry out four searches. For two of them you will be searching an online question and answer site. For the other two searches you will be searching a collection of photographs.

Before you start a search, I will read you the search topic description for that search. Please listen carefully, as I will not answer questions regarding the search topic once you start your search. You have up to 10 minutes for each search – please do these searches as you normally would. You are encouraged to think aloud as you are doing your searches, describing your thought processes, reactions and feelings as you are searching. The searches will be recorded using screen recording software.

We will review two of the search recordings. I will ask you questions about your search as we are reviewing the search recordings.

Do you have any questions?

Experiment Setup

Before you start on your searches I would like to familiarize you with the experiment setup. You will be carrying out your searches on this laptop, using the Firefox browser. Take note of this camera icon to the right of the browser. Clicking on it will take a screenshot and save it to the desktop.

This is the start screen for each search – to start your search, please enter your subject ID number, and click on the system you have been asked to use for the search. Do your search as you normally would and I will inform you when the 10 minutes are up. When you are done with a search, please press together these two keys (command-.).

We will proceed in the same fashion for the remaining searches.

Do you have any questions?

Then let’s get started.

Appendix B.3

Search Tasks

Search Tasks

Task 1

You are preparing flyers advertising events for Women's History Month. You want to find 5 photographs to use in the flyers, showing images of American women at work through the years. Save the 5 photographs you intend to use.

Use system _____

Task 2

You are interested in visiting Chicago for a weekend trip, and would like to find information about hotels, restaurants, and interesting things to do in the city. Save 3 pages you found most useful.

Use system _____

Task 3

You want to buy a new laptop computer and need to decide what kind to get. To help you make this decision, you would like to know what other people recommend, as well as their own experiences using different models of laptop computers. Save 3 pages you found most useful.

Use system _____

Task 4

You are taking a class on the history of cities. For your next homework assignment you have to present to the class historical images of New York City. You need to find 5 photographs for your presentation. Save the 5 photographs you intend to use.

Use system _____

Appendix B.4

Pre-Search Questionnaires

Participant ID: _____
Date: _____
System: _____

Pre-Search Questionnaire — Images

1. I have previously carried out Web searches for images on this topic.

Not at all	Rarely	Somewhat Frequently	Frequently	Very frequently
o	o	o	o	o

2. I have a clear idea of how I will proceed with this search.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
o	o	o	o	o

3. I know clearly what type of images I am looking for.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
o	o	o	o	o

Participant ID: _____
Date: _____
System: _____

Pre-Search Questionnaire — Text

1. I have previously carried out Web searches on this topic.

Not at all	Rarely	Somewhat Frequently	Frequently	Very frequently
o	o	o	o	o

2. I have a clear idea of how I will proceed with this search.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
o	o	o	o	o

3. I know clearly what type of information I am looking for.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
o	o	o	o	o

Appendix B.5

Post-Search Questionnaire

Participant ID: _____
Date: _____
System: _____

Post-Search Questionnaire

1. I thought this search task was easy.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
o	o	o	o	o

2. The results of this search were what I had expected to find.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
o	o	o	o	o

3. I was lost at some point during this search.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
o	o	o	o	o

4. I utilized all the available search features.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
o	o	o	o	o

5. I felt the search took longer than I expected.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
o	o	o	o	o

6. I am satisfied with the results of this search.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
o	o	o	o	o

Appendix B.6

Background Questionnaire

Participant ID: _____
Date: _____

Background Questionnaire

1. What is your major at the University of Michigan? _____

2. What year are you in your program?

☐ Freshman
☐ Sophomore
☐ Junior
☐ Senior
☐ Other, please specify _____

3. What is your age? _____

4. What is your gender?

☐ Female
☐ Male

5. On an average day, approximately how much time do you spend online from each of the following locations?

	None	<1 hour	1-2 hours	2-3 hours	3-4 hours	4-5 hours	>5 hours
School	☐	☐	☐	☐	☐	☐	☐
Work	☐	☐	☐	☐	☐	☐	☐
Home	☐	☐	☐	☐	☐	☐	☐
Other	☐	☐	☐	☐	☐	☐	☐

6. Which Web site or page do you typically start from when searching for **text information**?

_____

7. Please rate your ability to find information in **text form** on the Web.

Poor						Excellent
1	2	3	4	5	6	7

8. How often do you search the Web for information in **text form** for the following purposes? Please check all that apply.

	Everyday	3-4 times per week	1-2 times per week	< 1 time per week	Never
School	☐	☐	☐	☐	☐
Work	☐	☐	☐	☐	☐
Personal	☐	☐	☐	☐	☐
Other	☐	☐	☐	☐	☐

9. Which Web site or page do you typically start from when searching for **images**?

_____

10. Please rate your ability to find the **images** you are looking for on the Web.

Poor						Excellent
1	2	3	4	5	6	7

11. How often do you search for **images** on the Web for the following purposes? Please check all that apply.

	Everyday	3-4 times per week	1-2 times per week	< 1 time per week	Never
School	☐	☐	☐	☐	☐
Work	☐	☐	☐	☐	☐
Personal	☐	☐	☐	☐	☐
Other	☐	☐	☐	☐	☐

12. How often do you use the following sites?

	Everyday	3-4 times per week	1-2 times per week	Less than 1 time per week	Never
CiteULike	☐	☐	☐	☐	☐
Connotea	☐	☐	☐	☐	☐
Del.icio.us	☐	☐	☐	☐	☐
Digg	☐	☐	☐	☐	☐
Facebook	☐	☐	☐	☐	☐
Flickr	☐	☐	☐	☐	☐
Last.fm	☐	☐	☐	☐	☐
LibraryThing	☐	☐	☐	☐	☐
LiveJournal	☐	☐	☐	☐	☐
MySpace	☐	☐	☐	☐	☐
Photobucket	☐	☐	☐	☐	☐
Technorati	☐	☐	☐	☐	☐
Twitter	☐	☐	☐	☐	☐
Wikipedia	☐	☐	☐	☐	☐
Yahoo! Answers	☐	☐	☐	☐	☐
YouTube	☐	☐	☐	☐	☐
Other (_____)	☐	☐	☐	☐	☐

13. Which operating system do you use most frequently?

Windows	Macintosh	Linux	Unix	Other
☐	☐	☐	☐	☐

(If Other, please specify _____)

14. Which web browser do you use most frequently?

Firefox	Internet Explorer	Safari	Chrome	Opera	Other
☐	☐	☐	☐	☐	☐

(If Other, please specify _____)

Appendix B.7

Interview Guides

Interview Guide – Text

Starting the Interview with …

Now, let's look over your search. This was a search for [search task description].

What type of content or information did you have in mind when you started your search?

How did you intend to find these type of content or information?

For initial query formulation:

1. How difficult was it to come up with search terms at the start of your search?

For search results examination:

2. [If the subject selected a particular search result]: Why did you select this item to look at?
    a. What parts of the search results did you look at to help you decide what to click on?

    b. [For each of items mentioned in a] In what way was [item from *a*] helpful? (*Title, text snippet, tags, other*)

    c. [For items not mentioned in a] Why didn't you find them helpful? (*Title, text snippet, tags, other*)

3. [If the subject did not select a particular search result]: Why didn't you select anything? What information in the search results helped you decide?

4. [If the participant clicked on a tag]: Why did you click on the tag?

5. **What did you expect would happen when you clicked on the tag?**

6. [If the participant did not click on any tags]: Why didn't you try clicking on a tag?

