Image search: An investigation of factors affecting search behaviour of users

A thesis submitted for the degree of

Doctor of Philosophy

Rahayu A Hamid

Bachelor of Science (Computer Science), University of Technology Malaysia

Master of Science (Computer Science), University of Technology Malaysia

School of Science,

College of Science, Engineering and Health

RMIT University

April 2017

In the name of Allah, The Most Beneficent, The Most Merciful

I dedicate this thesis to:

my mother

Rehana Abd Rashid,

my children

Muhammad Danial, Nur Nadzirah, Nur Nabila, Nur Najwa, Nur Nadhira and Muhammad Darwish

Thank you Allah for everything.

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Rahayu A Hamid

School of Science

RMIT University

August 15, 2017

Acknowledgments

While offering my insubstantial gratitude to The Almighty for His Infinite Grace and Mercy, I would like to express my thanks to my supervisors, Associate Professor James A. Thom, Dr Seyed Mohammad Mehdi Tahaghoghi and Dr Dayang Nurfatimah Awang Iskandar, whose expertise, encouragement, understanding, and patience have made my graduate experience manageable and invaluable. I appreciate their enthusiasm in my research and their assistance in writing this thesis.

To my husband and my children, there are no words to express how much your presence gives me strength and courage to persevere. To my mother and my siblings, thank you for the endless prayers and words of support. I am also thankful to have understanding and helpful parent-in-laws.

I really appreciate Jen Anderson from SLC for conducting the Writing Circle, providing a platform and creating an environment to practice to be a better researcher and writer. To my writing circle buddies, thank you for reading and commenting on my writings. I benefited greatly from my statistics consultation sessions with James Baglin, assisting me in my analysis of results.

I would like to thank my PhD mates — Lorena, Husna, Sharin, Sargol, Lisa, Fahad and Amir for their advise and friendship. I also thank my friends, whom indirectly supported me throughout my PhD journey — Aini, Nuha, Naza, Kalsom and Linda . Last, but not least, I thank my working colleagues for always supporting and encouraging me during my struggles — Kak N, Alin, Su, Tipah, Fakri, and Asri.

I acknowledge that this research would not have been possible without the opportunity,

scholarship and study leave granted by the Universiti Tun Hussein Onn Malaysia.

Credits

Portions of the material in this thesis have previously appeared in the following publications:

- Hamid, R. A. & Thom, J. A., "Criteria That Have An Effect on Users While Making Image Relevance Judgments", Proceedings of the Fifteenth Australasian Document Computing Symposium, 2010, 76-83 (Chapter 3, Chapter 4 and Appendix A).

- Hamid, R. A., Thom, J. A. & Awang Iskandar, D. N. F., "Effects of relevance criteria and subjective factors on web image searching behaviour", Journal of Information Science, First published date: September-01-2016, Doi:10.1177/0165551516666968 (Chapter 3, Chapter 4 and Chapter 5).

Not included in the thesis, but during candidature the following additional paper was published:

- Hamid, R. A. & Thom, J. A., "RMIT at ImageCLEF 2011 Plant Identification", CLEF (Notebook Papers/Labs/Workshop), The Netherlands, September 2011.

This research was undertaken with a scholarship and study leave granted by Universiti Tun Hussein Onn Malaysia.

The thesis was written in the `TeXstudio` editor on Windows 7, and typeset using the LaTeX 2_ε document preparation system.

Note

Unless otherwise stated, all fractional results have been rounded to the displayed number of decimal figures.

Contents

List of Figures

List of Tables

Abstract

Searching for images can be challenging. How users search for images is governed by their information need. Nevertheless, in fulfilling their information need, users are often affected by subjective factors. These factors include topic familiarity, task difficulty, relevance criteria and satisfaction. This thesis focuses on three research questions exploring how image information needs together with these factors affect online web users' searching behaviour. The questions are:

1. *How does image information need affect the criteria users apply when selecting relevant images?*

2. *How do different factors in image retrieval affect users' image searching behaviour?*

3. *Can we identify image information needs solely from user queries?*

In addressing these challenges, we conducted both user studies and proxy log analysis to complement each other. User studies are conducted in a laboratory setting and the needs are artificial, while proxy log captures users' actual needs and behaviour in the wild. The main user study involved 48 students of various disciplines from RMIT University. In the

study, we represent image information needs as types of tasks. Data were collected from questionnaires and screen capture recordings. The questionnaire was used to collect data on criteria users find important when judging image relevance and perception on the effects of subjective factors to their searching. Screen capture recordings of their search activities were observed and time stamped to identify and measure search and retrieval behaviour. These measures were used to evaluate the effects of subjective factors on users' image search behaviour.

The results showed in judging image relevance, users may apply similar criteria, however, the importance of these criteria depend on the types of image. Similarly, ratings of users' perception of aspects of performing image search show they were task dependent and that effect of different aspects were related. Users were more affected by familiarity and satisfaction when performing difficult image search tasks. Results of correlation suggest that users' perception on aspects of performing image search did not always correspond with their actual search behaviour. However, for some subjective aspects of user search behaviour, we have identified particular objective measures that correlate well with that aspect.

The examination of users' queries in proxy logs, shows that users search for unambiguous images more frequently compared to conceptual images. Their sessions are short with two to three terms per query. When analysing queries from logs, we are actually making a guess of what users were searching for. However, by examining the way users modify/reformulate their queries may give an indication of their information need. Results show, that users frequently submit new queries or replace terms from their previous query rather than revising the query into more depth or breadth. Similar findings were found when compared with the user study

data, whereby users in both settings exhibit similarity in the number of queries, terms and reformulation type.

This thesis concludes that given similar image information needs, ordinary users make relevance judgements similar to specialised users (such as journalists, art historians and medical doctors) despite giving attention to different criteria of relevance. Moreover, only certain measures of search behaviour used in text retrieval are applicable to image retrieval due to the difference in judging the relevance of textual information and image. In addition, visual information needs can be better inferred when analysing series of queries and their reformulation within a search session.

Chapter 1

Introduction

Online searching has become an integral part of people's working and recreational lives. As a consequence, there is a need for searching to be time efficient and produce effective results that meet users' needs. While, online searching is not a trivial task (users do not always find what they were looking for during a search session), searching for images, is even more challenging, as an image can carry multiple representations of meaning. The growth of the Internet and advances in digital imaging technology have certainly led to an increase in the amount of digital images available for users and the increase in interest around the use of images to satisfy their visual information needs.

As an alternative to using a web image search engine, users can search through image hosting or image sharing websites such as Flickr, Picasa and Photobucket. Most of these sites offer features such as multiple views (thumbnails and slideshows), the capacity to classify photos into albums, and the capacity to add annotations (captions or tags) and comments. In addition, photos can be found through social networks such as Instagram, Facebook and

Twitter. Although the use of keywords is the most common method to search for images, these retrieval systems do not use the content of the images themselves, but use their surrounding text or manually added annotations.

Designing a digital image retrieval system that meet the needs of users remains challenging. When searching for images, users usually have a visual information need in mind. However, the question would be how do users describe the image. What users see in their mind are sometimes disconnected with what they write (search terms or queries) [Goodrum, 2005]. If users submit a query, which does not retrieve results of what they were looking for, they then engage in a query modification but most of the time without any clear search strategy.

Image retrieval systems are the mechanisms for conducting image searches. However, these systems have focused on retrieving best matched results rather than the actual needs of users [Fidel, 1997]. In order to meet the needs of users, the application of user-centred approach is desired. In a user-centred approach, users are made to feel that they, rather than the system are driving the search process. However, research on user-centred issues in image retrieval is still quite sparse.

1.1 Background

There are two methods to search for images: a keyword search or a content-based search. Similar to finding textual information, users can search for images using a search engine.

In the first method, users enter a single or multiple keywords, and the search engine returns a list of results based on keywords. Users need to examine the list of ranked results

to find the image that satisfies their visual information needs the most.

The other method of searching for images is by using the image content itself, known as Content-Based Image Retrieval (CBIR). An image example is given to a CBIR system, and the system retrieves similar or relevant images from its collection. However, the retrieval is based on low-level visual features, such as colour, shape, texture and spatial layout; in contrast to using high-level concepts, those interpreted from reading the image [Eakins, 2002]. Moreover, CBIR applies a system-centred approach, which focuses on image indexing for higher performance of retrieval.

Regardless of the technology supporting an image retrieval system, the problem remains fundamental whereby users in general do not assess images by visual features but by semantics, that is by the meaning of an image. This is commonly referred to as the *semantic gap* problem. Smeulders et al. [2000, p. 1353] defined the semantic gap 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". Semantic of an image refers to meaning of image as assigned by a user. However, other users may or may not refer to similar meaning of that image. Therefore, one approach in resolving the semantic gap problem—that occurs because of the mismatch between user queries and image semantics — is to study the nature of users visual information needs.

The design of a user-centred image retrieval system is based on identification and understanding of user searching behaviours. Although many studies on searching behaviour have been conducted in the area of text retrieval, these findings might not be applicable to image retrieval because of the basic difference between textual and visual information. Therefore,

it is necessary and desirable to examine image searching behaviours.

To understand the needs of users, query analysis from either transaction logs or user studies have been applied as an effective method. The majority of previous research on users' image needs investigated professionals using a specific image collection, such as photographs from news archives [Ornager, 1996; Markkula and Sormunen, 2000; Hung et al., 2005; Westman et al., 2008; McCay-Peet and Toms, 2009], art history collections [Hastings, 1999; Chen, 2001; Choi and Rasmussen, 2003], and image archives and libraries [Armitage and Enser, 1997]. However, there has been a shift in the research whereby it is now focusing on everyday online user image searches. Nevertheless, most of the online users in these studies were students [Goodrum and Spink, 2001; Cunningham et al., 2004; Pu, 2005; Cunningham and Masoodian, 2006; Chew et al., 2010; Yoon, 2011; Yoon and Chung, 2011; Huang and Kelly, 2013; Park et al., 2015]. Other studies relied on image search engine transaction log analysis [Jansen et al., 2000a; Pu, 2005; Tseng et al., 2009] and their results report on how users search for images.

Apart from needs, additional issues concerning users image search behaviours can be found in factors affecting image retrieval. Task type, task difficulty, topic familiarity, satisfaction and relevance judgements are some factors that have been studied. User studies related to image retrieval involve users completing or performing a number of assigned tasks; however, each of these studies uses a different definition of task type or classification [Choi and Rasmussen, 2003; Hung et al., 2005]. In studies that employed log analysis, the tasks were categorised according to the type of queries that the user submitted [Armitage and Enser, 1997; Chung and Yoon, 2011]. Indeed, there are visible similarities in the task classifi-

cation between user studies and log analysis, whereby it is characterised by largely descriptive aspects such as people, places, events, and objects in an image.

A previous study on user search behaviour has compared search strategies between novices and experts [Hölscher and Strube, 2000]. However, even when searchers are highly skilled, they may struggle to find what they are looking for. A user may have a specific image in mind but have difficulty expressing their needs in words. Studies have shown that when facing difficulty during searching, users issue more diverse queries resulting in longer completion time [Aula et al., 2010; Liu et al., 2010c]. Similarly, familiarity on certain topics can influence searching behaviour [Kelly and Cool, 2002; Qu et al., 2010].

Relevance is another factor that can have an effect on image searching behaviour. Relevance in an image is in fact difficult to define satisfactorily. A relevant image is one judged similar in the context of a query and relevance criteria, but it all depends on the person judging it and in what context the image is relevant. Selecting an image as an 'answer' to an image query requires the user to select the best image from the results list by comparing the results to each other [Cunningham and Masoodian, 2006] and suitable with the use of that image.

1.2 Research questions

Relevance failures in image search are due to causes such as poor initial classification, faulty query processing/ranking and also poor querying by the user. In this thesis, we are focusing on querying by the user. Specifically, we address the following research questions:

1. **How do visual information needs affect the criteria users apply when select-**

ing relevant images?

Like any information retrieval system, image retrieval systems are designed to provide users with relevant images. A good system should retrieve relevant images based on users' information needs. This question examines how users make decisions about the relevance of images relative to their search context and needs. Users' visual information needs are connected to the use and purpose of that need. In a study of image searching behaviour, Conniss et al. [2000] presented seven classes of image use. From the list, we selected *illustration* as a purpose for the context in the visual information needs of the controlled user experiment.

In answering this research question, relevance is concerned with users' perception on the importance of a criterion in selecting images for the resolution of their needs and not with images retrieved by the system. By using Batley's [1988] classification of visual information needs to represent task types and a selected set of criteria identified by previous researchers, we are looking into the importance in usage of relevance criteria for different visual information needs.

2. **How do different subjective factors in image retrieval affect users' image searching behaviour?**

This question addresses the issue of subjective factors such as topic familiarity, task difficulty, and satisfaction and how they affect the way users search for images. In previous studies, different methods and measures — either objective (number of queries and time spent on task) or subjective (user ratings) — have been used to analyse

the effect of these factors. In the same user experiment, we used these measures to gauge different aspects of subjective factors that affect search behaviour. Subsequently establishing the relationship between ratings on perception of factors affecting search and objective measures from actual search behaviour.

3. **Can visual information needs be identified solely from user queries?**

   The success or failure of a digital image retrieval system ultimately depends on whether or not it can really satisfy user needs. As more users perform image searches, it is important to explore how users construct their queries and retrieve images based on their visual information needs. Using the same classification of visual information needs as in the user study, we conducted an analysis of image search logs to elicit and infer users' visual information need from the queries they submitted and how the queries were reformulated.

   Using two sets of proxy log data, we consider (i) a set of queries and (ii) a single query that the user submits as one visual information need. Based on the assumption that a user has only one visual information need during an image search session, for the set of queries, we infer a significant change in query terms as a change in the information need. Analysis and comparison of results between the log and user study data were conducted to acquire characteristics of users' image search behaviour in both natural and controlled settings.

1.3 Contributions

This thesis builds on previous research about users' visual information needs for searching images and contributes to research about image search behaviour. By studying real users and simulated needs for visual information needs, this thesis concludes that:

1. *Given similar image information needs, ordinary users make relevance judgements similar to specialized users (such as journalists, art historians and medical doctors) despite giving attention to different criteria of relevance.*

2. *Only certain measures of search behaviour used in text retrieval are applicable to image retrieval due to the difference in judging relevance of textual information and image.*

3. *Visual information needs can be better inferred when analysing a series of queries and their reformulation within a search session.*

This is achieved by first, demonstrating users' selection of relevance criteria for relevance judgements. Visual information needs influence how users judge image relevance and the criteria they use in making that judgement. They may use similar criteria for different images, however the weight of each criterion is different.

Unlike users in a specialised domain, who rely on textual information as a criteria for judgement, this thesis finds it is not as important to online image users. Getting the right facts is crucial in a specialised domain. However, images on the web belong to very many different users and therefore the terms/keywords/meta data associated with the images are more diverse. Furthermore, the results show that the criteria that were previously used in specialised domain are applicable for web image searchers.

Using measures from users' actual search behaviour, this thesis establishes the relationship with their perception on the effect of subjective factors. Of the subjective factors, familiarity is not shown to affect users' searching behaviour. However, their behaviour is affected by task type, task difficulty and satisfaction. The examination of users' time-stamped search interaction is consistent with previous text retrieval studies which showed that users have a longer completion time when presented with a difficult task and that they issued more queries for the task. Nevertheless, compared to text search, image search users view considerably more result pages.

Moreover, this thesis exhibits results on image information needs from using a user study and log analysis that complement each other. Results show that users search for specific and nameable images more often compared to abstract and subjective images. As for query behaviour, this thesis demonstrates how users' approaches in querying are consistent in both natural (proxy log) and controlled (user study) settings.

It is plausible to identify what users were searching for from just a single query. This is particularly applicable for *Specific* needs that have no ambiguity and some cases of *Nameable* needs that can be expressed in keywords. However, inferring their information need is not as straightforward. By analysing sessions with at least two queries, this thesis shows that users' visual information needs can presumably be determined from their pattern of query reformulations.

1.4 Overview of thesis

In addressing the research questions raised in Section 1.2, this thesis is divided into seven chapters and organised as follows.

In Chapter 2, we set the scene for addressing all research questions by reviewing related and state of the art research on image search and retrieval. The review starts with approaches for image searching and is followed by visual information needs and how these needs are expressed using queries. Then, the chapter continues with a review of user search behaviour. Issues of web search and image search in particular are examined, giving attention to factors that affect search behaviour — task type, relevance, familiarity, difficulty and satisfaction.

In Chapter 3, we discuss the methodology used to answer the research questions. We justify conducting a user study to investigate users' image relevance judgements and the effects of subjective factors and while performing a search; and search log analysis to identify user information needs. In contrast to previous research, which has either analysed logs or conducted user studies, both log analysis and user studies were conducted from the same population, to complement each other.

In Chapter 4, we address the first research question. We provide details about the design of our user study, selection of relevance criteria, methods used in collecting data and study procedures. Results of users image relevance judgements were analysed in regards to their visual information needs (task type). Using a selected list of relevance criteria, the difference and importance in selection of relevance criteria for different visual information needs were analysed.

In Chapter 5, we discuss the second research question. Using data collected from the same

user experiment, we examine how users' perceptions of topic familiarity, task difficulty and satisfaction affect their image searching behaviour. Data from questionnaires and recording of observations were analysed. The observations were time stamped prior to analysis to derive objective measures. Using both objective and subjective measures, we discuss their correlation with regards to factors that affect users' searching behaviour.

In Chapter 6, we explore image queries from proxy logs to identify users' visual information needs. We investigate how these needs differ from each other by comparing the number of queries and terms used, and how the queries were modified. In addition, we compare the results between the proxy logs and user study after conducting similar methods of query and term analysis on the user study data.

In Chapter 7, we summarise the findings, conclude the present work and propose future work and open research questions based on the outcome and limitations of this research.

Chapter 2

Literature review

As the number of digital images increases enormously over the decade, users of various backgrounds are taking advantage in exploring new ways of accessing, sharing and manipulating these images. However, they soon realised that the process of locating a desired image in a large and varied collection can sometimes be frustrating. The problems of image retrieval are widely recognised, and a huge range of solutions have been proposed.

The chapter is organised as follows: approaches for image searching (Section 2.1), followed by image information needs and how these needs are expressed using queries (Section 2.2) and user search behaviour and factors that have an effect on search behaviour (Section 2.3). Section 2.4, summarises the chapter.

2.1 Image search and retrieval

A very large number of digital images have been made available and accessible due to the prevalence of digital imaging technology as well as the growth of the Internet. The increas-

ing volume of images can make the task of searching and retrieving images overwhelming for users.

2.1.1 Text-based image retrieval

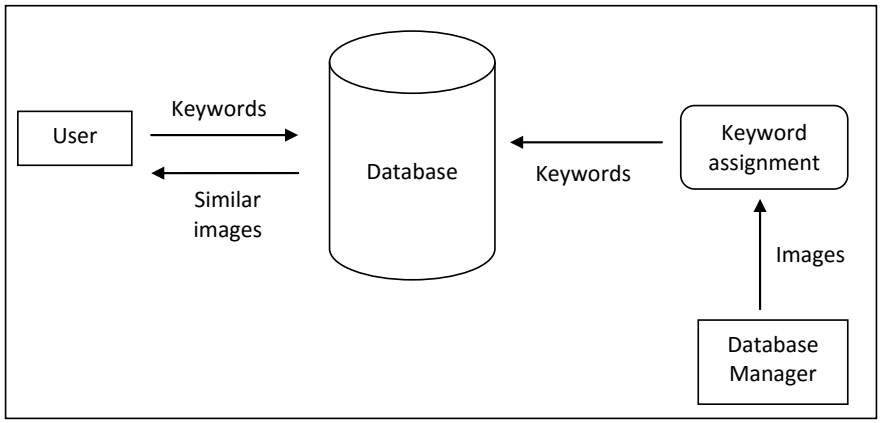

Figure 2.1: General framework of text-based image retrieval

In the late 1970s, image retrieval started by using a text-based database management system (DBMS). Prior to storing the images in the database, images were manually examined and annotated using keywords. This process is used to describe both image content and other metadata of the image such as image file name, image format, image size and image dimensions. The query process begins when the image retrieval system accepts a single or multiple keywords from the user. The keywords function as a search criterion whereby it is matched with the contents of the database to retrieve images associated with the same keywords, as shown in Figure 2.1.

However, there exist two major drawbacks of text-based image retrieval. First, is the

impractical amount of manual annotation to be done. Previously, annotation was performed by a small group of people. However, the last few years has seen the involvement of online users in tagging and labeling of images through the use of clickthrough data. Their involvement can be either explicit or implicit. Use of clickthrough data were found to be a workable alternative to labeling images [Ashman et al., 2009] and a reliable measure of relevance judgement [Smith et al., 2012]. Secondly, which is more crucial, is the rich content of the image themselves. Utilising text to describe the content of images is insufficient as image content is much richer than what can be expressed by a set of keywords. There is also an issue with subjectivity of human perception whereby people may perceive the same image differently [Rui et al., 1999].

To overcome these drawbacks, content-based image retrieval was proposed, whereby, images would be indexed by their own visual content, such as color and texture rather than keywords. Comprehensive surveys of early text-based image retrieval methods have been undertaken [Tamura and Yokoya, 1984; Chang and Hsu, 1992].

2.1.2 Content-based image retrieval

When referring to an image, what comes to mind is the phrase "a picture is worth a thousand words". But then, not everything in a picture or image can be described in text and not everything is described by text. Content-based image retrieval (CBIR), is a technology that helps to organise digital image archives by their visual content. A general framework of content based image retrieval is shown in Figure 2.2. The query process begins when a user submits an image example as a query to the image retrieval system. Feature extraction

is performed to extract a single or multiple low-level features such as colour, texture and shape. Similarity matching is then performed between the query image and images within the database to retrieve similar matching images.

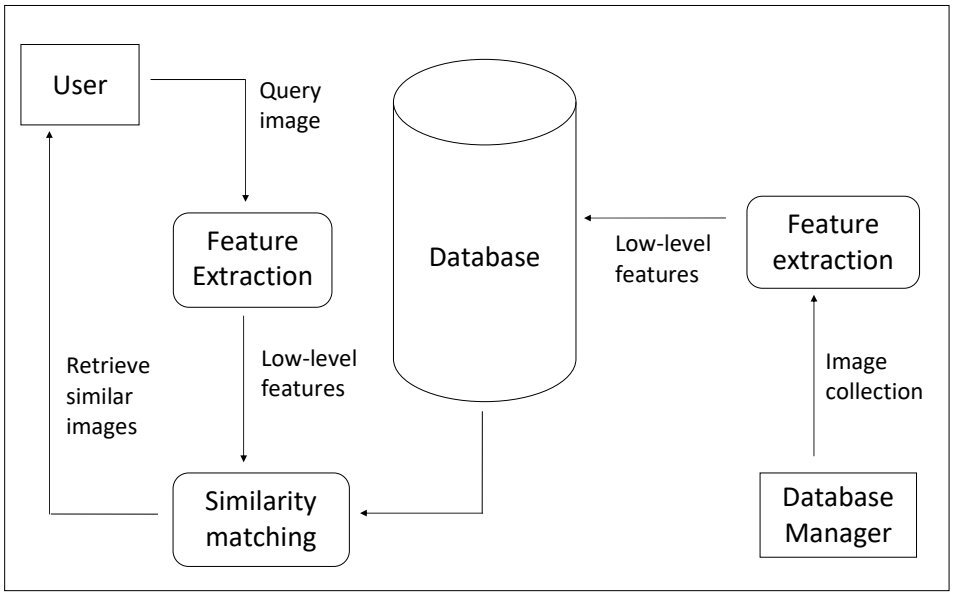

Figure 2.2: General framework of content-based image retrieval

CBIR draws many of its methods from the field of image processing and computer vision, and is regarded by some as a subset of that field. The fundamental difference of CBIR with these fields is that it focuses on retrieving images with specific features within a sizable collection. Nonetheless, it is far from perfect and, similar to keyword searching, the problems are due to the reliance on visual similarity for judging semantic similarity that may be problematic due to the *semantic gap* [Smeulders et al., 2000].

One of the key issues with any kind of image processing is the need to extract useful

information from the raw data (such as recognising the presence of particular shapes or textures) before any kind of reasoning about the images contents is possible. *Content-based* means that the content of the images themselves will be extracted for use in searching. Usually, a user would present an example image of what they are currently looking for. The system then performs feature extraction to extract low-level features within the image. Similarity matching will be carried out between the extracted low-level features with features within the database in order to find similar images.

Compared to text retrieval system, human interaction is much more vital in content-based retrieval system [Liu et al., 2007]. Humans tend to use high-level features (concepts), as expressed through keywords and text descriptors, to interpret images and measure their similarity. But, the features automatically extracted using computer vision techniques are mostly low-level features such as colour, texture, shape and spatial layout. Many algorithms have been designed to describe the various low-level features, yet they are unable to sufficiently model the rich content of an image [Mojsilovic and Rogowitz, 2001].

Based on the results of extensive experiments on CBIR systems, Zhou and Huang [2000] conclude that low-level features often fail to describe the high-level semantic concepts in the user's mind. Low-level features alone cannot consistently describe high-level semantic concepts. Nevertheless, using high-level annotation also poses problems as it depends upon the consistency of the annotation, and the consistency between the user and the annotation, and even the consistency among different users. The effect of semantics gap is avoided by using relevance feedback technique. Comprehensive surveys on the topic of CBIR have been published [Rui et al., 1999; Smeulders et al., 2000; Datta et al., 2008].

2.1.3 Image retrieval on the web

The web and search engines have revolutionised the way people gather information. Image search used to be done by professionals on specialised collections. An image search engine enables everyday online users to search for images. Current image search engines on the web rely on the keywords surrounding the images. Examples of popular web image search engines include Google Images, Bing Images and Yahoo Image Search. This method of searching is easy for the user, unfortunately, does not always guarantee desirable search results. This is mostly due to the ambiguous or irrelevant keywords used to describe the image.

Alternatively, there exist methods for content based image retrieval that require a user to submit a query image, and return images that are similar in content. The Google image search engine (https://images.google.com/)[1] for example, provides users four ways to search by image — 1) drag and drop, 2) upload an image, 3) copy and paste the URL of an image and 4) right-click and image on the web. It is useful in performing reverse image search, whereby using pictures as a query to find related images and other specific information from around the web [Klosowski, 2013].

2.2 Visual information needs

We would like to think that an image is always searched for a purpose; it can be either work related or for leisure. Therefore, an image information need may be a vague idea of what the user is searching for. A query is the exact search whereby they formulate terms and words into a precise search query that finds images that suit their needs. The ability for users

[1]From here on the brand name "Google Images" will be used to refer to this.

to express their information needs accurately and easily is crucial in any retrieval system. Image retrieval is no exception to this. Therefore, a good image retrieval system is measured by its intrinsic ability to retrieve results that suit the information needs of the user based on the query entered by the user.

An information need is always associated with a particular task that is translated into words and entered as a query in a search engine. In textual retrieval, Broder [2002] states that *need behind the query* are often not informational in nature and classify them into three classes — Navigational, Informational and Transactional. However, in image retrieval, the need is visual in nature and can be classified into various schemes and methods that we describe in the following section.

2.2.1 Categories of visual information needs

Categorisation or classification of image descriptions was first described by Panofsky [1955]. Since then, various image classification has been developed; either based on image themselves [Shatford, 1986], description of images [Hollink et al., 2004] or to be used for the purpose of image indexing [Jaimes and Chang, 1999]. Nevertheless, these classifications were based on perceived image content.

Shatford's [1986] image analysis was based on the perceived content of images searched by users as follows:

- **Specific**: Images referring to individually named person, group, thing, event, location, or action.
- **General**: Images referring to generic person, group, thing, event, place, condition,

or action.

- **Subjective**: Images that bear or evoke emotional or abstract concepts.

Whereas Batley [1988, p. 374] classifies visual information needs based on how the need is expressed as follows:

- **Specific**: "can be expressed in key words; can be expressed in a precise search statement; have no ambiguity; and deal with the concrete."
- **General/Nameable**: "can be expressed in key words; may result in unmanageably high recall (too many relevant items); and often have to be made more specific."
- **General/Abstract**: "are difficult to express in key words; may involve abstract concepts rather than concrete objects; and can be expressed verbally but not in a precise search statement."
- **General/Subjective**: "are difficult to express verbally; deal with emotional responses to a stimulus; cannot be expressed in a search statement; and depend upon characteristics of a scene as interpreted by individuals."

Studies that have used either of these two classifications include those by Choi and Rasmussen [2003], Hung et al. [2005], Cunningham and Masoodian [2006], Klavans et al. [2014], Park et al. [2015] and Göker et al. [2016].

2.2.2 Expressing needs through queries

Images are used extensively in many aspects of professional life, as well as emphasising that they can be used in a variety of different ways. However, little systematic research on image

use has yet been published. Most reported research has focused either on specific collections or on specific user types or populations. Examples include: newspaper image archives [Ornager, 1996; 1997; Markkula and Sormunen, 2000; Westman and Oittinen, 2006; McCay-Peet and Toms, 2009]; picture archives [Enser and McGregor, 1993]; medical archives [Keister, 1994; Sedghi et al., 2008]; other historical libraries and archives [Hastings, 1995; Armitage and Enser, 1997; Chen, 2001; McCay-Peet and Toms, 2009]; and creative professionals [Konkova et al., 2015]. Lately the area has seen work covering normal everyday users [Goodrum and Spink, 2001; Cunningham et al., 2004; Pu, 2005; Cunningham and Masoodian, 2006; Chew et al., 2010; Yoon, 2011; Yoon and Chung, 2011; Huang and Kelly, 2013; Park et al., 2015]. Although most of the users in these studies were students, it gives some insight into the needs of everyday users for images.

Image archives

Early works in query expression were mostly on image archives by professional users. Enser and McGregor [1993] categorised queries put to a large picture archive into those that could be satisfied by a picture of a unique person, object or event and those that could not. Uniquely identified subjects dominated the requests. Both categories — 'unique' and 'non-unique' — were subject to refinement in terms of time, location, action, event or technical specification. This work was further extended by Armitage and Enser [1997] in a study of seven picture libraries that sought to develop a general-purpose categorisation of user requests for still and moving visual images. An initial analysis led to the identification of four major types of query: image content; identification/attribution/provenance checking;

accessibility of image/artist of work; and miscellaneous. They analysed the image content queries in more depth, categorising them by named artist, known items, unique subject and non-unique subject.

Keister [1994] described the development of an automated still picture retrieval system at the National Library of Medicine (NLM). She analysed queries over a one-year period that showed that users did not ask for pictures in a consistent manner. She found that users who were picture professionals thought visually and used art and graphics jargon. Health professionals asked for images using subject oriected queries, such as "Do you have pictures of cholera?". She found that users use words to build a visual construct for an image that they know exists or one that they imagine would satisfy an information need.

Hastings [1995] investigated how art historians searched photographic and digital art images. She classified the major classes of queries: identification, subject, text, style, artist, category, compare and colour. Access points and computer manipulation functions used within the database varied with the level of complexity of the queries, ranging from the least complex queries — 'who', 'where' and 'when'; to the most complex queries — 'meaning', 'subject' and 'why'. The highest-level queries could often not be answered from the images or from available textual information and required access to full-text secondary subject resources.

Markkula and Sormunen [2000] conducted a field study on journalists using a digital newspaper photo archive. They found that journalist requests fell into four categories: concrete objects (that is, named persons, buildings or places); themes or abstractions interpretable from the photographs; background information on the image (such as documentary infor-

mation, specific news events and films and television programmes); and known photographs. Journalists preferred to search using proper name as it is easier than using words referring to object types or abstract themes.

Web image search

The nature of searches performed on specialised image collection may be very different from searching for images on a general-purpose image search engine[2]. Goodrum and Spink [2001] examined how people express their visual information needs in queries based on a study on the Excite Internet search engine. They provide information about image queries (number of queries and search terms per user), image search sessions (number of queries per user in a particular search session), and image terms (the rank and frequency distribution of search terms used). They found that a user on average commits 3.36 queries when searching for images on the Excite engine, each query containing an average of 3.74 terms. Most users frequently used words such as 'picture', 'photo' and 'image' to define their query as image information needs and adding refiners to further specify the visual request.

Using log data from VisionNEXT (a provider for image search services to the Chinese community) and Dreamer (once a web search engine in Taiwan), Pu [2005] examined the differences between web image and textual queries. They found that web users were likely to use short queries when searching for either textual or visual information. Results showed that image queries tend to have a higher level of specificity but more zero hits. However, image queries are unique and contained more refinements as compared to textual queries.

Cunningham and Masoodian [2006] studied college students image search behaviour by

[2]Henceforth by "search engine" we mean a general-purpose image search engine.

analyzing descriptions of their image needs. The results of their analysis showed that the most frequently searched image information needs were Specific, General/Nameable, General/Subjective and General/Abstract. They found that translating an image need into a textual query is challenging that resulted in more browsing rather than searching.

Yoon [2011] analyzed survey questionnaires of college student's image seeking behaviours and found that users when describing image needs used more generic terms. However, when searching using a query, they were more likely to use shorter queries and more specific terms. This is in accordance with the findings of Hollink et al. [2004], who suggested that people searching in a keyword-based search engine use more specific terms and less abstract and perceptual terms than people describing images in a more natural way.

Based on the assumption that a search engine has failed, Yoon and Chung [2011] analyzed questions from a social question-and-answer site. Their analysis showed that compared to queries submitted to a search engine, users ask questions that contains more diverse facets of image needs such as context of image needs, image attributes, and associated information related to the image.

Image use

The way users search for images depends also on the purpose for which they need the image [Batley, 1988; Fidel, 1997; Markkula and Sormunen, 2000; McCay-Peet and Toms, 2009]. Using an image in a context different from its original intended, may influence the terms chosen when describing a query. Conniss et al. [2000] have presented seven classes of image use:

1. **Illustration**: Images are used in conjunction with some accompanying media, as a means of representing what is being referred to.

2. **Information processing**: The use of the data contained within the image is of primary importance.

3. **Information dissemination**: The image is a stand-alone piece of information transmitted to someone else.

4. **Learning**: Gaining knowledge from the image content.

5. **Generation of ideas**: Images are used to provide inspiration or provoke thought patterns.

6. **Aesthetic value**: Images are simply required for decorations.

7. **Emotive/persuasive**: Images are used to stimulate emotions or convey a particular message.

Studies on visual information needs that has focused on the use of images as an illustration includes Markkula and Sormunen [2000]; Hung et al. [2005]; Fukumoto [2006]; McCay-Peet and Toms [2009].

2.2.3 Query formulation behaviour

The goal of a search is to find information that satisfies an information need. A search begins when a user enters a query and the system or search engine then returns a list of matching results. If the results are unsatisfactory, users often submit a new query or reformulate their

previous query. The retrieval is an interactive and iterative process. The process of modifying queries by a user is called query reformulations or query refinements.

In studying query reformulation behaviour, researchers have developed several reformulation classifications [Lau and Horvitz, 1999; Huang and Efthimiadis, 2009; Liu et al., 2010a]. Using an Excite server log, Lau and Horvitz [1999] created query reformulation classification to classify information goals according to identified ontologies. The four main classifications include New, Generalisation, Specialisation and Reformulation. This taxonomy was also used by He et al. [2002], Jansen et al. [2007] and Jansen et al. [2009] to automatically detect search boundaries within search logs. Later, Huang and Efthimiadis [2009], developed an extensive classification whereby they identified another ten reformulations in addition to the three classifications (except New) identified by Lau and Horvitz [1999]. In analyzing query reformulation for different task types, Liu et al. [2010a] extended Lau and Horvitz [1999]'s classification by considering the Repeat reformulation, where the query contains exactly the same terms but the format of these terms may be different.

Various researchers have studied query reformulation behaviour [Bruza and Dennis, 1997; Jansen and Spink, 2003; Mat-Hassan and Levene, 2005; Rieh and Xie, 2006; Jansen et al., 2007; Yoon and Chung, 2009; Jansen et al., 2009; Choi, 2010b; 2013]. Findings from these studies show that most online searchers rarely reformulate their query [Bruza and Dennis, 1997; Jansen and Spink, 2003; Mat-Hassan and Levene, 2005; Yoon and Chung, 2009]. Bruza and Dennis [1997] conducted an analysis in the occurence of query transformation in a search log and found that over 40% of queries were repeated queries and over 25% of queries were modified by either adding or deleting terms.

Jansen and Spink [2003] examined over one billion records of the FAST search engine to investigate user retrieval and viewing behaviour. Their analysis revealed that 53% of the users entered one query and about 54% of the users viewed only one page of results. Using log analysis, Mat-Hassan and Levene [2005] studied user clickthroughs to better understand search and nagivation behaviour. They found that more than 80% of users did not reformulate their query. Similarly, Yoon and Chung [2009] analysed the Excite 2001 Web search log for image query reformulation behaviour and found that 74% of the sessions included only initial queries.

The two most frequently used modification types are New query [Lau and Horvitz, 1999; Tseng et al., 2009] and Reformulation [Jansen et al., 2007; 2009; Choi, 2013]. The results surfaced various different strategies used in reformulating queries; such as, high occurences of query reformulation after Generalisation (17%) and Specialisation (32%) [Jansen et al., 2007]. Furthermore, users tend to replace query terms compared to adding or deleting [Tseng et al., 2009] and after submitting initial query, users tend to use Reformulation, Assistance and Specialisation [Jansen et al., 2009]. However, certain reformulations are better in improving users' current results, while other reformulations work best when the search is unsuccessful [Huang and Efthimiadis, 2009].