For document page examination:

7. What were the things that you looked at on this page? Why?

8. Do you think this page is a good fit for the search topic? If so, why do you think so? If not, why not?

9. What aspects of the page were important to you in making decisions about what to do next? Why?

10. Why did (didn't) you save it?
    a. What parts of the page helped you decide whether to save it not?

    b. [For each of items mentioned in a] On a scale of 1 to 5, where 1 is not important and 5 is very important, how important was [item from *a*]? *(Title, text, date, tags, other)*

11. [If the participant clicked on a tag]: Why did you click on the tag?

12. [if the participant did not click on tags]: Why didn't you try clicking on a tag?

13. Did you find the tags useful? Why or why not?

For query reformulation using the search box

14. What made you change your search terms?

15. How difficult was it to come up with these search terms?

16. How did you decide to change the search terms in the way you did?

17. What kind of information did you use to come up with these search terms? (*Title, text, tags, other*)

When there is no query reformulation

18. Why didn't you change your search terms during your search?

19. If you were to modify your search, how likely do you think you would use the following to get ideas for new search terms? (*Title, text, tags, other*)

Interview Guide – Images

Starting the Interview with …

Now, let's look over your search. This was a search for [search task description].

What types of photographs did you have in mind when you started your search? Could you describe them?

How did you intend to find these types of photographs?

For initial query formulation:

1. How difficult was it to come up with search terms at the start of your search?

For search results examination:

2. [If the subject selected a particular search result]: Why did you select this item to look at?
    a. What parts of the search results did you look at to help you decide what to click on?

    b. [For each of items mentioned in a] In what way was [item from *a*] helpful? (*Title, thumbnail, tags, other*)

    c. [For items not mentioned in a] Why didn't you find them helpful? (*Title, snippet, thumbnail, other*)

3. [If the subject did not select a particular search result]: Why didn't you select anything? What information in the search results helped you decide?

4. [If the participant clicked on a tag]: Why did you click on the tag?

5. **What did you expect would happen when you clicked on the tag?**

6. [If the participant did not click on any tags]: Why didn't you try clicking on a tag?

For document page examination:

7. What were the things that you looked at on this page? Why?

8. Do you think this page is a good fit for the search topic? If so, why do you think so? If not, why not?

9. What aspects of the page were important to you in making decisions about what to do next? Why?

10. Why did (didn't) you save it?
    c. What parts of the page helped you decide whether to save it not?

    d. [For each of items mentioned in a] On a scale of 1 to 5, where 1 is not important and 5 is very important, how important was [item from *a*]? *(Title, text, date, tags, other)*

11. [If the participant clicked on a tag]: Why did you click on the tag?

12. [if the participant did not click on tags]: Why didn't you try clicking on a tag?

13. Did you find the tags useful? Why or why not?

For query reformulation using the search box

14. What made you change your search terms?

15. How difficult was it to come up with these search terms?

16. How did you decide to change the search terms in the way you did?

17. What kind of information did you use to come up with these search terms? (*Title, photograph, text, tags, other*)

When there is no query reformulation

18. Why didn't you change your search terms during your search?

19. If you were to modify your search, how likely do you think you would use the following to get ideas for new search terms? (*Title, text, tags, other*)

Interview Guide

Concluding questions

1. What are some Web sites you visit frequently?

2. How often do you visit them?

3. Have you used tags on these sites? How did you use them?

4. (If said yes to 4) How were tags helpful? What kinds of tags were particularly helpful?

5. How familiar are you with tags?

6. How would you describe tags?

7. What do you think is their purpose?

8. Who do you think produces tags?

9. Why do you think they are producing tags?

10. Have you tagged items yourself? Yours or other people's? Why or why not do you tag items?

11. What are some library databases you have used?

12. Are you familiar with MTagger?

Thank you!

Appendix C

Example Transcripts

Transcript of Interview (Subject 40; Chicago Task)

[What follows is the transcript of the interview with Subject 40. Due to space limitations, only the part of the interview for the Chicago task is included.]

00:00 Speaker 1: You're doing for it.
00:01 Speaker 2: Okay.
00:03 S1: Okay. So let's look at which search this is. This is the Chicago one. So what type of contents or information did you have in mind when you started your search?
00:18 S2: I guess the first thing I thought about was finding travel, like travel prices. Like how much bus tickets would cost to Chicago? But that was like, are you talking about the first search or the second search?
00:34 S1: The second one, the second Chicago.
00:36 S2: Okay, the second one I was definitely thinking about hotel prices, like names of hotels that would be good to stay at. Restaurants, museums because you hear that Chicago has good museums so that's basically it. That's all that came to mind. [chuckle]
00:57 S1: So what was your plan for finding this type of information?
01:02 S2: That's a good question. Like, looking at this, like this there's only one bar so like for the that questions asked, did you utilize all the search options? I think, you know I assumed this was the only one but I didn't know if there were other ways to search. So basically, it was just to type in words that I thought might produce the results I was looking for.
01:27 S1: Okay. So, let's see, okay and since the start of your search, so how difficult was it to come up with these search terms to start your search?
01:41 S2: That's not difficult. Just Chicago attractions. Interesting places.
01:47 S1: Okay. So on the scale of 1 to 5 where 1 is not difficult and 5 is very difficult?
01:53 S2: I'd say like 2.
01:55 S1: 2. Okay so now let's look at what happened? Okay, can you walk me through what you're doing here? So are you reading mostly the titles the blue text or do gray text? [overlapping conversation]
02:10 S2: I'll show you. Yeah. I'm looking at the blue text and I guess like I would skip over Wisconsin obviously. Road trip from Chicago to Houston that didn't really apply so much. I thought that maybe with this Chicago, Madison, Milwaukee that there would be something about Chicago but as I read through that it didn't seem like there's a lot of information on Chicago. So, that's what that was.
02:35 S1: Okay, did you happen to notice any one of the green tags?
02:39 S2: I did. I did notice them. But I think I clicked on one at one point and it didn't come up with, I can't remember but I don't think it came up with anything that I was, that I thought it was useful, so.
02:56 S1: Okay so, well looking at, did you happen to notice the tags on this?
03:01 S2: Drury, road trip, Madison, Chicago. Yeah.
03:05 S1: So do you think that contributed to your clicking on this?
03:12 S2: No.
03:12 S1: No? Why was that?
03:18 S2: I guess the blue, the titles are what kind of what caught my attention first.
03:26 S1: Let's see. So what if instead of tags, what if there have been no Chicago in the tags, would you still have clicked on this thing?
03:36 S2: I see. I don't know. Probably I would've just because the heading said Chicago.
03:47 S1: Then you took a look and if I recall, you ended on not saving this one. Did you happen to notice the tags while you were here?
04:03 S2: I don't know if I did.
04:05 S1: Okay. Are you scrolling and looking at what people have to say?
04:05 S2: I think I could've researched these a lot more carefully. Probably that would've been better. [chuckled] But I just skimmed through them really fast. I think that there actually was some stuff on Chicago that I thought there wasn't but actually was.
04:28 S1: Okay, what made you change it to Chicago restaurants.
04:33 S2: I was hoping... I was just hoping to narrow it down and find Chicago restaurants.