2.3 Users' image search behaviour

Image searching, similar to textual information searching, is a living, evolving process of discovery — a conversation between a user and the web site. Unfortunately, this conversation is often filled with miscommunication. A typical keyword-based image search process tends

to be straightforward — the user types in a search and the search engine gives back matching results. Even though the process is correct, as a system-centred approach, it fails to take into consideration other factors such as user needs and behaviour.

Most studies on user search behaviour have been on textual information search. The classic model of information search was reformed to take into consideration that even experts sometimes have difficulty creating queries for their information needs [Belkin, 1980]. Users were not at fault when or if the system fails to retrieve the required information. This has led to studies on image search behaviour and sparked developments in image search interfaces. EGO [Urban and Jose, 2006] and FISH [Tandon et al., 2008] are two examples of user-centred image search systems.

A considerable amount of literature has been published on image search behaviour by understanding the types of search that users perform, and how they perform these searches. Similar to text, image search behaviour studies fall into one or more of the following three categories [Jansen and Spink, 2006]: (1) those that primarily use transaction-log analysis [Goodrum and Spink, 2001; Jörgensen and Jörgensen, 2005; Chung and Yoon, 2010; Hollink et al., 2011; Palotti et al., 2016], (2) those that incorporate users in a laboratory survey or other experimental setting [Markkula and Sormunen, 2000; Goodrum, 2005; Cunningham and Masoodian, 2006; Sedghi et al., 2008; Yoon, 2011; Huang and Kelly, 2013; Lu and Jia, 2014; Konkova et al., 2015; Park et al., 2015; Göker et al., 2016], and (3) those that examine issues related to or affecting web searching behaviour [Choi and Rasmussen, 2003; Hung et al., 2005; Choi, 2010a; 2013; Khashman et al., 2013]. This thesis considers all three categories. The following section discusses the factor of interest that affects web image searching behaviour.

2.3.1 Factors that affect search behaviour

Previous studies [Markkula and Sormunen, 2000; Choi and Rasmussen, 2003; Hung et al., 2005; Cunningham and Masoodian, 2006; Choi, 2010a] found user's searching behaviour is affected by several factors including task type, relevance judgement criteria, task difficulty, topic familiarity and satisfaction. Most of these factors have been extensively studied in textual information search behaviour. The following sections, review these factors for both text and image search behaviours.

Task type

In textual information retrieval, tasks are classified as closed or open-ended. Closed tasks require specified factual answers, while an open-ended task, do not. Both tasks, however, can have a single or multiple answers. Many studies have classified tasks along other dimensions and examined their effects on users search behaviours.

For example, Kellar et al. [2007] looked at four types of tasks: fact-finding, information gathering, browsing, and transactions, and examined how users navigated and interacted with their web browser. They found that information gathering was the most complex task, in that participants spent more time completing it, viewed more pages, and used the web browser functions most heavily. Similarly, Gwizdka and Lopatovska [2009] used two types of task, fact finding and information gathering, to investigate the role of subjective factors in the information search process.

Aula et al. [2010] conducted user studies using only closed informational tasks (which required a single, unambiguous answer) and found that users wrote longer queries, used

more queries per session and spent more time on the results page when performing difficult search tasks. Liu et al. [2010c] used both closed and open-ended tasks and their results were consistent with findings of Aula et al. [2010]. Regardless of the type used, most tasks in information search were based on the three intent classes identified by Broder [2002].

However, in image retrieval, tasks are classified differently [Shatford, 1986; Batley, 1988; Jaimes and Chang, 1999; Hollink et al., 2004]. Fidel [1997] states that there are in general two poles for retrieval tasks:

- the Data Pole — focuses on the retrieval of images for the information that the image includes.
- the Objects Pole — focuses on the retrieval of the image as an object.

Searching for images is ambiguous as compared to searching for documents that match a particular textual information need. An image can be searched either for the information it contains or as an object itself. Nevertheless, image task types were created based on classifications of visual information needs.

Markkula and Sormunen [2000] examine the search behaviour of journalists by addressing different types of common search topics in journalistic illustration tasks. They state that general search topics resulted in multiple sessions, various queries, and heavy browsing. In contrast, specific needs resulted in only a few queries and browsing sessions. They observed that the test population also tried to convert general needs in photo retrieval into more concrete queries. Choi and Rasmussen [2003] studied user queries to detect image attributes that are important for retrieval and to identify characteristics of user queries for images.

Based on a scheme developed by Batley [1988] the search needs of a test population were classified. Their results showed that most search needs are general or nameable needs and specific needs.

Using the same scheme, Cunningham and Masoodian [2006] analyzed descriptions of casual image information needs. Their results showed that Specific needs were the most dominant compared to other needs. Differences in results between the two studies were presumably due to the nature of the image search task. Users in the study by Choi and Rasmussen [2003] were focused on American history, whereas the users in the study by Cunningham and Masoodian [2006] were not focused on a particular topic. Hung et al. [2005] created three search tasks based on Shatford's classification to investigate users' image relevance judgements. Their results showed that when judging relevance for general and subjective images, users relied more on personal feelings and textual information. Conversely, users depended on features of objects in photos for judging relevant specific images.

Nevertheless, there are some image search studies that use other classifcation for their image search tasks. For example, Tseng and Tjondronegoro [2010] used exploratory and retrieval tasks in three web image domains to investigate users' query reformulation behaviour. Their results showed that an exploratory type of task involved more browsing, while users introduced more unrelated terms when reformulating their queries.

Relevance judgement criteria

In information retrieval, relevance is not stated, but implied [Saracevic, 2007]. It is expressed through the use of relevance criteria in evaluating whether to obtain and use or discard

information. Information may assume different meanings to different people and assume different meanings to the same person at different times. However, Barry and Schamber [1998, p. 219] studying both text and multimedia retrieval give evidence suggesting that "a finite range of (relevance) criteria exists and that these criteria are applied consistently across types of information users, problem situations, and source environments".

Empirical studies have been conducted in which image relevance criteria were either elicited directly from users or selected from previous studies. Saracevic [2007] identified these studies as "clues to research". The clues represent artifacts of the search process and the criteria used by the subjects are the attributes that describe these clues. These studies investigated a wide range of criteria and came up with different lists and classifications. For example:

- *accuracy, depth and scope, clarity, recency* [Barry, 1994];
- *presentation quality, currency, reliability, accuracy* [Schamber, 1991].
- *authority, accessibility, interesting, topicality, quality* [Hirsh, 1999];

Although each of the studies were widely varied, they made similar observations about the relevance criteria, which can be generalised as follows —

- searchers use the same criteria but assign different weights to these criteria;
- the importance of these criteria changes with task, progress in task over time, and varies by some categorisation or class of user; and
- criteria may interact with each other [Saracevic, 2007].

Research studies that explore users' relevance judgement on image retrieval are conducted by applying specific information needs and then identifying relevance criteria utilised by the users while making relevance inference. The focus is on criteria users apply while thinking of what is or is not relevant, and to what degree it may be relevant. Hirsh [1999] investigated children's relevance criteria as they performed text searches for a class assignment and found that most of the children based their relevance of textual information on topicality.

Choi and Rasmussen [2002] investigated criteria applied by users when judging image relevance. The criteria were selected from previous information retrieval studies as they showed a considerable degree of overlap despite differences in terminology. They found that topicality was the most important criteria and that the importance of the criteria changes during searching. Hung et al. [2005] also elicited relevance criteria from users performing three image search tasks. From the study, they identified three core criteria (typicality, emotion and aesthetic) used across the three searches and differences in making judgements for each search.

The results of these studies show that similar to findings by Xu and Chen [2006], users apply other criteria beyond topicality in making relevance judgements and that an overlap exists between criteria elicited directly from users and criteria that have been previously suggested in the literature. But most importantly, they confirm that relevance judgements involve users' perception of information, at a certain point in time, and based on their information need situation [Borlund, 2003].

Buerger [2010] identified relevance criteria in image search by building a conceptual model on clusters of factors used to assess relevance of a media object. His findings, similar to

those of Westman and Oittinen [2006], showed relevance assessment were most affected by three factors: Abstract and Affective Features, Information and Topicality, and Visual and Compositional Features.

Task difficulty

Task difficulty has been defined as a subjective perception assessed by task doers [Li and Belkin, 2008] and identified as a factor that influences users' search behaviour and performance. Researchers have used various approaches in evaluating task difficulty, such as varying task types [Liu et al., 2010c; Qu et al., 2010; Liu et al., 2010b; 2011], multiple levels of difficulties [Aula et al., 2010; Liu et al., 2010c] and assessing difficulty before and after searching [Liu et al., 2011]. In most previous research, difficulty is related to "the amount of skill and effort required to complete a task, and the likelihood of success" [Wildemuth et al., 2014, p. 1129].

Previous user studies have compared search strategies between novices and experts [Hölscher and Strube, 2000]. However, even highly skilled searchers sometimes struggle to find what they are seeking. Byström [2002] found that when searching on a difficult task, more queries were issued.

Two studies by Aula et al. [2010] identified changes in user behaviour; when given a difficult search task, users would reformulate their queries into questions, increase usage of advanced operators, and view result pages for longer. Liu et al. [2010c] applied behavioural signals to predict when users are facing difficulty during searching; when performing a difficult or closed task, users have significantly longer dwell time.

Cunningham and Masoodian [2006] conducted a study with the aim to better understand users' everyday image information seeking behaviour by analysing descriptions of their image related information needs. They showed that people tend to struggle when searching for General/Nameable or General/Abstract types of images and that users preferred to browse rather than search for images.

Topic familiarity

Users of textual information retrieval systems employ a variety of strategies when searching for information. One factor that can directly influence how searchers go about their information finding task is the level of familiarity with a search topic. It is a factor that has been found to affect users' search behaviours and performance.

Kelly and Cool [2002] investigated the relationship between topic familiarity and two types of search behaviours — reading time and search efficacy — and showed that as familiarity with a topic increases, searching efficiency increases and reading time decreases. Wen et al. [2006] examined the effect of users' topic familiarity on the use of resources and relevance criteria by searchers; users tended to use more generic and fewer specialised resources when searching for an unfamiliar topic. Qu et al. [2010] investigated the effect of task type and topic familiarity on users' search behaviour; the less people knew about the topic, the more effort (longer time and more queries) they would require in completing the search task.

A study was conducted by Choi [2010a] to examine effects of topic familiarity of users performing three self identified image search task on the web. Her findings showed that topic familiarity affects users search tactics and how they assess relevance. Topic familiarity

seems to influence users' level of satisfaction, usefulness and confidence in assessing image relevance. Subsequent work by Choi [2010b] showed that familiarity with a search topic does not influence query reformulation. Smith et al. [2012] consider topic familiarity using clickthrough as a proxy for relevance. Clickthrough from users who are able to visually identify a topic would be included in high-relevance results sets and vice versa. Their study shows that users with high topic familiarity have higher chances of successfully completing a search.

Satisfaction

Tessier et al. [1977, p. 383] defined satisfaction as "ultimately a state experienced inside the users head". Measuring satisfaction can be challenging as users may still express satisfaction with their results and the overall performance of a system even when in actuality, the results showed to them were poor [Hildreth, 2001]

Previous research on satisfaction mostly focused on system evaluation and effectiveness [Al-Maskari et al., 2006; Al-Maskari and Sanderson, 2010]. As there were no comprehensive studies on factors influencing satisfaction, Al-Maskari and Sanderson [2010] investigated the relationship between user satisfaction and the following four factors: system effectiveness, user effectiveness, user effort, and user characteristics. Their results showed that system effectiveness has a significant influence on satisfaction, and that having to put in more effort and time in searching will decrease ones satisfaction.

Gwizdka and Lopatovska [2009] investigated the role of satisfaction as a subjective factor in textual information search. Collecting subjective user feelings and perceptions using ques-

tionnaires before and after each search task, they found, for example, that positive emotions before a search were linked with positive mood after searching but not necessarily better search effectiveness.

Choi [2010a] examined students' image searching processes on the web to identify factors that influence the behaviours. Results for one of the factors, searching expertise, showed that searchers with lower expertise have the tendency to rate lower satisfaction with their search results. In another study, Choi [2010b] investigated the extent of users' satisfaction with their search results and found that those who reformulate the queries less frequently were more satisfied with search results presented to them. A possible reason could be because the search was a natural search task that relates to the user and not an artificially created task.

2.4 Summary

In this chapter, we reviewed literatures related to our work. Approaches for image search and retrieval, categorisation of visual information need, query expression and reformulation behaviour, and factors that affect users' search behaviour were described. Discussions in this chapter provided us with a selection of ideas in answering the research questions. The coming chapters will include some of the literature discussed here and the approach in relation to the research questions. In particular, web image search, categorisation of visual information needs and use of images. Users' visual information needs and their intended use of the image influence their querying and searching behaviour. The effect of subjective factors such as relevance criteria, difficulty, familiarity and satisfaction are measured through analysis of their query reformulation, perception of the factors and actual searching behaviour.

Chapter 3

Methodology

In this research, effects of relevance criteria and subjective factors on users' image searching behaviour is studied by two major methods, namely a user study and proxy log analysis. The two methods complement each other allowing some insights to be confirmed.

The chapter is organised as follows: approaches for analysing search behaviour (Section 3.1), followed by details of the user study (Section 3.2), and the proxy log analysis (Section 3.3). Finally, Section 3.4 provides a summary of the methodology.

3.1 Evaluation approaches

Most researchers of search behaviour have either conducted a user study or performed a transaction log analysis. The following researchers have conducted user studies of image search behaviour — Hirsh [1999], Markkula and Sormunen [2000], Chen [2001], Choi and Rasmussen [2002], Hung et al. [2005], Westman and Oittinen [2006], McCay-Peet and Toms [2009], Choi [2010a], Tseng and Tjondronegoro [2010], Buerger [2010], Yoon [2011] and Choi

[2013]. In contrast, researchers of image search behaviour who have used log analysis include Jansen et al. [2000a], Goodrum and Spink [2001], Pu [2005], Jansen [2008], Pu [2008], Tseng et al. [2009], Yoon and Chung [2009], and Hollink et al. [2011].

During an image search session, users make one or more queries to describe their image needs. These queries can be captured either as part of a controlled user study or by extracting them from web proxy logs. Analysis of queries have been used to identify different types of needs [Armitage and Enser, 1997; Choi and Rasmussen, 2003; Jansen et al., 2008; Huang and Kelly, 2013; Göker et al., 2016] and to better understand users' search behaviour [Choi and Rasmussen, 2002; Hung et al., 2005; Aula et al., 2010; Choi, 2010a; Tseng and Tjondronegoro, 2010].

User studies are widely used in research that focuses on understanding users' needs and search behaviour. A user study is conducted when a researcher wants to explore, describe or explain a particular phenomenon [Kelly, 2009]. In the area of information search behaviour, user studies are conducted in either natural or laboratory settings. The type of setting used is determined by what the research is trying to achieve.

User studies conducted in a natural setting means that researchers will observe users' behaviour as they go about performing their daily or routine search activities. Although the behaviour is more natural, comparison between participants is difficult as the researcher has little control over the setting. Conversely, in a laboratory, users are asked to perform search tasks that are controlled by the researchers. This type of user study is often used in identifying the effect of one or more variables. Nonetheless, its main drawback is that the behaviour is artificial, and does not represent real life and is less generalisable.

Search details on users of search engines are recorded in their logs. The details can be used to discover useful knowledge about users' behavioural patterns. Web usage mining [Srivastava et al., 2000] or web analytics [Jansen, 2009] are techniques used in discovering this knowledge within the logs. However, in the area of information retrieval, it is known as transaction log analysis. Jansen [2006] defines a transaction log as an electronic record of interactions that have occurred between a system and users of that system. The aim of analysing a transaction log is to gain a better understanding of an information retrieval system, its users and the interaction between users and the system.

Transaction log analysis methods include analysis of Web system logs, analysis of search engine logs and proxy logs. A key benefit of transaction log analysis is that the logs are routinely generated by information systems and servers. Another benefit of log analysis is that it enables researchers to study and track a system and its users over a long period of time. Furthermore, it provides the researchers with a discreet way to observe users and the system. In addition, the sample size is considerably larger than those obtained from a user study.

Nevertheless, it is not without limitations. Complete understanding of interactions is difficult, as information regarding the context of users' search goals and their thoughts are not explicit within an interaction. Moreover, identifying individual users solely based on IP address can be unreliable as it is as IP addresses are often dynamic and may represent multiple clients or multiple users on the same client. Privacy is an additional concern, and IP addresses are often required to be purged from such logs before they are made available for analysis. Furthermore, specifying user sessions is difficult. Each researcher must decide

how much time of inactivity denotes a new session, and whether a search session can span multiple days or take place from more than one computer. Users cannot be asked whether they found what they were searching for, making it difficult to determine a successful or unsuccessful use of the system. Despite these limitations, log analysis provides useful clues about search behaviour and their interactions with a search system.

The problem with past work on image search behaviour is only conducting either user studies or log analysis. Even if the two were conducted, the users were of different groups or population. Therefore, we carried out the research in two ways — a user study and proxy log analysis. Prior to the user study, we conducted a pilot study to explore and understand users' behaviour when performing image search. The goal is to identify criteria important to a user when they perform image search and their perception on factors that affect their search behaviour. Details of the pilot study are in Appendix A. Limitations identified in the pilot study were taken into consideration while designing the main user study. Log analysis was conducted to enable us to analyse users' natural image searching behaviour and perform comparison with data from the controlled user study.

3.2 User study

The first research question examines how users make decisions about the relevance of images relative to their search context and needs, while the second question addresses the issue of subjective factors such as topic familiarity, task difficulty, satisfaction and how they affect the way users search for images. Therefore, we employed a user study in answering both these research questions.

3.2.1 Data collection methods

Data collection is a vital part in any information retrieval evaluation. Researchers often use a mixture of different methods to gather data. These methods include logging, questionnaires, interviews, observations, think-aloud protocols, eye tracking and crowdsourcing. Table 3.1 shows the mixture of data collection methods used by previous researchers in user studies of image search behaviour.

Table 3.1: Data collection methods used in user studies in image search behaviour

Author(s)	Interview	Logging	Observation	Questionnaire	Think-aloud
Batley [1988]	✓			✓	
Hirsh [1999]	✓				
Markkula and Sormunen [2000]	✓		✓		
Chen [2001]	✓			✓	
Choi and Rasmussen [2002]	✓			✓	
Hung et al. [2005]	✓			✓	
Fukumoto [2006]			✓		
Westman and Oittinen [2006]	✓		✓	✓	
McCay-Peet and Toms [2009]	✓			✓	
Choi [2010a]	✓		✓	✓	
Tseng and Tjondronegoro [2010]		✓	✓	✓	✓
Buerger [2010]				✓	
Yoon [2011]				✓	
Choi [2013]				✓	✓

Interviews are often used as a technique to elicit answers to open questions. Interviews allow researchers to get more individualised responses from participants and to clarify the meanings of word or other ambiguities. However, in order to conduct the analysis, information gathered from interviews have to be transcribed.

Logging is one of the oldest and most common methods for collecting data in information retrieval evaluations. It is a useful method for capturing users' natural search behaviours as it records interaction between a system and users of that system. The term logging is differentiated from transaction log analysis. Logging refers to client-side logging whereby users' search interaction was logged on their local machine, for example by using a custom-built Web browser [Kellar et al., 2007] or an external program [Gwizdka and Lopatovska, 2009]. Data collected from logging is more comprehensive as opposed to transaction log analysis. However, it can be expensive and difficult to implement.

Observation can be conducted either in real-time or at play-back time. In real-time observation, a researcher sits close to or follows the subject and watches on while they perform searching activities. Conversely, a play-back time observation is conducted by using a video camera or screen capture software.

Questionnaires are the most frequently used method of collecting data as they allow for quick and direct capture of data based on the participants' responses. Questionnaires can consist of closed or open questions or a mixture of both and can be administered at different stages in the study. In closed questions, users are given a specific set of response, often resulting in quantitative data. However, an open question gives users the opportunity to express themselves freely, producing qualitative data.

Think-aloud protocols involve participants thinking and talking aloud while performing search tasks. Participants are required to verbalise anything and everything that they

looked at, thought of and felt during the search process. During a think-aloud process, the researcher needs to be objective and take notes of everything that the users are saying.

Eye tracking is an approach that captures human processing of visual stimuli. It records users' eye movements such as their gaze (where one is looking), fixation (area of focus) and scan path (motion of the eye) in real-time.

Crowdsourcing is a method of obtaining large scale needed services, ideas, or content by requesting contributions from a large group of people, especially from an online community. Estellés-Arolas and González-Ladrón-de Guevara [2012, p. 194] concluded that "the crowd refers to a group of individuals whose characteristics of number, heterogeneity and knowledge will be determined by the requirements of the crowdsourcing initiative". Studies in information retrieval have used crowdsourcing primarily to collect annotations. Annotations can be in the form of eliciting criteria, relevance assessment and snippet evaluation.

3.2.2 Design of user study

Figure 3.1 shows the overall structure of our user study. The user study was performed in a controlled setting whereby users were provided with a simulated work scenario of visual information needs. They would then submit the queries on Google Images and select relevant images for that scenario. We did not have any control over the results returned by Google Images. Nor, did we apply further control for any personalisation that maybe applied by Google Images. Due to this, possible variation in search results between participants may exist.

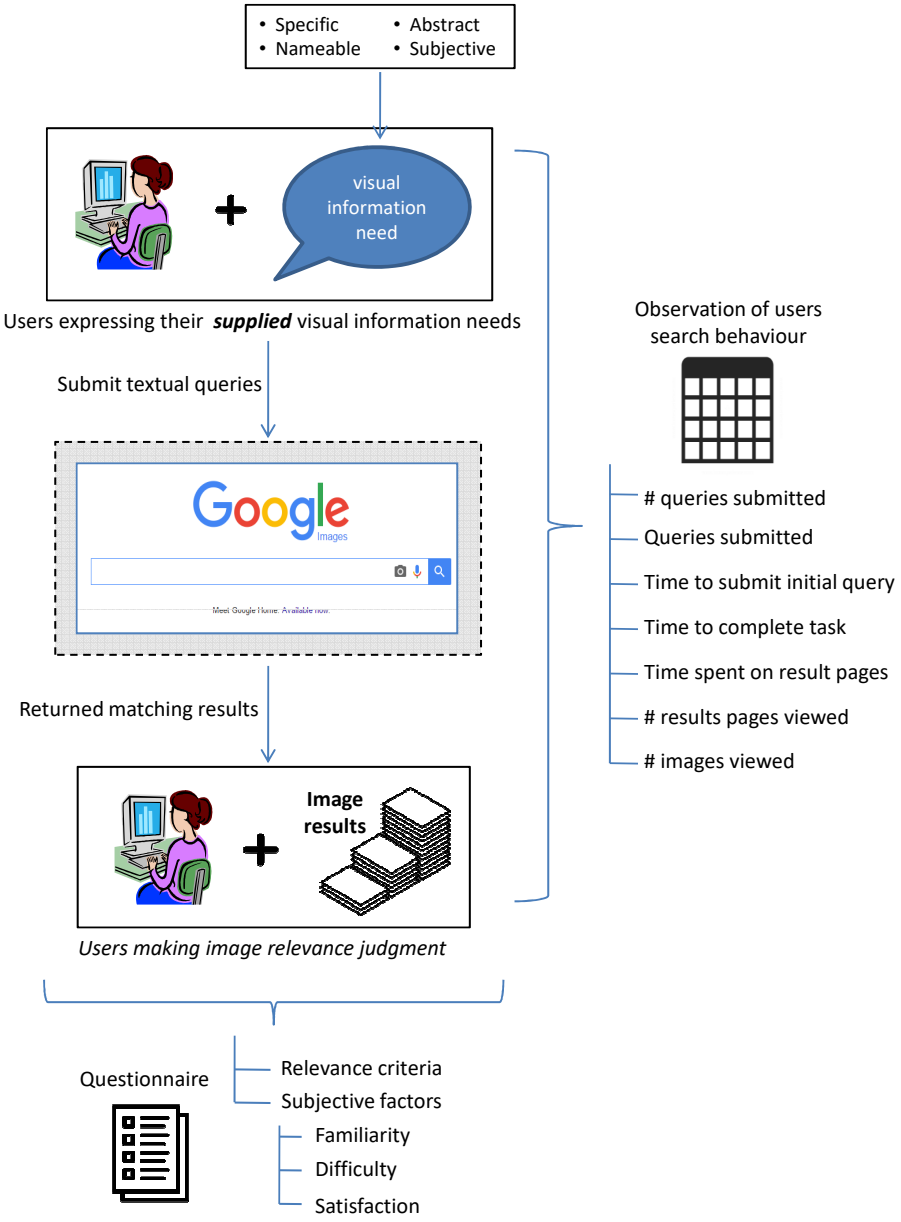

Figure 3.1: Overall structure of each task in user study

The user study collected two primary forms of data: 1) questionnaires that user's filled out about their searching backgrounds and response to the tasks they completed, and 2) recordings of users search sessions as they completed each task. Questionnaires were chosen as a primary survey instrument because they allow for quick and direct capture of data based on the subject's responses [Kelly et al., 2008]. In both the pilot and the main user study, the questionnaire was divided into two parts; demographic and post-task questionnaire. The demographic questionnaire was used to elicit background information about the participants including previous experience with image search (there is a slight risk this may bias their subsequent behaviour). The post-task questionnaire was used to gather feedback about the users' perception of relevance criteria and subjective factors that affect searching for each task they performed.

Both the demographic and post-task questionnaires consisted of closed questions, enabling us to perform statistical comparisons, with the exception of one open question at the end of the post-task questionnaire. The open questions provides participants with the discretion to further comment and elaborate on any other aspects of the search or questions that they have previously answered. Participants were required to answer the post-task questionnaire after completing each search task. Details and analysis of the questionnaires will be discussed in Chapters 4 and 5.

Although we could have conducted an interview in place of the post-task questionnaire, we decided not to. As each participant needed to perform four image search tasks, conducting and interview would unduly prolong the duration of experiment. In addition, we would need to be able to maintain consistency in the nature and order of the questions. Further-

more, transcribing would be laborious and users may not necessarily express different or new comments, but simply using different words [Kelly et al., 2008]. In addition, we did not use think-aloud protocol as to avoid distracting the participants during their search sessions.

Another option to be considered is crowdsourcing as we would be able the collect a lot of information in a significantly less time. However, we decided that it is not the most appropriate method; for example, in identifying criteria participants' find important when judging relevance of images that they have searched for. In this situation, we wanted participants to judge the selection of relevance criteria for images based on given information needs rather than asking for relevance judgements on given images. Quality control is also a major problem when using crowdsourcing. When making judgements, assessors tend to cheat to make more money with less effort; resulting in unreliable data [Zhu and Carterette, 2010].

In collecting data of search behaviour, we selected play-back observation in favour of either real-time observation or eye tracking. In a play-back observation via screen capture recordings, we capture what is happening on screen, so we have the luxury of time for analysis. Observation starts as soon as the participant begins their search task and ends upon the completion of that task. Moreover, we are able to manually time-stamp the recordings to indicate search and retrieval activities. These include the time when a participant submits an initial query, time spent on search results pages, time taken to complete a task, number of queries submitted, number of results pages viewed and number of images viewed (as shown in Figure 3.1). In addition to time-based measures, we observe participants searching tactics, for example, query creation and modification. Data obtained from observations were compared with questionnaire data to gauge the relationship between participants' perception

of performing image search with their actual search behaviour. Details and analysis of the observations are discussed in Chapter 5.

We did not use real-time observations or observe users performing real tasks because of time constraint, plus we have a specific type of task that we are studying. Likewise, we did not use eye tracking since we were more interested in how participants perform searches compared to where on the screen they looked at while searching.

3.2.3 Sample and population

All human studies involve a pool of human participants. In the user studies, we used convenience sampling as a method to recruit participants. The rationale for choosing students as the sampled population is that they regularly interact with the World Wide Web and information technologies and are thus a good potential sample of those who interact with images online. Similarly, with the log analysis, the sampled populations are students (and staff) as the logs used were extracted from the RMIT University's proxy servers. So, there may be a slight difference on the background of the sample.

Having an adequate sample size would allow one to accept or refute the hypothesis given the appropriate choice of statistical test. Larger sample sizes tend to give more reliable findings. It is desirable to recruit as many participants as possible for any user study, so as to increase the reliability of the results. However, a larger number of participants would mean longer data collection time, and difficulty finding the required number of participants.

In the pilot study (Appendix A), we used a small sample size. Although there were some interesting results, it was not considered reliable. Therefore, in order to instill confidence

in the findings of the main study, we used power analysis [Cohen, 1992] as a method to determine an appropriate sample size based on statistical tests that we intended to perform with the data we collected.

3.2.4 Constructing image information need

The most crucial aspects of information search and retrieval is the user's information need. This need forms the basis of their search activities and relevance judgements. Researchers [Vakkari and Hakala, 2000; Taylor et al., 2007] have shown that users often face difficulty when communicating their information needs and expressing them in words. These researchers also have shown that as people learn more about their information needs, their relevance assessment behaviours change.

Information needs can be characterised in terms of task and topic. These terms (information need, task and topic) are sometimes used interchangeably in information retrieval literature. Here, we define and distinguish amongst them to clarify what is being studied. A *task* represents the goal and purpose that a user wants to accomplish by searching. The *topic* represents the subject area that is the focus of the task. This combination of task and topic forms the *information need*. To understand users' image search behaviour, we created image information needs using a *work task scenario* [Borlund and Ingwersen, 1997]. The work task scenario acts as an (artificial) image information need that provides the user with context that facilitates searching.

In the pilot study, we used Shatford's [1986] image analysis as the basis for constructing the information needs. This groups images into three categories, namely Specific, General

and Subjective. However, for the main study, we adopted Batley's [1988] visual information needs — Specific, General/Nameable, General/Abstract and General/Subjective. The rationale behind the shift was because Shatford's categorisation focuses on searched content, whereas Batley [1988] views the information need at a higher level, encompassing expression of needs and not just image content. Moreover, Batley's categorisation was much clearer in distinguishing between abstract and subjective needs than Shatford's categorisation. Following this, throughout the thesis, Batley's visual information needs will be referred to as *Specific*, *Nameable*, *Abstract* and *Subjective*.

Within the information need, we define the image's intended use. Conniss et al. [2000] have identified seven classes of image use. The findings of Chung and Yoon [2011] showed that illustration was the most obvious use of images. For this reason, illustration was selected to be used in the user study. Therefore, each task was associated with one image information need and illustration as the intended use of that need. Once the tasks has been established (image information need with intended use), a survey for topics that matched the task were conducted (details can be referred to in Appendix B). Given a total of 40 topics, people were asked to match each of the topics with the most appropriate information need. The purpose of the survey was to have an indication of the topics classification and avoid bias towards our own understanding and interpretation. Based on the results from 15 people, four topics were selected for each task that mutually corresponds with both users and our classification.

3.3 Proxy log analysis

The third research question is on inferring visual information needs from queries. Therefore, we analysed image proxy logs to infer users' natural visual information needs from queries they submitted to the web image search engine. Proxy log analysis were used as it is "an unobtrusive method of collecting significant amounts of searching data on a sizable number of system users" [Jansen, 2006, p. 408]. Figure 3.2 shows an overview of the method for proxy log analysis. With the proxy log analysis, we were focusing on 1) users' visual information needs and 2) the expression and exchanges of communication (queries) between the user and the searching system in a natural setting. For example, a user may submit a query. The system may respond with a results page. The user almost never clicks on a URL, but on a result that they believe is relevant, and that is hyperlinked to a specific URL. Therefore, in using a log analysis, we were interested in the visual information need a user has and their interaction, specifically in terms of query submission and modification as a mechanical expression of underlying visual information needs or motivations and their changes.

The proxy log data in this research was extracted from RMIT University's proxy server with assistance of the university's Information Technology Services (ITS). The log data used were collected over a period of one month during semester time. Based on Figure 3.2, image search logs from the Bing and Google image search engines were analysed. In the user study, users were provided with artificial visual information needs, however, in the logs, visual information needs are unknown. In the logs, we did not have any control of what the users searched, the results returned by the image search engine or the results that they viewed. Prior to analysis, some pre-processing steps were performed (details provided

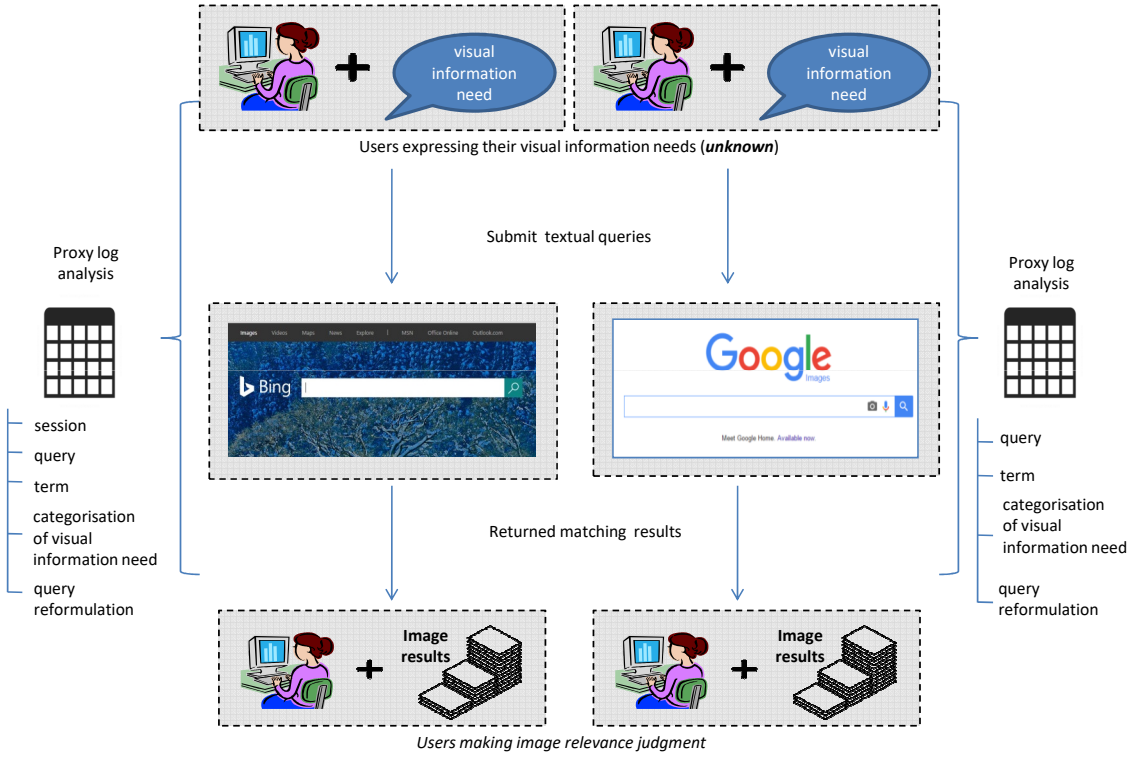

Figure 3.2: Structure of proxy log analysis

in Chapter 6). Next, user sessions and queries were identified within each session. The queries were then manually categorised according to Batley's [1988] visual information needs. Analysis performed on the proxy log data includes session, query and term level analysis, users' image information needs as expressed by their queries and users' query reformulation. Last, comparisons of findings from the user study with the proxy logs were performed. Details and results of this analysis are discussed in Chapter 6.

3.4 Summary

In this chapter, we presented methods used by past researchers of search behaviour. Details of the methodology used in investigating users' image searching behaviour were provided. In particular, the methodology was designed with the aims to: 1) analyse users' perception of relevance criteria for image information needs, 2) gauge factors affecting image search behaviour and 3) deduce image information needs from queries. Analyses were conducted in both a natural and a controlled setting. Multiple sources of data were used in ensuring the validity of results obtained and reliability of conclusions made.

Chapter 4

Image relevance judgements

According to Saracevic [2007], relevance is not stated, but implied. Relevance of an image is implied through the use of relevance criteria. However, relevance is subjective and dependent on various factors. Images convey multiple and different messages. Searching and judging an image as relevant depends on the user, their information needs and when the judgement is made. In this chapter, we address the first research question — How does visual information needs affect selection of criteria for image relevance judgements? We describe a user study conducted to identify the importance of relevance criteria based on visual information needs.

The chapter is organised as follows: details of the study design (Section 4.1), followed by selection of relevance criteria (Section 4.2), data collection method (Section 4.3) and study procedure (Section 4.4). Description of the participants and their topic selection in the main study (Section 4.5) is followed by results on users' image relevance judgements (Section 4.6). Section 4.7 summarises the chapter.

4.1 Experimental design

In the main user study, we employed a within-subject design [Kelly, 2009] whereby each participant needs to perform an equal number of search tasks. We chose a within-subject design because we wanted to compare results of participants on the different types of task. The types of task were based on Batley [1988] four visual information needs. For each task type, users were given four topics to choose from. The rationale behind choice of topics is to ensure that the results were due to task type effect and not due to topic effect.

A simulated real work task situation [Borlund and Ingwersen, 1997] was used to place the participants in a work task scenario. We tailored the simulated work scenarios to the participants to ensure realism of the evaluation [Borlund and Schneider, 2010]. The scenario allows them to fashion the visual information needs of each task as if they were performing an actual image search, as well as providing the search context and the basis for relevance judgements.

Determining the number of participants depends on the following:

- statistical test performed,
- significance criterion used in the test, and
- the expected magnitude of the effect of interest found in the population.