04:39 S1: Did you happen to notice any of the related tags on top?
04:43 S2: I think I did. I'm pretty sure I did but I didn't, tags you know it's a, it's like a really good idea but for some reason I'm not like as accustomed to like clicking on them. So.
05:00 S1: Okay so this one ["In the Loop Chicago Romantic Restaurants"] I guess the title had mentioned restaurants.
05:05 S2: Right.
05:07 S1: And okay did you happen to look at the tags for that one?
05:13 S2: I think... I mean, it's probably, I feel like it's the second thing, cause I don't wanna read like all of the fine print. So I'll look at the bold and then look at the tags and if the tags are there and the tags correspond to what the bold blue says then I'll click on it. Yeah.
05:34 S1: Okay.
[pause]
05:43 S1: Okay and then you started reading the replies. Did you happen to notice when this was posted?
05:50 S2: No. I didn't. No.
05:52 S1: All I see was this, fairly recently so I just kind of thought it's okay.
05:58 S2: Okay.
05:59 S1: Was there anything specific you're looking for or are you just skimming?
06:02 S2: I was kind of looking for names until I get something like I noticed this person listed several a lot of different and I thought some of these would be probably like I'm not so much, I don't really have any interest in going to a really expensive fancy restaurant. But I just thought it would be good to have names cause I didn't know how I don't know what exactly kind of restaurant, like it was as sweet so.
06:26 S1: Okay so you saved that one.
06:29 S2: Yeah.
06:29 S1: And now at this point, okay I'm trying to find hotels [chuckle]
06:34 S2: Right.
06:36 S1: Wait instead of hotels do you want to Chicago museums.
06:40 S2: Yeah, I think I searched hotels earlier something and I didn't... I don't know why I did that actually. I think my mind was just set on museum so.
06:47 S1: Okay.
06:49 S2: Yeah.
06:53 S1: And so you're continuing to look at the blue text? Okay. So, do you happen to notice this one for example, "Chicago for an atypical tourist"?
07:13 S2: I did see that. I thought about clicking on that but I am I'm not a techie or I don't consider myself really one so that's why I didn't click on it.
07:25 S1: Okay. So, techie, was it, did you see it in the tags or in the text?
07:33 S2: I saw it just actually, yeah, I see it... Well, I thought "Chicago for an atypical tourist", I thought that's the one that I would click on but then like once I made the decision that I was going to click on it then I decided reading the description and I've realized it didn't fit. So...
07:47 S1: Okay. Let's see, continuing... Oh, what made you click on that one "Christmas in Chicago?"
07:58 S2: I just thought that in spite of the fact that it's only Christmas, I mean if they still like drop some like good names of restaurants or whatever attractions and that would, even though it's Christmas, I'm not necessarily going for Christmas like, yeah.
08:14 S1: Okay so you happen to read some of the gray text there then?
08:19 S2: Let me see, yeah I think after I decided to click on it I did read some of that.
08:31 S1: Okay and this one. Okay. And did you happen to notice any other tags at any point?
08:41 S2: I don't think for this one I did.
08:44 S1: Okay this is actually a very long question and they list quite a lot of hotels.
08:53 S2: Yeah. I think they started they start talking about like London hotels and then I'm like okay. [chuckle]
09:08 S1: Okay and then people are giving their opinions of the various hotels.
09:21 S2: It's kind of a neat program that allows you to go back and see what people have typed. It's that like ordinarily on computers or is does it make you have to download in?

09:32 S1: In the latest version of the Mac OS it's actually built in now. Okay so then you keep looking and let's see, yeah you ended up saving that one. What made you save it?
09:47 S2: I think that, let me see, I can't remember what it was I just maybe a [09:51] ____ scroll down.
09:54 S1: Why we can't go back to see what you're looking at?
10:01 S2: Yeah, I think it was just, I think people were talking about specific hotels and I thought might be interesting to stay at.
10:08 S1: Okay. So, you saved that one, so now let's see you found hotels. What else is next? [chuckle] Well you have restaurants, you have hotels. And what made you put Chicago navy pier.
10:27 S2: I think I saw something, I've heard Chicago, it's in Chicago I don't really know. I have only lived in Chicago once, so I don't know but...
10:37 S1: So, did you put in navy pier because it was something you thought up yourself or you saw something?
10:42 S2: It's something I thought of myself, yeah.
10:45 S1: Okay.
10:44 S2: It was too specific, it felt so.
10:47 S1: So you changed it, oh what made you change it to aquarium?
10:56 S2: [chuckle] I don't know, it's really about the same level of [10:56] ____ I guess but...
10:59 S1: It's museum.
11:00 S2: Right. I thought maybe that would, I heard there's this big aquarium in Chicago so that's why.
11:08 S1: Okay, so on a scale of one to five where one is not difficult and five is very difficult, how difficult would you rate coming up with these search terms?
11:19 S2: It wasn't, I don't think it was like two maybe, it was just something in my head so.
11:27 S1: Well there's only one [chuckle] but did you happen to look at any of the tags for this one?
11:33 S2: I think I did, but I felt like if "chicago aquarium" was really specific, then shedd, because it's Shedd Aquarium, I felt like that would probably be too specific, but it didn't occur to me to click on the tags.
11:50 S1: On the other hand, let's see, did looking at, seeing these tags did they like contribute to your clicking on that ...?
11:59 S2: I think it probably probably reaffirmed it's like a good idea to click on it because it had those terms in it on Chicago.
12:12 S1: Okay. And then you went and after that... And actually you kind of had... Yeah you ended up saving this one. So, what made you save it?
12:22 S2: I just had a good list of things so... Shedd Aquarium, Adler Planetarium, all of these things, and then someone said that they're all right next to each other, which I thought kind of useful information if I were planning to go there so...
12:36 S1: Oh wait. Somebody even mentions which bus to take. That's good.
12:39 S2: Yeah.
12:41 S1: Okay. So that was that search. Let's see... So did you consider clicking on any of the tags that you saw?
12:52 S2: No. I mean I thought... Is it fairly useful? Like when... Should I have? I feel that's something I just didn't really but it probably is fairly useful.
13:08 S1: No. I mean, you should have just done what you normally do when searching.
13:11 S2: Yeah.
13:11 S1: So what do you think would have happened if you'd clicked on say, the chicago tag?
13:17 S2: I feel like I would have gotten just a wide variety of results, possibly having to do with Chicago, and possibly not having to do with Chicago, because I know of some of the searches, if anything not very specific will just be a lot of other things so.
13:39 S1: So let's see... So what if you'd entered... So you think you'd be... Do you think you would have gotten the same or different results from putting in "chicago" in the search box or just clicking on "chicago"?
13:57 S2: Oh, I think... I don't know. I think my guess is I would've had gotten same results.
14:03 S1: Okay. So that was that for that search.
[The interview continued on to the New York City task]

[What follows is the part of the interview asking about general experience with tags on the Web]

27:35 S1: Okay. So now I do have some like general questions... Okay. What are some websites that you visit frequently and this is any kind of website?
27:51 S2: Google. I mean, Gmail definitely. [27:58] ____ ITS Webmail and LiveJournal. That's really... That's about it. The New York Times once [28:08] ____. That's really...
28:11 S1: Okay. How frequently when you say you visit them?
28:15 S2: Several times day. Once a day. I mean, it depends on which one. I think, I'd check my email like... [chuckle]
28:20 S1: Multiple.
28:21 S2: [chuckle] A lot. I try like six times a day or something like that. New York Times is probably once a day, LiveJournal once a day and... Also say Google, Gmail... I think Gmail... Probably like three times a day and then Google... I'd say three or five times a day.
28:45 S1: ____?
28:47 S2: Yeah.
28:49 S1: Have you noticed tags on any of these sites?
28:52 S2: On LiveJournal, I noticed them. I think they're really handy actually but not [28:59] ____ so much for Google and not so much for New York Time. Pretty much only LiveJournals, the only place like I sort of use tags. So...
29:09 S1: Why is it that they're handy on LiveJournal?
29:14 S2: You can organize and choose according like by topic or... It just makes it easier to find if you ever want to go back for your journal and come up with entries that are like related to a specific topic or have a... Yeah.
29:38 S1: So have you used text in looking for somebody else's LiveJournal entry?
29:42 S2: Yeah. That's stuff works... It's that... People are kind of erratic though like I know I won't always use tags on my entries and a lot of people don't use that like really religiously but it is helpful, yeah.
29:59 S1: How are they helpful? Like if you're looking at somebody else's entry?
30:03 S2: You can, I mean you can like it arranges them sporadically so if you need to look at someone's journals generally you're just looking at someone's journal. But yeah, it just takes you to a bunch of entries that are loosely related somehow. That's unique.
30:26 S1: When looking at tags, were there tags that you felt were particularly helpful or particularly not helpful?
30:34 S2: I kind of looked at the stuff that are useful to me. If someone has a tag called life, you know like I don't really know which... I'm generally not looking, like that didn't seem like a helpful one but if its work, you know, music, concerts, that's a little bit more specific, and it depends and I think it just varies on the content of whatever the person's writing about and what they tend to write about.
31:13 S1: You've mentioned tags in relation to Google and also New York Times. Have you noticed them there?
31:19 S2: I think I've noticed them but... And I've seen this like CNN.com to these tags but a number... I don't know, I've tried them before and it seems just like a very, like hit or miss the things that you get are sort of hit or miss. I always feel like there must be a more specific... There must be a more... Yeah, more specific way of finding the things you want to find.
[pause]
31:53 S1: How familiar would you say you are with tags?
31:56 S2: Not... I mean, not very. It varies.
31:59 S1: So on a scale of one to five, where one is not familiar to 5 is very familiar?
32:05 S2: Probably, I'd say I know the rudiments of tags, so 2.
32:14 S1: You said you know the rudiments of tags, so how would you describe tags?
32:18 S2: Just sort of, I don't know what the term, there's probably a term for them, like a handle or something like a link that brings you to a page where websites or entries or articles are sort of grouped according to that, according to their tags. So you can find articles that are on weather if you click the weather tag and reserved.
32:46 S1: What do you think then is the purpose of tags?

32:50 S2: I feel like it's... It's just to help... It's to help people organize information and sort through information which makes it easier to find certain things because there's an alternative. It's like an alternative search tool if you don't want to, for whatever reason, you don't like just typing in something to Google and it gives an alternative.
33:19 S1: Okay, going back to LiveJournal. When you clicked on that tag, what did you get back? Are you getting back things that have that tag or things that, you know, continue that term anywhere in the entry?
33:37 S2: It's things that have like that have that been intentionally assigned to that tag. So it doesn't matter if, it has to open to assigned that tag.
33:47 S1: So, if I had a tag that's favorite song, even though like I had journal entry that said something about that favorite song. If it doesn't have that tag, it's not?
34:00 S2: It will not turn up anyway. Yeah.
34:06 S1: Thinking about on all these, like places on the web that you've seen tags, who do you think is tagging things?
34:13 S2: I feel with LiveJournal, it's the user but these websites like New York Times, I'm sure that's their stuff like that's their web stuff or their computer stuff, kind of going through and doing that. But yeah I would assume like LiveJournal would probably be the exception or blogging sites are probably the exception where people can tag it themselves but.
34:40 S1: In LiveJournal, have you tagged somebody else's, something that wasn't yours?
34:46 S2: Yes. I think they've got a tag called Memory. I don't know if that's a tag but there's an option to add someone else's entry to like a memories thing that when you click on your memories they can go, that person's entry can go in that range. I don't know if that's a tag per se. I'm not sure. But I don't generally speaking, though it's just the websites that... It's that what you find that's already been tagged.
35:21 S1: Are you in any LiveJournal community?
35:23 S2: I think, yeah.
35:25 S1: Okay. Do any of them have... Do they tag?
35:28 S2: Yeah.
35:30 S1: Have you posted to any of these communities or tag anything there?
35:34 S2: I haven't tried anything but I have posted some communities, yeah.
35:39 S1: Okay. In LiveJournal it's... In communities you can tag even though even though it's not yours.
35:46 S2: Yeah.
35:47 S1: Okay. Why do think people are tagging things?
35:52 S2: It just makes it more, I mean it makes the... I think it makes... Makes it more understandable, it makes it a lot more cohesive. It makes it less random. It's just helpful to have those 'cause it's... Yeah. I mean, especially if you use LiveJournal a lot. Like if people just... If you just update it randomly and sporadically, like there's not much use if you don't use it a lot. But yeah, and it just makes it more understandable.
36:26 S1: Okay. That's about all the questions I have to ask. No, wait. Do you know how to tag YouTube or Flickr?
[pause]
36:35 S1: Let's see. Have you noticed tags on YouTube and Flickr?
36:40 S2: Yes, I have. And sometimes they help, sometimes they don't. It's sort of... Yeah. And I don't really... I don't use them a lot but sometimes I do. They seem to be marginally helpful I think. Sometimes they work. Sometimes they don't.
36:58 S1: Could you elaborate a little bit?
37:00 S2: I think I have clicked on a tag that maybe something like the video. Yeah. [laughter] How do I say this? Like if there's an actor and actress and like a clip of the... And like a trailer or something and there's a tag that has the actor's name and I'll click on it and if I want to find more clips with the actor in it.
37:29 S1: So this is on YouTube?
37:31 S2: I think. Yeah. I want to say that's... I want to say two thumbs up but I'm not sure yet.
37:37 S1: Okay. So why would you click on the tag instead of say typing in the actor's in the search box?

37:45 S2: I would rather... Yeah. I mean, just out of habit, that's what I usually do. I usually do that. I usually type it in.
37:54 S1: Okay. Those are all the questions I have about tagging. So now I have some questions about library use. What are some library databases you have used recently?
38:03 S2: I use, I used the Search Tools. I don't know if that's a database, but the Search Tools I actually use pretty often to find. If I'm going to Askwith, if I want to check out a movie from Askwith, I usually see if the library has it. The same thing with books, I'll see if the library has it. But in the past, it's been kind of a problem because I'll go and type in the title of a book and it will say that it's on the shelf, but I get there and it's not on the shelf. But I think I've talked to someone and figured that out, it's usually being held elsewhere. It's, you know, it's checked in but it's just not there. But yeah, I use that. Yeah.
38:44 S1: Are you familiar with MTagger?
38:47 S2: I... Yeah. I know the name but I don't know how to use it and I don't use it. So...
38:53 S1: Okay. Well, thank you. Those are all the questions I have.
38:56 S2: Okay.

Bibliography

Abbas, J. (2007). In the margins: Reflections on scribbles, knowledge organization, and access. *Knowledge Organization, 34*(2), 72-77.

Agger, M. (2009). When the hive mind works, it's a beautiful thing. Retrieved September 1, 2011, from http://www.slate.com/id/2211694/

Al-Khalifa, H. S., & Davis, H. C. (2006). Folksonomies versus Automatic Keyword Extraction: An Empirical Study. *IADIS International Journal on Computer Science and Information Systems*, *1*(2), 132-143.