We performed power analysis in determining the sample size that would enable us to detect statistical significance accurately and reliabily in the results. For each statistical test with a given significance criterion and expected effect, there is a difference in the number of samples required for the results to be deemed statistically significant. Therefore, considering

the statistical tests that we planned to perform which include descriptive statistics, Chi-Square analysis, Friedman's test, repeated measures ANOVA and correlation analysis; a significance criterion of 0.05 and expecting to observe a large effect size (a power of 0.8), we arrived at the largest number, that is 48 as the required number of participants [Cohen, 1992].

As participants were expected to perform more than one task that is similar in nature, they may experience effects in learning order. Therefore, we structure the order of the tasks using a mathematical factorial design. We derive 24 permutations of task order for the four task types. Each permutation occurred twice to conform to the 48 participants needed in the study. Assignment of participants to permutations was done randomly.

4.2 Selection of relevance criteria

In Chapter 2, we reviewed previous research on image relevance and the criteria users apply when making judgements. Previous research elicited criteria users identified as important. Studies on use of relevance criteria were mostly conducted in specialised domain (art history, journalism, medical) and using specific image collection. No other studies have been conducted on users using online web image search. In the study, we are not attempting to elicit new relevance criteria, but rather, we are investigating the importance and difference of a selected set of relevance criteria for the image information needs that we have previously defined.

As there exist overlap of criteria between these specialised domains, we decided to select relevance criteria identified by Choi and Rasmussen [2002] and Hung [2006] to be used in

the study. Seven criteria (*topicality, accuracy, suggestiveness, completeness, appeal of information, technical attributes of images* and *textual description*) were selected from Choi and Rasmussen [2002]. We only selected these criteria because they are applicable for all search tasks and not just historical tasks (*time frame* and *novelty*). From Hung [2006] we selected six criteria (*topicality, composition, consequence, emotion, interest* and *text*). These criteria were selected as they were the core criteria elicited from users when making image relevance judgements for different types of image search tasks. Other criteria were not chosen as we did not want to confuse the participants as some criteria can be similar (*symbol, context* and *implication*) or too specific (*facial expression*).

Of the thirteen criteria selected from the two studies, three criteria were overlapping. Therefore, we applied a total of ten relevance criteria for the user study and adapted them in the post-task questionnaires. The relevance criteria and their brief definitions are listed in Table 4.1. We selected these ten criteria because they are applicable for all search tasks including online web image search and are not limited to a specialised domain. Nevertheless, participants in the study were assigned information needs similar to users in a specialised domain.

4.3 Data collection methods

Previously, in Chapter 3, we identified two suitable data collection methods to use in the user studies: (i) a paper-based questionnaire and (ii) observation via screen capture. In this chapter, we are only analysing the questionnaire (a copy of the questionnaire is in Appendix E) and the screen capture is analysed in the next chapter. There were two types of question-

Table 4.1: Relevance criteria applied in this study

Relevance Criteria	Definition
Topicality	The image relates to the user's task.
Accuracy	The picture accurately represents what the user is looking for.
Suggestiveness	The image generates new ideas or new insights for the user.
Appeal of information/interest	The picture is interesting and appealing to the user.
Completeness	The image contains all the required details the user is looking for.
Technical attributes of images	Supplementary information of the image e.g. resolution, size, colour, dimension, angle etc.
Emotion	Contain emotional context telling what is happening in the image.
Textual information	Words associated with the image.
Consequence	Identifying the results and effect of the topic (either of human or object sides).
Composition	Identifying the way in which the parts of elements in the photo are arranged.

naires: a demographic and a post-task questionnaire. The demographic questionnaire was used to collect background information about the participants experience with web search prior to undertaking any search tasks which include questions such as frequency of searching, reasons for searching, preferred search engine and search expertise. The questionnaire included questions for both textual and image searches.

In the post-task questionnaire, participants were asked to rate different aspects of their search, such as relevance criteria used when making judgements on images, topic familiarity, search satisfaction, task difficulty and overall performance on a 5-point scale (from Strongly Disagree to Strongly Agree). The post-task questionnaire allowed us to collect data and have a better understanding of participants' perception of relevance and subjective factors that

affect searching for each task they performed. We included one open question at the end of the questionnaire to provide participants who may want to make additional comment or elaborate on any other aspects of the search or questions that they have previously answered.

4.4 Experimental procedure

To maintain consistency and validity of the data, each participant followed the same experimental procedure. The experiment was conducted confidentially and identity details of participants were not retained, so that responses could not be traced back to individual participants. They were met separately one at a time. As required by the RMIT University Ethics Committee, an introductory session was conducted whereby all participants were given a plain language statement (Appendix C) explaining what the study is about, how it is going to be carried out, the risks or disadvantages of participation, the use of information gathered, and their rights as a participant. Once they have understood the document and decided to continue participating, they were asked to sign a consent form (Appendix D).

Next, the participants were asked to complete a pre-search questionnaire (Appendix E) to collect demographic and background information. This is followed by a training session using Google Images to familiarise participants on performing the actual search tasks; specifically how to record their search interaction and where to save their images. Then, they could begin the actual first search task by recording their search interaction themselves.

They were instructed to select only one of four topics for a given task type (see Tables 4.2 - 4.5), then search using Google Images and save four images from among the search results into a predefined folder. The images saved should be, in their mind, the most appropriate for

the chosen topic. Saving images implies that participants judge the saved images as relevant to their topic. After the participants have selected a topic, they will begin their task by first starting the recording of their search interaction. In the course of the search, participants were allowed to submit as many separate queries as needed to find four relevant images for their task.

Table 4.2: List of topics for Specific task

Topic	Scenario
Yellowstone National Park	Imagine you are a designer with the responsibility of designing leaflets for the Yellowstone National Park. These leaflets will consist of a body of text interspersed with images. Your task is to search and save 4 images appropriate for the leaflets
Mercedes Benz SL Class	Your team is responsible for designing a brochure for the Mercedes Benz SL Class models. Your task is to search and save 4 images that could be used in the brochure.
Taj Mahal	For your assignment on 'Seven Wonders of the Middle Ages', you have chosen The Taj Mahal as the topic of your report. Your task is to search and save 4 images that are appropriate for the report
Harley Davidson Cruiser Bike	You are writing a special issue article on Harley Davidson Cruiser Bike models. Your task is to search and save 4 images that are appropriate for the article.

During the experiment, users were asked to start a new session on Google for each task type. We define a new session by closing the browser window upon finishing a task and opening a new window for each of the consecutive tasks. We did not apply any further control for any personalisation that may be applied by the search engine. However, in hindsight, using a new private window for each task is probably better as a method to lessen the effect of personalisation.

In addition, they were allowed to delete any images that they had previously saved if they

Table 4.3: List of topics for Nameable task

Topic	Scenario
Immigrants	You are writing an article on 'Immigration Nation'. Your task is to search and save 4 images on immigrants that would be appropriate for the article.
Kimono	As a design student, you are required to present a poster on the Japanese traditional garment, Kimono. Your task is to search and save 4 images that would be suitable for your poster.
Athlete	You are preparing a blog entry on 'Becoming a Great Athlete'. Your task is to search and save 4 images that would make your entry more interesting.
Lighthouse	In one of your assignments as an architecture student, you are required to give a presentation on lighthouses. Your task is to search and save 4 images that would be suitable for your presentation.

Table 4.4: List of topics for Abstract task

Topic	Scenario
Economic Unrest	Your editor wants you to write an article on economic unrest which will be interspersed with images. Your task is to search and save 4 images that are related to the article.
Logical Thinking	Your group is preparing a presentation on logical thinking. Your task is to search and save 4 images that would be suitable for the presentation.
Urban Development	You and your classmates are preparing a report on urban development. Your task is to search and save 4 images that would further explain and increase understanding on the topic.
Financial Security	Your company is distributing a free booklet entitled 'Keys to Financial Security'. Your task is to search and save 4 images that would be suitable for use in the booklet.

changed their mind about the suitability of a particular image. After saving the required number of images, participants need to end their recording and proceed to answering the post-task questionnaire. The steps of performing searches and answering the questionnaire were repeated until all four tasks were completed. We included a final open question for participants to comment or raise issues, for example, copyright of images. However, no responses were received.

Table 4.5: List of topics for Subjective task

Topic	Scenario
Hope	Imagine you are taking a photography course. For your assignment on capturing emotions, you are required to search and save 4 images that best capture the emotion 'hope'.
Joy	For your art assignment, you were asked to set up a mini gallery entitled 'Joy'. You are required to search and save 4 images that you could use for the assignment.
Curiosity	You are writing a psychology report on curiosity. You are required to search and save 4 images that would be appropriate for the report.
Adversity	You are taking a class on creative writing for which you need to write a fictional story on overcoming adversity. Your task is to search and save 4 suitable images that would accompany your story.

4.5 Profile of participants and topics

As stated in Section 4.1, we recruited 48 students to volunteer as participants in the user study. Table 4.6 shows the profiles of participants. Most of the participants (87.5%) were postgraduate students from RMIT who were approached and recruited via posters (Appendix F), electronic forums, mailing lists, and face-to-face recruitment after lecture sessions. They come from a variety of discipline ranging from computer science, engineering, business and mathematics. Being students, it was expected that they would perform web information search on a daily basis. In contrast to information search, fewer than 7% of participants regard themselves as experts in image searching and most users (60%) search for images on a weekly basis (Figure 4.1).

Table 4.6: Participants profiles

Gender	18 males; 30 females
Level of study	Undergraduate = 6; Master's = 6; PhD = 36
Discipline	CSIT = 30; Other = 18
Expertise in image searching	Novice = 21; Intermediate = 24; Expert = 3

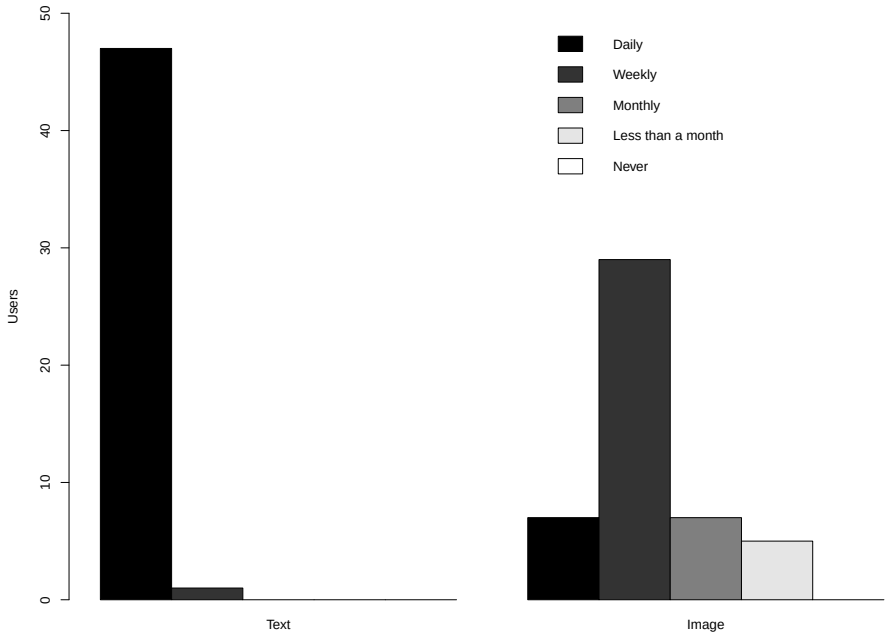

Figure 4.1: Frequency of searching on the web

Gender was not a major factor analysed in this study. However, given the unbalanced number between male and female participants, we analysed gender in relation to relevance criteria before discussing the results in Section 4.6. Similarly, expertise in image search was not a major factor considered in the study. Nevertheless, as the number of users with image search expertise is higher compared to non-expert image searchers, we analysed search expertise in relation to aspects of performing image search in Chapter 5.

Google was the preferred search engine for both information and image search and as shown in Figure 4.2, the reasons most participants gave for search engine preference were volume, speed and ease of use. For image search in particular, we asked participants about their familiarity with other image search engines such as Bing Images and Yahoo Images. Figure 4.3 suggests that as most users preferred Google for information search, they continue

to use Google for everything else. Another explanation would be that they use Google because they were not familiar with the other image search engines.

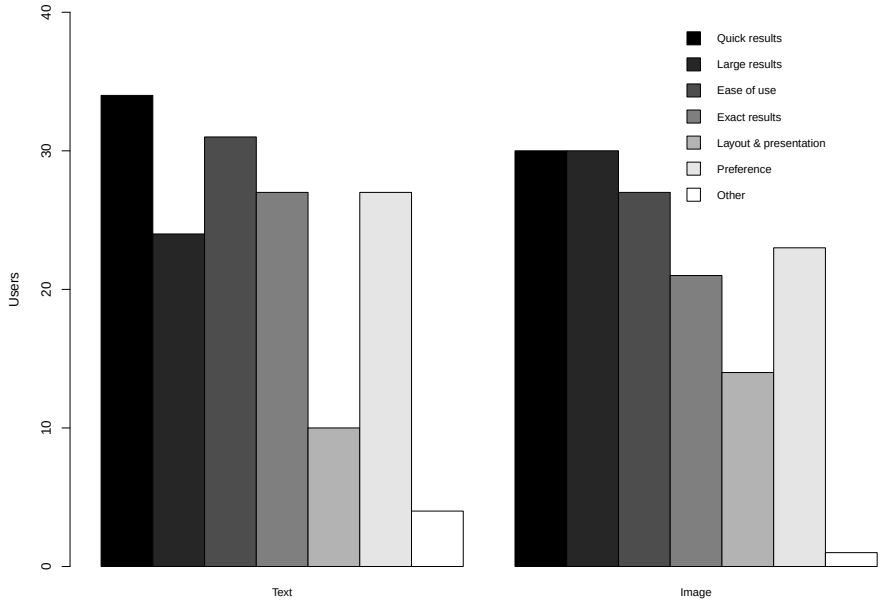

Figure 4.2: Reasons for using a particular search engine

Based on Broder's taxonomy [2002], informational needs were the most frequent reasons of searching followed by navigational and transactional as indicated in Figure 4.4. The top three informational needs are related to finding information on a particular topic. However, finding links to the information they were searching for were less preferred. In the context of image use, Figure 4.5 revealed that illustration is the main reason users search for images which is consistent with findings by Chung and Yoon [2011]. Because the focus was on the context of image use, it is not surprising that none of the participants reported searching for images purely for amusement.

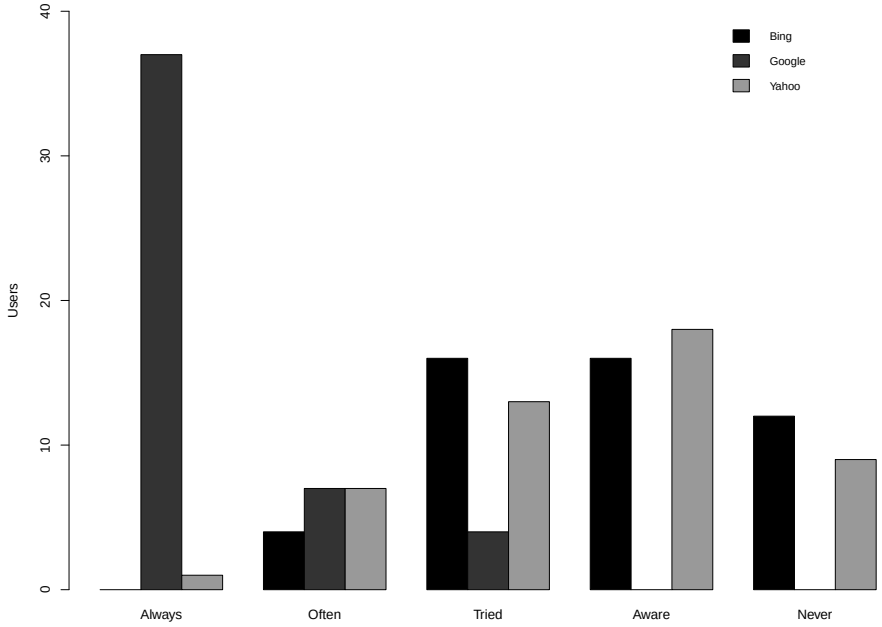

Figure 4.3: Usage familiarity of image search engine

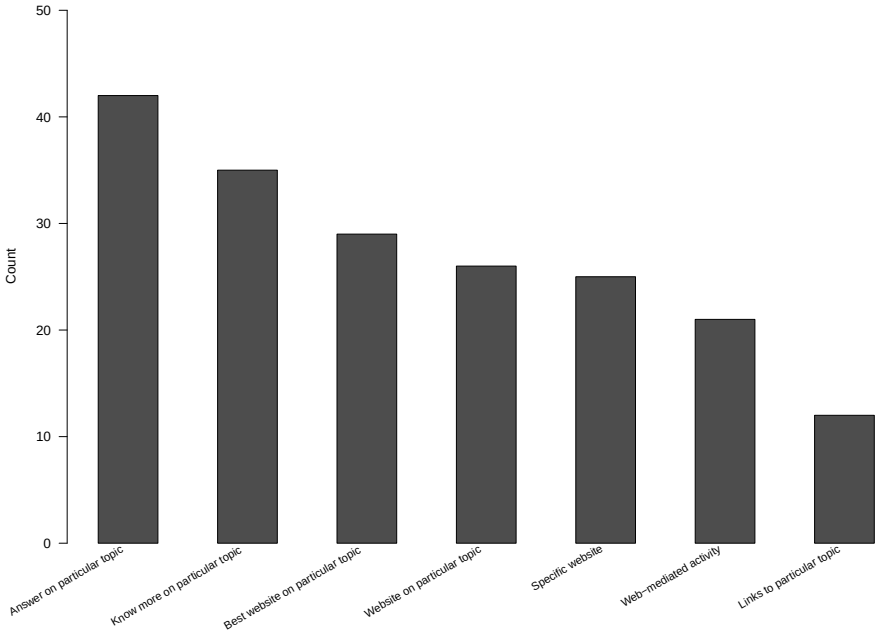

Figure 4.4: Reasons for searching information

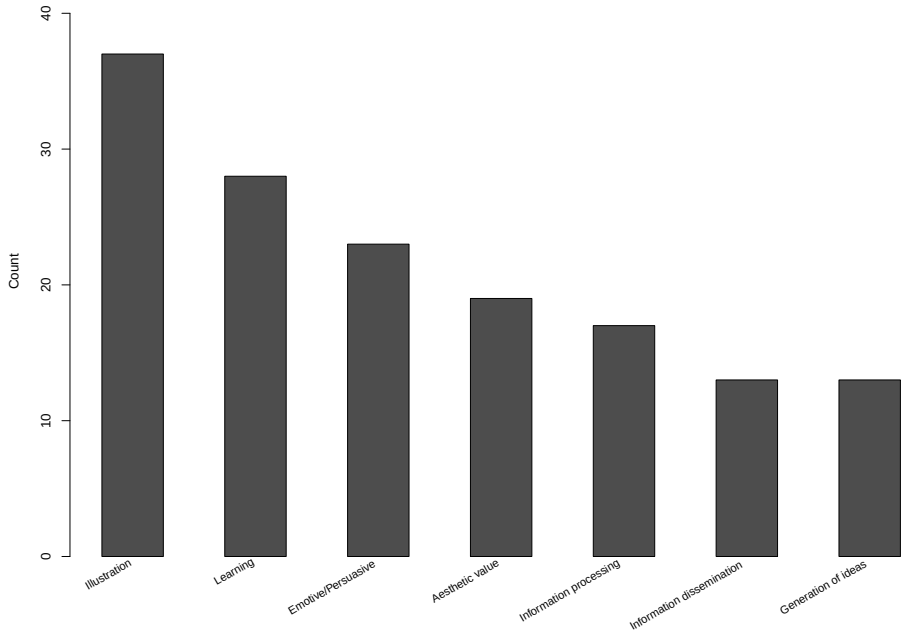

Figure 4.5: Reasons for searching images in the context of image use

For every task, participants were given four topics to choose from. Table 4.7 displays the number of participants, grouped by topics for all task. For the Specific and Nameable task, topics can be classified into two groups. In the Specific tasks, topics were grouped as place/location and object while in the Nameable tasks, were person and object.

4.6 Users' judgements on criteria for image relevance

During a search session, the system may retrieve pages of image results that matched a query. However, deciding a relevant image from the results is in the hands of the user. Judging the relevance of an image involves multiple and different criteria. Previous researchers have elicited and identified various relevance criteria used when making image relevance judgements. These researchers [Hirsh, 1999; Choi and Rasmussen, 2002; Crystal and Greenberg,

Table 4.7: Number of participants per topic for each task

Task type	Topic	Number of participants	Total
Specific	Yellowstone National Park	15	48
	Mercedes Benz S Class	13	
	Taj Mahal	17	
	Harley Davidson Cruiser Bike	3	
Nameable	Immigrants	5	48
	Kimono	21	
	Athlete	9	
	Lighthouse	13	
Abstract	Economic Unrest	8	48
	Logical Thinking	14	
	Urban Development	18	
	Financial Security	8	
Subjective	Hope	24	48
	Joy	19	
	Curiosity	5	
	Adversity	0	

2006; Hung, 2006; Xu and Chen, 2006; Sedghi et al., 2008; Kim and Oh, 2009] unanimously found that topicality is the most widely used criteria in image search in numerous domains. In this section, we present and discuss the results obtained from the main user study.

We apply the ten relevance criteria identified by Choi and Rasmussen [2002] and Hung [2006] to four image search tasks. Using its definition, each criterion was rephrased as a question in the post-task questionnaire as can be seen in Table 4.8. Participants were asked to rate their response on a 5-point scale: Strongly Disagree (1), Disagree (2), Neutral (3), Agree (4), and Strongly Agree (5).

Regarding the unbalanced number between male and female participants, we compared relevance criteria with gender in Table 4.9. For criteria that participants frequently rated as important, the means of these criteria were closely similar between males and females.

Table 4.8: Relevance criteria rephrased as questions

Relevance Criteria	Question
Topicality	The images I selected were relevant to the search topic.
Accuracy	The images I selected were an accurate representation of what I was looking for on the search topic.
Suggestiveness	The images I selected gave me new ideas or new insights about the search topic.
Appeal of information/interest	The images I selected were interesting in regards to the search topic.
Completeness	The images I selected contained the kinds of details I could use to clarify important aspects of the search topic.
Technical attributes of images	The images I selected had technical attributes (such as colour, perspective or angle) that were important to me for this search topic.
Emotion	The images I selected evoked an emotional response in me regarding the search topic.
Textual information	The images I selected had useful text descriptions on the search topic.
Consequence	The images I selected contained consequences or implications of the search topic.
Composition	The images I selected have a strong visual impact regarding the search topic.

Welch's t-test (variance not assumed to be equal) found that there is no significant difference between frequency of male and female participants in rating the importance of each relevance criteria.

Figure 4.6 shows the frequency of user rating on relevance criteria for each task type. We analysed quantitative data from the questionnaire using descriptive statistics to identify criteria users' find important when making image relevance judgements.

In addition, we performed a Chi-Square goodness-of-fit test to examine statistical signifi-

Table 4.9: Frequency of male and female participants in rating the importance of each relevance criteria

Relevance criteria	Males Mean (SD)	Females Mean (SD)
Topicality	3.67 (0.69)	3.83 (0.38)
Accuracy	3.39 (0.70)	3.53 (0.63)
Suggestiveness	2.83 (1.25)	2.78 (1.07)
Appeal of information	3.11 (1.02)	3.30 (0.88)
Completeness	2.67 (1.14)	3.27 (0.78)
Technical attributes	3.06 (1.00)	2.87 (1.22)
Emotion	2.61 (1.42)	2.30 (1.32)
Textual information	1.17 (1.22)	2.27 (1.44)
Consequence	3.00 (1.19)	2.33 (1.21)
Composition	3.33 (0.97)	3.23 (0.90)

cant differences in the attitude of participants in regards to rating the importance of certain criterion when making relevance judgements. The p-value is calculated based on two categories which are (i) *important* — combination of Strongly Agree and Agree, and (ii) *not important* — combination of Strongly Disagree, Disagree and Neutral. We group a neutral response as disagree as it indicates that the participant is not positive towards a statement [Sturgis et al., 2014]. In this study, we adopted a minimum significance level of $p < 0.05$. Table 4.10 shows the descriptive statistics, participants' rating and Chi-Square's p-value on the importance of relevance criteria for each search task. Results for each topic by type of task are shown in Tables 4.11 to 4.14.

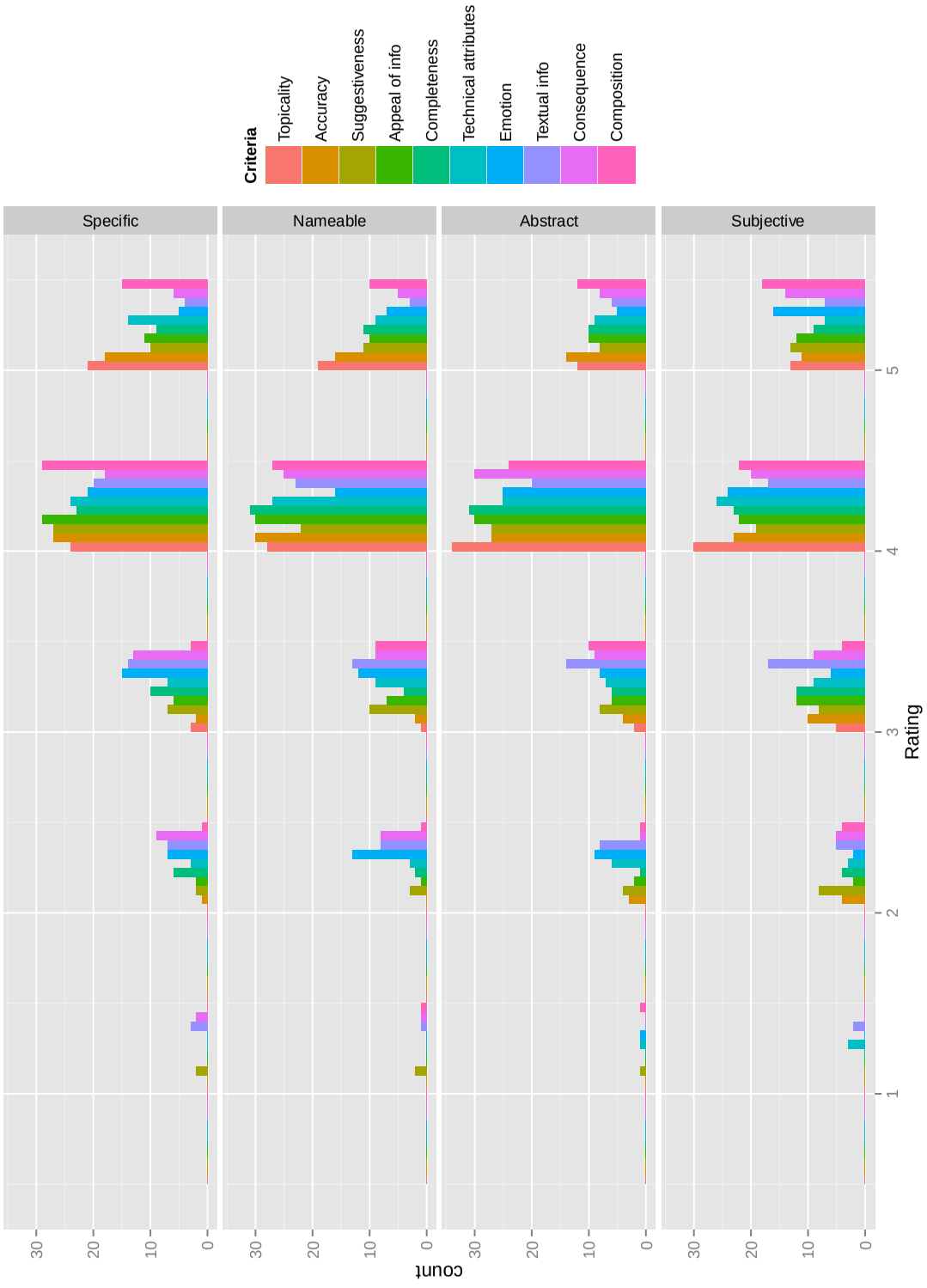

Figure 4.6: Frequency of user ratings on relevance criteria by types of task

By comparing the mean ratings of each relevance criteria across search tasks, it is obvious that *Topicality* is the criteria participants find most important. This is apparent from the consistently high means value of *Topicality* across all task types. Moreover, the results show variation in participants' rating of relevance criteria between types of task. This indicates that the importance of relevance criteria varies between task types whereby higher mean values suggest that it is more widely seen as important when making relevance judgements.

Participants rate both *Topicality* and *Accuracy* as important across all image task types; however, the number of participants' rating both *Topicality* and *Accuracy* as important was not as high for the Subjective task. *Suggestiveness*, *Appeal of information*, *Completeness*, *Technical attributes of image* and *Composition* are other criteria participants used for all task types. As expected, participants rate that Emotion was more often used as a criterion for Subjective tasks (mean=4.125). Images are depicted by objects, and for a Subjective task, the response that users perceive from viewing the object is more important than the object itself.

Participants performed the image search tasks using a text-based web search engine by submitting textual queries. Therefore, the returned results will be images that are described by that text. However, for all task types, participants considered *Textual information* less important when making image relevance judgements. The reason could be that the textual description does not always represent the image that users were looking for or they were more focused on the image content. This finding is in contrast to findings from specialised domains [Choi and Rasmussen, 2003; Hung et al., 2005; Sedghi et al., 2008] whereby users rely on textual information before making image relevance judgments. One possible reason is

*Table 4.10: Descriptive statistics, number of participants' rating and Chi-Square's p-value on the importance of relevance criteria for all search tasks. p-values in **bold** indicate statistical significance*

Relevance criteria	Statistics	Specific	Nameable	Abstract	Subjective
Topicality	μ	4.375	4.375	4.208	4.167
	σ	0.606	0.531	0.504	0.595
	#agree	45	47	46	43
	p-value	**0.0001**	**0.0001**	**0.0001**	**0.0001**
Accuracy	μ	4.292	4.292	4.083	3.854
	σ	0.651	0.544	0.794	0.875
	#agree	45	46	41	34
	p-value	**0.0001**	**0.0001**	**0.0001**	**0.0039**
Suggestiveness	μ	3.854	3.771	3.771	3.771
	σ	0.945	1.016	0.905	1.036
	#agree	37	33	35	32
	p-value	**0.0002**	**0.0094**	**0.0015**	**0.0209**
Appeal of information	μ	4.021	4.021	4.000	3.917
	σ	0.729	.0668	0.715	0.821
	#agree	40	40	40	34
	p-value	**0.0001**	**0.0001**	**0.0001**	**0.0039**
Completeness	μ	3.729	4.063	4.042	3.771
	σ	0.917	0.697	0.651	0.857
	#agree	32	42	41	32
	p-value	**0.0209**	**0.0001**	**0.0001**	**0.0209**
Technical attributes of image	μ	4.021	3.875	3.729	3.646
	σ	0.838	0.789	0.984	1.021
	#agree	38	36	34	33
	p-value	**0.0001**	**0.0005**	**0.0039**	**0.0094**
Emotion	μ	3.500	3.354	3.500	4.125
	σ	0.875	1.041	0.989	0.789
	#agree	26	23	30	40
	p-value	0.5637	0.7728	0.0833	**0.0001**
Textual information	μ	3.313	3.396	3.521	3.458
	σ	1.035	0.917	0.967	1.010
	#agree	24	26	26	24
	p-value	1	0.5637	0.5637	1
Consequence	μ	3.354	3.521	3.938	3.896
	σ	1.062	0.923	0.665	0.951
	#agree	24	30	38	34
	p-value	1	0.0833	**0.0001**	**0.0039**
Composition	μ	4.208	3.917	3.896	4.125
	σ	0.651	0.821	0.861	0.890
	#agree	44	37	36	40
	p-value	**0.0001**	**0.0002**	**0.0005**	**0.0001**

Table 4.11: Users' rate of agreement for each relevance criterion based on topics for Specific task

Relevance Criteria	Statistics	Yellowstone National Park	Mercedes Benz S Class	Taj Mahal	Harley Davidson Cruiser Bike
Topicality	μ	4.467	4.231	4.412	4.333
	σ	0.516	0.725	0.618	0.577
	#agree	15	11	16	3
Accuracy	μ	4.400	4.154	4.353	4.000
	σ	0.507	0.899	0.606	0.000
	#agree	15	11	16	3
Suggestiveness	μ	4.067	3.462	3.941	4.000
	σ	0.458	1.266	0.966	0.000
	#agree	14	7	14	2
Appeal of information	μ	4.000	4.154	4.000	3.667
	σ	0.655	0.899	0.707	0.577
	#agree	12	11	15	2
Completeness	μ	3.800	3.769	3.765	3.000
	σ	0.676	1.166	0.970	0.000
	#agree	10	9	13	0
Technical attributes of image	μ	3.867	4.385	3.941	3.667
	σ	0.834	0.650	0.966	0.577
	#agree	11	12	13	2
Emotion	μ	3.333	3.692	3.471	3.667
	σ	0.976	1.032	0.717	0.577
	#agree	6	8	10	2
Textual information	μ	3.800	2.923	3.176	3.333
	σ	0.676	1.256	1.074	0.577
	#agree	10	5	8	1
Consequence	μ	3.667	2.923	3.412	3.333
	σ	0.816	1.320	1.064	0.577
	#agree	9	5	9	1
Composition	μ	4.133	4.231	4.294	4.000
	σ	0.516	0.927	0.588	0.000
	#agree	14	11	16	3

results of web image search are presented as visual thumbnails, therefore, textual information is not salient and therefore ignored when judging relevance [Tjondronegoro et al., 2009].

Table 4.12: Users' rate of agreement for each relevance criterion based on topics for Nameable task

Relevance Criteria	Statistics	Immigrants	Kimono	Athelete	Lighthouse
Topicality	μ	4.000	4.524	4.222	4.385
	σ	0.707	0.512	0.441	0.506
	#agree	4	21	9	13
Accuracy	μ	3.800	4.524	4.000	4.308
	σ	0.447	0.512	0.000	0.480
	#agree	4	21	8	13
Suggestiveness	μ	3.800	4.048	3.000	3.846
	σ	0.447	0.865	1.581	0.689
	#agree	4	16	4	9
Appeal of information	μ	3.800	4.095	3.778	4.154
	σ	0.447	0.625	0.972	0.555
	#agree	4	18	6	12
Completeness	μ	4.000	3.952	4.111	4.231
	σ	0.000	0.669	0.782	0.832
	#agree	5	18	7	12
Technical attributes of image	μ	3.400	4.000	3.667	4.000
	σ	0.548	0.775	0.866	0.816
	#agree	2	19	4	11
Emotion	μ	3.800	3.286	3.556	3.154
	σ	0.447	1.007	1.130	1.214
	#agree	4	8	5	6
Textual information	μ	3.800	3.429	3.111	3.385
	σ	0.447	1.121	0.782	0.768
	#agree	4	12	3	7
Consequence	μ	3.600	3.429	3.778	3.462
	σ	0.894	1.028	0.833	1.050
	#agree	4	12	7	7
Composition	μ	3.600	3.905	3.889	3.769
	σ	0.548	0.944	1.054	0.515
	#agree	3	16	6	12

Overall, from the ten selected criteria, not all were used by participants when judging image relevance. The results showed that users use more criteria when judging image relevance for Subjective and Abstract tasks as compared to Specific and Nameable tasks. A

Table 4.13: Users' rate of agreement for each relevance criterion based on topics for Abstract task

Relevance Criteria	Statistics	Economic Unrest	Logical Thinking	Urban Development	Financial Security
Topicality	μ	4.375	4.071	4.278	4.125
	σ	0.518	0.616	0.461	0.354
	#agree	8	12	18	8
Accuracy	μ	4.375	4.143	3.833	4.250
	σ	0.518	0.864	0.924	0.463
	#agree	8	12	13	8
Suggestiveness	μ	3.500	3.714	3.833	4.000
	σ	1.309	0.994	0.786	0.535
	#agree	6	9	13	7
Appeal of information	μ	4.125	3.786	4.000	4.250
	σ	0.641	0.893	0.686	0.463
	#agree	7	9	16	8
Completeness	μ	4.250	3.857	4.056	4.125
	σ	0.463	0.663	0.802	0.354
	#agree	8	10	15	8
Technical attributes of image	μ	3.250	3.714	4.056	3.500
	σ	1.156	0.994	0.725	1.195
	#agree	5	9	16	4
Emotion	μ	4.250	3.143	3.333	3.750
	σ	0.463	1.167	0.970	0.707
	#agree	8	7	10	5
Textual information	μ	3.125	3.643	3.500	3.625
	σ	0.835	0.929	0.985	0.916
	#agree	3	9	9	5
Consequence	μ	4.125	3.929	3.833	4.000
	σ	0.354	0.730	0.707	0.756
	#agree	8	10	14	6
Composition	μ	4.375	4.000	3.667	4.000
	σ	0.744	0.877	0.970	0.535
	#agree	7	9	13	7

possible explanation is that in the Subjective and Abstract tasks, participants were looking for new and interesting images of the search topic whereas in the Specific and Nameable

Table 4.14: Users' rate of agreement for each relevance criterion based on topics for Subjective task

Relevance Criteria	Statistics	Hope	Joy	Curiosity
Topicality	μ	4.125	4.211	4.200
	σ	0.612	0.631	0.447
	#agree	21	17	5
Accuracy	μ	3.708	4.105	3.600
	σ	0.955	0.809	0.548
	#agree	15	16	3
Suggestiveness	μ	3.833	3.947	2.800
	σ	1.049	0.970	0.837
	#agree	17	14	1
Appeal of information	μ	3.875	4.000	3.800
	σ	0.850	0.882	0.447
	#agree	16	14	4
Completeness	μ	3.917	3.789	3.000
	σ	0.881	0.713	1.000
	#agree	18	12	2
Technical attributes of image	μ	3.958	3.474	2.800
	σ	0.690	1.172	1.304
	#agree	20	11	2
Emotion	μ	4.208	4.158	3.600
	σ	0.779	0.834	0.548
	#agree	21	16	3
Textual information	μ	3.667	3.211	3.400
	σ	1.007	1.084	0.548
	#agree	15	7	2
Consequence	μ	4.000	3.947	3.200
	σ	0.885	1.026	0.837
	#agree	19	13	2
Composition	μ	4.042	4.316	3.800
	σ	0.955	0.749	1.095
	#agree	20	18	4

tasks, participants were looking for particular images. Images depict objects and for Subjective and Abstract tasks, the response that users perceived from viewing the object is more important than the object itself [Choi and Rasmussen, 2002; Göker et al., 2016]. A possible

*Table 4.15: Results of Friedman's test on rating differences of relevance criteria across search tasks (Values in **bold** indicate statistical significance)*

Relevance Criteria	χ^2	p
Topicality	**9.503**	**0.023**
Accuracy	**9.884**	**0.020**
Suggestiveness	0.469	0.926
Appeal of information	0.757	0.860
Completeness	0.757	0.860
Technical attributes of images	6.518	0.089
Emotion	**28.374**	**0.000**
Textual information	0.600	0.896
Consequence	**12.955**	**0.005**
Composition	5.149	0.161

explanation is that in Subjective and Abstract tasks, relevance of an image are complex or unclear; therefore requiring additional criteria to make judgements [Crystal and Greenberg, 2006].

For each criterion that participants have rated as important, we performed the Friedman test which is a non-parametric repeated measure ANOVA to identify differences in their importance between the different types of task. The Friedman test was selected because of the ordinal scale used in rating relevance. Out of the ten criteria, only four had overall statistically significant differences across search tasks. These criteria were *Topicality*, *Accuracy*, *Emotion* and *Consequence* as presented in Table 4.15.

Although the Friedman test have identified four statistically significant different criteria, it does not reveal where the differences occur between the tasks. For that reason, we ran a post-hoc test, a Wilcoxon signed-rank test on the four criteria to determine where the differences were likely to occur. As there were four task types, we compared between six combinations

*Table 4.16: Results of Wilcoxon signed-rank test for differences on importance of relevance criteria between tasks (Values in **bold** indicates statistical significance, $p < 0.0083$)*

Combinations	Topicality	Accuracy	Emotion	Consequence
Specific to Nameable	Z = 0.000 p = 1.000	Z = -0.030 p = 0.976	Z = -1.167 p = 0.243	Z = -1.008 p = 0.313
Specific to Abstract	Z = -1.617 p = 0.106	Z = -1.389 p = 0.165	Z = -0.206 p = 0.837	Z = -2.945 p = **0.003**
Specific to Subjective	Z = -1.995 p = 0.046	Z = -2.578 p = 0.010	Z = -3.919 p <**0.001**	Z = -2.345 p = 0.019
Nameable to Abstract	Z = -1.999 p = 0.046	Z = -1.724 p = 0.085	Z = -1.008 p = 0.313	Z = -2.450 p = 0.014
Nameable to Subjective	Z = -1.995 p = 0.046	Z = -2.797 p = **0.005**	Z = -3.995 p <**0.001**	Z = -1.927 p = 0.054
Abstract to Subjective	Z = -0.577 p = 0.564	Z = -1.647 p = 0.100	Z = -3.416 p = **0.001**	Z = -0.390 p = 0.696

of the tasks. Comparison combinations included: 1) Specific to Nameable, 2) Specific to Abstract, 3) Specific to Subjective, 4) Nameable to Abstract, 5) Nameable to Subjective, and 6) Abstract to Subjective. With the multiple comparisons, a Bonferroni correction was used on the results to obtain a new significance level. The new level is calculated by dividing the initial value with the number of combinations i.e. $0.05/6 = 0.0083$. The resulting p-value was then compared with the new significance level to show occurence of differences. Results of the Wilcoxon signed rank test appear in Table 4.16.

Even though *Topicality* was a statistically significant criterion, results of the post-hoc analysis was unable to detect any significant difference on its importance between task type. As for *Accuracy*, a statistically significant difference occurred between Nameable and Subjective tasks. Specifically, when judging image relevance, more participants rated that they selected images that were an accurate representation of a Nameable task (Z = -2.797, p =

0.005) compared to a Subjective task. Accuracy of a Nameable image is more easily identified and unambiguous compared to a Subjective image.

Participants collectively agreed that *Emotion* is important in the Subjective task. This is conclusive from the results through which significant differences were detected between 3 combinations of task types — Specific to Subjective (Z = -3.919, p = 0.000), Nameable to Subjective (Z = -3.995, p = 0.000), and Abstract to Subjective (Z = -3.416, p = 0.001) tasks. Surprisingly, even though Subjective task has the highest number of ratings for the criterion *Emotion* (4.125 in Table 4.5), many participants still rated that they select *Emotion* as well when judging relevance for Specific, Nameable and Abstract images.

For *Consequence*, despite having overall importance across task types, a significant difference was only detected between Specific and Abstract tasks. It seems that participants selected images that contained consequences or implications of an Abstract task more often compared to a Specific task (Z = -2.945, p = 0.003).

4.7 Summary

In this chapter, we discussed the design and conduct of the main user experiment. Specifically, the chapter focuses on results of users' judgement of image relevance. By using criteria identified in previous research, results confirmed that users apply different criteria when making relevance judgements for different task types. Several criteria were commonly used in all task types. However, criteria such as *Emotion* and *Consequence* were more important for Subjective and Abstract tasks. Further analysis showed that users may apply the same criteria; however, the importance for these criteria differs between the task types and affirms

that different search tasks affect how users' judge image relevance. When ranking image search results, search engines should take into consideration the subtleties of an image as users judge relevance not just by the associated information (tags/description, size, etc.) but also by what is conveyed within the image itself.

Chapter 5

Factors affecting image search behaviour

Factors that affect users' search behaviour can be categorised as either *contextual*, for example attributes of the searchers and their needs [Choi, 2010a], or *subjective* such as familiarity, difficulty and satisfaction [Gwizdka and Lopatovska, 2009]. By assigning users to four image search tasks and asking them about their search experience, we address the second research question — determining the effect of different subjective factors on users' image searching behaviour.

The chapter is organised as follows: details of the subjective factors and how their effects are measured in the study (Section 5.1), followed by results on users' perception on effect of the factors in relation to performing image search (Section 5.2) and users' actual behaviour while performing the searches (Section 5.3). Next is the results of correlation that examine whether users' actual search behaviour reflects their perception of the factors (Section 5.4)

and Section 5.5, summarises the chapter.

5.1 Defining and assessing factors

Text and image are two different forms of information. However, as users search for both text and images online using keywords, their searching behaviour is affected by similar factors. Much research on factors affecting search — including task type, topic familiarity, difficulty and satisfaction — has been done in text search and retrieval. However, little research has been done on factors affecting image search.

The first factor, *task type* refers to the kind of image a user is currently looking for. Users' attention on an image is in its perceived content regardless of categorisation [Westman, 2009]. Therefore, in the study, we vary task types using the four visual information needs identified by Batley [1988].

The second factor, *topic familiarity* relates to users' current level of knowledge on a particular search topic. We limit the effect of topic familiarity by asking users to select only one topic from a list of four topics in each task. We assess topic familiarity through users' rating on whether they have selected a familiar topic for each task type.

The next factor is *task difficulty*. Task difficulty is assessed by users; and researchers have used various approaches in evaluating task difficulty, such as varying task types [Qu et al., 2010; Liu et al., 2010b;c; 2011], multiple levels of difficulties [Aula et al., 2010; Liu et al., 2010c] and assessing difficulty before and after searching [Liu et al., 2011]. In this work, we assess difficulty through users' self-assessment of ease of creating the initial query and ease of performing varying task types.

The last factor is *satisfaction*. We interpret satisfaction as a sense of achievement that one experiences. Satisfaction is usually evaluated at the system level or user level. The research is towards understanding user search behaviour and therefore we are more interested on satisfaction at the user level. Satisfaction is assessed at various points during the search session, which includes satisfaction related to image results retrieved by the search engine, order of image results and overall search session.

5.2 Users' perception on factors affecting search behaviour

We assess the effect of factors using two different methods — questionnaires and observations. In this section, we report and analyse results from the questionnaire data. The factors are assessed based on various aspects of performing search. Participants were asked to rate these aspects in relation to their searching experience while completing the tasks on a 5-point Likert scale (from Strongly Disagree to Strongly Agree).

A total of twelve questions on the various aspects of performing search are listed in the questionnaire. Seven of the questions, specifically questions 11, 12, 16, 17, 18, 19 and 22 are adapted and rephrased from an exit questionnaire used by Kelly et al. [2008]. These questions are the same as the questions used in the pilot study. The remaining five questions are newly added questions based on limitations, findings and comments from participants in the pilot study. One question (question 13) is on topic familiarity and the remaining questions (14, 15, 20 and 21) are on satisfaction.

Each task in the main study is designed as a simulated work task scenario and participants may or may not be familiar with some of the topics provided. Question 13 determines whether

participants' have an idea of images that would satisfy the requirement of the search topic regardless of whether they are familiar or not with the chosen topic. Satisfaction can vary at various points. Therefore, the additional questions try to identify satisfaction at various points during an image search session. Question 14 is a continuation of question 13 to find out about participants' satisfaction by comparing their initial ideas with the image results retrieved by the search engine.

Selecting and making decisions on images can be challenging, as participants from the pilot study pointed out. They commented that images can describe a topic in multiple ways and be repetitive at the same time (refer Appendix A). Question 15 tries to find out if participants change their mind assuming they were not satisfied after looking at images retrieved by the search engine. The participants noted that they were changing their mind about the relevance of an image. Therefore, we note that question 15 is a negative measure of satisfaction in terms of making relevance judgments. As for question 20, it reflects on the effect of participants changing their mind in their final decision regarding images. Lastly, question 21 takes into account the whole process of seaching and making decisions on images.

In the following subsections, we present results of the twelve questions. The results are grouped by factors. The quantitative data is analysed using descriptive statistics. Chi-Square goodness-of-fit analysis was performed to examine statistical significant differences in the attitudes of participants in regards to rating perception of factors affecting search. We select a significance level of 0.05 and calculate the p-value based on combined categories of Strongly Agree with Agree and Strongly Disagree with both Neutral and Disagree. Results for each topic by task type can be found in Appendix G.

Apart from the descriptive statistics, we conduct further analysis to investigate the effect of task type in relation to participants' ratings of other factors affecting search behaviour. We conduct the Friedman test to examine the differences between task types for the participants' ratings of the aspects of performing search. For aspects that show significant differences, we conduct a post-hoc test (Wilcoxon signed-rank test) to identify where the differences are likely to occur. We make comparisons on six combinations of task types and apply a Bonferroni correction, resulting in a new significant level set at p <0.0083.

5.2.1 Topic familiarity

For each task, participants were asked two questions on aspects of performing search that relates to topic familiarity:

11. **I was familiar with the topic that I choose for this search task**.

13. **I had an idea of the kind of images that would satisfy the requirements of the topic before starting the search**.

Figure 5.1 shows the frequency distribution of participants' ratings on the two questions for all tasks. From the bar graph, we can see that participants had a slightly higher rate of familiarity on the Nameable task compared to the other three tasks. Although users may rate Strongly Disagree and Disagree on topic familiarity, we accept that they selected the topic they were most familiar with compared to other topics in the list.

Data in Table 5.1 show that participants' mean rating for familiarity with Nameable topics is 3.54. For question 13, having an idea of image, more than half of participants for all task types were in agreement that they have some idea in their mind on the kind of images

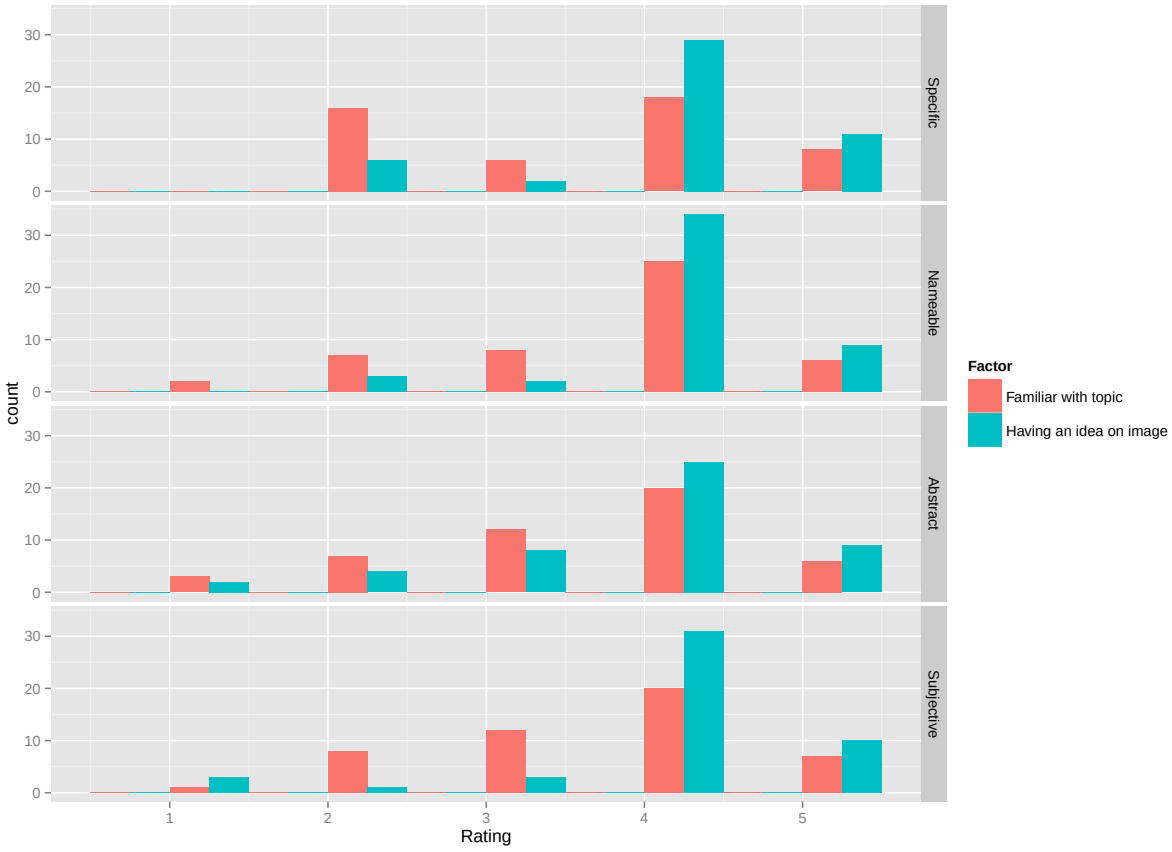

Figure 5.1: Frequency distribution of participants' ratings on perception of topic familiarity by types of task (1 = Strongly Disagree and 5 = Strongly Agree)

they were looking for. The assumption is that, as the topics were given in the form of a simulated work task scenario, participants were able to build and develop their ideas using the context provided.

Meanwhile, Table 5.2 presents results of Friedman's test. The results show that no significant differences were detected in participants' ratings on perception of effects of topic familiarity on image search for all types of task.

*Table 5.1: Descriptive statistics, number of participants' ratings in agreement, and Chi-Square's p-value on perception of agreement with topic familiarity for all search tasks (Values in **bold** indicate statistical significance, $p < 0.05$)*

Aspects of topic familiarity	Task type	μ (σ)	#agree	p-value
Familiar with search topic	Specific	3.38 (1.12)	26	0.564
	Nameable	3.54 (1.03)	31	**0.043**
	Abstract	3.40 (1.09)	26	0.564
	Subjective	3.50 (1.01)	27	0.387
Having an idea of image	Specific	3.94 (0.89)	27	**0.000**
	Nameable	4.02 (0.79)	40	**0.000**
	Abstract	3.73 (1.01)	34	**0.004**
	Subjective	3.92 (0.97)	33	**0.000**

Table 5.2: Results of Friedman's test on rating for aspects of topic familiarity

Aspects of topic familiarity	χ^2	p-value
Familiar with search topic	0.19	0.979
Having an idea on image	4.76	0.190

5.2.2 Task difficulty

For task difficulty, participants were asked two questions:

12. **I was able to create queries for the topic of the search task easily**.

22. **I found overall, the search task was easy to perform**.

Figure 5.2 shows the frequency distribution of participants' ratings on aspects of task difficulty for all tasks. The bar graph shows a decrease in Agree and Strongly Agree ratings for the Subjective task indicating that participants were facing difficulty with the task. Values in Table 5.3 show agreement among most participants that it was easier to create queries for Specific tasks (4.15) compared to Subjective tasks (3.48). Using keywords to express a concrete need that has no ambiguity is far less challenging compared to a need that deals

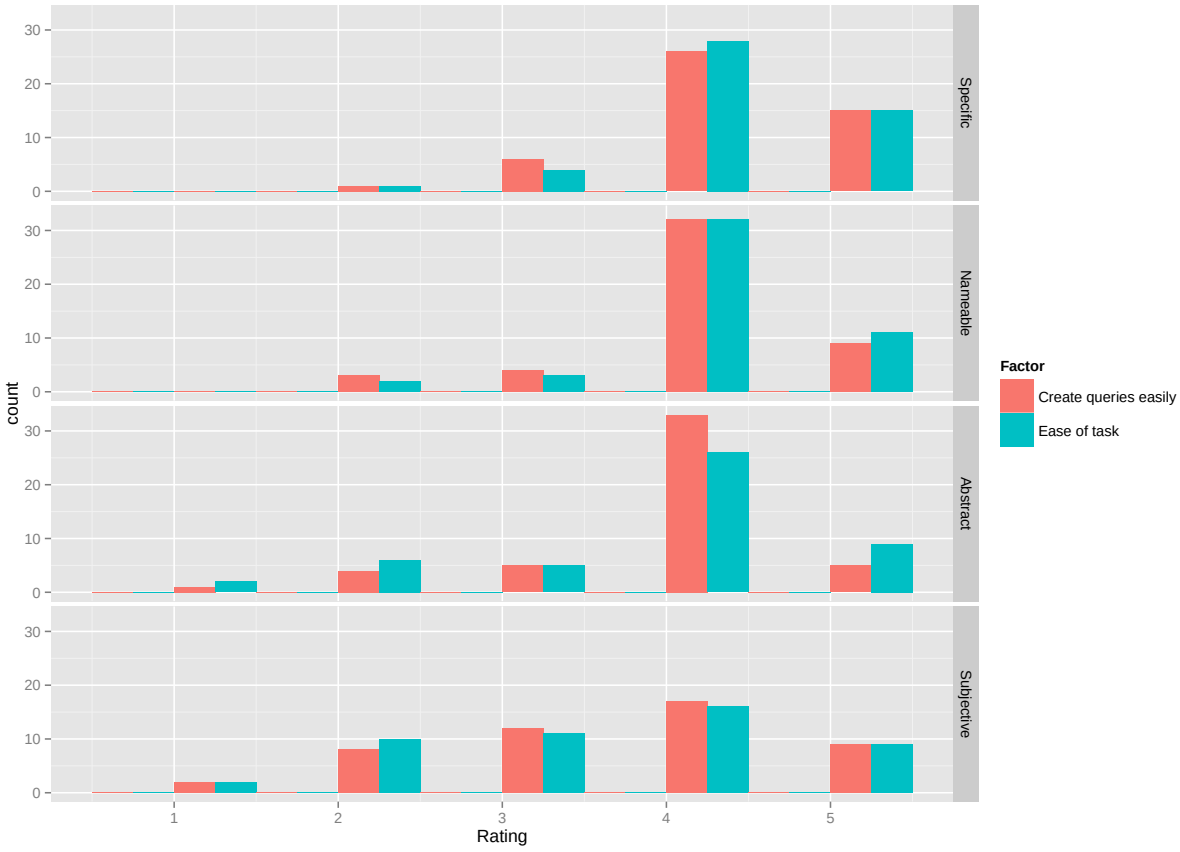

Figure 5.2: Frequency distribution of participants' ratings on perception of task difficulty by types of task (1 = Strongly Disagree and 5 = Strongly Agree)

with emotional responses. Similarly, participants agreed that Specific tasks were easier to perform than Subjective tasks.

Results of Friedman's test (Table 5.4) show that there were significant differences in participants' ratings between task types for both ability to create queries easily and ease of performing task. After conducting the post-hoc test (Wilcoxon signed rank test), we identified the occurrence of differences as shown in Table 5.5. For ability to create queries easily, significant differences in agreement occurred between Specific and Subjective tasks. Participants' rating on ability to create queries easily were higher for Specific tasks compared

to Subjective tasks (Z = -3.088, p = 0.002). The reason is they have a clear and well-defined image information need for a Specific task that makes it easier to create queries.

*Table 5.3: The number of participants' ratings in agreement, descriptive statistics and Chi-Square's p-value on perception of agreement with task difficulty for all search tasks (Values in **bold** indicate statistical significance, $p < 0.05$)*

Aspects of task difficulty	Task type	μ (σ)	#agree	p-value
Able to create queries easily	Specific	4.15 (0.71)	41	**0.000**
	Nameable	3.98 (0.73)	41	**0.000**
	Abstract	3.77 (0.83)	38	**0.000**
	Subjective	3.48 (1.11)	26	0.564
Ease of performing search	Specific	4.19 (0.67)	43	**0.000**
	Nameable	4.08 (0.68)	43	**0.000**
	Abstract	3.71 (1.05)	35	**0.002**
	Subjective	3.47 (1.15)	25	0.773

Significant differences in rating on ease of performing search were found when comparing Specific to Subjective tasks and Nameable to Subjective tasks. In both cases, participants rated that the Subjective task was harder to perform than the Specific (Z = -3.720, p <0.001) and the Nameable (Z = -1.607, p <0.001) tasks. This shows that similar to information search, difficulty in image search is dependent on the type of task.

5.2.3 Satisfaction

In the questionnaire after each task, participants were asked about various aspects of satisfaction:

*Table 5.4: Results of Friedman's test for aspects of task difficulty with an overall statistically significant difference ($p < 0.05$) across search tasks are indicated in **bold***

Aspects of task difficulty	χ^2	p
Able to create queries easily	**11.111**	**0.011**
Ease of performing search	**21.129**	**0.000**

*Table 5.5: Results of Wilcoxon signed-rank test for differences on aspects of task difficulty (Values in **bold** indicates statistical significance, $p < 0.0083$)*

Combinations	Able to create queries easily	Ease of performing search
Specific to Nameable	Z = -1.325, p = 0.185	Z = -0.716, p = 0.474
Specific to Abstract	Z = -2.391, p = 0.017	Z = -2.546, p = 0.011
Specific to Subjective	Z = -3.088, p = **0.002**	Z = -3.720, p <**0.001**
Nameable to Abstract	Z = -1.198, p = 0.231	Z = 2.268, p = 0.023
Nameable to Subjective	Z = -2.423, p = 0.015	Z = -1.607, p <**0.001**
Abstract to Subjective	Z = -1.715, p = 0.086	Z = -1.607, p = 0.108

14. **I found that images retrieved by the search engine matched my initial idea of what would satisfy the requirements of the search task.**

15. **I frequently changed my mind on the images that I was looking for.**

16. **I was satisfied with the images presented to me.**

17. **I was satisfied with the order of the images that were presented to me.**

18. **I believe I have seen all possible images that would satisfy the requirements of the search task.**

19. **I am very satisfied with my search results.**

20. **I saved images that matched my initial idea of what would satisfy the requirements of the search task.**

21. **I am very satisfied with my search interaction.**

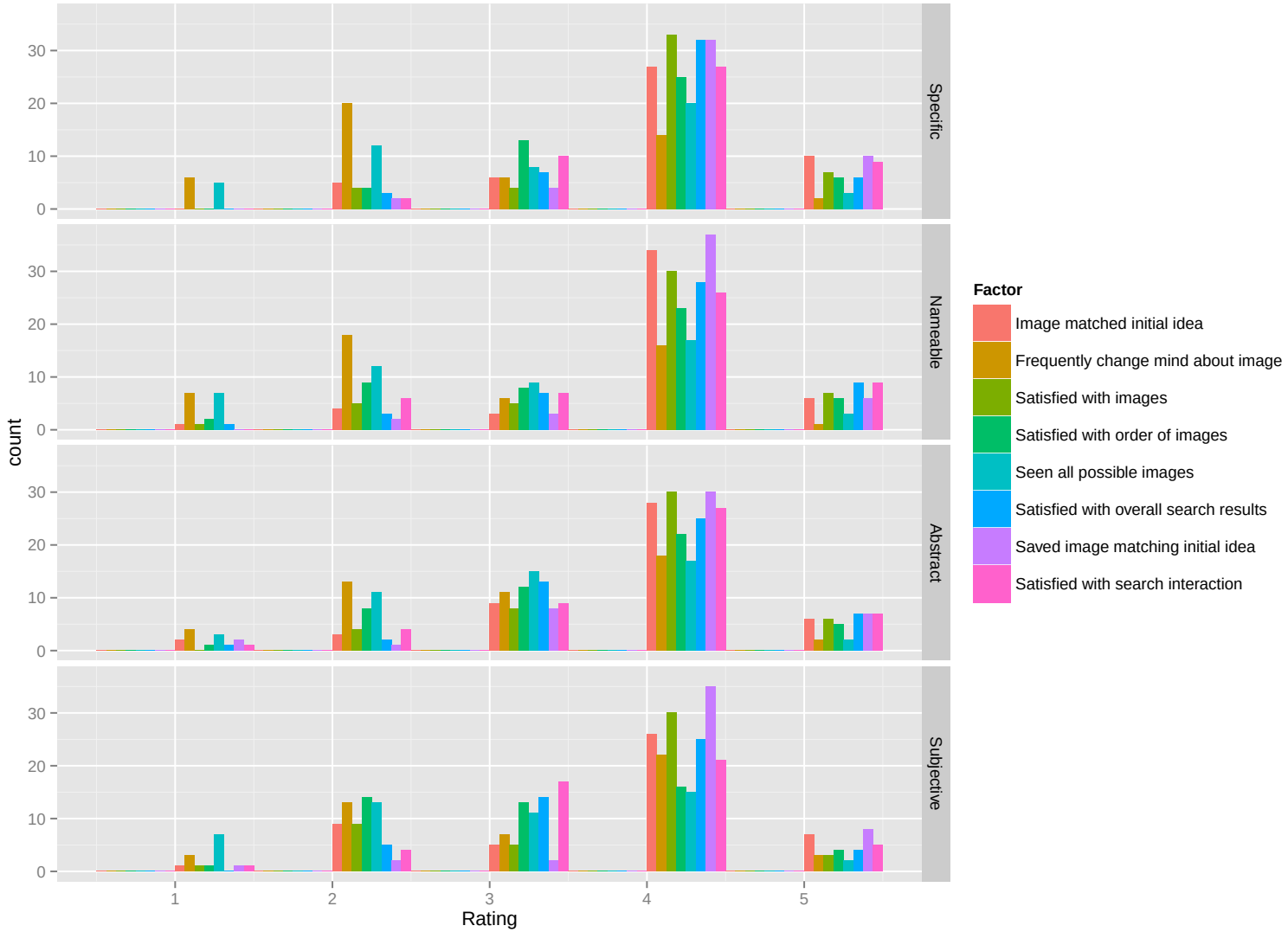

Figure 5.3: Frequency distribution of participants' ratings on perception of satisfaction by types of task (1 = Strongly Disagree and 5 = Strongly Agree)

Figure 5.3 shows the fluctuation in frequency of participants' ratings on agreement with aspects of satisfaction for all tasks. This suggests that level of satisfaction changes during searching. Results in Table 5.6 show that satisfaction is much harder to achieve in Subjective tasks. In the first aspect that relates to satisfaction, participants agree that images retrieved by the search engine matched their initial idea more closely for Specific tasks (3.88). The mean ratings showed that they were changing their minds more frequently for Abstract (3.02) and Subjective (3.19) tasks as the images conveys multiple and different messages.

Although participants may be satisfied with the images presented to them, they were not always satisfied with the order in which the images were presented. This is visible from the lower mean values and ratings of agreement for all task types. When searching for images, users expect that relevant images would be ranked higher in the list. Lu and Jia [2014] have shown that most users view image search results at 'top-centre', 'top-left' and 'centre-centre' position. Therefore, if relevant images were ranked lower in the list, users need to view more results pages in order to find them. When facing such difficulties, they reformulated their queries significantly more, in order to retrieve better results but taking a longer time to complete. So apart from search engine capabilities, users' efforts affect their search satisfaction [Choi, 2010a].

In the case of this happening, participants were asked whether they have viewed all possible images to complete the task. Clearly, for the Subjective task, they disagree that they were able to view all possible images (2.83). Even when they had an idea of images they were looking for, they faced difficulty expressing that idea in the form of a query. When the idea was not expressed clearly, they were not satisfied with the retrieved images. It is more

*Table 5.6: The number of participants' ratings in agreement, descriptive statistics and Chi-Square's p-value on perception of satisfaction for all search tasks (Values in **bold** indicate statistical significance, $p < 0.05$)*

Aspects of satisfaction	Task type	μ (σ)	#agree	p-value
Image matched initial idea	Specific	3.88 (0.87)	37	**0.000**
	Nameable	3.83 (0.83)	40	**0.000**
	Abstract	3.69 (0.93)	34	**0.004**
	Subjective	3.60 (1.03)	33	**0.009**
Frequently change mind about images	Specific	2.71 (1.15)	16	**0.021**
	Nameable	2.71 (1.15)	17	**0.043**
	Abstract	3.02 (1.08)	20	0.248
	Subjective	3.19 (1.10)	25	0.773
Satisfied with images	Specific	3.90 (0.75)	40	**0.000**
	Nameable	3.77 (0.91)	37	**0.000**
	Abstract	3.79 (0.77)	36	**0.000**
	Subjective	3.52 (0.95)	33	**0.009**
Satisfied with order of images	Specific	3.69 (0.80)	31	**0.043**
	Nameable	3.46 (1.07)	29	0.149
	Abstract	3.46 (0.97)	27	0.387
	Subjective	3.17 (1.02)	20	0.248
Seen all possible images	Specific	3.08 (1.16)	23	0.773
	Nameable	2.94 (1.21)	20	0.248
	Abstract	3.08 (1.01)	19	0.149
	Subjective	2.83 (1.16)	17	**0.0433**
Satisfied with overall search results	Specific	3.85 (0.72)	38	**0.000**
	Nameable	3.85 (0.88)	37	**0.000**
	Abstract	3.73 (0.84)	32	**0.021**
	Subjective	3.58 (0.79)	29	0.149
Saved image that matched initial idea	Specific	4.04 (0.68)	42	**0.000**
	Nameable	3.98 (0.60)	43	**0.000**
	Abstract	3.81 (0.87)	37	**0.000**
	Subjective	3.98 (0.76)	43	**0.000**
Satisfied with search interaction	Specific	3.90 (0.75)	36	**0.000**
	Nameable	3.79 (0.90)	35	**0.002**
	Abstract	3.73 (0.89)	34	**0.004**
	Subjective	3.52 (0.88)	26	0.564

practical to submit a new query that would retrieve a new set of relevant images rather than continue viewing a list of irrelevant images.

*Table 5.7: Results of Friedman's test for aspects of satisfaction with an overall statistically significant difference ($p < 0.05$) across search tasks are indicated in **bold***

Aspects of satisfaction	χ^2	p
Image matched initial idea	2.528	0.470
Frequently change mind about image	**8.062**	**0.045**
Satisfied with images	5.253	0.154
Satisfied with order of images	**9.272**	**0.026**
Seen all possible images	5.659	0.129
Satisfied with overall search results	7.723	0.052
Saved images that matched initial idea	4.584	0.205
Satisfied with search interaction	**12.094**	**0.007**

Participants' dissatisfaction was also visible from their ratings on overall search results and search interaction. However, it is surprising to see that participants were saving images that matched their initial idea even though their mean rating in "frequently changing their mind" about images were the highest.

From the eight aspects of performing search that relate to satisfaction, only three showed statistically significant difference as shown in Table 5.7. They were frequently changing minds about images, satisfied with order of images and satisfied with search interaction. However, results of post-hoc test (Table 5.8) show that for the aspect, frequently changing minds about images, there were no significant differences detected between the types of task, suggesting that users often rethink about the relevance of images and it is not task dependent.

In satisfaction with order of images, participants rated that they were less satisfied with images in Subjective tasks as compared to Specific tasks (Z = -3.463, p = 0.001). Using a text-based image search engine, images are retrieved based on their textual annotation and

images for Subjective tasks can have multiple annotations. These annotations together with methods of indexing and ranking may influence how images are ranked in result pages.

*Table 5.8: Results of Wilcoxon signed-rank test for differences on aspects of satisfaction (Values in **bold** indicates statistical significance, $p < 0.0083$)*

Combinations	Frequently change mind about images	Satisfied with order of images	Satisfied with search interaction
Specific to Nameable	Z = -0.041 p =0.967	Z = -1.150 p = 0.250	Z = -0.782 p = 0.434
Specific to Abstract	Z = -1.878 p = 0.060	Z = -1.417 p = 0.156	Z = -1.360 p = 0.174
Specific to Subjective	Z = -2.159 p = 0.031	Z = -3.463 p = **0.001**	Z = -2.751 p = **0.006**
Nameable to Abstract	Z = -1.841 p = 0.066	Z = -0.108 p =0.914	Z = -0.502 p = 0.616
Nameable to Subjective	Z = -2.083 p = 0.037	Z = -1.864 p =0.062	Z = -2.017 p = 0.044
Abstract to Subjective	Z = -0.686 p = 0.493	Z = -2.013 p = 0.044	Z = -1.537 p = 0.124

For satisfaction of search interaction, participants rated that they were less satisfied with the interaction while performing a Subjective task compared to a Specific task (Z = -2.751, p = 0.006). They might not find exactly what they were looking for, despite putting in a lot of effort in the search.

Searching can be a challenging and frustrating task for both novices and experts [Hölscher and Strube, 2000]. Image search expertise is not a major factor considered in this research. However, in Chapter 4, we noted that, there is a difference in the number of participants who considered themselves as novices and experienced image searchers. Therefore, we performed Welch's t-test (variance not assumed to be equal) to identify whether there is a difference between participants image search experience in rating different aspects of performing search.

Table 5.9: Impact of participants' image search expertise in rating aspects of performing search

Aspects of performing search	Novice Mean (SD)	Intermediate/Expert Mean (SD)
Familiar with search topic	2.33 (1.32)	2.26 (1.02)
Able to create queries easily	3.14 (0.85)	2.96 (0.98)
Having an idea of images	3.48 (0.81)	3.15 (1.10)
Image matched initial idea	2.81 (1.12)	3.15 (1.06)
Frequently change mind about images	1.81 (1.36)	1.48 (1.34)
Satisfied with images*	2.71 (1.15)	3.30 (0.72)
Satisfied with order of images	2.62 (1.28)	1.93 (1.41)
Seen all possible images	1.33 (1.39)	1.89 (1.60)
Satisfied with overall search results	2.48 (1.33)	3.11 (1.05)
Saved image that matched initial idea	3.38 (0.97)	3.48 (0.70)
Satisfied with search interaction	2.52 (1.44)	2.89 (1.22)
Ease of performing search	2.71 (1.06)	3.30 (0.72)

Asterisk (*) indicates statistical significance (Welch's t-test $p < 0.05$).

The results in Table 5.9 show that novices and experienced searchers have more or less the same spread in rating. Except for whether users were satisfied with images, there is no significant difference between participants' image search expertise in rating aspects of performing search.

5.3 Time-stamped search interaction of participants

The effect of subjective factors while searching can be detected from participants' search behaviour. In the main user study, we observe participants search behaviour by a mechanism of screen capture software to examine whether their behaviour reflects the answers given in the questionnaire. The observations enabled us to gather information towards understanding effects of different factors on image search behaviour.