Ames, M., & Naaman, M. (2007). Why We Tag: Motivations for Annotation in Mobile and Online Media. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 971-980).

Anick, P. (2003). Using terminological feedback for web search refinement: a log-based study. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 88-95).

Anick, P., & Kantamneni, R. G. (2008). A longitudinal study of real-time search assistance adoption. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 701-702).

Armitage, L. H., & Enser, P. G. B. (1997). Analysis of user need in image archives. *Journal of Information Science, 23*(4), 287-299.

Aurnhammer, M., Hanappe, P., & Steels, L. (2006). Integrating Collaborative Tagging and Emergent Semantics for Image Retrieval. In *Proceedings of the 15th international conference on World Wide Web Tagging Workshop*.

Balatsoukas, P., & Ruthven, I. (2010). The use of relevance criteria during predictive judgment: An eye tracking approach. *Proceedings of the American Society for Information Science and Technology*, *47*(1), 1-10.

Barry, C. L. (1998). Document Representations and Clues to Document Relevance. *Journal of the American Society for Information Science, 49*(14), 1293-1303.

Beaulieu, M. (1997). Experiments on interfaces to support query expansion. *Journal of Documentation, 53*(1), 8-19.

Belkin, N. J., Oddy, R., & Brooks, H. (1982). ASK for information retrieval. *Journal of Documentation*, 38, 61-71.

Belkin, N. J. (1993). Interaction with Texts: Information Retrieval as Information-Seeking Behavior. In *Information retrieval '93. Von der Modellierung zur Anwendung* (pp. 55-66). Konstanz: Universitaetsverlag Konstanz.

Belkin, N. J. (1996). Intelligent information retrieval: whose intelligence? In J. Krause, M. Herfurth, & J. Marx (Eds.), *Proceedings of the 5th International Symposium for Information Science (ISI '96)* (pp. 25-31). Konstanz: Universitätsverlag Konstanz:.

Belkin, N. J. (2000). Helping People Find What They Don't Know. *Communications of the ACM, 43*(8), 58-61.

Belkin, N. J., Cool, C., Kelly, D., Lin, S.-J., Park, S. Y., Perez-Carballo, J. et al. (2001). Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval. *Information Processing & Management, 37*(3), 403-434.

Berendt, B., & Hanser, C. (2007). Tags are not Metadata, but "Justs More Content" - to Some People. In *Proceedings of the International Conference on Weblogs and Social Media (ICSWM07)*.

Bischoff, K., Firan, C. S., Nejdl, W., & Paiu, R. (2008). Can all tags be used for search? In *Proceeding of the 17th ACM conference on Information and knowledge management* (pp. 193-202).

Borlund, P. (2000). Experimental Components for the Evaluation of Interactive Information Retrieval Systems. *Journal of Documentation, 56*(1), 71-90.

Borlund, P. (2003). The Concept of Relevance in IR. *Journal of the American Society for Information Science and Technology, 54*(10), 913-925.

Bray, T. (2005). Still Wondering About Tags. Retrieved September 3, 2011, from http://www.tbray.org/ongoing/When/200x/2005/04/10/Do-Tags-Work

Brooks, C. H., & Montanez, N. (2006). Improved Annotation of the Blogosphere via Autotagging and Hierarchical Clustering. In *Proceedings of the 15th World Wide Web Conference (WWW06)*.

Bruza, P., McArthur, R., & Dennis, S. (2000). Interactive Internet search: Keyword, directory and query reformulation mechanisms compared. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 280-287).

Capra, R., Marchionini, G., Oh, J. S., Stutzman, F., & Zhang, Y. (2007). Effects of structure and interaction style on distinct search tasks. In *Proceedings of the 7th ACM/IEEE joint conference on Digital libraries* (pp. 442-451).

Carman, M. J., Baillie, M., & Crestani, F. (2008). Tag data and personalized information retrieval. In *Proceedings of the 2008 ACM Workshop on Search in Social Media* (pp. 27-34).

Chen, H., & Dumais, S. (2000). Bringing order to the Web: automatically categorizing search results. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 145-152).

Choi, Y., & Rasmussen, E. M. (2002). Users' Relevance Criteria in Image Retrieval in American History. *Information Processing and Management, 38*, 695-726.

Clarke, C. L. A., Agichtein, E., Dumais, S., & White, R. W. (2007). 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* (pp. 135-142).

Collins, B. (2009). How a Minnesota man blew the whistle on the New York Times. Retrieved September 7, 2011 from http://minnesota.publicradio.org/collections/special/columns/news_cut/archive/2009/07/fakery_in_the_times.shtml

Conniss, L. R., Ashford, A. J., & Graham, M. E. (2000). Information seeking behaviour in image retrieval: VISOR I final report. The Library and Information Commission, United Kingdom.

Cool, C., Belkin, N. J., Frieder, O., & Kantor, P. (1993). Characteristics of texts affecting relevance judgments. In *Proceedings of the 14 National Online Meeting* (pp. 77-84).

Cosijn, E., & Ingwersen, P. (2000). Dimensions of Relevance. *Information Processing & Management, 36*(4), 533-550.

Cosley, D., Baxter, J., Lee, S., Alson, B., Nomura, S., Adams, P. et al. (2009). A tag in the hand: supporting semantic, social, and spatial navigation in museums. In *Proceedings of the 27th international conference on Human factors in computing systems* (pp. 1953-1962).

Cosley, D., Lewenstein, J., Herman, A., Holloway, J., Baxter, J., Nomura, S. et al. (2008). ArtLinks: fostering social awareness and reflection in museums. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 403-412). ACM.

Crystal, A., & Greenberg, J. (2006). Relevance criteria identified by health information users during Web searches. *Journal of the American Society for Information Science and Technology, 57*(10), 1368-1382.

Cunningham, S. J., Bainbridge, D., & Masoodian, M. (2004). How People Describe Their Image Information Needs: A Grounded Theory Analysis of Visual Arts Queries. In *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries (JCDL '04)* (pp. 47-48).

Cunningham, S. J., & Nichols, D. M. (2008). How People Find Videos. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries* (pp. 201-210).

Cunningham, S. J., Reeves, N., & Britland, M. (2003). An ethnographic study of music information seeking: implications for the design of a music digital library. In *Proceedings of the 2003 Joint Conference on Digital Libraries* (pp. 5-17).

Datta, R., Joshi, D., Li, J., & Want, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, *40*(2), Article No. 5.

Davis, I. (2005). Why Tagging is Expensive. Retrieved September 4, 2011, from http://blogs.talis.com/panlibus/archives/2005/09/why_tagging_is.php

Dochy, F., Segers, M. & Buehl, M. M. (1999). The Relation Between Assessment Practices and Outcomes of Studies: The Case of Research on Prior Knowledge. *Review of Educational Research, 69* (2), 145-186.

Drabenstott, K. M. (2003). Do nondomain experts enlist the strategies of domain experts? *Journal of the American Society for Information Science and Technology, 54*(9), 836-854.

Dumais, S., Cutrell, E., & Chen, H. (2001). Optimizing Search by Showing Results in Context. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 277-284).

Dziadosz, S., & Chandrasekar, R. (2002). Do thumbnail previews help users make better relevance decisions about web search results? In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieva*l (pp. 365-366).

Elkins, J. (1999). *The Domain of Images*. Ithaca, New York: Cornell University Press.