Recordings on observations of participants search interaction were manually time-stamped

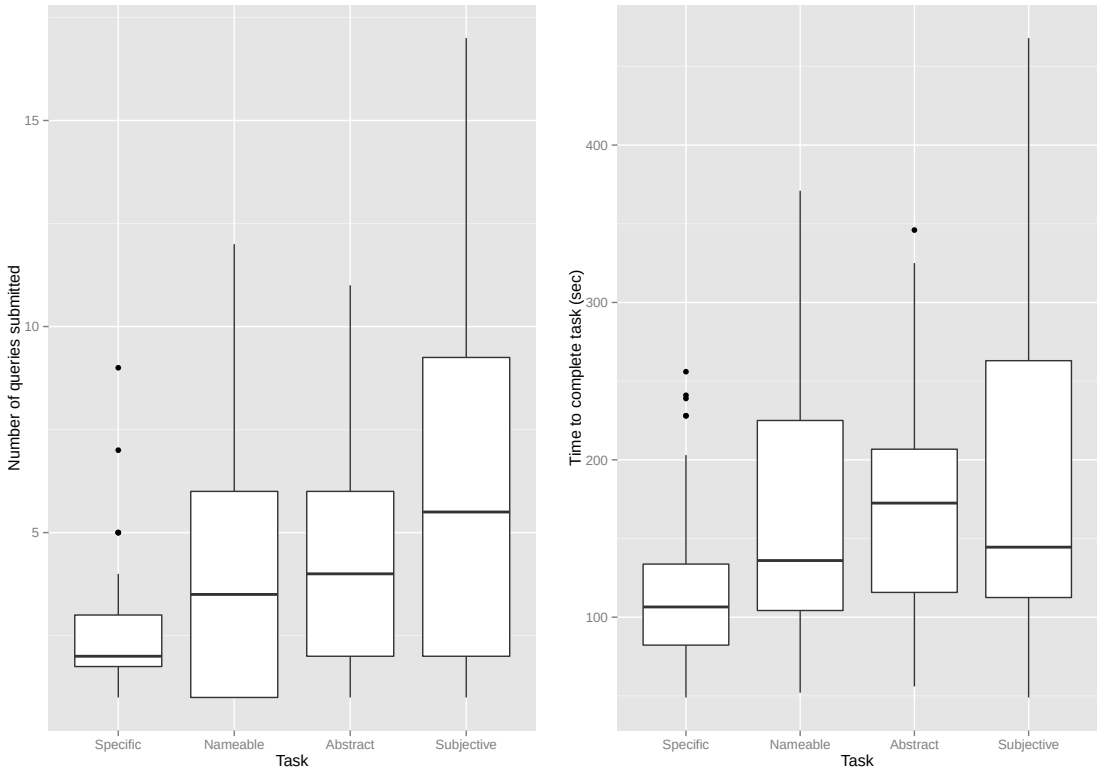

Figure 5.4: Example of measures of user search behaviour: number of queries submitted (left) and time taken to complete task (right)

and examined to distinguish between search and retrieval activity. This includes the time spent viewing image results and the time taken for a participant to complete each task. A *view* is defined as when the participant either hovers the cursor on the result image until the image meta data is displayed or views the image in a new browser tab. Figure 5.4 shows examples of two measures extracted from the screen capture recordings. Both figures suggest that measures in search behaviour increase as the difficulty of tasks increases.

Descriptive statistics of the measures for all types of tasks are listed in Table 5.10. The results show that there is an increase in measures of search behaviour between task types from Specific to Subjective tasks. This suggests that difficulty is increasing from Specific tasks

Table 5.10: Overall mean and standard deviation of users' search interaction for all task types

User search behaviour	Statistics	Specific	Nameable	Abstract	Subjective
Time to submit initial query (s)	μ	9.63	12.48	10.35	7.69
	σ	6.56	17.26	7.84	6.16
Number of queries submitted	μ	2.70	4.10	4.30	6.30
	σ	1.64	2.95	2.65	4.26
Number of results pages viewed	μ	9.30	14.70	14.90	14.50
	σ	5.65	11.14	9.93	12.51
Number of images viewed	μ	11.60	16.30	17.30	16.10
	σ	8.07	11.41	10.55	9.10
Time spent on results pages (s)	μ	108.29	148.40	158.19	171.56
	σ	52.57	74.87	70.35	88.38
Time to complete task (s)	μ	117.92	160.88	168.54	179.25
	σ	53.96	76.86	71.65	89.68

to Subjective tasks. Previous findings from text retrieval have shown that task completion time, reading time and number of queries submitted decreases with familiarity. However, based on the results, increase in familiarity does not necessarily decrease these measures of search behaviour. For detailed results by types of task, refer to Appendix H.

From Table 5.10, we can see that participants took an average of 117.92 seconds to complete a Specific task compared to 179.25 seconds for a Subjective task. Looking at their search interaction, one would assume that if they were familiar with the search topic, they would be able to easily issue an initial query (refer Table 5.1). However, this is not always necessarily the case.

Data from the questionnaire showed that participants were more familiar with topics for the Nameable task compared to topics of other task types. Interestingly, it took them an average of 12.48 seconds to submit their initial query which was longer than the other types

of task. This indicates that searching and judging images is different to text even though the method of searching (use of keywords) is similar. Cunningham and Masoodian [2006] found that searchers often struggle to express Nameable needs as queries.

Examining the time taken to complete a search task, we found that participants face difficulty in completing the Subjective task with an average time of 179.25 seconds compared to the Specific tasks (117.92 seconds). The time recorded reflects the questionnaire data whereby participants rated that it is the most difficult task to perform. This is consistent with findings from previous text retrieval studies which showed that users have a longer completion time when presented with a more difficult task [Aula et al., 2010; Liu et al., 2010c].

From the table, we can see an increase in the number of queries as the task becomes more difficult. Typically, the queries that participants formulate for an easy task contains the main facet from the task description. Even though the same approach can be used for an Abstract and Subjective task, the results may not be satisfactory as it covers a broad area and the representation can be varied.

There are two obvious approaches participants used when searching for images. First, participants submitted only one query, going back and forth down the list, viewing all images. The percentage of sessions that contain a single query is 21.9% and it is used most often for searching Nameable images as shown in Table 5.11.

Second, participants submitted multiple queries and viewed the retrieved results. With this approach, when unsatisfied or unsuccessful, they reformulated their query to retrieve a new set of results. The process is repeated until, in their mind, they have fulfilled the

Table 5.11: Distribution of queries for each type of task

Number of queries	**Task type**			
	Specific	Nameable	Abstract	Subjective
1	12	14	11	5
2	15	4	4	8
3	10	6	4	3
4	6	4	7	3
>= 5	5	20	22	29

task's requirements. The more time they took in judging relevance, the longer it will take for them to complete the task. When users changed their mind, viewing more images extended the search session, however the length of time depends on the type of images they were looking for and the variety of images they were presented with. The percentage of sessions where participants submits multiple queries is 78.1% and it is used most often for searching Subjective images as shown in Table 5.11.

In addition, we conducted repeated measures ANOVA on the time-stamped data to investigate significant difference in the measures between types of task. Results in Table 5.12 show that except for time to submit initial query, the measures were significantly influenced by the types of task performed. Therefore, we performed a post-hoc test whereby we compare each task type with every other task. The Bonferroni correction (p <0.0083) was used on the p-value to identify where the significant difference between tasks is likely to occur.

By looking at the means (μ in Table 5.10), we can see that participants submitted significantly more queries for Nameable tasks compared to Specific tasks ($p = 0.002$) but not compared to Abstract tasks ($p = 1.000$). Participants submitted significantly more queries for Abstract tasks compared to Specific tasks ($p = 0.0003$). Finally, participants submit-

*Table 5.12: Results of repeated measures ANOVA on measures of search behaviour across search tasks (Values in **bold** indicate statistical significance)*

Time stamped search behaviour	F	p
Number of queries submitted	**17.123**	**<0.001**
Number of results pages viewed	**9.537**	**<0.001**
Number of images viewed	**6.972**	**<0.001**
Time spent on results pages	**11.904**	**<0.001**
Time to complete task	**11.119**	**<0.001**

ted significantly more queries for Subjective tasks compared to Specific ($p = 0.000002$) or Nameable ($p = 0.005$) or Abstract ($p = 0.005$) tasks.

Results of the post-hoc test did not detect any significant difference on the number of results pages viewed for Nameable tasks when compared to both Abstract and Subjective tasks. However, the number of result pages viewed was significantly more for Abstract tasks compared to Specific tasks ($p = 0.004$) but not compared to Subjective tasks ($p = 0.147$). The number of result pages viewed was significantly more for Subjective tasks compared to Specific tasks ($p = 0.000002$).

There is also no significant difference in the number of images viewed for Nameable tasks when compared to Specific, Abstract or Nameable tasks. The number of images viewed was significantly more for Abstract tasks compared to Specific tasks ($p = 0.0001$) but not compared to Nameable or Subjective tasks (both $p = 1.000$). Number of images viewed was significantly more for Subjective tasks compared to Specific tasks ($p = 0.010$).

The time spent on result pages was significantly longer for Nameable tasks compared to Specific tasks ($p = 0.0003$) but not compared to Abstract ($p = 1.000$) or Subjective tasks ($p = 0.380$). Participants spent significantly more time on result pages for Abstract

tasks compared to Specific tasks ($p = 0.000004$) but not compared to Subjective tasks ($p = 1.000$). The time spent on result pages was significantly longer for Subjective tasks compared to Specific tasks ($p = 0.00004$).

Lastly, time to complete the task was significantly longer for Nameable tasks compared to Specific tasks ($p = 0.0003$) but not compared to Abstract ($p = 1.000$) or Subjective tasks ($p = 0.886$). Participants spent significantly more time completing search for Abstract tasks compared to Specific tasks ($p = 0.000005$) but not compared to Subjective tasks ($p = 1.000$). The time to complete the task was significantly longer for Subjective tasks compared to Specific tasks ($p = 0.00009$).

Analysis of results shows that users' search behaviour was significantly different when completing difficult tasks as compared to easy tasks. Users were submitting more queries, viewing more results pages, viewing more images and spending more time on results pages. All of these measures are the process that one goes through to find and judge an image that satisfies their information need. Repetitions of these search behaviours results in longer completion time.

5.4 Relationship between aspects of performing search and time-stamped search behaviour

This section examines the direction and strength of the relationship between users' ratings on perceived attitudes of factors affecting search and their actual searching behaviour. Ordinal data (i.e., Strongly Disagree (1), Disagree (2), Neutral (3), Agree (4), and Strongly Agree (5)) were collected for perception of factors affecting search. Therefore, to determine the

correlation between these variables, we calculated Spearman's rank correlation coefficients. There are various guidelines for the interpretation of a correlation. Cohen [1988] has noted that the criteria are in some ways arbitrary and should not be observed too strictly. Twelve aspects of performing searches were used to assess the three subjective factors (familiarity, difficulty and satisfaction) affecting image search behaviour.

5.4.1 Topic familiarity

For the aspect of *familiarity with the search topic*, we were expecting a negative correlation for most measures of user search behaviour. However, only a few were negative and it was too small and insignificant. Although participants may be familiar with a search topic, this does not mean it would be easier for them to create an initial query. As we have highlighted in the previous section, participants were facing difficulty creating queries even though they rate the topic as familiar.

One possible reason may be that they used a keyword from the description of the work task to create a query. Previous studies have shown that users often have difficulty expressing their needs. This difficulty is a contributing factor to the number of queries they submitted, or they simply disliked the results of previous queries. Obviously, the more queries users submit, the more results pages they would have to view.

The analysis shows that familiarity does not correlate well with the number of images viewed. Users view images not just because of familiarity but most importantly, because they are making judgements on the relevance of that image for the given context of the task they were completing. Similarly, familiarity is not a direct indication that users would take

Table 5.13: Spearman rank correlation, ρ, values of topic familiarity on participants' search behaviour for all types of task

Task type	Time to submit initial query	Number of queries submitted	Number of result pages viewed	Number of images viewed	Time spent on result pages	Time to complete task
Familiar with search topic						
Specific	0.100	0.018	-0.095	-0.095	-0.034	0.013
Nameable	-0.198	0.201	0.138	0.138	-0.012	-0.088
Abstract	-0.027	0.234	0.100	0.100	0.064	0.054
Subjective	-0.084	-0.137	*-0.315*	*-0.315*	-0.201	-0.187
Having an idea of image						
Specific	0.011	0.214	-0.088	0.000	0.016	0.020
Nameable	0.036	-0.134	-0.220	0.002	-0.222	-0.201
Abstract	-0.064	-0.017	-0.070	0.086	-0.082	-0.073
Subjective	-0.029	*-0.403*	*-0.419*	*-0.328*	-0.449	*-0.442*

Note:
Values in **bold** indicate large correlation, in *italics* indicate medium correlation

less time to complete the search. Comparing the correlation of familiarity across all measures for all task types, we can say that participants were most affected by topic familiarity in the Subjective task as there is a consistent negative correlation for all measures of their search behaviour.

We associate *having an idea of image* with topic familiarity because we expect that if participants were familiar with a topic, they would have an idea of images relevant to the topic. However, significant negative correlation exists for only the Subjective task. Data from Table 5.1 shows that most users agreed that they have an idea on the kind of images they were looking for, but as with topic search familiarity, they were facing a difficulty expressing that idea in the form a query. When the idea is not expressed clearly, users iteratively reformulate their queries to get better results.

Users then went through the result pages, looking for images that were similar to their ideas. It is not surprising that participants were viewing more images for the Subjective

task. The length of this process and eventually the search itself depends on whether their idea and the image results match. However, there was no significant correlation between having images that matched their initial idea with the number of images they viewed.

5.4.2 Task difficulty

For correlation between users' ability to create queries easily and measures of search behaviour, a few significant negative correlations exist (Table 5.14). How fast participants submit an initial query is not a clear indicator of their ability to create queries easily. However, users submitting more queries may be a sign that they were facing difficulty in creating effective queries, particularly for the Subjective task. Consequently, this affects the amount of time they spent on result pages as they needed to view results pages and images for each query they have submitted, prolonging their search duration.

Table 5.14: Spearman rank correlation, ρ, values of task difficulty on participants' search behaviour for all types of task

Task type	Time to submit initial query	Number of queries submitted	Number of result pages viewed	Number of images viewed	Time spent on result pages	Time to complete task
Able to create queries easily						
Specific	-0.047	-0.199	*-0.336*	-0.263	-0.247	-0.220
Nameable	-0.267	0.196	-0.054	-0.074	-0.160	-0.248
Abstract	0.125	-0.201	*-0.308*	-0.021	-0.152	-0.112
Subjective	-0.114	**-0.576**	*-0.495*	*-0.385*	**-0.536**	**-0.538**
Ease of performing search						
Specific	-0.046	-0.209	*-0.342*	-0.145	*-0.317*	-0.303
Nameable	-0.076	0.008	-0.274	-0.048	-0.227	-0.217
Abstract	-0.055	-0.156	-0.216	-0.135	-0.257	-0.233
Subjective	0.060	**-0.609**	*-0.489*	-0.220	**-0.572**	**-0.567**

Note:
Values in **bold** indicate large correlation, in *italics* indicate medium correlation

With regards to the aspect *ease of performing search*, the negative correlations with the

measures from participants' search behaviour were not as significant as anticipated. Similar to satisfaction, efforts exhibited from participants' search behaviour correlate with ease of performing search especially for the Subjective task. When facing difficulties, users would reformulate their queries to retrieve better results. This process is done iteratively until they are satisfied with the results, thus making it longer to complete.

5.4.3 Satisfaction

Satisfaction is a subjective factor that is difficult to measure. We attempted to find the correlation of different levels of satisfaction from participants search behaviour. It might be expected that if the search engine retrieved *images that matched users' initial ideas*, users would view fewer result pages. However, there was no significant correlation between having an image match the initial idea with number of images they viewed. This suggests that although the returned image results matched their ideas, it is not a factor in judging relevance. Clearly, for Subjective images, having an image that matched the initial idea meant that participants viewed significantly fewer result pages, spent less time on them and completed the task faster.

Making decisions about images was not easy — participants *frequently changed their minds about the images* they were looking for. One indication of them changing their mind is from the queries they submitted. Table 5.15 shows significant medium positive correlations for three task types demonstrating that participants frequently change their minds about images. Another indicator is from the number of images they viewed. Again, there are significant medium positive correlations for all task types. Viewing more images extended

Table 5.15: Spearman rank correlation, ρ, values of satisfaction with participants' search behaviour for all types of task

Task type	Time to submit initial query	Number of queries submitted	Number of result pages viewed	Number of images viewed	Time spent on result pages	Time to complete task
Image matched initial idea						
Specific	-0.103	0.045	-0.157	-0.046	-0.021	-0.026
Nameable	-0.071	-0.030	-0.103	0.151	-0.017	-0.098
Abstract	-0.053	-0.219	-0.172	0.077	-0.165	-0.185
Subjective	-0.140	-0.257	*-0.322*	-0.185	*-0.364*	*-0.369*
Frequently change mind about images						
Specific	-0.010	*0.426*	0.044	*0.366*	0.268	*0.308*
Nameable	0.009	*0.318*	0.221	*0.340*	0.297	0.170
Abstract	-0.190	0.248	0.248	0.165	0.218	0.233
Subjective	-0.101	*0.328*	0.271	*0.311*	*0.318*	0.231
Satisfied with images						
Specific	-0.102	-0.072	*-0.391*	-0.014	-0.192	*-0.361*
Nameable	-0.186	0.229	-0.170	0.138	-0.042	-0.250
Abstract	-0.182	-0.097	-0.206	-0.022	-0.229	-0.072
Subjective	-0.203	-0.294	*-0.356*	-0.158	*-0.352*	-0.221
Satisfied with order of images						
Specific	-0.044	-0.206	-0.067	-0.067	-0.206	-0.217
Nameable	-0.088	0.072	*0.310*	*0.310*	0.072	0.002
Abstract	-0.134	-0.051	0.121	0.121	-0.051	-0.068
Subjective	-0.151	-0.234	0.067	0.067	-0.234	-0.243
Seen all possible images						
Specific	-0.038	0.065	-0.011	-0.011	-0.100	-0.125
Nameable	-0.267	0.025	-0.166	0.027	-0.085	-0.140
Abstract	-0.100	-0.066	-0.259	0.019	-0.167	-0.199
Subjective	0.004	-0.055	-0.083	0.063	-0.092	-0.099
Saved images that matched initial idea						
Specific	0.083	0.081	-0.108	0.027	-0.109	-0.104
Nameable	-0.281	-0.056	-0.237	-0.221	*-0.405*	*-0.460*
Abstract	-0.058	-0.224	-0.270	-0.131	*-0.332*	*-0.318*
Subjective	-0.113	*-0.452*	**-0.510**	*-0.386*	**-0.516**	**-0.513**
Satisfied with overall search results						
Specific	-0.061	-0.067	*-0.379*	-0.280	-0.257	-0.257
Nameable	*-0.360*	0.236	-0.221	-0.011	-0.198	-0.286
Abstract	-0.143	-0.074	-0.121	-0.040	-0.173	-0.197
Subjective	0.036	-0.298	-0.238	0.064	-0.268	-0.275
Satisfied with search interaction						
Specific	-0.165	-0.203	*-0.342*	-0.123	-0.278	*-0.304*
Nameable	-0.204	-0.052	-0.202	0.029	-0.184	0.237
Abstract	-0.197	-0.103	-0.151	-0.024	-0.207	-0.233
Subjective	0.009	*-0.433*	*-0.320*	-0.113	*-0.391*	*-0.396*

Note:
Values in **bold** indicate large correlation, in *italics* indicate medium correlation

the search session, however the length of time depends on the type of images they were looking for and the variety of images they were presented with.

Participants who were not *satisfied with the images presented* to them would submit more queries to retrieve better results, or view more result pages. Next is the *order in which the results were presented*. It could be expected that unsatisfied users would submit more queries, but in fact there is no strong correlation between these variables. Surprisingly for Nameable tasks, participants' satisfaction increased as they viewed more result pages. One possibility is that images that they found relevant appeared near the top of result pages. Nevertheless, there is no strong or significant correlation between seeing all possible images with any of the search behaviour for all task types.

For the aspect *saved images that matched initial idea*, there were quite a few significant correlations. Users gave more effort in the amount of time spent viewing results pages and completing the Nameable, Abstract and Subjective tasks in order to save images that matched their initial idea. With regards to *satisfaction with overall search results*, participants submitted more queries on the Subjective task as an indication of dissatisfaction with overall search results, while they viewed more result pages when unsatisfied with search results for the Specific task. As for *satisfaction with search interaction*, we can see that participants' satisfaction decreases with increases in the need to submit queries, view more result pages, view more images, spend time on result pages or spend time to complete the task. This suggests that apart from search engine capabilities, the amount of effort they gave affects their search satisfaction as well.

Comparing all the different aspects of performing searches and their correlations with

user search behaviour, we have shown that for certain aspects, there are search behaviours that correlate more significantly than others. This indicates that time-based measures can be used to gauge some aspects or factors that affect searching behaviour, for example, number of queries and task completion time has been repeatedly used as a measure of both familiarity and difficulty for textual information search tasks [Kelly and Cool, 2002; Aula et al., 2010; Liu et al., 2010c; Qu et al., 2010; Singer et al., 2013] — however this depends on the type of search task.

In image search, users may issue more queries because they want to further diversify their pool of image results before making a decision. Likewise, longer completion time may mean that users are not satisfied and taking their time in making relevance judgments. From Tables 5.13 to 5.15, we can see that different factors affect task type differently. The Subjective task was most affected by these factors. Therefore, a single measure of users' search behaviour is not a conclusive measure on the effect of subjective factor on image searching.

5.5 Summary

In this chapter, we discussed the examination of users' search interaction and time stamped the interactions to determine distinct search and retrieval activity. A comparison was conducted between users' search interaction and their responses to the post-task questionnaires. It seems that users' perception on aspects of performing image searches does not always correspond with their actual search behaviour. Relating these subjective factors to the objective measures is difficult because it is affected by task type. The results show that there are some aspects of the search, such as difficulty, that are clearly measurable by examining

users' search interaction, while others may need different methods of measurement. Indeed, aspects such as familiarity and satisfaction might only be measured subjectively. Therefore, search engine operators are recommended to create a mechanism that uses these measures of search behaviour to assist users in fulfilling their image information needs rather than settling or stopping halfway.

Chapter 6

Inferring image information needs from queries

Examination of users' image searching behaviour aims at providing an understanding of how users search for images. One of the most effective methods is the analysis of search interaction data collected in search engines' transaction logs [Jansen et al., 2008]. People mine logs because they capture actual user behaviour. From the logs, researchers are able to unobtrusively record large amounts of user-system interaction in a natural setting over significant periods of time [Jansen et al., 2008]. Researchers who have used search logs to gain understanding of users' image needs and searching behaviour include Markkula and Sormunen [2000], Jansen et al. [2000a], Pu [2005], and Tseng et al. [2009]. Mining information from search logs is the primary method used by large commercial search engines to characterise and analyse the needs and behaviour of their users [Silvestri, 2010]. Furthermore, it contributes to the understanding of the needs and behaviour of their users. Therefore, in this chapter,

we address the third research question of inferring users' visual information needs through their queries.

The chapter is organised as follows: description of the data collection (Section 6.1), followed by details on log pre-processing were performed (Section 6.2) and data analysis (Section 6.3). Next, is the discussion of the results on Google image search (Section 6.4), Bing image search (Section 6.5) and from the user study (Section 6.6). Followed by a comparison of the three data sets (Section 6.7). Section 6.8, summarises the chapter.

6.1 Data collection

A log is a file that records the interaction between a system and the users of that system. Logs have been used to enhance the efficiency and effectiveness of a system [Orlando and Silvestri, 2009]. Unlike other logs, a proxy log records access to pages all over the net. Proxy servers can be configured to log information about user requests, and the responses provided to those requests by servers on the Internet. Each log entry reveals information about the client making the request, the date and time of the request, and the identifier of the object requested.

The proxy log data in this research was extracted from RMIT University's proxy server with assistance of the university's Information Technology Services (ITS). The log extraction process were conducted in accordance to approved conditions as outlined by the university's Ethics Committee. Specifically, the data used were collected over a period of one month during part of one semester (1st October to 31st October 2011)[1].

[1]Entries from 30th September were also included but later removed during preprocessing stage during preprocessing for convenient alignment to a single calendar month.

From the request, we have acquired two sets of proxy log data — a large collection of user interactions with Google Images and a much smaller collection of user interactions with Bing Images. The IP addresses anonymisation on both the Google and Bing data was conducted by RMIT's ITS. Unfortunately, the Google data was totally anonymised as opposed to pseudo-anonymised. Details of the initial number of entries in both sets of log data are listed in Table 6.1.

Table 6.1: Information on logs analyzed

Search engine log	Number of entries
Google Images	1 622 328
Bing Images	15 467

6.1.1 Google log

From the proxy log data, we were particularly interested in entries that indicate interaction of users searching for images. Firstly, patterns of Google Images URLs that indicate searching for images were identified. The URLs patterns included URLs of search queries, search results pages that the user views and images that they might have clicked from the results pages. The URLs were identified by examining the changes of URLs when performing image searches. Based on the patterns, RMIT's ITS used the following given awk script to extract all entries corresponding to Google Images searches from the proxy log.

```
awk '/www.google.com.au\/img[hp|res]/||/www.google.com.au\/search/&&
/tbm=isch/ {print $0}' logfiles
```

After the entries have been extracted, their IP adressess were anonymised by RMIT ITS

using a Perl script from the CPAN forum (Appendix I). The script uses the IP::Anonymous module and a private key of any random number between 0 to 255 to anonymise the IP addresses. However, the lack of user information, that is IP Address, does not allow us to perform session identification. Shown below is an example entry extracted from the Google proxy logs.

```
2011-10-31 00:32:35 164 x.x.x.x 0 0 0 OBSERVED No-Authentication-URLs
http://www.google.com.au/search?q=edmond+and+corrigan&oe=utf-8&rls=
org.mozilla:en-GB:official&client=firefox-a&um=1&ie=UTF-8&hl=en&tbm=
isch&source=og&sa=N&tab=wi&biw=1920&bih=1006&sei=%20GeytTpuHIMSAmQXLlrzRDg
200 TCP_NC_MISS GET image/jpeg http t1.gstatic.com 80 /images
?q=tbn:ANd9GcSV__WPvR75aiHeZwN9P578CUHMjdSiTZlo5R-ZsxHGes88LjyVc4v5xdDp 0
"Mozilla/5.0 (Windows; U; Windows NT 5.1; en-GB; rv:x.x.x.x) Gecko/20110920
Firefox/3.6.23 (.NET CLR 3.5.30729)" x.x.x.x 8709 756 0
```

6.1.2 Bing log

Similarly, the patterns of URLs that indicate image searching were identified for Bing Images and we requested RMIT's ITS to use the following grep command to extract all image associated entries from Bing Images within the proxy log.

```
grep http://www.bing.com/images/search?
```

Again, the IP adresses of the entries were anonymised by ITS using the same Perl script. An example of an entry from the Bing log is given below.

```
2011-10-01 16:55:17 186 2.157.151.209 200 TCP_HIT 8121 559 GET http
ts1.mm.bing.net 80 /images/thumbnail.aspx ?q=1239089551416&id=
afadf06d4d640dccf49b844d501b70c3&url=http%3a%2f%2fguti.bitacoras.com
%2fwp-content%2fuploads%2f2004%2f10%2fborlandc.png - - DIRECT
141.10.163.9 image/jpeg http://www.bing.com/images/search?q=
c%2b%2b+builder+5&FORM=BIFD "Mozilla/5.0 (Windows NT 5.1; rv:7.0.1)
Gecko/20100101 Firefox/7.0.1" OBSERVED "none" - 236.37.157.60  -
```

6.1.3 User study data

The user study data used were from the same user study in Chapters 4 and 5. Users' information needs from the user study were collected through observations of participants' search interaction. Queries submitted by each participant for all task types were manually extracted from the screen capture recordings.

6.2 Data pre-processing

Prior to performing analysis, we conducted some pre-processing on the logs. The steps involved in the pre-processing stage are depicted in Figure 6.1. We begin by first, identifying and parsing the proxy log field format. RMIT's proxy server uses the Blue Coat field format summarised in Table 6.2. However, the order of field format in the logs sometimes varies.

Next, we perform the data cleaning step as entries in the logs may contain corrupted data due to errors when logging the data. Cleaning was performed by removing entries/lines

Table 6.2: Blue Coat proxy log field format

Field	Description
`date`	Date at which transaction completed
`time`	GMT time in HH:MM:SS format
`time-taken`	Time taken (in miliseconds) to process the request
`c-ip`	IP address of the client (IP anonymised)
`sc-status`	Protocol status code from appliance to client
`s-action`	What type of action did the appliance take to process the request
`sc-bytes`	Number of bytes sent from appliance to client
`cs-bytes`	Number of bytes sent from client to appliance
`cs-method`	Request method used from client to appliance
`cs-uri-scheme`	Scheme from the 'log' URL
`cs-host`	Hostname from the client's request URL. If URL rewrite policies are used, this field's value is derived from the 'log' URL
`cs-uri-port`	Port from the 'log' URL
`cs-uri-path`	Path from the 'log' URL. Does not include query
`cs-uri-query`	Query from the 'log' URL.
`cs-username`	Relative username of a client authenticated to the proxy (i.e. not fully distinguished)
`cs-auth-group`	One group that an authenticated user belongs to. If a user belongs to multiple groups, the group logged is determined by the Group Log Order configuration specified in VPM. If Group Log Order is not specified, an arbitrary group is logged. Note that only groups referenced by policy are considered
`s-hierarchy`	How and where the object was retrieved in the cache hierarchy
`s-supplier-name`	Hostname of the upstrean host (not available for a cache hit)
`rs(Content-Type)`	Response header: Content-Type
`cs(Referer)`	Request header: Referer
`cs(User-Agent)`	Response header: User-Agent
`sc-filter-result`	Content filtering result: Denied, Proxied or Observed
`cs-category`	Single content category of the request URL (a.k.a. sc-filter-category)
`x-virus-id`	Identifier of a virus if one was detected
`s-ip`	IP address of the appliance on which the client established its connection
`s-sitename`	The service type used to process the transaction

Figure 6.1: Log pre-processing steps

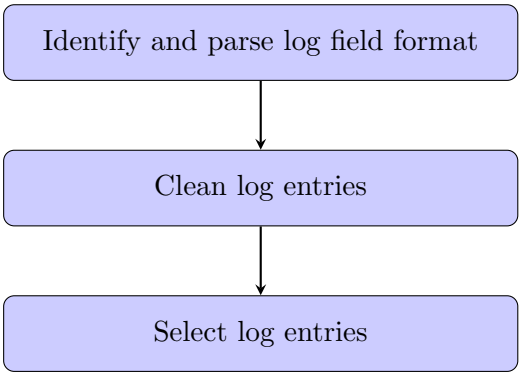

Table 6.3: Number of log entries after pre-processing with number of actual queries

Search engine log	Number of entries	Number of queries
Google Images	824 840	2 980
Bing Images	14 818	105

which either do not fit the pattern of data in each of the field or were not actually queries. We then filter the data by date to select entries only from 1st October to 31st October 2011. Table 6.3 shows the number of entries analysed after pre-processing with respective number of actual queries.

Number of entries in the table refers to URLs patterns which included URLs of search queries, search results pages that the user views and images on the results pages. However, low numbers of actual queries were due to most of the entries being generated by the returned image results list and not queries submitted by users. Details of query analysis will be discussed in Sections 6.4.1 and 6.5.1.

6.3 Data analysis

The third research question asks whether we can identify users' information needs solely from their queries. Therefore, in identifying those needs from their queries, we must first identify user sessions and queries within each session. In identifying user sessions and inferring needs, we adopt the session, query and term definition of Jansen et al. [2000b].

6.3.1 Session analysis

For the purposes of this study, a *session* is defined as a series of requests from a single IP address with not more than 30 minutes passing between individual requests. If more than 30 minutes passes between requests from that IP address, the next request from that address marks the beginning of a separate user session. Thirty minutes was chosen as a cut off time after experimenting with two cut off times: 5 and 30 minutes.

Using a cut off time of five minutes, we noticed that the same requests from the same IP address would be in a separate session. On the other hand, a cut off time of 30 minutes would include significant difference in requests from the same IP address. Accordingly, as we had previously made the assumption that a user has only one image information need during a search session, we acknowledge that these differences could be an indication that users have had a change in their information needs. *Queries* are defined as the complete strings that are submitted in the search box. While, parts of the queries separated by whitespace are called *terms*.

As the Blue Coat log format does not have any cookies, we used the `c-ip` and `cs(User-Agent)` field to identify and distinguish sessions. For each entry with the same

IP address in the Bing log, we examine if the user agent is identical or not. Entries with identical IP address and user agent were considered as a session while entries with a different user agent were considered a separate session. Furthermore, for entries with identical IP address and user agent, we review their cut off time to differentiate sessions. If more than 30 minutes passes between the previous and subsequent entry, the subsequent entry is considered as the start of a new session. Session analyses were conducted only on the Bing log data because the Google log was totally anonymised.

6.3.2 Query analysis

After identifying and distinguishing user sessions, we proceed by extracting queries submitted in each session. To extract the queries submitted by the user in each session, we analysed entries in the `cs(Referer)` and `cs-uri-query` field of the logs. In analysing users' queries, we accept what users have entered, including misspellings. We do however, disregard capitalisation and remove any use of punctuation such as comma, colon, semicolon and hyphen.

Once information on sessions, queries and terms have been identified, we allocated a significant amount of time to manually categorise the information needs (queries in a session) and queries according to Batley's [1988] visual information needs. The categorisation is solely based on the queries, as we have no knowledge of the user's actual search context. Query analysis was conducted on all data: Google log, Bing log and user study data. However, query categorisation was only performed on the Google and Bing log data as queries from the user study were already categorised.

6.3.3 Query reformulation

Analysis on query reformulation is important as queries are the primary expression of users' information needs. Users reformulate their queries to make the result set larger, smaller or more diverse. However in this research, query reformulation is also used to indicate possible changes in information needs. The changes may be identified by how the query is modified and the terms(s) used in expressing the new information need.

Similar to the previous works about query reformulation type [Lau and Horvitz, 1999; Tseng et al., 2009], we adopted five reformulation categories based on the common and different search terms used in two successive queries: *New*, *Add*, *Remove*, *Replace*, and *Reformulate*.

- ***New (N)***: A query for a topic not previously searched for by this user within the scope of the data set. The initial query was classified in this category. Q_i and Q_{i+1} do not contain any common terms.

- ***Add (A)***: A query on the same topic as the previous query, but seeking more specific information than the previous query; one or more keywords have been added to the query and disregards word order. Q_i and Q_{i+1} contain at least one term in common; Q_{i+1} contains more terms than Q_i (depth).

- ***Remove (R)***: A query on the same topic as the previous query, but seeking more general information than the previous query; one or more keywords have been deleted from the query. Word order is disregarded. Q_i and Q_{i+1} contain at least one term in common; Q_{i+1} contains fewer terms than Q_i (breadth).

- ***Replace (Rp)***: Q_i and Q_{i+1} contain at least one term in common and at least one different term.

- ***Reformulate (Rf)***: A query on the same topic that can be viewed as neither a generalisation nor a specialisation, but a reformulation of the prior query, including the following cases:

  - changing the order of words (Repeat: Q_i and Q_{i+1} contain exactly the same terms; the order of these terms may be different),
  - changing spelling and/or verb tense,
  - changing words from singular to plural or plural to singular.

Analysis of query modification was conducted on the Bing log and on the user study data.

6.4 Results of analysis on Google log

6.4.1 Query- and term-level analysis

As mentioned in Section 6.1, the IP addresses in the Google log entries were completely anonymised. Therefore, we were only able to perform query and term level analysis. From the log entries, we identified 2980 queries. From these queries, 2525 were distinct and 111 of the queries are non-English queries; 64 of which were non-English characters/alphabets. The non-English queries were in Chinese, Korean, Japanese, Vietnamese, Arabic, Malay/Indonesian and French. As the analysis focuses on English queries, all 111 non-English queries were excluded. Exclusion was done because a non-English term may not necessarily represent one single term in English.

Table 6.4: Query and terms data of the Google log

Statistics	All queries	Distinct queries
Number of queries	2980	2525
Total number of terms	7670	6549
Minimum number of terms	1	1
Maximum number of terms	14	14
Average number of terms per query	2.6	2.6

The analysis shows that users used 6549 terms with an average of 2.6 terms per query (Table 6.4). The average number of terms per query is lower compared to findings of Goodrum and Spink [2001] (3.74 terms per query) and Jansen et al. [2004] (4 terms per query). Figure 6.2 shows the percentage of different query lengths for the Google log. It can be seen that queries with two terms has the highest percentage (40.3%) followed by queries with three terms (24.2%) and single term query (16.4%).

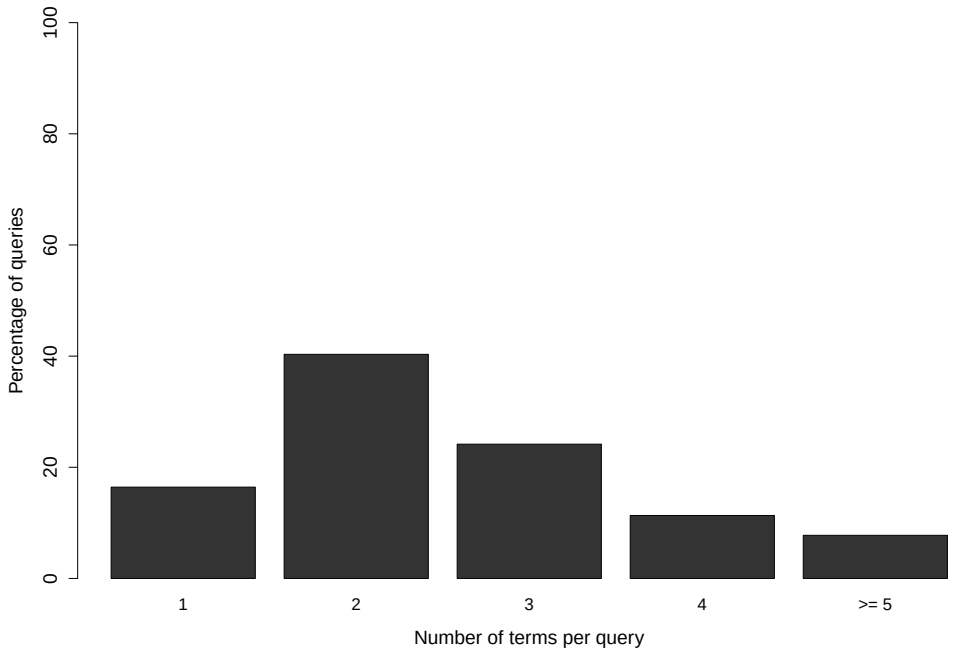

Figure 6.2: Percentage of terms per query in Google log

6.4.2 Image information needs

The objective of the third research question is to infer users' image information needs from their queries. Therefore, queries were manually classified according to image information needs as identified by Batley [1988]. With the Google log, we only classify individually identified distinct queries. We observed that 30.8% of queries were Specific, 64.3% were Nameable, 4.2% were Abstract and less than 1% were Subjective needs (Table 6.5). The highest number of queries in the log were of Nameable needs compared to Specific needs, which is a contrast to the findings of Cunningham and Masoodian [2006] and Chung and Yoon [2010]. However, it is interesting that Abstract needs have the highest average number of terms per query compared to other information needs. Table 6.6 lists the top 10 queries in the Google log for each category of image information need.