Enser, P. G. B. (1993). Query analysis in a visual information retrieval context. *Journal of Document and Text Management, 1*(1), 25-52.

Enser, P. G. B. (1995). Progress in Documentation: Pictorial Information Retrieval. *Journal of Documentation*, 126-170.

Enser, P. G. B., Sandom, C., & Lewis, P. H. (2005). Surveying the Reality of Semantic Image Retrieval. In S. Bres & R. Laurini (Eds.), *VISUAL 2005*, LNCS 3736 (pp. 177-188). Berlin Heidelberg: Springer-Verlag.

Fallows, Deborah. (2008). Search Engine Use. Retrieved September 4, 2011, from http://www.pewinternet.org/~/media//Files/Reports/2008/PIP_Search_Aug08.pdf.pdf

Fast, K. V., & Campbell, D. G. (2004). "I still like Google": University student perceptions of searching OPACs and the web. In *Proceedings of the 67th ASIS&T Annual Meeting* (pp. 138-146).

Fidel, R. (1997). The Image Retrieval Task: Implications for the Design and Evaluation of Image Databases. *The New Review of Hypermedia and Multimedia*, 181-199.

Foster, J. (2006). Collaborative Information Seeking and Retrieval. In *Annual Review of Information Science and Technology* (pp. 329-356). Information Today.

Freyne, J., Farzan, R., Brusilovsky, P., Smyth, B., & Coyle, M. (2007). Collecting Community Wisdom: Integrating Social Search and Social Navigation. In *Proceedings of the 12th international conference on Intelligent user interfaces* (pp. 52-61).

Fukumoto, T. (2006). An analysis of image retrieval behavior for metadata type image database. *Information Processing & Management, 42*(3), 723-728.

Geisler, G., & Burns, S. (2007). Tagging Video: Conventions and Strategies of the YouTube Community. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries* (p. 480).

Gilbert, E., Karahalios, K., & Sandvig, C. (2008). The network in the garden: an empirical analysis of social media in rural life. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1603-1612). ACM.

Golder, S. A., & Huberman, B. A. (2006). Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science, 32*(2), 198-208.

Goodrum, A. (2005). I Can't Tell You What I Want, But I'll Know It When I See It. *Reference and User Services Quarterly, 45*(1), 46-53.

Goodrum, A. A., Bejune, M. M., & Siochi, A. C. (2003). A state transition analysis of image search patterns on the Web. In CIVR 2003 (pp. 281-290). Berlin: Springer-Verlag.

Goodrum, A., & Spink, A. (2001). Image searching on the Excite Web search engine. *Information Processing & Management, 34*(2/3), 295-311.

Guy, M., & Tonkin, E. (2006). Folksonomies: Tidying up Tags? *D-Lib Magazine, 12*(1).

Halpin, H., Robu, V., & Shepherd, H. (2007). The complex dynamics of collaborative tagging. In *Proceedings of the 16th international conference on World Wide Web* (pp. 211-220).

Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social Bookmarking Tools (I): A General Review. *D-Lib Magazine, 11*(4).

Hare, J. S., Lewis, P. H., Enser, P. G. B., & Sandom, C. J. (2006). Mind the gap: another look at the problem of the semantic gap in image retrieval. In *Proceedings of SPIE, the International Society for Optical Engineering* (pp.75-86).

Heckner, M., Neubauer, T., & Wolff, C. (2008). Tree, funny, to_read, google: what are tags supposed to achieve? a comparative analysis of user keywords for different digital resource types. In *Proceeding of the 2008 ACM workshop on Search in social media* (pp. 3-10).

Heymann, P., Koutrika, G., & Garcia-Molina, H. (2007). Fighting Spam on Social Web Sites: A Survey of Approaches and Future Challenges. *IEEE Internet Computing*, *11*(6), 36-45.

Heymann, P., Koutrika, G., & Garcia-Molina, H. (2008). Can social bookmarking improve web search? In *Proceedings of the international conference on Web search and web data mining* (pp. 195-206).

Hildreth, C. (1989). *Intelligent Interfaces and Retrieval Methods for Subject Searching in Bibliographic Retrieval Systems*. Cataloging Distribution Service, Library of Congress: Washington, DC.

Hirsh, S. G. (1999). Children's Relevance Criteria and Information Seeking on Electronic Resources. *Journal of the American Society for Information Science, 50*(14), 1265-1283.

Holley, R. (2010). Tagging Full Text Searchable Articles: An Overview of Social Tagging Activity in Historic Australian Newspapers August 2008 — August 2009. *D-Lib Magazine*, *16*(1/2).

Hollink, L., Schreiber, A. T., Wielinga, B. J., & Worring, M. (2004). Classification of User Image Descriptions. *International Journal of Human-Computer Studies, 61*, 601-626.

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006). Information Retrieval in Folksonomies: Search and Ranking. In Y. Sure & J. Domingue (Eds.), *The Semantic Web: Research and Applications* (Vol. 4011, pp. 411-426). Berlin: Springer.

Huang, C.-K., Chien, L.-F., & Oyang, Y.-J. (2003). Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology, 54*(7), 638-649.

Hunter, J., Khan, I., & Gerber, A. (2008). Harvana: harvesting community tags to enrich collection metadata. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries* (pp. 147-156).

Höchstötter, N., & Lewandowski, D. (2009). What users see – Structures in search engine results pages. *Information Sciences*, In press.

Ingwersen, P. (1992). *Information Retrieval Interaction*. London, UK: Taylor Graham.

Ingwersen, P., & Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context*. Dordrecht, The Netherlands: Springer.

Inskip, C., Butterworth, R., & MacFarlane, A. (2008). A study of the information needs of the users of a folk music library and the implications for the design of a digital library system. *Information Processing & Management, 44*(2), 647-662.

Inskip, C., MacFarlane, A., & Rafferty, P. (2008). Content or Context? Searching for Musical Meaning in Task-based Interactive Information Retrieval. In *Proceedings of the second international symposium on Information interaction in context: IIiX 2008* (pp. 72-74).

Inskip, C., MacFarlane, A., & Rafferty, P. (2010). Creative professional users' musical relevance criteria. *Journal of Information Science,* 36(4), 517-529.

Jacob, E. K. (2004). Classification and Categorization: A Difference that Makes a Difference. *Library Trends, 52*(3), 515-540.

Jaimes, A. (2006). Human Factors in Automatic Image Retrieval System Design and Evaluation. In S. Santini, R. Schettini, & T. Gevers (Eds.), *Proceedings of SPIE-IS&T -- Volume 6061, Internet Imaging VII* (Vol. 6061).

Janes, J. W. (1994). Other people's judgments: A comparison of users' and others' judgments of document relevance, topicality, and utility. *Journal of the American Society for Information Science, 45*(3), 160-171.

Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management, 42*(1), 248-263.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing and Management, 36*(2), 207-227.

Jansen, B. J., Zhang, M., & Zhang, Y. (2007). Brand awareness and the evaluation of search results. In *Proceedings of the 16th international conference on World Wide Web* (pp. 1139-1140). New York, NY: ACM.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56-65). ACM.

Joho, H., & Jose, J. M. (2008). Effectiveness of additional representations for the search result presentation on the web. *Information Processing and Management, 44*(1), 226-241.

Jörgensen, C., & Jörgensen, P. (2005). Image Querying by Image Professionals. *Journal of the American Society for Information Science and Technology, 56*(12), 1346-1359.

Kahneman, D., & Snell, J. (1990). Predicting Utility. In H. J. Einhorn & R. M. Hogarth (Eds.), *Insights in decision making* (pp. 295-310). Chicago: University of Chicago Press.

Kahneman, D., & Snell, J. (1992). Predicting a changing taste: Do people know what they will like? *Journal of Behavioral Decision Making, 5*(3), 187-200.

Keister, L. H. (1994). User Types and Queries: Impact on Image Access Systems. In R. Fidel, T. B. Hahn, E. M. Rasmussen, & P. J. Smith (Eds.), *Challenges in Indexing Electronic Text and Images* (pp. 7-22). Medford, NJ: Learned Information.

Kelly, D., Murdock, V., Yuan, X.-j., Croft, W. B., & Belkin, N. J. (2002). Features of documents relevant to task- and fact- oriented questions. In *Proceedings of the eleventh international conference on Information and knowledge management* (pp. 645-647).

Kelly, D., & Teevan, J. (2003). Implicit feedback for inferring user preference: a bibliography. *ACM SIGIR Forum, 37*(2), 18-28.

Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval, 3*(1-2), 1-224.

Kerstetter, D., & Cho, M.-H. (2004). Prior knowledge, credibility and information search. *Annals of Tourism Research*, *31*(4), 961-985.

Kherfi, M. L., Ziou, D., & Bernardi, A. (2004). Image Retrieval from the World Wide Web: Issues, Techniques, and Systems. *ACM Computing Surveys, 36*(1), 35-67.

Kim, K.-S. (2001). Information-seeking on the Web: Effects of user and task variables. *Library & Information Science Research*, *23*(3), 233-255.

Kim, Y.-M. (2010). Social Tags in Text and Image Search. In *Proceeding of the third symposium on Information interaction in context* (pp. 353-358).

Kipp, M. E. I. (2006). Exploring the Context of User, Creator and Intermediary Tagging. In *Proceedings of the Information Architecture Summit*.

Kipp, M. E. I. (2007a). @toread and Cool: Tagging for Time, Task and Emotion. In *Proceedings of the Information Architecture Summit*.

Kipp, M. E. I. (2007b). Patterns in Tagging: Collaborative Classification Practices in Social Bookmarking Tools. In *Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIST)*.

Koenemann, J., & Belkin, N. J. (1996). A case for interaction: A study of interactive information retrieval behavior and effectiveness. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 205-212).

Koshman, S., Spink, A., & Jansen, B. J. (2006). Web Searching on the Vivisimo Search Engine. *Journal of the American Society for Information Science and Technology, 57*(14), 1875-1887.

Kuhlthau, C. C. (2004). *Seeking Meaning: a process approach to library and information services*. Westport, Connecticut: Libraries Unlimited.

Kwasnik, B. (1989). How a personal document's intended use or purpose affects its classification in an office. In *Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 207-210).

Kwasnik, B. (1991). The importance of factors that are not document attributes in the organisation of personal documents. *Journal of Documentation*, *47*(4), 389-398.

Lan, W.-C. (2002). From Document Clues to Descriptive Metadata: Document Characteristics Used by Graduate Students in Judging the Usefulness of Web Documents. Unpublished dissertation, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

Lange, P. G. (2008). Publicly Private and Privately Public: Social Networking on YouTube. *Journal of Computer-Mediated Communication*, *13*(1), 361-380.

Lerman, K. (2007). User Participation in Social Media: Digg Study. In *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops* (pp. 255-258).

Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications, 2*(1), 1-19.

Lewandowski, D. (2008). The retrieval effectiveness of web search engines: considering results descriptions. *Journal of Documentation, 64*(6), 915-937.

Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G. et al. (2008). Eye Tracking and Online Search: Lessons Learned and Challenges Ahead. *Journal of the American Society for Information Science and Technology, 59*(7), 1041-1052.

Macgregor, G., & McCulloch, E. (2006). Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Library Review, 55*(5), 291-300.

Malone, T. W. (1983). How Do People Organize Their Desks? Implications for the Design of Office Information Systems. *ACM Transactions on Office Information Systems*, *1*(1), 99-112.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval.* Cambridge University Press.

Markey, K. (1986). *Subject Access to Visual Resources Collections*. New York: Greenwood Press.

Markkula, M., & Sormunen, E. (1998). Searching for photos - journalists' practices in pictorial IR. Paper presented at the The Challenge of Image Retrieval Research Workshop, University of Northumbria at Newcastle, UK.

Markkula, M., & Sormunen, E. (2000). End-User Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive. *Information Retrieval, 1*(4), 259-285.

Marlow, C., Naaman, M., boyd, d., & Davis, M. (2006). Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. In *Proceedings of the 15th international conference on World Wide Web, Collaborative Web Tagging Workshop*.

Marshall, C. (1997). Annotation: from paper books to the digital library. In *Proceedings of the second ACM international conference on Digital libraries* (pp. 131-140).

Marshall, C. C., & Bly, S. (2004). Sharing encountered information: digital libraries get a social life. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 218-227).

Mathes, A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Retrieved September 5, 2011, from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

McDonald, S., & Tait, J. (2003). Search strategies in content-based image retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 80-87).

Mizzaro, S. (1997). Relevance: The Whole History. *Journal of the American Society for Information Science, 48*(9), 810-832.

Morrison, P. J. (2008). Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web. *Information Processing and Management, 44*(4), 1562-1579.

Negoescu, R. A., & Gatica-Perez, D. (2008). Analyzing Flickr Groups. In *Proceedings of the 2008 international conference on Content-based image and video retrieval* (pp. 417-426).

Nov, O., Naaman, M., & Ye, C. (2008). What drives content tagging: the case of photos on Flickr. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1097-1100).

Novotny, E. (2004). I don't think I click: A protocol analysis study of use of a library online catalog in the Internet age. *College and Research Libraries, 65*(6), 525-537.

OCLC. (2009). *Online Catalogs: What Users and Librarians Want*. OCLC: Dublin, Ohio.

Ørnager, S. (1997). Image Retrieval: Theoretical Analysis and Empirical User Studies on Accessing Information in Images. In *Proceedings of the ASIS Annual Meeting* (pp. 202-211).

Park, C. W. & Lessig, V. P. Familiarity and Its Impact on Consumer Decision Biases and Heuristics. (1981). *Journal of Consumer Research, 8*, 223-230.

Rader, E., & Wash, R. (2008). Influences on tag choices in del.icio.us. In Proceedings of the *ACM 2008 conference on Computer supported cooperative work* (pp. 239-248). New York, NY: ACM.

Rainie, L. (2007). 28% of Online Americans have Used the Internet to Tag Content. Retrieved September 8, 2011, from http://www.pewinternet.org/~/media//Files/Reports/2007/PIP_Tagging.pdf.pdf

Rao, A. R., & Monroe, K. B. (1988). The Moderating Effect of Prior Knowledge on Cue Utilization in Product Evaluations. *Journal of Consumer Research*, *15*(2), 253-264.

Rasmussen, E. M. (1997). Indexing Images. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology* (Vol. 32, pp. 169-196). Medford, NJ: Information Today.

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology, 53*(2), 145-161.

Rieh, S. Y., & Belkin, N. J. (2000). Interaction on the Web: Scholars' Judgment of Information Quality and Cognitive Authority. In *Proceedings of the 63rd Annual Meeting of the American Society for Information Science* (pp. 25-38).