Table 6.5: Distinct queries from Google log by image information needs

Statistics	Specific	Nameable	Abstract	Subjective
Number of distinct queries	778	1624	105	18
Total number of terms	2279	3848	366	56
Minimum number of terms	1	1	1	2
Maximum number of terms	14	13	11	6
Average number of terms per query	2.9	2.4	3.5	3.1

We noted that the proportions of 1-, 2-, 3- and 4-word queries might be different because these are image queries. For example, "Melbourne cup" is a single conceptual unit and is distinct from both "Melbourne" and "cup". In contrast, "cute kitten" is a combination of two single concepts. Therefore, it could be argued that "Melbourne cup" is equivalent to a single-word query. However, in the analysis, we considered terms in a query as separate conceptual units.

Table 6.6: Top 10 queries from Google log for each category of image information needs

Specific		Nameable	
Frequency	Query	Frequency	Query
7	*alice in wonderland*	7	*couples silhouette*
5	*federation square*	7	*glass*
5	*yarra trams*	6	*basketball players in action*
3	*ali farahani*	5	*library*
3	*burnie tasmania*	5	*sand*
3	*gordon matta clark*	4	*ballerina*
3	*mary longrigg*	4	*clouds*
3	*maurizio cattelan drawings*	4	*designer rugs hotel*
3	*melbourne cup*	4	*glass design*
3	*seriphos island*	4	*gym people*

Abstract		Subjective	
Frequency	Query	Frequency	Query
3	*earthy and delight*	2	*beautiful interior with acrylic paint*
3	*pathway kindergarten*	2	*beautiful scientific posters*
2	*existing*	2	*cute bike*
2	*humidity in a box*	1	*cute baby animals*
2	*have harp will travel*	1	*cute cartoon pigs*
1	*7 secret success*	1	*cute kitten*
1	*absolute power*	1	*cute pigs*
1	*all the way through evening*	1	*cute puppy*
1	*centre of gravity*	1	*best bulding*
1	*dont waste my time*	1	*best san serif fonts*

6.5 Results of analysis on Bing log

6.5.1 Session-, query- and term-level analysis

Unlike the Google log, the entries in the Bing log have had their IP address pseudo-anonymised, which enabled session identification. Therefore, from the analysis of the Bing log, we have identified 49 user sessions. Users of these sessions submitted a total of 105 queries (96 distinct queries) with an average of 2.1 queries per session (Table 6.7), which is lower than the average session length reported by Goodrum and Spink [2001] (3.36 queries per session), Jansen

et al. [2004] (4 queries per session), and Tjondronegoro et al. [2009] (2.8 queries per session). However, in more than half of the sessions, users submitted only a single query (Figure 6.3). Furthermore, there is a consistent decline in percentage of sessions as the number of queries increases.

Table 6.7: Session and query data of the Bing log

Statistics	Bing logs
Number of sessions	49
Number of queries	105
Minimum number of queries	1
Maximum number of queries	24
Average number of queries per session	2.1

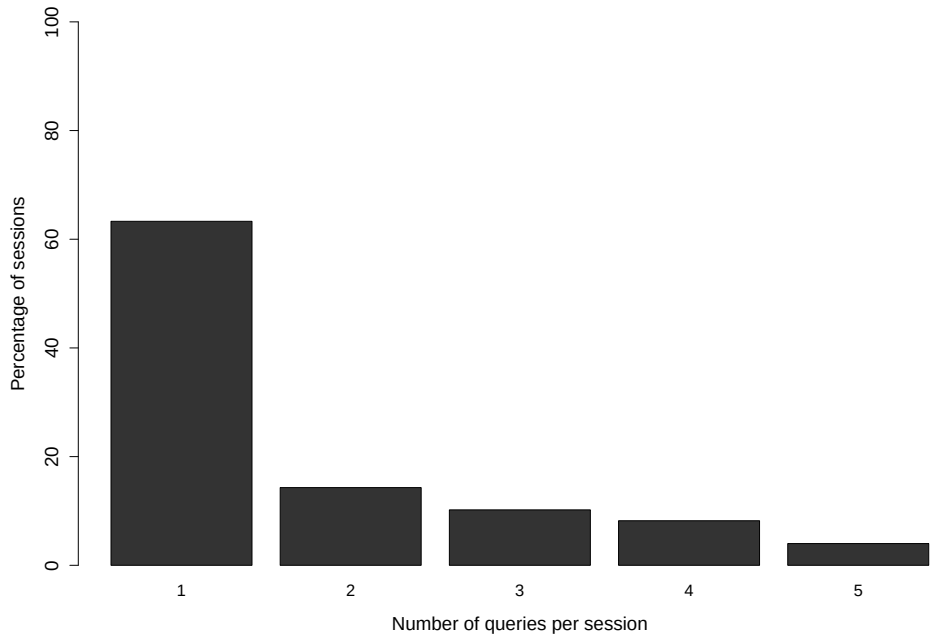

Figure 6.3: Percentage of queries per sessions in Bing log

Query level analysis was then performed on the Bing log. Similar to the Google log, two non-English queries were excluded and the remaining 103 were analysed. The analysis

showed that users used 251 terms with an average of 2.4 terms per query (Table 6.8). Figure 6.4 displays the percentage of queries of different lengths in the Bing log.

Table 6.8: Query and terms data of the Bing log

Statistics	Bing log
Number of queries	103
Total number of terms	251
Minimum number of terms	1
Maximum number of terms	5
Average number of terms per query	2.4

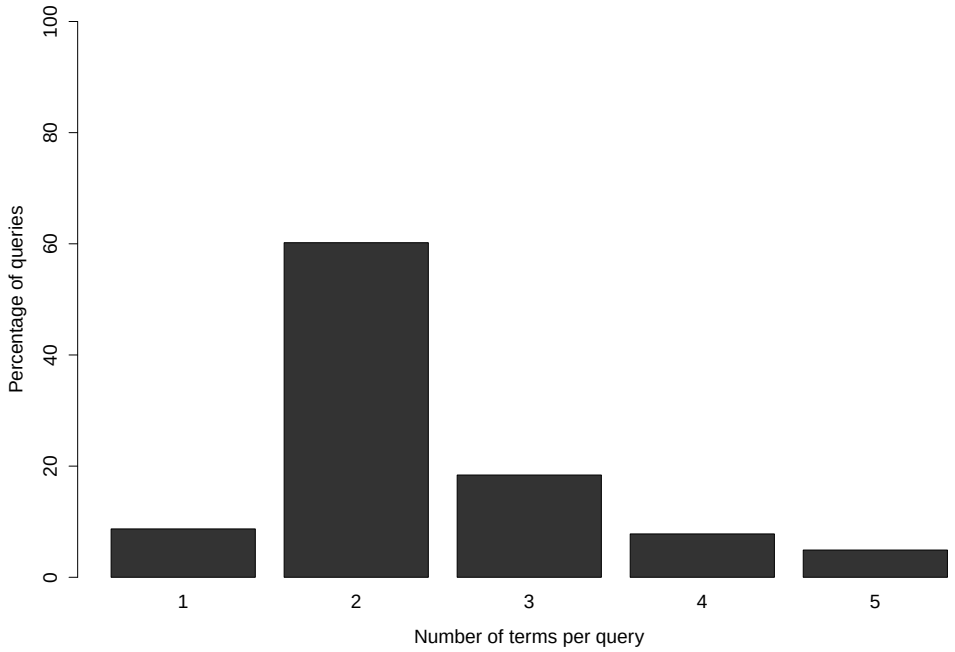

Figure 6.4: Percentage of terms per query in Bing log

6.5.2 Image information needs

In classifying sessions into image information needs, we looked at the queries submitted within the session. As we have no knowledge of their search context, we are manually

inferring users' needs based on their queries and matching it with the definition of Batley's [1988] classification. Results of the classification is shown in Table 6.9.

Table 6.9: Sessions from Bing log by category of image information needs

Statistics	Specific	Nameable	Abstract	Subjective
Number of sessions	20	27	1	1
Number of queries	55	45	1	2
Minimum number of queries	1	1	1	2
Maximum number of queries	24	4	1	2
Average number of queries per session	2.8	1.7	1.0	2.0

In the table, we can see that the sessions were dominated by Nameable and Specific needs. As expected, the highest percentage in number of queries submitted were for Specific needs (53.3%), followed by Nameable (43.7%), Abstract and Subjective needs (1% and 2% respectively). There were hardly any difference in the minimum number of queries. However the Specific needs has a particularly high value for maximum number of queries.

Figure 6.5 shows the percentage of queries per session for Specific and Nameable tasks. Abstract and Subjective needs were excluded to avoid being misleading due to the extremely limited number of sessions. Looking at the figure, we can see that more than 70% of users sessions are short, consisting of either one or two queries.

From an analysis of query terms used for the four image information needs, we found that on average users used two to three terms in searching for images (Table 6.10). A look at the queries for Specific needs, revealed that most are concerned with named entities, particularly specific persons. Example of queries with a Specific need are *pippa middleton*, *steve jobs*, *mother teresa*.

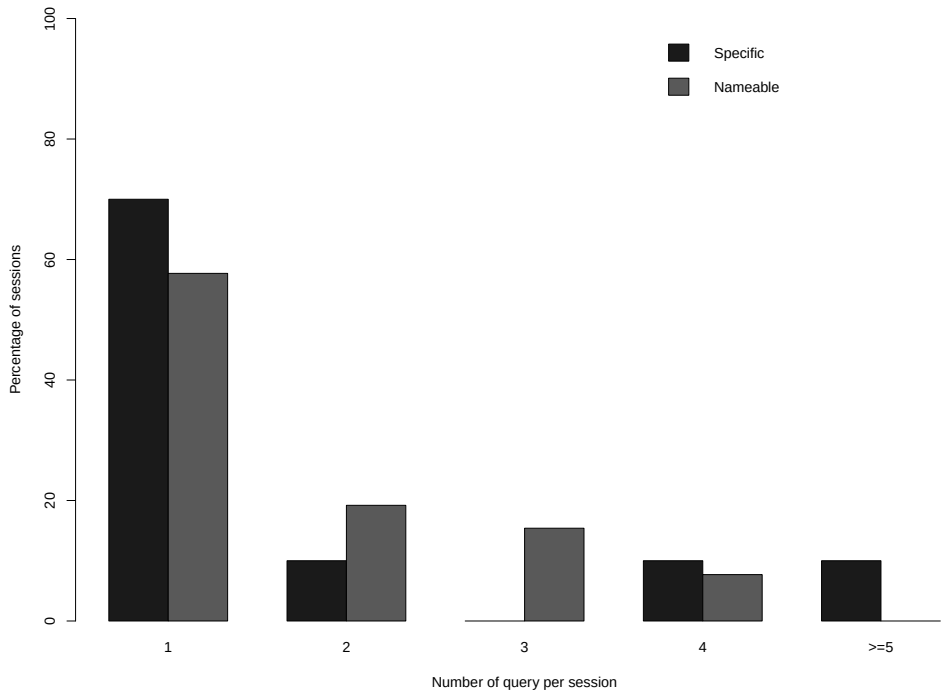

Figure 6.5: Percentage of queries per sessions in Bing log by category of image information needs

6.5.3 Query reformulation

Analysis on reformulation done in the log, shows that there were 103 query reformulations in total. The most frequent reformulation type used is *New* (77.7%), followed by *Replace* (9.7%), *Add* (7.8%) and *Remove* and *Reformulate* respectively (2.9%). In Table 6.11, we can see that Specific needs has the highest number of reformulation (53.4%). The difference is discernible if compared to Abstract and Subjective needs. Nevertheless, the number of queries in both Abstract and Subjective needs were substantially low.

Out of the 49 sessions identified, only 39% of sessions have at least two queries, while the rest were single-query sessions. Sessions with at least two queries were analysed and classified by looking at their terms. Possible changes in needs were identified based on

Table 6.10: Queries from Bing log by category of image information needs

Statistics	Specific	Nameable	Abstract	Subjective
Number of queries	55	45	1	2
Total number of terms	127	115	3	6
Minimum number of terms	1	1	3	2
Maximum number of terms	5	5	3	4
Average number of term per query	2.3	2.6	3	3

Table 6.11: Reformulation types of Bing log by category of image information needs

Category	Specific	Nameable	Abstract	Subjective
New	48	30	1	1
Add	1	6	0	0
Remove	1	1	0	1
Replace	3	7	0	0
Reformulate	2	1	0	0

changes of terms from successive queries. The following are examples of sessions with (a) no changes in information needs, and (b) a change in information needs.

(a) no changes in information needs

- images of cooktown australia (N) → cooktown accommodation (Rp) → cook town australia (Rp) → flight ot cooktown (Rp)
- dont forget eruv tavshilin (N) → eruv tavshilin (R)
- c++ builder 5 (N) → c++ builder 6 (Rp)
- hair styles (N) → up hair styles (A) → hair Updos (Rp)

(b) a change in information needs

- forbo flooring stone looks (N) → floating white desk (N)
- flowers (N) → nexus wallpaper (N) → nexus desctop (Rp)

- equine supplies (N) → bird poultry supplies (Rp) → bird feed (R) → types of bird feed (A)
- digestive system of the rat (N) → Rats Internal Organs (Rp) → pancreas location (N) → pancreatic duct location (Rp)

Clearly, the *New* reformulation type is used when there is a change in information need. Moreover, changes can be expressed by using the *Replace* reformulation type. For example, queries in the session below:

- equine supplies → bird poultry supplies → bird feed → types of bird feed

The term *equine* was replaced by *bird poultry* as an indication that the need has changed. This suggests that a search session may start with one clear and distinct information need, but during the session that need may evolve or change slightly. However, a change of image information need can be difficult to discern. Perhaps in this example, the change of term is actually a correction of information need (Rf would be more appropriate then Rp). Nevertheless, looking at the queries, change of information need can occur in two possible situations: 1) a definite change of type of information need, for example from a Nameable to a Specific need and 2) a change/correction of topic within an information need as in the *equine* example.

6.6 Results of analysis of user study data

In Chapters 4 and 5, we analysed user study data in order the answer the first and second research questions. Here, the same data is analysed and results are presented in relation

to how users express their needs through queries based on the provided image information needs.

6.6.1 Image information needs

Data from the user study consists 192 sessions with 48 sessions for each of the information needs. Users submitted in total, 833 queries for the four information needs. As shown in Table 6.12, the Subjective needs has the highest number of queries, maximum number of queries submitted and average number of queries per session. A Subjective need is more difficult to express using keywords which accounts for a higher average number of queries per session.

Table 6.12: Session and query data from user study by category of image information needs

Statistics	Specific	Nameable	Abstract	Subjective
Number of sessions	48	48	48	48
Number of queries	127	196	207	303
Minimum number of queries	1	1	1	1
Maximum number of queries	9	12	11	17
Average number of queries per session	2.6	4.1	4.3	6.3

Detailed statistics on the percentage of queries per session can be seen in Figure 6.6. From the figure, we can see that a single query session is dominated by Nameable needs (29%) followed by Specific (25%), Abstract (23%) and Subjective (10%) needs. Therefore, it is no surprise that the Subjective need has the highest percentage of more than five queries per session compared to the other needs. In the Specific needs, the percentage of queries per session decreases as the number of queries increases from three queries onwards.

Meanwhile, statistics from analysis of query terms by image information needs for the user study data is shown in Table 6.13. From the table, it is visible that the Specific need

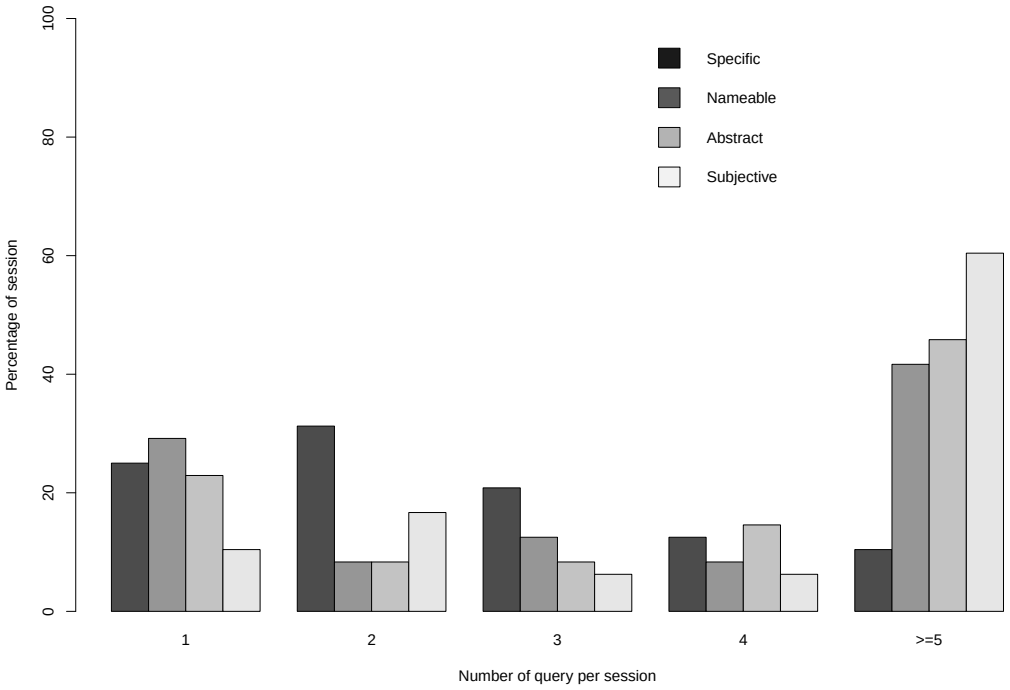

Figure 6.6: Percentage of queries per session in user study data by category of image information needs

has the highest average number of terms per query (3.7) compared to the other three image information needs. A possible explanation is that the topics used in the user study require users to submit at least two query terms. However, users performing the Subjective need may require a longer search session but do not necessarily use more query terms.

Table 6.13: Term data from user study by category of image information needs

Statistics	Specific	Nameable	Abstract	Subjective
Number of queries	127	196	207	303
Total number of terms	465	456	568	701
Minimum number of terms	1	1	1	1
Maximum number of terms	11	8	9	8
Average number of term per query	3.7	2.3	2.7	2.3

Analysis on the percentage of terms per query can be seen in Figure 6.7. The analysis is

consistent with data in Table 6.13 whereby 60% of user queries used only one or two terms. Moreover, it has the lowest percentage of queries with more than five terms.

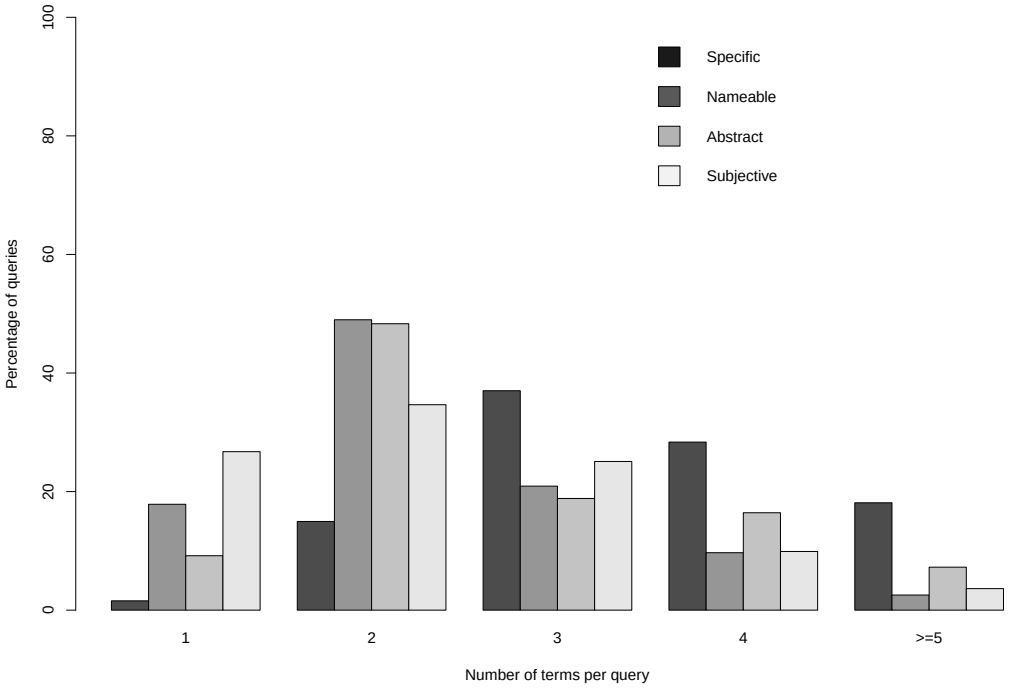

Figure 6.7: Percentage of terms per query in user study data by category of image information needs

6.6.2 Query reformulation

Analysis on reformulation done by participants in the user studies, shows that there were a total of 833 query reformulations (641 query reformulation if all the initial queries were excluded). Table 6.14, shows the numbers of reformulations issued by users according to image information needs. The highest overall number of reformulation is for Subjective (36.4%) needs, followed by Abstract (24.8%), Nameable (23.5%) and Specific (15.2%) needs. *New* (41.4%), *Replace* (32%) and *Add* (19.5%) were the most used reformulation type across

all information needs (Figure 6.8).

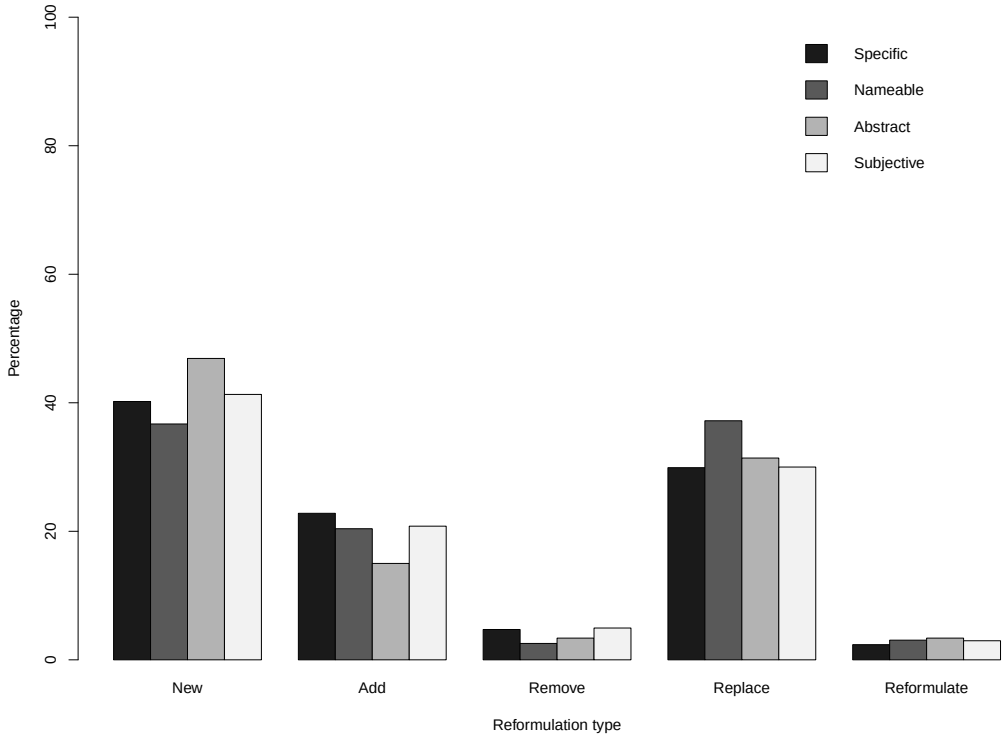

Figure 6.8: Percentage of query reformulation by category of image information needs

From the user study data, 78% of the sessions contains at least two queries while the remaining 22% were single query sessions. Similar to the Bing log, sessions with at least two queries were analysed. However, as the user study have been designed with an associated information need, in the analysis of query reformulation we focus primarily on the reformulation patterns users apply in fulfilling the needs.

Looking at Table 6.14, we could say that for all image information needs, initial queries were predominantly followed by the reformulation types — *Add*, *Replace* and *New*. For example, in the Specific need, 18 users used the *Add* reformulation type to make the search results more precise, followed by the use of *Replace* to take into account the use of synonyms.

Table 6.14: Reformulation types of user study data by category of image information needs

Category	Specific	Nameable	Abstract	Subjective
New	51	72	97	125
Add	29	40	31	63
Remove	6	5	7	15
Replace	38	73	65	91
Reformulate	3	6	7	9

6.7 Characteristics of users' image search behaviour

In Sections 6.4 and 6.5, we presented results of analysis on two collections of proxy log data, that is, the Google log and the Bing log. Additionally, we included analysis of queries from the user study data. In this section, we compare these results to associate the characteristics between users performing image searches in both natural and controlled settings.

The proxy log is an example of data in a natural setting whereby we have queries without the knowledge on the context or information need of these queries. Conversely, the user study is considered as a controlled setting since users submit queries based on artifically created information needs. Comparisons were done as follows (1) comparison at session, query and term levels, (2) comparion at session, query and term levels based on image information needs, and (3) comparison on query reformulation type.

A comparison on average number of queries per session were done between user study data and the Bing log showed that users in the controlled setting submitted more queries (4.3 queries per session) compared to users in the natural setting (2.1 queries per session). A possible reason is that users in the controlled setting were assigned search tasks that had been designed with clearly defined image information needs. Moreover, the task also specifies the purpose of the searched images.

As for average number of terms per query (Figure 6.9), both users in natural and controlled setting submitted short queries, mostly consisting of two to three terms. However, the average number of terms per query in both settings was slightly lower than the findings from previous research [Goodrum and Spink, 2001; Jansen et al., 2004].

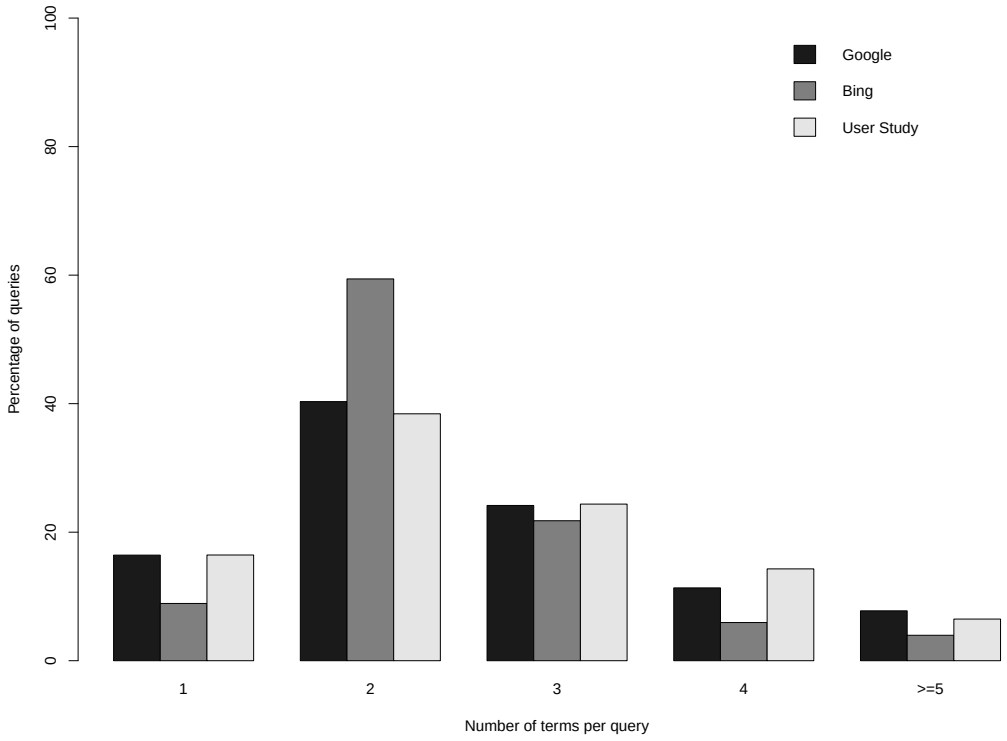

Figure 6.9: Comparison on terms per query for all three data sets

Analysing the proxy log data by category of information needs shows that in both logs, Abstract and Subjective needs were being searched much less often (Figure 6.10). In the Google log, Nameable needs were searched more frequently (64.3%) compared to Specific needs in the Bing log (53.3%). Nevertheless the difference may be the result of how the classification was performed. In the Google log, each query was classified independently while in the Bing log, queries in a session were classified.

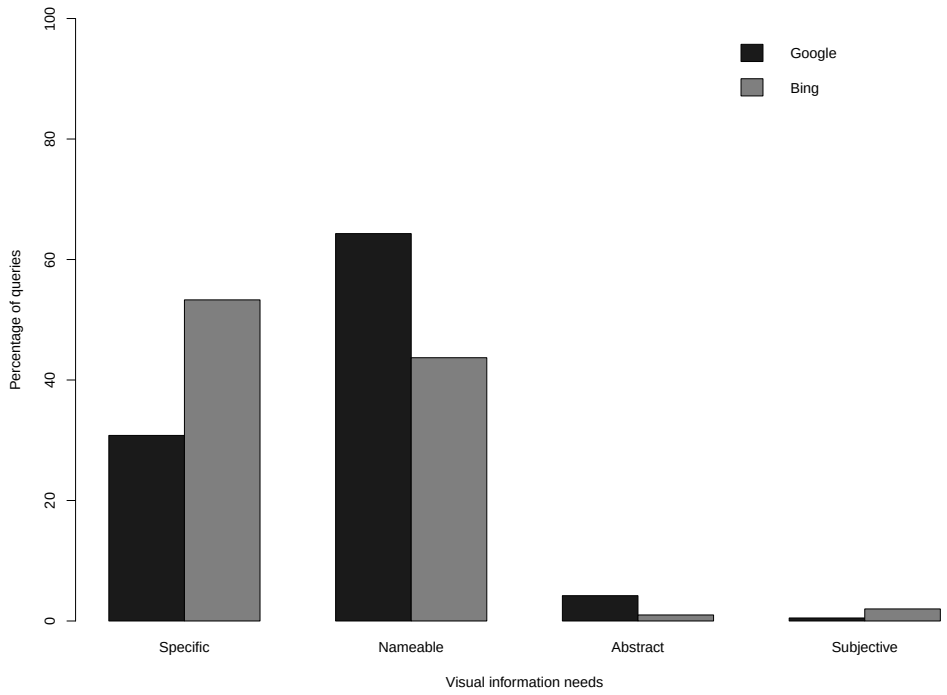

Figure 6.10: Frequency of search in log data by category of information need

The average number of queries per session and terms per query by information needs for both settings was consistent with earlier comparisons — users in the controlled setting submitted more queries compared to others in the natural setting, and users in both setting submitted short queries. However, for the Specific need, the average number of queries per session in the controlled setting is slightly lower (2.6 queries per session) compared to the natural setting (2.8 queries per session), while the average number of terms per query in the controlled setting was much higher (3.7 terms per query) than in the natural setting (2.3 terms per query). One possible reason is the topics in the work task scenario provided for the Specific needs in the user study requires users to submit longer queries. On the other hand, although users in the controlled setting submitted between four to seven queries for the Nameable, Abstract and Subjective needs, they do not necessarily submit longer queries.

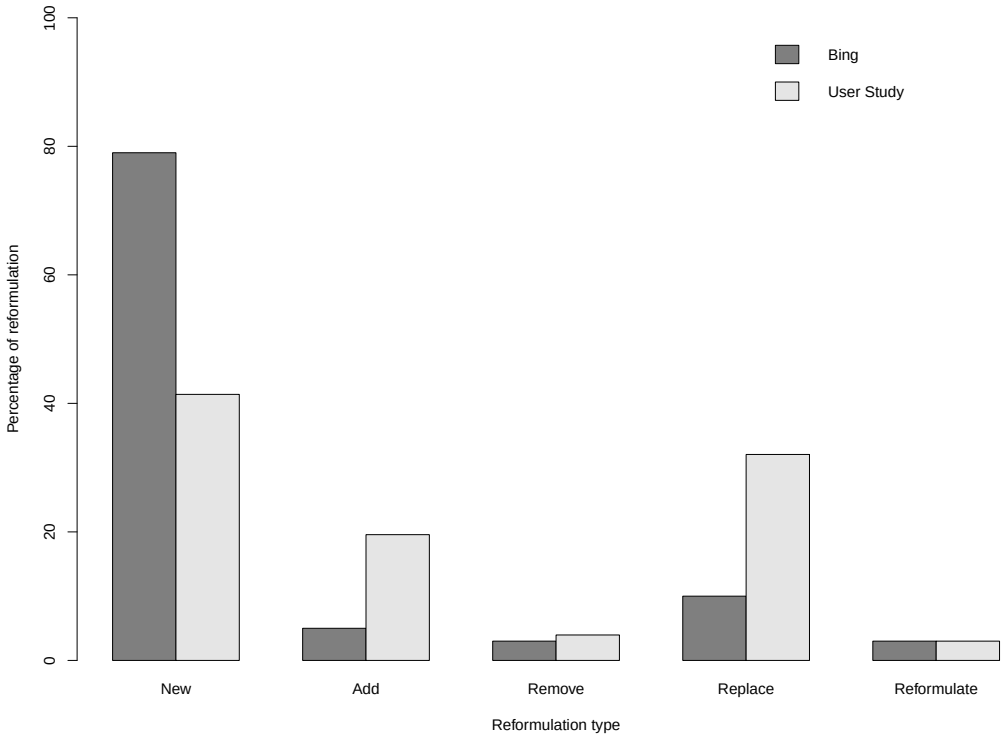

Figure 6.11: Comparison on reformulation type

Lastly, comparison on the reformulation types used, shows that in both settings, users frequently modify their queries using the *New*, *Replace* and *Add* type (Figure 6.11). Queries can be used to identify what users were searching for, but inferring their information need is not as straightforward. However, the sequence and type of query modification used may provide hints on the possible type of visual information need.

6.8 Summary

In this chapter, we discuss the work on proxy logs to infer users' information needs from their queries. Results of the analysis on two proxy logs showed that based on their queries, users mostly searched for Specific and Nameable images compared to Abstract and Subjective

images. Although a single query can be used, a search session with at least two queries gives a better idea of a user's information need. Query reformulation type can also give an indication of a user's need. Analysis of reformulation type shows that users frequently add and replace terms compared to other reformulation types. Moreover, a comparison was made between the behaviour of users searching in the wild (from the logs) and in a controlled setting (user study). Similar findings were found when comparing results from the proxy log data with the user study data. This suggests that users' information needs can be identified from the queries they submitted and how they reformulated them. Patterns of query reformulation could be analysed to further determine the type of information needs. Given the average number of queries submitted by users, search engine engineers could utilize their queries to identify query reformulation patterns which would help in further determining the type of information needs.

Chapter 7

Conclusion

In this thesis, we have conducted a user study and proxy logs analysis to answer research questions on factors that affect users' image searching behaviour. Specifically, the thesis looked at how visual information needs affect the use of criteria when judging image relevance; perception of factors on searching behaviour; and whether these needs can be inferred from user queries.

The chapter is organised as follows: summary of findings (Section 7.1- Section 7.3), followed by directions for future research (Section 7.4). Section 7.5, summarises the chapter.

7.1 Image relevance judgements

The first research question investigated how visual information needs affect the use of relevance criteria when making image relevance judgements. We designed a within-subjects user study and used the visual information needs identified by Batley [1988]. We selected ten relevance criteria from Choi and Rasmussen [2002] (*topicality*, *accuracy*, *suggestiveness*,

completeness, *appeal of information*, *technical attributes of images* and *textual description*) and Hung [2006] (*composition*, *consequence* and *emotion*).

We analysed ratings on relevance criteria by 48 users performing four online image search tasks using Google Images. The findings reveal interesting facts about the criteria that ordinary users apply when choosing relevant images. When making image relevance judgements, users employ multiple criteria — and they apply more criteria when judging relevance for Subjective and Abstract tasks compared to Specific and Nameable tasks. They may use the same criteria for different types of task; however, the importance of each criterion depends on the task and the topic of that task. For example, comparing all four tasks, participants rated *Accuracy* as least important in the Subjective task, but agreed that *Emotion* was an important criterion.

Unlike users in a specialised domain, who rely on textual information as criteria for judgement, only 50% of participants considered it as an important criterion. Getting the right facts is crucial in a specialised domain. However, images on the web belong to very many different users and therefore the terms, keywords and metadata associated with the images are more diverse. Furthermore, the results show that criteria which were previously used in specialised domain are applicable for ordinary web searchers and the importance of each criterion is dependent of the visual information needs.