Rui, Y., Huang, T. S., & Chang, S.-F. (1999). Image Retrieval: Current Techniques, Promising Directions, and Open Issues. *Journal of Visual Communication and Image Representation, 10*, 39-62.

Ruthven, I. (2005). Integrating Approaches to Relevance. In A. Spink & C. Cole (Eds.), *New Directions in Cognitive Information Retrieval* (pp. 61-80). Dordrecht, Netherlands: Springer.

Ruthven, I. (2008). Interactive Information Retrieval. *Annual Review of Information Science and Technology*, 42, 43-91.

Ruthven, I., Baillie, M., Azzopardi, L., Bierig, R., Nicol, E., Sweeney, S. et al. (2008). Contextual factors affecting the utility of surrogates within exploratory search. *Information Processing & Management, 44*(2), 437-462.

Ruthven, I., Tombros, A., & Jose, J. M. (2001). A Study on the Use of Summaries and Summary-based Query Expansion for a Question-Answering Task. Paper presented at the 23rd BCS European Annual Colloquium on Information Retrieval Research.

Saracevic, T. (1975). Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science. *Journal of the American Society for Information Science, 26*(6), 321-343.

Saracevic, T. (1997). The stratified model of information retrieval interaction: Extension and applications. In *Proceedings of the American Society for Information Science* (pp. 313-327).

Saracevic, T. (2007a). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology, 58*(13), 1915-1933.

Saracevic, T. (2007b). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology, 58*(13), 2126-2144.

Savolainen, R., & Kari, J. (2006). User-defined relevance criteria in web searching. *Journal of Documentation, 62*(6), 685-707.

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A Re-examination of Relevance: Toward a Dynamic, Situational Definition. *Information Processing and Management, 26*(6), 755-776.

Schatz, B. R., Johnson, E. H., Cochrane, P. A., & Chen, H. (1996). Interactive term suggestion for users of digital libraries: using subject thesauri and co-occurrence lists for information retrieval. In *Proceedings of the first ACM international conference on Digital libraries* (pp. 126-133).

Sen, S., Harper, F. M., LaPitz, A., & Riedl, J. (2007). The Quest for Quality Tags. In *Proceedings of the 2007 international ACM conference on Supporting group work* (pp. 361-370).

Sen, S., Lam, S., K. (Tony), Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J. et al. (2006). tagging, communities, vocabulary, evolution (N. 4-8, Trans.). In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 181-190). Banff, Alberta, Canada: ACM.

Shneiderman, B., Byrd, D., & Croft, W. B. (1997). Clarifying Search: A User-Interface Framework for Text Searches. *D-Lib Magazine* Retrieved September 8, 2011, from http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html

Slone, D. J. (2005). A bird's eye view of cross-platform web interaction. *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, *61*(5), 657-669.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12), 1349-1380.

Sparck Jones, K. (Ed.). (1981). *Information retrieval experiment*. London: Butterworths.

Spink, A. (1997). Study of interactive feedback during mediated information retrieval. *Journal of the American Society for Information Science, 48*(5), 382-394.

Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From E-sex to E-commerce: Web search changes. *IEEE Computer, 35*(3), 133-135.

Sriram, S., Shen, X., & Zhai, C. (2004). A session-based search engine. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 492-493).

Tang, R., & Solomon, P. (2001). Use of relevance criteria across stages of document evaluation: On the complementarity of experimental and naturalistic studies. *Journal of the American Society for Information Science and Technology, 52*(8), 676-685.

Taylor, A. G. (2006). *Introduction to Cataloging and Classification* (10th ed.). Westport, Connecticut: Libraries Unlimited.

Thom-Santelli, J., Muller, M. J., & Millen, D. R. (2008). Social Tagging Roles: Publishers, Evangelists, Leaders. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1041-1044).

Thom-Santelli, J., Cosley, D., & Gay, G. (2010). What do you know?: experts, novices and territoriality in collaborative systems. In *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 1685-1694). ACM.

Tombros, A., Ruthven, I., & Jose, J. M. (2005). How users assess Web pages for information seeking. *Journal of the American Society for Information Science and Technology, 56*(4), 327-344.

Tombros, A., & Sanderson, M. (1998). Advantages of query biased summaries in information retrieval. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 2-10). New York, NY: ACM.

Toms, E. G., Freund, L., Kopak, R., & Bartlett, J. C. (2003). The effect of task domain on search. In *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative Research*.

Toms, E. G., & Latter, C. (2007). How consumers search for health information. *Health Informatics Journal, 13*(3), 223-235.

Uitdenbogerd, A., & Zobel, J. (1999). Melodic matching techniques for large music databases. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)* (pp. 57-66).

Vander Wal, T. (2007). Folksonomy. Retrieved September 6, 2011, from http://vanderwal.net/folksonomy.html

Vander Wal, T. (2005). Explaining and Showing Broad and Narrow Folksonomies. Retrieved September 6, 2011, from http://personalinfocloud.com/2005/02/explaining_and_.html

Voorhees, E. M., & Harman, D. K. (Eds.). (2005). *TREC: Experiment and Evaluation in Information Retrieval*. Cambridge, MA: The MIT Press.

Vélez, B., Weiss, R., Sheldon, M. A., & Gifford, D. K. (1997). Fast and effective query refinement. In *Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 6-15). New York, NY: ACM.

Wactlar, H. D., Christel, M. G., Gong, Y., & Hauptmann, A. G. (1999). Lessons Learned from Building a Terabyte Digital Video Library. *Computer, 32*(2), 66-73.

Wash, R., & Rader, E. (2007). Public Bookmarks and Private Benefits: An Analysis of Incentives in Social Computing. In *Proceedings of the American Society for Information Science and Technology* (pp. 1-13).

White, R. W., Jose, J. M., & Ruthven, I. (2003). A task-oriented study on the influencing effects of query-biased summarisation in web searching. *Information Processing & Management, 39*(5), 707-733.

White, R. W., & Ruthven, I. (2006). A Study of Interface Support Mechanisms for Interactive Information Retrieval. *Journal of the American Society for Information Science and Technology, 57*(7), 933-948.

Wood, S. L. & Lynch, Jr., J. G. (2002). Prior Knowledge and Complacency in New Product Learning. *Journal of Consumer Research, 29*, 416-426.

Woodruff, A., Faulring, A., Rosenholtz, R., Morrison, J., & Pirolli, P. (2001). Using Thumbnails to Search the Web. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 198-205).

Xie, H. I. (2008). *Interactive Information Retrieval in Digital Environments*. Hershey, PA: IGI Publishing.

Yanbe, Y., Jatowt, A., Nakamura, S., & Tanaka, K. (2007). Can Social Bookmarking Enhance Search in the Web? In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries* (pp. 107-116).

Yang, M., & Marchionini, G. (2004). Exploring users' video relevance criteria – a pilot study. In *Proceedings of the American Society for Information Science and Technology* (pp. 229-238).

Yee, M. M. (1991). System design and cataloging meet the user: User interfaces to online public access catalogs. *Journal of the American Society for Information Science, 42*(2), 78-98.

Zamir, O., & Etzioni, O. (1999). Grouper: a dynamic clustering interface to Web search results. *Computer Networks, 31*(11-16), 1361-1374.

Zhang, Y., Jansen, B. J., & Spink, A. (2009). Time series analysis of a Web search engine transaction log. *Information Processing and Management, 45*(2), 230-245.