7.2 Factors affecting image search behaviour

The second research question was addressing effect of subjective factors on users' image searching behaviour. As part of the same user study conducted for research question one,

we studied the effect of task type, familiarity, difficulty and satisfaction. The question was addressed in two parts — participants rating their perception on different aspect of performing search which included the previously mentioned factors and observation of users' actual search behaviour while completing all the search tasks.

Analysis of users' ratings showed that the effect of subjective factors were task dependent. For instance, 41 participants agreed they were able to create queries easily for the Specific tasks compared to only 26 participants for the Subjective tasks. Participants also agreed that they changed their mind more frequently while performing the Subjective tasks compared to the other three tasks.

Examination of users' time-stamped search interactions show that their behaviour was consistent with previous text retrieval studies whereby users have a longer completion time when presented with a difficult task and that they issued more queries for the task. Nevertheless, compared to text search, image search users viewed considerably more result pages. Of the factors that affect search behaviour, familiarity was not shown to affect users' searching behaviour. However, their behaviour is affected by task type, task difficulty and satisfaction.

The Spearman's rank correlation coefficient ρ suggests that users' perception of aspects of performing image searches does not necessarily correspond with their actual search behaviour. Relating these subjective factors to the objective measures is difficult because it is task dependent. The results showed that there are some aspects of image search that were clearly measurable by examining users' search interaction. However, it is not sufficient to determine the effect of subjective factors in image searching using only one measure of search behaviour. Indeed, aspects such as satisfaction might only be measured subjectively.

7.3 Image information needs

The third research question examines whether users' queries can be used to deduce their visual information needs. We performed log analysis on two sets of image search logs (Google and Bing) extracted from the RMIT University proxy log. We included analysis of queries submitted by participants in the user study to compare the difference between users in natural and controlled settings. We analysed the queries and performed the following comparisons — (1) comparison at session, query and term levels; (2) comparison at session, query and term levels based on image information needs; and (3) comparison on query reformulation type.

The findings showed that users searched for images for Specific and Nameable needs more frequently compared to Abstract and Subjective needs. Queries can be used to identify what users were searching for, but inferring the information need can be difficult due to missing context. However, analysis of users' sequence of query reformulation suggests that their information needs can be elicited from how they reformulate their queries or the reformulation type used. Users' exhibits different patterns of query reformulation for different types of visual information need. For example, users searching for Specific needs used the Add and Replace reformulation type more often, while users searching for Subjective needs used the same reformulation type with the addition of the New reformulation type. Last but not least, users in both natural and controlled settings exhibit similar search behaviour in terms of querying behaviour and query reformulation.

7.4 Opportunities for future research

This research has addressed a number of research questions, however many opportunities for further research remain. First, we were unable to explore users' tendency in selecting a criteria when judging image relevance. We propose to perform factor analysis on the relevance criteria to identify whether the selection of particular criteria was dependent of another criterion. Based on the results, users applied similar relevance criteria in all task types. Therefore, it would be interesting to find out the interaction between criteria when judging image relevance.

Second, each factor was not explored extensively. For example, we controlled the effect of familiarity by asking participants to select from a list of four, the topic they were most familiar with. Alternatively, to further analyse the effect of topic familiarity, participants could be asked to select a topic that they are very familiar with and another topic they are unfamiliar with.

Third, only some general behaviours were analysed, such as the number of queries submitted, time spent on result pages, and task completion time. The method of data collection may limit the search behaviour that we were able to observe, and hence correlate with the subjective factors. Eye tracking could be used as a method to collect data related to what users are looking at on the screen while searching for images.

Fourth, the experiment was carried out separately for each participant, which meant that it was possible that there were variations in the search results, especially when personalisation of results is taken into account. A future experiment that locked in a reduced set of search results for a limited number of queries might help establish whether search result

variation affected the outputs of the experiment. In addition, the number of male and female participants was unbalanced. Although gender is not a factor of interest in the study, there exist possible differences in results. Nonetheless, a Welch t-test performed to test differences in rating importance of relevance criteria by gender showed no significant difference.

Fifth, there is an increased chance of Type 1 error due to the number of statistical tests performed on the data. Nevertheless, we feel that the results are still valid as we have performed power analysis to determine the likelihood of achieving statistical significance. Therefore, future work should involve a larger sample size and an increased confidence level to reduce the chance of Type 1 error.

Lastly, the log extracts were not of sufficiently large size that would allow us to analyse the different visual information needs. Therefore, future work should involve a larger dataset so as to allow a better understanding of user needs and relative frequency of each query category.

Other suggestions on areas for future research are as follows:

- exploring other contexts of image use to see if there is a difference in users' searching behaviour;
- comparing the user satisfaction with image search for a range of querying methods, including standard text query methods and non-text query;
- exploring the dynamics of users' image search behaviour (for example, number of queries per session) by comparing the results in the thesis with a future user experiment and log analysis that takes into account the evolution of image search engines;

- proposing an in-depth analysis of proxy log data as it records access to pages all over the Internet. From the proxy log, we can analyse how users continue to search for images apart from just using a search engine;

- performing analyses on proxy log data before conducting a user study. The results show that users mostly searched for Specific and Nameable images. Therefore, conducting the proxy log analysis first would give insights to the topics that were commonly searched by online users for the two visual information needs. A better work task scenario that more realistically represents what users actually search for can be created to be used in the user study.

Despite of these limitations, the results of this research contributes to the understanding of ordinary users' image search behaviour in terms of use of relevance criteria, effect of subjective factors on search behaviour and inferring visual information needs through query reformulation.

7.5 Concluding remarks

In this thesis we have discussed online web users' image search behaviour. In particular, examination of the relevance criteria used in judging image relevance for four different visual information needs. Using the visual information needs, effects of subjective factors on users' image searching behaviour were analysed. Users' visual information needs were then inferred based on their query reformulation from proxy log data.

The research was completed by first conducting a user study and then an analysis of proxy log data. Results from the user study showed the importance of relevance criteria

when making judgements and how users' were affected by subjective factors. Meanwhile, the proxy log analysis showed what users actually search for and their query reformulations. Nevertheless, both showed that users either in a controlled or natural setting, exhibit similar image searching behaviour.

Analyses of the proxy logs, found that users rarely search for Abstract and Subjective images. The comparison of queries from user experiment and proxy log data, showed similar querying pattern. Although a search context from the proxy log data is not known, users' visual information needs may be inferred from their queries and how they reformulate them.

Appendix A

Pilot study on criteria for image relevance judgements and effect of subjective factors

A.1 Experimental design and procedure

The pilot study used a within-subjects experimental design [Kelly, 2009]. We recruited 12 people as volunteers to participate in our study as the subjects of the experiments. All of them were either undergraduate or postgraduate students from RMIT who were approached and recruited via posters, electronic forums and face-to-face recruitment after lecture sessions. The participants were met one at a time, each on a separate occasion. The experiment was conducted anonymously, so that responses could not be traced back to individual participants.

Three types of image search tasks were created based on Shatford's image analysis [Shat-

ford, 1986]. These include specific, general and abstract image search tasks.

- ***Specific Task***: You are interested in entering a World Cup 2010 contest. One of the contest conditions is that you have to find 6-8 images that best depicts the 2006 World Cup final match in Germany. Your task is to make a selection from a large collection of images from the World Wide Web and save those that in your opinion would most effectively fulfil the contest's condition.

- ***General Task***: As a fashion design student, you are required to create a portfolio showcasing the traditional fabrics of different cultural heritages. Your portfolio will include several different traditional fabrics and one of them is entitled "Timeless Songket". Your task is to make a selection from a large collection of images from the web and save 6-8 images that in your opinion would most effectively highlight its uniqueness.

- ***Abstract Task***: You and your classmates are preparing a report on the topic 'Justice and Equality' and your task is to make a selection from a large collection of images from the World Wide Web and save 6-8 images those that in your opinion would most effectively illustrate the meaning of 'justice'.

For each participant, our procedure are as follows:

1. an introductory orientation session;

2. a pre-search questionnaire;

3. a training session to familiarise the participant on how the task was to be performed;

4. a written instruction for the first task;

5. a search session in which the participant perform the first task;

6. a post-session questionnaire about the first task;

7. steps 4 to 6 were repeated for the remaining two tasks;

8. a final exit questionnaire.

We used a simulated real work task situation [Borlund and Ingwersen, 1997] to place our participants in a work task scenario. This scenario allows the participants to fashion their information needs in the same manner as they would when performing an actual search session. In determining the order of tasks which the participants were to perform, we employed a mathematical factorial design with two users for each of the six permutations of the three tasks. The experiment used Google Images[1] search engine to perform image search and retrieval.

Data for the study was collected through questionnaires. Questionnaires were used as it was found to be more effective for users to communicate their response as compared to interview [Kelly et al., 2008]. According to Kelly et al. [2008], although users may express more ideas, many of these ideas are similar; they seem to be repeating it rather than providing new ideas. The pre-search questionnaire was used to collect participant's prior experience with image search such as frequently used search engines, search frequency, and search expertise. There were two types of relevance criteria questionnaires: the post-session and the exit questionnaire.

[1] http://images.google.com.au

A.2 Method of data collection

Data for the study was collected through questionnaires and participants' screen capture recordings. Questionnaires were used as it was found to be more effective for users to communicate their response as compared to interview [Kelly et al., 2008]. According to Kelly et al. [Kelly et al., 2008], although users may express more ideas, many of these ideas are similar; they seem to be repeating it rather than providing new ideas. The pre-search questionnaire was used to collect participant's prior experience with image search such as frequently used search engines, search frequency, and search expertise. There were two types of relevance criteria questionnaires: the post-session and the exit questionnaire.

The post-session questionnaire have two sets of closed-ended questions. The first set, asked participants to rate their agreement on the reasons they selected images for the search task that they had just performed based on a selected set of relevance criteria while the second set asked to rate other aspects of the task such as topic familiarity, ease of navigation and result satisfaction. Finally, open-ended questions were used in the exit questionnaire to collect information regarding the users' whole search experience and any other issues that may have an effect on how they judge image relevance such as what justifies a relevant image, what makes judging relevance difficult (if any) and how to make it easier.

A.3 Results

The experiment was carried out over several weeks and during that time, Google Images changed the way they present image search results. These changes include removing the metadata below the image and having it pop up whenever the user put the cursor on it,

which creates a mosaic of images and an infinite scrolling page that presents up to 1000 results per "page" [Hachman, 2010]. Only three participants performed their search using the old search interface, while the remaining nine participants performed the tasks using the new interface.

A.3.1 Relevance judgement criteria

Quantitative data from the post-session questionnaires were analyzed using descriptive statistics by assigning numerical values for each agreement rating. This is to determine the average scores of each criteria for relevance judgements and to measure how widely spread the scores were.

In order to examine whether there are statiscally significance differences in the attitudes of the participants in regards to the importance of certain criteria while making image relevance judgements, a Chi-Square analysis was done. The p-value is calculated based on two categories which are (i) combination of strongly agree and agree, and (ii) combination of strongly disagree, disagree and neutral/undecided. For the purpose of this study, it was decided to adopt a minimum significance level of $p<0.05$. Table A.1 shows the mean value of each relevance criteria for the three search tasks.

A.3.2 Effects of subjective factors on image search behaviour

As mentioned in the previous section, users were asked to rate different aspects of their search and retrieval, such as familiarity, satisfaction and overall performance. The questions, as listed below, was adapted from an exit questionnaire by Kelly et al. [2008].

Table A.1: The mean, standard deviation, number of users' in agreement and Chi-Square's p-value for each relevance criteria across search tasks

Relevance criteria	Statistics	Specific Task	General Task	Abstract Task
Topicality	μ	4.83	4.42	4.17
	σ	0.39	0.51	0.94
	# agree	12	12	10
	p-value	**0.0005**	**0.0005**	**0.0209**
Accuracy	μ	4.5	3.75	3.83
	σ	0.67	0.62	1.27
	# agree	11	10	9
	p-value	**0.0039**	**0.0209**	0.0832
Suggestiveness	μ	3.5	3.42	4.17
	σ	1.17	0.9	0.72
	# agree	7	6	10
	p-value	0.5637	1.0000	**0.0209**
Appeal of information	μ	4	4.08	3.92
	σ	0.85	1.08	0.79
	# agree	10	10	10
	p-value	**0.0209**	**0.0209**	**0.0209**
Completeness	μ	4.08	3.83	4.08
	σ	0.9	0.94	0.79
	# agree	10	8	9
	p-value	**0.0209**	0.2482	0.0832
Technical attributes of image	μ	4.25	4.42	3.33
	σ	0.62	0.67	1.23
	# agree	11	11	6
	p-value	**0.0039**	**0.0039**	1.0000
Emotion	μ	4.25	3.5	4
	σ	0.96	1.31	1.21
	# agree	10	7	9
	p-value	**0.0209**	0.5637	0.0832
Textual information	μ	3.58	3.75	3.58
	σ	1.38	1.29	1.44
	# agree	8	8	8
	p-value	0.2482	0.2482	0.2482
Consequence	μ	3.08	3	3.75
	σ	1.16	0.85	1.06
	# agree	6	4	8
	p-value	1.0000	0.2482	0.2482
Composition	μ	4.08	4.42	4.25
	σ	1.14	0.9	0.87
	# agree	11	10	11
	p-value	**0.0039**	**0.0209**	**0.0039**

1. I was familiar with the topic of the search task.

2. I was able to create a query for the task easily.

3. I was able to navigate the search results easily.

4. I was satisfied with the images presented to me.

5. I was satisfied with the order the images were presented to me.

6. I believe that I have seen all the possible images that satisfy the task's requirement.

7. I am very satisfied with my search results.

8. Overall, the task was easy to perform.

For each question, users rated their response on a 5-point scale : strongly disagree (1), disagree (2), neutral (3), agree (4), and strongly agree (5). We analysed the questionnaire data using descriptive statistical analysis to determine how users' rate different aspects of search and retrieval for each type of task. The users' ratings are shown in Table A.2.

A.3.3 Time-stamps of users' search interaction

Next, we proceed to examine the recordings of users' search interactions. The recordings were manually time stamped to determine distinct search and retrieval activity. These include the time when a user submits a query, the time a user views, saved and closed an image,and the time taken for a user to complete the task. In this study, we define a view as either when the user hovers the cursor on the result image until the image meta data is displayed or displays

Table A.2: The mean, standard deviation, number of users' in agreement and Chi-Square's p-value for each relevance criteria across search tasks

Aspects of search	Statistics	Specific Task	General Task	Abstract Task
Topic familiarity	μ	3.42	2.58	2.83
	σ	1.00	1.51	1.11
	# agree	7	5	5
	p-value	0.564	0.564	0.564
Create query	μ	4.42	3.33	3.58
easily	σ	0.51	0.98	0.79
	# agree	12	6	7
p-value	0.005	1.000	0.564	
Navigate results	μ	4.17	3.83	4
easily	σ	0.58	0.72	0.85
	# agree	11	10	10
	p-value	0.004	0.021	0.021
Satisfied with	μ	3.58	3.5	2.92
images	σ	1.38	1.24	1.24
	# agree	8	8	5
	p-value	0.248	0.248	0.564
Satisfied with order	μ	2.92	3.25	2.67
of images	σ	1.31	0.97	1.30
	# agree	5	5	4
	p-value	0.564	0.564	0.248
Viewed all possible	μ	3.33	2.92	2.92
images	σ	1.50	1.16	1.24
	# agree	8	4	5
	p-value	0.248	0.248	0.564
Satisfied with	μ	3.75	3.5	3.42
search results	σ	1.22	1.17	1.08
	# agree	9	7	7
	p-value	0.083	0.564	0.564
Ease of task	μ	4.33	3.83	3.67
	σ	0.89	1.03	0.78
	# agree	11	9	8
	p-value	0.004	0.083	0.248

Table A.3: Overall mean and standard deviation of users' search interaction for all task types

User search behaviour	Statistics	Specific	General	Abstract
Time to submit	μ	26.78	17.44	25.11
initial query (s)	σ	19.10	10.31	19.94
Number of	μ	4.33	4.33	6.89
queries submitted	σ	3.46	3.12	4.23
Number of results	μ	24.56	24.11	33.56
pages viewed	σ	18.80	13.23	16.34
Number of	μ	32.78	34.33	39.33
images viewed	σ	22.54	16.33	22.52
Time spent on	μ	479.89	480.22	794.22
results pages (s)	σ	209.75	111.15	362.81
Time to	μ	559.11	512.89	850.44
complete task (s)	σ	248.52	113.09	384.12

the image in a new tab. By examining each user's search interaction, we were able to derive Table A.3 that summarises the main facets of the interaction.

Appendix B

Topic survey for information needs used in main user study

Batley's (1988) classification of visual information needs

Specific	can be expressed in key words; can be expressed in a precise search statement; have no ambiguity; and deal with the concrete.
General/Nameable	can be expressed in key words; may result in unmanageably high recall (number of items retrieved); and often have to be made more specific.
General/Abstract	are difficult to express in key words; may involve abstract concepts rather than concrete objects; and can be expressed verbally but not in a precise search statement.
General/Subjective	are difficult to express verbally; deal with emotional responses to a stimulus; cannot be expressed in a search statement; and are dependent on characteristics of a scene as interpreted by the individual.

Based on the above classification of visual information needs, please mark (X) on the most appropriate visual information needs for each of the following topics.

	Topics	Specific	General/ Nameable	General/ Abstract	General/ Subjective
1	heaven on earth	☐	☐	☐	☐
2	joy	☐	☐	☐	☐
3	marathon	☐	☐	☐	☐
4	funny faces	☐	☐	☐	☐
5	hope	☐	☐	☐	☐
6	Black Saturday	☐	☐	☐	☐
7	Mount Fuji, Japan	☐	☐	☐	☐
8	sunset	☐	☐	☐	☐
9	friendship	☐	☐	☐	☐
10	high performance cars	☐	☐	☐	☐
11	lighthouse	☐	☐	☐	☐
12	Melbourne skyline	☐	☐	☐	☐
13	Harley Davidson motorcycles	☐	☐	☐	☐
14	iceberg	☐	☐	☐	☐
15	financial security	☐	☐	☐	☐

	Topics	Specific	General/ Nameable	General/ Abstract	General/ Subjective
16	curiosity	☐	☐	☐	☐
17	immigrant	☐	☐	☐	☐
18	New Year's Eve	☐	☐	☐	☐
19	economic unrest	☐	☐	☐	☐
20	kimono	☐	☐	☐	☐
21	fighting spirit	☐	☐	☐	☐
22	fruit picking	☐	☐	☐	☐
23	wonderland	☐	☐	☐	☐
24	The Taj Mahal	☐	☐	☐	☐
25	renaissance	☐	☐	☐	☐
26	urban development	☐	☐	☐	☐
27	Albert Einstein	☐	☐	☐	☐
28	science fiction	☐	☐	☐	☐
29	innocence	☐	☐	☐	☐
30	Wimbledon 2011	☐	☐	☐	☐
31	logical thinking	☐	☐	☐	☐
32	prosperity	☐	☐	☐	☐
33	The Legend of Zelda	☐	☐	☐	☐
34	nature	☐	☐	☐	☐
35	a walk to remember	☐	☐	☐	☐
36	adversity	☐	☐	☐	☐
37	Yellowstone National Park	☐	☐	☐	☐
38	excitement	☐	☐	☐	☐
39	athlete	☐	☐	☐	☐
40	cute babies	☐	☐	☐	☐

Table B.1: Results of topic survey for information needs with the highest number of votes.

Topics	Classification of needs(by user)	# votes receiver
heaven on earth	Abstract/Subjective	6
joy	Subjective	9
marathon	Specific	8
funny faces	Specific/Subjective	4
hope	Subjective	8
Black Saturday	Specific	10
Mount Fuji, Japan	Specific	15
sunset	Specific	9
friendship	Abstract	8
high performance cars	Specific	8
lighthouse	Nameable	10
Melbourne skyline	Specific	10
Harley Davidson motorcyles	Specific	11
iceberg	Nameable	9
financial security	Abstract	7
curiosity	Subjective	9
immigrant	Nameable	12
New Year's Eve	Specific	7
economic unrest	Abstract	10
kimono	Specific	9
fighting spirit	Subjective	6
fruit picking	Specific/Nameable	6
wonderland	Specific/Nameable	4
The Taj Mahal	Specific	14
renaissance	Abstract	5
urban development	Abstract	8
Albert Einstein	Specific	13
science fiction	Nameable/Abstract	5
innocence	Abstract/Subjective	6
Wimbledon 2011	Specific	11
logical thinking	Abstract	11
prosperity	Subjective	8
The Legend of Zelda	Specific	10
nature	Abstract	6
a walk to remember	Subjective	6
adversity	Subjective	8
YellowStone National Park	Specific	13
excitement	Subjective	7
athlete	Nameable	11
cute babies	Subjective	6

Appendix C

Plain language statement

RMIT UNIVERSITY

School of Computer Science and IT

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2348
Fax +61 3 9662 1617
www.rmit.edu.au

Invitation to Participate in a Research Project

PROJECT INFORMATION STATEMENT

Project Title: Relevance Feedback for Content-Based Image Retrieval

Investigators:

- Ms Rahayu A.Hamid (PhD student, School of CS&IT, RMIT University),
- Dr James Thom (Assoc Professor, School of CS&IT, RMIT University),
- Dr Seyed Mohammad Mehdi (Saied) Tahaghoghi (Senior Program Manager, Microsoft),
- Dr Dayang Nurfatimah Awg Iskandar (Lecturer, Faculty of CS&IT, Universiti Malaysia Sarawak).

Dear student,

You are invited to participate in a research project being conducted by School of Computer Science & IT, RMIT University. This information sheet describes the project in straightforward language, or 'plain English'. Please read this sheet carefully and be confident that you understand its contents before deciding whether to participate. If you have any questions about the project, please ask one of the investigators above.

Who is involved in this research project? Why is it being conducted?

This study is being conducted as part of my PHD research project. We are interested in understanding people's behaviour when performing image search. We are conducting the research to identify factors that might be important to a user when they perform image search. Findings from the study will be used to enhance the image search process in order to minimise the users' effort.

Why have you been approached?

We are approaching you to be involved as a voluntary participant in this study. Any students who have experience conducting web and/or image search can be participants in this project.

What is the project about? What are the questions being addressed?

An image is often very subjective. It can mean different things to different people and sometimes, finding the image that matches our information need is not that easy. We are interested in understanding how people search for images and what are the important factors that might affect their decision on selecting a particular image based on their information need.

If I agree to participate, what will I be required to do?

The study will take approximately 1 hour. You as the participant will be asked to:

- complete a background and user experience questionnaire,
- perform a training task,
- perform 4 image search tasks, in which your search interactions while performing these tasks will be recorded, and
- answer a post-task questionnaire after completing each task.

The data is totally anonymous and you will not need to identify yourself (such as; name and ID) at any stage.

What are the risks or disadvantages associated with participation?

Any participation will remain voluntary and anonymous. You will be assigned an anonymous ID, so your personal information will not be attached to any data analysis or publication results. There is no risk to you for being involved in this particular study apart from normal hazards of computer use.

What are the benefits associated with participation?

Participation in this study may not benefit you directly. However, the knowledge obtained from your participation will help us to better understand how users conduct image search and identify some factors that might be important when conducting image search and making decision on image relevance.

What will happen to the information I provide?

All the information collected during the study will be kept strictly confidential and you will remain anonymous. Data will be accessible to members of the investigative team and will be used in publications related to the research in an anonymous fashion. The data will be kept secure for a period of five years from end of PhD research and then destroyed. The outcomes of this experiment will be used for my PhD research. Due to the nature of the experiment, you will be asked to sign a consent form for your participation.

What are my rights as a participant?

Your participation is this study is voluntarily and as a participant; you have the following rights at any time:

- The right to withdraw your participation at any time, without prejudice.
- The right to have any unprocessed data withdrawn and destroyed, provided it can be reliably identified, and provided that so doing does not increase any risk for you.
- The right to have any questions answered at any time.

Please keep in mind that we are trying to understand human behaviour and have no intention of evaluating you, the participant.

Whom should I contact if I have any questions?

Any member of the investigative team listed at the beginning of this plain language statement may be contacted at any time. Any complaints about the conduct of this research project can be made to the Executive Officer, RMIT Human Research Ethics Committee, see http://www.rmit.edu.au/rd/hrec_complaints

Yours sincerely,

Ms Rahayu A.Hamid, MSc
Dr James Thom, PhD
Dr S.M.M. Tahaghoghi, PhD
Dr Dayang Nurfatimah Awg Iskandar, PhD

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001.
Details of the complaints procedure are available at:
http://www.rmit.edu.au/rd/hrec_complaints

Appendix D

Consent form

RMIT UNIVERSITY

School of Computer Science and IT

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2348
Fax +61 3 9662 1617
www.rmit.edu.au

Prescribed Consent Form For Persons Participating In Research Projects Involving Interviews, Questionnaires or Disclosure of Personal Information

Portfolio		Science, Engineering and Technology	
School of		CS&IT	
Name of participant:			
Project Title:		Relevance Feedback for Content-Based Image Retrieval	
Name(s) of investigator	(1)	Ms Rahayu A.Hamid	Phone:
	(2)	Dr James Thom	Phone:
	(3)	Dr Seyed Mohammad Mehdi (Saied) Tahaghoghi	Phone:
	(4)	Dr Dayang Nurfatimah Awg Iskandar	Phone:

1. I have received a statement explaining the interview/questionnaire involved in this project.

2. I consent to participate in the above project, the particulars of which - including details of the interviews or questionnaires - have been explained to me.

3. I authorise the investigator or his or her assistant to interview me or administer a questionnaire.

4. I acknowledge that:

   a) Having read Plain Language Statement, I agree to the general purpose, methods and demands of the study.
   b) I have been informed that I am free to withdraw from the project at any time and to withdraw any unprocessed data previously supplied.
   c) The project is for the purpose of research and/or teaching. It may not be of direct benefit to me.
   d) The privacy of the personal information I provide will be safeguarded and only disclosed where I have consented to the disclosure or as required by law.
   e) The security of the research data is assured during and after completion of the study. The data collected during the study may be published, and a report of the project outcomes will be provided to RMIT University. Any information which will identify me will not be used.

Participant's Consent

Participant: _____ Date : _____
(Signature)

Witness: _____ Date: _____
(Signature)

Participants should be given a photocopy of this consent form after it has been signed.

Any complaints about your participation in this project may be directed to the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT, GPO Box 2476V, Melbourne, 3001. The telephone number is (03) 9925 2251.
Details of the complaints procedure are available from the above address.

Appendix E

Instructions and questionnaires

The questionnaire consists of:

- Opening remarks and information of study
- Pre-search questionnaire that collects demographic information
- Description of the training task
- Instructions of task, followed by a list of questions

An investigation of users' image search behaviour

Thank you for agreeing to participate in this research. By participating, you are making a valuable contribution. We, the researchers, appreciate your input, and thank you for sharing your experiences and thoughts.

The study will take approximately 1 hour. You as the participant will be asked to:

- sign a consent form,
- complete a background and user experience questionnaire,
- perform a training task,
- perform 4 image search tasks, in which your steps while performing these tasks will be recorded, and
- answer a post-task questionnaire after completing each task.

To preserve your anonymity, we ask you **NOT** to place your name, or any identifying information anywhere on the survey. The confidentiality of your responses will be protected at all times, and the data will be destroyed after completion of the PhD research study. The results of this study will be reported anonymously.

If you have any concerns or require further information, you may do so by emailing the researchers directly. If you have an ethical issue or complaint about the study, you can contact the Executive Officer, RMIT Human Research Ethics Committee, Research & Innovation, RMIT University, GPO Box 2476, Melbourne, 3001. The telephone number is (03) 9925 2251. Details of the complaints procedure are available from the above address.

Investigators:

- Ms Rahayu A.Hamid, (rahayu.ahamid@student.rmit.edu.au)
- Dr James Thom, (james.thom@rmit.edu.au)
- Dr Seyed Mohammad Mehdi (Saied) Tahaghoghi, (Saied.Tahaghoghi@microsoft.com)
- Dr Dayang Nurfatimah Awg Iskandar, (dnfaiz@fit.unimas.my)

Pre-search Questionnaire

In order to better understand and interpret your searching behaviour, please tick (√) on the answer for the following questions regarding your background information.

1. Gender:

- ☐ Male
- ☐ Female

2. Current level of study:

- ☐ Undergraduate
- ☐ Master
- ☐ PhD

3. Area of study/discipline:

- ☐ Computer Science & Information Technology
- ☐ Other:_____

4. How often do you search for information on the World Wide Web?

- ☐ Daily
- ☐ Weekly
- ☐ Monthly
- ☐ Less than once a month
- ☐ Never

5. What are your most frequent reasons for searching for information? (You may select more than one reason)

- ☐ I want to get to a specific website that I already have in mind
- ☐ I want a good website on a particular topic but I don't have a specific website in mind
- ☐ I want to perform some web-mediated activity, for example shopping and downloading
- ☐ I want to find answers to questions on a particular topic
- ☐ I want to know more about a particular topic
- ☐ I want to find a website which is a collection of links to other sites regarding a particular topic
- ☐ I want the best website regarding a particular topic
- ☐ Other:_____

6. Which search engine do you most frequently use to search for information?

_____

7. Why do you choose to use a particular search engine? (You may select more than one reason)

- [ ] It returns results quickly
- [ ] It returns a large number of results
- [ ] It is easy to use
- [ ] Its results are often exactly what I want
- [ ] I like the layout and presentation
- [ ] I have been using it for a long time and I haven't thought seriously about changing
- [ ] Other: _____

8. How often do you search for **images** on the World Wide Web?

- [ ] Daily
- [ ] Weekly
- [ ] Monthly
- [ ] Less than once a month
- [ ] Never

9. In the context of image use, what are your most frequent reasons for searching for **images**? (You may select more than one reason)

- [ ] I want to use images as a means of representing accompanying text
- [ ] I want to use images as a means of analysing information
- [ ] I want to use images as a means of distributing information
- [ ] I want to use images as a means of gaining knowledge
- [ ] I want to use images as a means of providing inspiration
- [ ] I want to use images as a means of decoration
- [ ] I want to use images as a means of conveying a message or emotion
- [ ] Other: _____

10. At what level do you consider your expertise in searching for **images**?

- [ ] Novice
- [ ] Intermediate
- [ ] Expert

11. How familiar are you with the following **image** search engines?

	Bing Images	Google Images	Yahoo Images
Never heard of it	☐	☐	☐
Aware of it but never used it	☐	☐	☐
Have tried using it	☐	☐	☐
Used it quite often	☐	☐	☐
Use it all the time	☐	☐	☐

12. Have you used any other **image** search engines apart from Bing, Google and Yahoo?

☐ Yes, please specify: _____

☐ No

13. Why do you choose to use a particular **image** search engine? (You may select more than one reason)

☐ It returns results quickly

☐ It returns a large number of results

☐ It is easy to use

☐ Its results are often exactly what I want

☐ I like the layout and presentation

☐ I have been using it for a long time and I haven't thought seriously about changing

☐ Other: _____

Training Task

This task is just to familiarise you with the experimental set up. While undertaking this training task, please refer to the “Study Procedure” booklet.

Imagine you are a designer with the responsibility for the design of leaflets on various subjects for the Scottish Tourist Board. These leaflets will consist of a body of text interspersed with 2 images selected on the basis of appropriateness for the leaflets. Your task is to search from a large collection of images from the World Wide Web, and save those that in your opinion would most effectively support the given theme which is ‘The scenic splendour of the Scottish countryside in Autumn and Winter’.

Task 1

1. To start your recording, please click ‘Save As’ in the record interface. Save the recording in the corresponding folder of the task that you are currently performing, i.e. '**Task 1**'.

2. Choose and tick (√) a topic for the task from the following list:

☐ Imagine you are a designer with the responsibility of designing leaflets for the Yellowstone National Park. These leaflets will consist of a body of text interspersed with images. Your task is to search and save 4 images appropriate for the leaflets.

☐ Your team is responsible for designing a brochure for the Mercedes-Benz SL-Class models. Your task is to search and save 4 images that could be used in the brochure.

☐ For your assignment on “Seven Wonders of the Middle Ages”, you have chosen The Taj Mahal as the topic of your report. Your task is to search and save 4 images that are appropriate for the report.

☐ You are writing a special issue article on Harley-Davidson Cruiser bike models. Your task is to search and save 4 images that are appropriate for the article.

3. Complete the task using Google Images search engine.

4. Save all your images in the same folder, i.e. '**Task 1**'.

5. After you have finished the task, please end/stop the recording by clicking the red square on the tray icon and proceed to the questionnaire.

Post-task questionnaire for Task 1

Please tick (√) on the appropriate level of agreement for each of the following statements.

Statements	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. The images I selected were relevant to the search topic.					
2. The images I selected were an accurate representation of what I was looking for on the search topic.					
3. The images I selected gave me new ideas or new insights about the search topic.					
4. The images I selected were interesting in regards to the search topic.					
5. The images I selected contained the kinds of details I could use to clarify important aspects of the search topic.					
6. The images I selected had technical attributes (such as colour, perspective or angle) that were important to me for this search topic.					
7. The images I selected evoked an emotional response in me regarding the search topic.					
8. The images I selected had useful text descriptions on the search topic.					
9. The images I selected contained consequences or implications of the search topic.					
10. The images I selected have a strong visual impact regarding the search topic.					
11. I was familiar with the topic that I choose for this search task.					
12. I was able to create queries for the search task easily.					
13. I had an idea of the kind of images that would satisfy the requirements of the topic before starting the search.					
14. I found that images retrieved by the search engine matched my initial idea of what would satisfy the requirements of the search task.					
15. I frequently changed my mind on the images that I was looking for.					
16. I was satisfied with the images presented to me.					
17. I was satisfied with the order the images were presented to me.					
18. I believe I have seen all possible images that would satisfy the requirements of the search task.					
19. I am very satisfied with my search results.					
20. I saved images that matched my initial idea of what would satisfy the requirements of the search task.					
21. I am very satisfied with my search interaction.					
22. I found overall, the search task was easy to perform.					

If you wish to elaborate on any of your answers or any other aspect of the task, please write your comments here.

Task 2

1. To start your recording, please click ‘Save As’ in the record interface. Save the recording in the corresponding folder of the task that you are currently performing, i.e. '**Task 2**'.

2. Choose and tick (√) a topic for the task from the following list:

   ☐ You are writing an article on “Immigration Nation”. Your task is to search and save 4 images on immigrants that would be appropriate for the article.

   ☐ As a design student, you are required to present a poster on the Japanese traditional garment, Kimono. Your task is to search and save 4 images that would be suitable for your poster.

   ☐ You are preparing a blog entry on “Becoming a Great Athlete”. Your task is to search and save 4 images that would make your entry more interesting.

   ☐ In one of your assignments as an architecture student, you are required to give a presentation on lighthouses. Your task is to search and save 4 images that would be suitable for your presentation.

3. Complete the task using Google Images search engine.

4. Save all your images in the same folder, i.e. '**Task 2**'.

5. After you have finished the task, please end/stop the recording by clicking the red square on the tray icon and proceed to the questionnaire.

Post-task questionnaire for Task 2

Please tick (√) on the appropriate level of agreement for each of the following statements.

Statements	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. The images I selected were relevant to the search topic.					
2. The images I selected were an accurate representation of what I was looking for on the search topic.					
3. The images I selected gave me new ideas or new insights about the search topic.					
4. The images I selected were interesting in regards to the search topic.					
5. The images I selected contained the kinds of details I could use to clarify important aspects of the search topic.					
6. The images I selected had technical attributes (such as colour, perspective or angle) that were important to me for this search topic.					
7. The images I selected evoked an emotional response in me regarding the search topic.					
8. The images I selected had useful text descriptions on the search topic.					
9. The images I selected contained consequences or implications of the search topic.					
10. The images I selected have a strong visual impact regarding the search topic.					
11. I was familiar with the topic that I choose for this search task.					
12. I was able to create queries for the search task easily.					
13. I had an idea of the kind of images that would satisfy the requirements of the topic before starting the search.					
14. I found that images retrieved by the search engine matched my initial idea of what would satisfy the requirements of the search task.					
15. I frequently changed my mind on the images that I was looking for.					
16. I was satisfied with the images presented to me.					
17. I was satisfied with the order the images were presented to me.					
18. I believe I have seen all possible images that would satisfy the requirements of the search task.					
19. I am very satisfied with my search results.					
20. I saved images that matched my initial idea of what would satisfy the requirements of the search task.					
21. I am very satisfied with my search interaction.					
22. I found overall, the search task was easy to perform.					

If you wish to elaborate on any of your answers or any other aspect of the task, please write your comments here.

Task 3

1. To start your recording, please click ‘Save As’ in the record interface. Save the recording in the corresponding folder of the task that you are currently performing, i.e. '**Task 3**'.

2. Choose and tick (√) on a topic for the task from the following list:

☐ Your editor wants you to write an article on economic unrest which will be interspersed with images. Your task is to search and save 4 images that are related to the article.

☐ Your group are preparing a presentation on logical thinking. Your task is to search and save 4 images that would be suitable for the presentation.

☐ You and your classmates are preparing a report on urban development. Your task is to search and save 4 images that would further explain and increase understanding on the topic.

☐ Your company is distributing a free booklet entitled “Keys to Financial Security”. Your task is to search and save 4 images that would be suitable for use in the booklet.

3. Complete the task using Google Images search engine.

4. Save all your images in the same folder, i.e. '**Task 3**'.

5. After you have finished the task, please end/stop the recording by clicking the red square on the tray icon and proceed to the questionnaire.

Post-task questionnaire for Task 3

Please tick (√) on the appropriate level of agreement for each of the following statements.

Statements	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. The images I selected were relevant to the search topic.					
2. The images I selected were an accurate representation of what I was looking for on the search topic.					
3. The images I selected gave me new ideas or new insights about the search topic.					
4. The images I selected were interesting in regards to the search topic.					
5. The images I selected contained the kinds of details I could use to clarify important aspects of the search topic.					
6. The images I selected had technical attributes (such as colour, perspective or angle) that were important to me for this search topic.					
7. The images I selected evoked an emotional response in me regarding the search topic.					
8. The images I selected had useful text descriptions on the search topic.					
9. The images I selected contained consequences or implications of the search topic.					
10. The images I selected have a strong visual impact regarding the search topic.					
11. I was familiar with the topic that I choose for this search task.					
12. I was able to create queries for the search task easily.					
13. I had an idea of the kind of images that would satisfy the requirements of the topic before starting the search.					
14. I found that images retrieved by the search engine matched my initial idea of what would satisfy the requirements of the search task.					
15. I frequently changed my mind on the images that I was looking for.					
16. I was satisfied with the images presented to me.					
17. I was satisfied with the order the images were presented to me.					
18. I believe I have seen all possible images that would satisfy the requirements of the search task.					
19. I am very satisfied with my search results.					
20. I saved images that matched my initial idea of what would satisfy the requirements of the search task.					
21. I am very satisfied with my search interaction.					
22. I found overall, the search task was easy to perform.					

If you wish to elaborate on any of your answers or any other aspect of the task, please write your comments here.

Task 4

1. To start your recording, please click ‘Save As’ in the record interface. Save the recording in the corresponding folder of the task that you are currently performing, i.e. '**Task 4**'.

2. Choose and tick (√) on a topic for the task from the following list:

   ☐ Imagine you are taking a photography course. For your assignment on capturing emotions, you are required to search and save 4 images that best capture the emotion “hope”.

   ☐ For your art assignment, you were asked to set up a mini gallery entitled “Joy”. You are required to search and save 4 images that you could use for the assignment.

   ☐ You are writing a psychology report on curiosity. You are required to search and save 4 images that would be appropriate for the report.

   ☐ You are taking a class on creative writing for which you need to write a fictional story on overcoming adversity. Your task is to search and save 4 suitable images that would accompany your story.

3. Complete the task using Google Images search engine.

4. Save all your images in the same folder, i.e. **'Task 4**'.

5. After you have finished the task, please end/stop the recording by clicking the red square on the tray icon and proceed to the questionnaire.

Post-task questionnaire for Task 4

Please tick (√) on the appropriate level of agreement for each of the following statements.

Statements	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1. The images I selected were relevant to the search topic.					
2. The images I selected were an accurate representation of what I was looking for on the search topic.					
3. The images I selected gave me new ideas or new insights about the search topic.					
4. The images I selected were interesting in regards to the search topic.					
5. The images I selected contained the kinds of details I could use to clarify important aspects of the search topic.					
6. The images I selected had technical attributes (such as colour, perspective or angle) that were important to me for this search topic.					
7. The images I selected evoked an emotional response in me regarding the search topic.					
8. The images I selected had useful text descriptions on the search topic.					
9. The images I selected contained consequences or implications of the search topic.					
10. The images I selected have a strong visual impact regarding the search topic.					
11. I was familiar with the topic that I choose for this search task.					
12. I was able to create queries for the search task easily.					
13. I had an idea of the kind of images that would satisfy the requirements of the topic before starting the search.					
14. I found that images retrieved by the search engine matched my initial idea of what would satisfy the requirements of the search task.					
15. I frequently changed my mind on the images that I was looking for.					
16. I was satisfied with the images presented to me.					
17. I was satisfied with the order the images were presented to me.					
18. I believe I have seen all possible images that would satisfy the requirements of the search task.					
19. I am very satisfied with my search results.					
20. I saved images that matched my initial idea of what would satisfy the requirements of the search task.					
21. I am very satisfied with my search interaction.					
22. I found overall, the search task was easy to perform.					

If you wish to elaborate on any of your answers or any other aspect of the task, please write your comments here.

Appendix F

Recruitment poster

RMIT UNIVERSITY

School of Computer Science and IT

GPO Box 2476V
Melbourne VIC 3001
Australia

Tel. +61 3 9925 2348
Fax +61 3 9662 1617
www.rmit.edu.au

Volunteers needed!

User Study on Image Searching Behaviours

We would like to invite you to participate in a research project that aims to understand how different factors can either individually or collectively affect users' image searching behaviour.

The study explores image information needs and factors that are important to users when performing image search. Participants will be asked to evaluate their search interaction in a questionnaire after completing each image search task.

If you are an undergraduate or post-graduate student enrolled at RMIT and interested in participating, please email Rahayu A Hamid at rahayu.ahamid@student.rmit.edu.au for more details.

Appendix G

Users' rating of agreement on perception of performing image search according to task type

Table G.1: Users' rate of agreements on perception of performing image search based on topics for Specific task

Aspects of performing image search	Statistics	Yellowstone National Park	Mercedes Benz S Class	Taj Mahal	Harley Davidson Cruiser Bike
Familiar with search topic	μ	3.133	3.538	3.471	3.333
	σ	1.246	1.198	1.007	1.155
	# agree	6	8	10	2
Able to create queries easily	μ	4.133	4.154	4.118	4.333
	σ	0.516	0.801	0.857	0.577
	# agree	14	10	14	3
Having an idea on image	μ	3.533	4.308	4.235	2.667
	σ	1.060	0.630	0.437	1.155
	# agree	10	12	17	1
Image matched initial idea	μ	3.667	3.923	4.059	3.667
	σ	0.724	1.038	0.899	0.577
	# agree	10	10	15	2
Frequently change mind about images	μ	2.533	2.462	2.882	3.667
	σ	0.834	1.127	1.317	1.528
	# agree	3	3	8	2
Satisfied with images	μ	3.800	4.000	3.941	3.667
	σ	0.414	1.000	0.827	0.577
	# agree	12	11	15	2
Satisfied with order of images	μ	3.600	3.923	3.647	3.333
	σ	0.507	1.038	0.862	0.577
	# agree	9	10	11	1
Seen all possible images	μ	2.667	3.231	3.235	3.667
	σ	1.175	1.363	1.033	0.577
	# agree	5	6	10	2
Satisfied with search results	μ	3.600	4.000	3.941	4.000
	σ	0.507	1.000	0.659	0.000
	# agree	9	11	15	3
Saved image that matched initial idea	μ	3.867	4.385	4.000	3.667
	σ	0.834	0.506	0.612	0.577
	# agree	11	13	16	2
Satisfied with search interaction	μ	3.800	4.077	3.882	3.667
	σ	0.561	0.954	0.781	0.577
	# agree	11	10	13	2
Satisfied with search results	μ	4.200	4.154	4.235	4.000
	σ	0.676	0.899	0.562	0.000
	# agree	13	11	16	3

Table G.2: Users' rate of agreements on perception of perforiming image search based on topics for Nameable task

Aspects of performing image search	Statistics	Immigrants	Kimono	Athelete	Lighthouse
Familiar with search topic	μ	3.600	3.286	3.889	3.692
	σ	0.894	1.231	0.782	0.855
	# agree	4	11	6	10
Able to create queries easily	μ	3.400	3.905	4.111	4.231
	σ	0.894	0.831	0.601	0.439
	# agree	3	17	8	13
Having an idea on image	μ	4.000	4.000	4.333	3.846
	σ	0.000	0.837	0.500	0.689
	# agree	5	18	9	11
Image matched initial idea	μ	3.400	4.095	3.778	3.615
	σ	0.894	0.700	0.833	0.961
	# agree	3	19	7	11
Frequently change mind about images	μ	3.400	2.714	2.667	2.462
	σ	0.894	1.231	1.323	0.967
	# agree	3	7	4	3
Satisfied with images	μ	3.200	3.952	3.444	3.923
	σ	0.837	0.805	1.236	0.760
	# agree	2	18	6	11
Satisfied with order of images	μ	3.400	3.667	3.000	3.462
	σ	0.894	1.017	1.323	1.050
	# agree	3	14	5	7
Seen all possible images	μ	3.000	3.476	2.111	2.615
	σ	1.000	0.981	1.269	1.261
	# agree	2	12	1	5
Satisfied with search results	μ	3.200	4.143	3.333	4.000
	σ	1.095	0.573	1.225	0.707
	# agree	3	19	5	10
Saved image that matched initial idea	μ	3.400	4.095	4.111	3.923
	σ	0.894	0.436	0.601	0.641
	# agree	3	20	8	12
Satisfied with search interaction	μ	3.200	4.000	3.333	4.000
	σ	1.095	0.707	1.118	0.816
	# agree	3	18	3	11
Satisfied with search s result	μ	3.400	4.333	4.111	3.923
	σ	0.894	0.658	0.333	0.641
	# agree	3	19	9	12

Table G.3: Users' rate of agreements on perception of performing image search on topics for Abstract task

Aspects of performing image search	Statistics	Economic Unrest	Logical Thinking	Urban Development	Financial Security
Familiar with search	μ	3.875	3.286	3.222	3.500
topic	σ	0.991	1.139	1.215	0.756
	# agree	6	6	9	5
Able to create queries	μ	4.000	3.500	3.778	4.000
easily	σ	0.535	1.160	0.808	0.000
	# agree	7	9	14	8
Having an idea on image	μ	4.250	3.357	3.667	4.000
	σ	0.707	1.008	1.188	0.535
	# agree	7	6	14	7
Image matched initial	μ	4.125	3.786	3.278	4.000
idea	σ	0.991	0.699	1.127	0.000
	# agree	7	9	10	8
Frequently change mind s	μ	2.750	3.000	3.167	3.000
about image	σ	1.282	0.961	1.098	1.195
	# agree	3	5	9	3
Satisfied with images	μ	4.125	3.714	3.611	4.000
	σ	0.641	0.914	0.778	0.535
	# agree	7	10	12	7
Satisfied with order	μ	3.875	3.286	3.278	3.750
of images	σ	0.641	0.994	1.127	0.707
	# agree	6	5	9	7
Seen all possible images	μ	3.125	3.143	2.944	3.250
	σ	0.835	1.231	0.873	1.165
	# agree	3	7	5	4
Satisfied with search	μ	4.000	3.571	3.500	4.250
results	σ	0.756	0.756	0.985	0.463
	# agree	6	8	10	8
Saved image that	μ	4.250	3.714	3.556	4.125
matched initial idea	σ	0.463	0.611	1.149	0.641
	# agree	8	9	13	7
Satisfied with search n	μ	4.125		3.444	4.125
interactio	σ	0.641	0.842	1.042	0.641
	# agree	7	10	10	7
Satisfied with search	μ	4.125	3.714	3.333	4.125
results	σ	0.354	1.139	1.283	0.354
	# agree	8	9	10	8

Table G.4: Users' rate of agreements on perception of perforiming image search based on topics for Subjective task

Aspects of performing image search	Statistics	Hope	Joy	Curiosity
Familiar with search topic	μ	3.333	3.842	3.000
	σ	1.090	0.834	1.000
	# agree	12	13	2
Able to create queries easily	μ	3.333	3.684	3.400
	σ	1.239	1.003	0.894
	# agree	11	12	3
Having an idea on image	μ	3.583	4.263	4.200
	σ	1.139	0.653	0.447
	# agree	19	17	5
Image matched initial idea	μ	3.417	3.737	4.000
	σ	1.060	1.098	0.000
	# agree	15	13	5
Frequently change mind about images	μ	3.250	3.053	3.400
	σ	1.113	1.177	0.894
	# agree	13	9	3
Satisfied with images	μ	3.333	3.895	3.000
	σ	0.963	0.658	1.414
	# agree	14	16	3
Satisfied with order of images	μ	2.958	3.369	3.400
	σ	0.999	1.065	0.894
	# agree	8	9	3
Seen all possible images	μ	2.750	3.053	2.400
	σ	1.152	1.177	1.140
	# agree	8	8	1
Satisfied with search results	μ	3.500	3.737	3.400
	σ	0.834	0.806	0.548
	# agree	15	12	2
Saved image that matched initial idea	μ	3.875	4.105	4.000
	σ	0.900	0.658	0.000
	# agree	20	18	5
Satisfied with search n interactio	μ	3.333	3.789	3.400
	σ	1.007	0.713	0.548
	# agree	12	12	2
Satisfied with search results	μ	3.458	3.421	3.200
	σ	1.179	1.121	1.304
	# agree	13	10	1

Appendix H

Time-stamps of users' search interaction according to task type

Table H.1: Users' search interaction by topics for Specific task

Aspects of performing image search	Statistics	Yellowstone National Park	Mercedes Benz S Class	Taj Mahal	Harley Davidson Cruiser Bike
Time to submit initial query	μ	10.867	11.769	6.000	14.667
	σ	7.520	5.847	2.828	12.503
Number of queries submitted	μ	1.933	2.615	3.294	2.667
	σ	0.884	2.181	1.490	2.082
Number of images viewed	μ	11.067	10.462	11.941	17.000
	σ	5.418	5.739	9.397	19.157
Time spent on results pages	μ	113.400	91.462	117.588	103.000
	σ	58.032	46.061	52.623	60.630
Time to complete task	μ	124.231	103.231	123.588	117.667
	σ	61.926	47.034	53.767	56.695

Table H.2: Users' search interaction by topics for Nameable task

Aspects of performing image search	Statistics	Immigrants	Kimono	Athelete	Lighthouse
Time to submit initial query	μ	9.600	9.810	26.444	8.231
	σ	8.204	8.841	34.692	5.449
Number of queries submitted	μ	4.600	3.286	6.222	3.692
	σ	3.782	2.194	3.232	3.066
Number of images viewed	μ	21.800	15.619	13.667	16.923
	σ	11.100	10.879	7.041	14.818
Time spent on results pages	μ	193.400	129.286	189.000	133.846
	σ	41.040	67.696	96.596	67.727
Time to complete task	μ	203.000	139.095	215.444	142.077
	σ	36.586	69.140	94.494	67.702

Table H.3: Users' search interaction by topics for Abstract task

Aspects of performing image search	Statistics	Economic Unrest	Logical Thinking	Urban Development	Financial Security
Time to submit initial query	μ	12.375	8.929	11.111	9.125
	σ	9.531	4.922	10.070	4.324
Number of queries	μ	6.000	3.786	4.611	2.875
	σ	2.268	2.940	2.660	1.553
Number of images viewed	μ	17.375	14.571	20.889	14.000
	σ	8.035	10.931	12.199	6.211
Time spent on results pages	μ	156.250	162.571	169.611	126.750
	σ	62.057	100.595	53.301	46.705
Time to complete task	μ	168.625	171.500	180.722	135.875
	σ	63.392	102.500	52.691	49.470

Table H.4: Users' search interaction by topics for Subjective task

Aspects of performing image search	Statistics	Hope	Joy	Curiosity
Time to submit initial query	μ	9.958	5.263	6.000
	σ	7.636	3.088	2.345
Number of queries	μ	5.750	6.316	9.000
	σ	4.204	4.243	4.416
Number of images viewed	μ	17.125	12.105	26.400
	σ	8.941	6.975	8.961
Time spent on results pages	μ	187.417	140.789	212.400
	σ	95.469	77.983	62.500
Time to complete task	μ	197.375	146.053	218.400
	σ	96.611	78.942	60.575

Appendix I

IP anonymisation script

```
#!/usr/bin/perl -wT

#Sample script to anonymize all IPv4 address looking text in log files

use strict;

$|=1;


use IP::Anonymous;

open FILE, ">output.txt" or die $!;

my @key = (67); #(you can choose any random values between 0 to 255)

my $obj = new IP::Anonymous(@key);


while(defined(my $line=<>))

{
```

```
chomp $line;

if($line =~ /\d{1,3}(?:\.\d{1,3}){3}/)

{

$line =~ s/(\d{1,3}(?:\.\d{1,3}){3})/$obj->anonymize($1)/eg;

}

print FILE $line."\n";

}

close FILE;
```

Bibliography

A. Al-Maskari and M. Sanderson. A review of factors influencing user satisfaction in information retrieval. *Journal of the American Society for Information Science and Technology*, 61(5):859–868, 2010.

A. Al-Maskari, P. Clough, and M. Sanderson. Users' effectiveness and satisfaction for image retrieval. In *Proceedings of the LWA 2006 Workshop*, pages 84–88, Hildesheim, October 2006. Lernen - Wissensentdeckung - Adaptivität.

L. H. Armitage and P. G. B. Enser. Analysis of user need in image archives. *Journal of Information Science*, 23(4):287–299, 1997.

H. Ashman, M. Antunovic, C. Donner, R. Frith, E. Rebelos, J.-F. Schmakeit, G. Smith, and M. Truran. Are clickthroughs useful for image labelling? In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, pages 191–197, September 2009.

A. Aula, R. M. Khan, and Z. Guan. How does search behavior change as search becomes

more difficult? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 35–44, New York, USA, April 2010. ACM.

C. L. Barry. User-defined relevance criteria: An exploratory study. *Journal of the American Society for Information Science*, 45(3):149–159, 1994.

C. L. Barry and L. Schamber. Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing & Management*, 34(2–3):219–236, 1998.

S. Batley. Visual information retrieval: Browsing strategies in pictorial database. In *Proceeding of the 12th International Online Information Meeting*, pages 373–380, December 1988.

N. J. Belkin. Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5:133–143, 1980.

P. Borlund. The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10):913–925, 2003.

P. Borlund and P. Ingwersen. The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53(3):225–250, 1997.

P. Borlund and J. W. Schneider. Reconsideration of the simulated work task situation: A context instrument for evaluation of information retrieval interaction. In *Proceedings of the Third Symposium on Information Interaction in Context*, IIiX '10, pages 155–164, New York, USA, August 2010. ACM.

A. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.

P. D. Bruza and S. Dennis. Query ReFormulation on the Internet: Empirical data and the hyperindex search engine. In *Computer-Assisted Information Searching on Internet*, RIAO '97, pages 488–499, Paris, France, France, June 1997. Le Centre De Hautes Etudes Internatinales D'Informatique Documentaire.

T. Buerger. A model of relevance for reuse-driven media retrieval. In *Proceedings of the 12th International Workshop of the Multimedia Metadata Community, the 2nd Workshop focusing on Semantic Multimedia Database Technologies 2010*, pages 41–54, Saarbrcken, Germany, December 2010. CEUR-WS.

K. Byström. Information and information sources in tasks of varying complexity. *Journal of the American Society for Information Science and Technology*, 53(7):581–591, 2002.

S. K. Chang and A. Hsu. Image information systems: Where do we go from here? *IEEE Transactions on Knowledge and Data Engineering*, 4(5):431–442, 1992.

H.-L. Chen. An analysis of image queries in the field of art history. *Journal of the American Society Information Science Technology*, 52(3):260–273, 2001.

B. Chew, J. A. Rode, and A. Sellen. Understanding the everyday use of images on the Web. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 102–111, New York, USA, October 2010. ACM.

Y. Choi. Effects of contextual factors on image searching on the Web. *Journal of the American Society for Information Science and Technology*, 61(10):2011–2028, 2010a.

Y. Choi. Investigating variation in querying behavior for image searches on the web. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–10, 2010b.

Y. Choi. Analysis of image search queries on the web: Query modification patterns and semantic attributes. *Journal of the American Society for Information Science and Technology*, 64(7):1423–1441, 2013.

Y. Choi and E. Rasmussen. Searching for images: The analysis of users queries for image retrieval in American history. *Journal of the American Society For Information Science and Technology*, 54(6):498–511, 2003.

Y. Choi and E. M. Rasmussen. Users' relevance criteria in image retrieval in American history. *Information Processing & Management*, 38(2):695–726, 2002.

E. Chung and J. Yoon. An exploratory analysis on unsuccessful image searches. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–2, 2010.

E. Chung and J. Yoon. Image needs in the context of image use: An exploratory study. *Journal of Information Science*, 37(2):163–177, 2011.

J. Cohen. *Statistical Power Analysis for the Behavioral Sciences (2nd Edition)*. Lawrence Erlbaum Associates, Inc., New Jersey, 1988.

J. Cohen. A Power Primer. *Psychological Bulletin*, 112(1):155–159, 1992.

L. R. Conniss, A. J. Ashford, and M. E. Graham. Information seeking behavior in image retrieval: Visor I Final Report. Technical report, Institute for Image Data Research, Newcastle upon Tyne, 2000.

A. Crystal and J. Greenberg. Relevance criteria identified by health information users during Web searches. *Journal of the American Society for Information Science and Technology*, 57(10):1368–1382, 2006.

S. Cunningham and M. Masoodian. Looking for a picture: an analysis of everyday image information searching. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries 2006*, JCDL '06, pages 198–199, New York, USA, June 2006. ACM.

S. J. Cunningham, D. Bainbridge, and M. Masoodian. How people describe their image information needs: A grounded theory analysis of visual arts queries. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '04, pages 47–48, New York, USA, June 2004. ACM.

R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):5:1–5:60, 2008.

J. P. Eakins. Towards intelligent image retrieval. *Pattern Recognition*, 35(1):3–14, 2002.

P. G. B. Enser and C. G. McGregor. Analysis of visual information retrieval queries. British Library Research & Development Report, 1993.

E. Estellés-Arolas and F. González-Ladrón-de Guevara. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2):189–200, 2012.

R. Fidel. The image retrieval task: Implications for the design and evaluation of image databases. *New Review of Hypermedia and Multimedia*, 3(1):181–199, 1997.

T. Fukumoto. An analysis of image retrieval behavior for metadata type image database. *Information Processing & Management*, 42(3):723–728, 2006.

A. Göker, R. Butterworth, A. MacFarlane, T. S. Ahmed, and S. Stumpf. Expeditions through image jungles: The commercial use of image libraries in an online environment. *Journal of Documentation*, 72(1):5–23, 2016.

A. Goodrum and A. Spink. Image searching on the Excite Web search engine. *Information Processing & Management*, 37(2):295–311, 2001.

A. A. Goodrum. I can't tell you what I want, but I'll know it when I see it: Terminological disconnects in digital image reference. *Reference & User Services Quarterly*, 45(1):46–53, 2005.

J. Gwizdka and I. Lopatovska. The role of subjective factors in the information search process. *Journal of the American Society for Information Science and Technology*, 60(12): 2452–2464, 2009.

M. Hachman. Google images gets revamped interface, more relevant results. http://www.pcmag.com/article2/0,2817,2366736,00.asp, 2010.

R. A. Hamid and J. A. Thom. Criteria that have an effect on users while making image relevance judgments. In *Proceedings of the Fifteenth Australasian Document Computing Symposium*, pages 76–83, December 2010.

R. A. Hamid, J. A. Thom, and D. N. F. A. Iskandar. Effects of relevance criteria and

subjective factors on web image searching behaviour. *Journal of Information Science*, 0 (0):0165551516666968, 0. doi: 10.1177/0165551516666968.

S. K. Hastings. Query categories in a study of intellectual access to digitized art images. In *Proceedings of the ASIS Annual Meeting*, volume 32, pages 3–8. ERIC, 1995.

S. K. Hastings. Evaluation of image retrieval systems: Role of user feedback. *Library Trends*, 48(2):438–452, 1999.

D. He, A. Göker, and D. J. Harper. Combining evidence for automatic web session identification. *Information Processing & Management*, 38(5):727–742, 2002.

C. R. Hildreth. Accounting for users' inflated assessments of on-line catalogue search performance and usefulness: An experimental study. *Information Research*, 6(2), 2001. URL http://InformationR.net/ir/6-2/paper101.html.

S. G. Hirsh. Childrens relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science*, 50(14):1265–1283, 1999.

L. Hollink, A. T. Schreiber, B. J. Wielinga, and M. Worring. Classification of user image descriptions. *International Journal of Human-Computer Studies*, 61(5):601–626, 2004.

V. Hollink, T. Tsikrika, and A. P. de Vries. Semantic search log analysis: A method and a study on professional image search. *Journal of the American Society for Information Science and Technology*, 62(4):691–713, 2011.

C. Hölscher and G. Strube. Web search behavior of Internet experts and newbies. *Computer Networks*, 33(1-6):337–346, 2000.

J. Huang and E. N. Efthimiadis. Analyzing and evaluating query reformulation strategies in web search logs. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 77–86, New York, USA, November 2009. ACM.

K. Huang and D. Kelly. The daily image information needs and seeking behavior of chinese undergraduate students. *College and Research Libraries*, 74(3):243–261, 2013.

T.-Y. Hung. *Search Strategies For Image Retrieval in The Field of Journalism.* PhD thesis, School of Communication, Information and Library Studies, Rutgers University, 2006.

T.-Y. Hung, C. Zoeller, and S. Lyon. Relevance judgments for image retrieval in the field of journalism: A pilot study. In *Proceedings of the 8th International Conference on Asian Digital Libraries: Implementing Strategies and Sharing Experiences*, ICADL'05, pages 72–80, Berlin, Heidelberg, December 2005. Springer-Verlag.

A. Jaimes and S.-F. Chang. Conceptual framework for indexing visual information at multiple levels. In *Electronic Imaging*, pages 2–15. International Society for Optics and Photonics, January 1999.

B. Jansen. Searching for digital images on the web. *Journal of Documentation*, 64(1):81–101, 2008.

B. Jansen and A. Spink. An analysis of web document retrieved and viewed. In *4th International Conference on Internet Computing*, pages 65–69, Las Vegas, Nevada, June 2003.

B. Jansen and A. Spink. How are we searching the World Wide Web? A comparison of nine

search engine transaction logs. *Information Processing & Management*, 42(1):248–263, 2006.

B. J. Jansen. Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research*, 28(3):407–432, 2006.

B. J. Jansen. Understanding user-web interactions via web analytics. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–102, 2009.

B. J. Jansen, A. Goodrum, and A. Spink. Searching for multimedia: analysis of audio, video and image Web queries. *World Wide Web*, 3(4):249–254, 2000a.

B. J. Jansen, A. Spink, and T. Saracevic. Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing & Management*, 36(2): 207–227, 2000b.

B. J. Jansen, A. Spink, and J. O. Pedersen. The effect of specialized multimedia collections on web searching. *Journal of Web Engineering*, 3(3-4):182–199, 2004.

B. J. Jansen, A. Spink, and B. Narayan. Query modifications patterns during web searching. In *Information Technology, 2007. ITNG'07. Fourth International Conference on*, pages 439–444. IEEE, April 2007.

B. J. Jansen, D. L. Booth, and A. Spink. Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management*, 44(3): 1251–1266, 2008.

B. J. Jansen, D. L. Booth, and A. Spink. Patterns of query reformulation during Web searching. *Journal of the American Society for Information Science and Technology*, 60 (7):1358–1371, 2009.

C. Jörgensen and P. Jörgensen. Image querying by image professionals: Research Articles. *Journal of the American Society for Information Science and Technology*, 56(12):1346–1359, 2005.

L. H. Keister. User types and queries: Impact on image access systems. *Challenges in Indexing Electronic Text and Images*, pages 7–22, 1994.

M. Kellar, C. Watters, and M. Shepherd. A field study characterizing Web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology*, 58(7):999–1018, 2007.

D. Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(1–2):1–224, 2009.

D. Kelly and C. Cool. The effects of topic familiarity on information search behavior. In *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital libraries*, JCDL '02, pages 74–75, New York, USA, July 2002. ACM.

D. Kelly, D. J. Harper, and B. Landau. Questionnaire mode effects in interactive information retrieval experiments. *Information Processing & Management*, 44(1):122–141, 2008.

N. Khashman, J. Dorey, and E. Ménard. If I could read your mind: An exploration of online image searcher's behaviors. In *Proceedings of the 76th ASIS&T Annual Meeting:*

Beyond the Cloud: Rethinking Information Boundaries, ASIST '13, pages 105:1–105:4, Silver Springs, MD, USA, November 2013. American Society for Information Science.

S. Kim and S. Oh. Users' relevance criteria for evaluating answers in a social Q&A site. *Journal of the American Society for Information Science and Technology*, 60(4):716–727, 2009.

J. L. Klavans, R. LaPlante, and J. Golbeck. Subject matter categorization of tags applied to digital images from art museums. *Journal of the Association for Information Science and Technology*, 65(1):3–12, 2014.

T. Klosowski. Clever uses for reverse image search. http://lifehacker.com/clever-uses-for-reverse-image-search-473032092, 2013.

E. Konkova, A. MacFarlane, and A. Göker. Analysing creative image search information needs. In *ISKO UK biennial conference Knowledge Organization making a difference. The impact of knowledge organization on society, scholarship and progress*. ISKO UK, July 2015.

T. Lau and E. Horvitz. Patterns of search: Analyzing and modeling web query refinement. In J. Kay, editor, *UM99 User Modeling: Proceedings of the Seventh International Conference*, pages 119–128. Springer, Vienna, 1999.

Y. Li and N. J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management*, 44(6):1822–1837, 2008.

C. Liu, J. Gwizdka, J. Liu, T. Xu, and N. J. Belkin. Analysis and evaluation of query

reformulations in different task types. In *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47*, ASIS&T '10, pages 17:1–17:10, Silver Springs, MD, USA, October 2010a. American Society for Information Science.

J. Liu, M. J. Cole, C. Liu, R. Bierig, J. Gwizdka, N. J. Belkin, J. Zhang, and X. Zhang. Search behaviors in different task types. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, JCDL '10, pages 69–78, New York, USA, June 2010b. ACM.

J. Liu, C. Liu, J. Gwizdka, and N. J. Belkin. Can search systems detect users' task difficulty?: Some behavioral signals. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 845–846, New York, USA, July 2010c. ACM.

J. Liu, C. Liu, X. Yuan, and N. J. Belkin. Understanding searchers' perception of task difficulty: Relationships with task type. *Proceedings of the American Society for Information Science and Technology*, 48(1):1–10, 2011.

Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262–282, 2007.

W. Lu and Y. Jia. An eye-tracking study of user behavior in web image search. In D.-N. Pham and S.-B. Park, editors, *PRICAI 2014: Trends in Artificial Intelligence: 13th Pacific Rim International Conference on Artificial Intelligence*, pages 170–182. Springer International Publishing, Cham, 2014.

M. Markkula and E. Sormunen. End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval*, 1(4):259–285, 2000.

M. Mat-Hassan and M. Levene. Associating search and navigation behavior through log analysis. *Journal of the American Society for Information Science and Technology*, 56(9): 913–934, 2005.

L. McCay-Peet and E. Toms. Image use within the work task model: Images as information and illustration. *Journal of the American Society for Information Science and Technology*, 60(12):2416–2429, 2009.

A. Mojsilovic and B. Rogowitz. Capturing image semantics with low-level descriptors. In *Proceedings 2001 International Conference on Image Processing*, volume 1, pages 18–21. IEEE, September 2001.

S. Orlando and F. Silvestri. Mining query logs. In M. Boughanem, C. Berrut, J. Mothe, and C. Soule-Dupuy, editors, *Advances in Information Retrieval: 31th European Conference on IR Research, ECIR 2009*, pages 814–817. Springer, Berlin, Heidelberg, 2009.

S. Ornager. The newspaper image database: empirical supported analysis of users' typology and word association clusters. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '96, pages 212–218, New York, USA, July 1996. ACM.

S. Ornager. Image retrieval: Theoretical and empirical user studies on accessing information

in images. In *ASIS 97: Proceedings of the 60th ASIS Annual Meeting*, volume 34, pages 202–211, 1997.

J. Palotti, A. Hanbury, H. Müller, and C. E. Kahn Jr. How users search and what they search for in the medical domain. *Information Retrieval Journal*, 19(1-2):189–224, 2016.

E. Panofsky. *Meaning in the Visual Arts. Papers in and on Art History.* Doubleday Anchor, New York, NY, 1955.

J. Y. Park, N. O'Hare, R. Schifanella, A. Jaimes, and C.-W. Chung. A large-scale study of user image search behavior on the web. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 985–994, New York, USA, April 2015. ACM.

H.-T. Pu. A comparative analysis of web image and textual queries. *Online Information Review*, 29(5):457–467, 2005.

H.-T. Pu. An analysis of failed queries for web image retrieval. *Journal of Information Science*, 34(3):275–289, 2008.

P. Qu, C. Liu, and M. Lai. The effect of task type and topic familiarity on information search behaviors. In *Proceedings of the Third Symposium on Information Interaction in Context*, IIiX '10, pages 371–376, New York, USA, August 2010. ACM.

S. Y. Rieh and H. I. Xie. Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information Processing & Management*, 42(3):751–768, 2006.

Y. Rui, T. S. Huang, and S.-F. Chang. Image Retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10(1):39 – 62, 1999.

T. Saracevic. 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, 2007.

L. Schamber. Users criteria for evaluation in a multimedia environment. In *Proceedings of the 54th Annual Meeting of the American Society for Information Science*, pages 27–31, October 1991.

S. Sedghi, M. Sanderson, and P. Clough. A study on the relevance criteria for medical images. *Pattern Recognition Letter*, 29(15):2046–2057, 2008.

S. Shatford. Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly*, 6(3):39–62, 1986.

F. Silvestri. Mining query logs: Turning search usage data into knowledge. *Found. Trends Inf. Retr.*, 4(1—2):1–174, 2010.

G. Singer, U. Norbisrath, and D. Lewandowski. Ordinary search engine users carrying out complex search tasks. *Journal of Information Scicience*, 39(3):346–358, 2013.

A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

G. Smith, C. Brien, and H. Ashman. Evaluating implicit judgments from image search clickthrough data. *Journal of the American Society for Information Science and Technology*, 63(12):2451–2462, 2012.

J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan. Web usage mining: discovery and applications of usage patterns from Web data. *ACM SIGKDD Explorations Newsletter*, 1 (2):12–23, 2000.

P. Sturgis, C. Roberts, and P. Smith. Middle alternatives revisited: How the neither/nor response acts as a way of saying "I don't know'? *Sociological Methods & Research*, 43(1): 15–38, 2014.

H. Tamura and N. Yokoya. Knowledge based image analysis image database systems: A survey. *Pattern Recognition*, 17(1):29–43, 1984.

P. Tandon, P. Nigam, V. Pudi, and C. V. Jawahar. FISH: A practical system for fast interactive image search in huge databases. In *Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval*, CIVR '08, pages 369–378, New York, USA, July 2008. ACM.

A. R. Taylor, C. Cool, N. J. Belkin, and W. J. Amadio. Relationships between categories of relevance criteria and stage in task completion. *Information Processing & Management*, 43(4):1071–1084, 2007.

J. Tessier, W. Crouch, and P. Atherton. New measures of user satisfaction with computer based literature searches. *Special Libraries*, 68(11):383–389, 1977.

D. Tjondronegoro, A. Spink, and B. J. Jansen. A study and comparison of multimedia Web searching: 1997–2006. *Journal of the American Society for Information Science and Technology*, 60(9):1756–1768, 2009.

L.-C. J. Tseng and D. Tjondronegoro. Effect of topic domain and task type on web image searching. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI '10, pages 348–351, New York, USA, November 2010. ACM.

L.-C. J. Tseng, D. W. Tjondronegoro, and A. H. Spink. Analyzing web multimedia query reformulation behavior. In *Proceedings of the 14th Australasian Document Computing Symposium*, pages 118–125, Sydney, Australia, December 2009.

J. Urban and J. M. Jose. EGO: A personalized multimedia management and retrieval tool. *International Journal of Intelligent Systems*, 21(7):725–745, 2006.

P. Vakkari and N. Hakala. Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*, 56(5):540–562, 2000.

L. Wen, I. Ruthven, and P. Borlund. The effects on topic familiarity on online search behaviour and use of relevance criteria. In M. Lalmas, A. MacFarlane, S. Rüger, A. Tombros, T. Tsikrika, and A. Yavlinsky, editors, *Advances in Information Retrieval: 28th European Conference on IR Research, ECIR 2006*, pages 456–459. Springer, Berlin, Heidelberg, 2006.

S. Westman. Image users' needs and searching behaviour. In A. Göker and J. Davies, editors, *Information Retrieval*, pages 63–83. John Wiley & Sons, Ltd, Chichester, UK, 2009.

S. Westman and P. Oittinen. Image retrieval by end-users and intermediaries in a journalistic work context. In *Proceedings of the 1st International Conference on Information Interaction in Context*, IIiX, pages 102–110, New York, USA, October 2006. ACM.

S. Westman, A. Lustila, and P. Oittinen. Search strategies in multimodal image retrieval. In *Proceedings of the Second International Symposium on Information Interaction in Context*, IIiX '08, pages 13–20, New York, USA, October 2008. ACM.

B. Wildemuth, L. Freund, and E. G. Toms. Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation*, 70 (6):1118–1140, 2014.

Y. C. Xu and Z. Chen. Relevance judgment: What do information users consider beyond topicality? *Journal of the American Society for Information Science and Technology*, 57 (7):961–973, 2006.

J. Yoon. Searching images in daily life. *Library & Information Science Research*, 33(4): 269–275, 2011.

J. Yoon and E. Chung. Image query reformulation over different search stages. *Proceedings of the American Society for Information Science and Technology*, 46(1):1–5, 2009.

J. Yoon and E. Chung. Understanding image needs in daily life by analyzing questions in a social Q & A site. *Journal of the American Society for Information Science and Technology*, 62(11):2201–2213, 2011.

X. S. Zhou and T. S. Huang. CBIR: From low-level features to high level semantics. In

Proceedings of the SPIE, Image and Video Communication and Processing, pages 426–431, January 2000.

D. Zhu and B. Carterette. An analysis of assessor behavior in crowdsourced preference judgments. In *SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation*, pages 17–20, 2010.