Relevance judgment during health information seeking on a discussion forum

Pian, Wenjing

2016

Pian, W. (2016). Relevance judgment during health information seeking on a discussion forum. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/67016>

<https://doi.org/10.32657/10356/67016>

Nanyang Technological University

Downloaded on 26 Apr 2023 21:29:59 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**RELEVANCE JUDGMENT DURING HEALTH INFORMATION
SEEKING ON A DISCUSSION FORUM**

PIAN WENJING

WEE KIM WEE SCHOOL OF COMMUNICATION AND INFORMATION

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2015

Abstract

More and more people are seeking and accessing health information on the Internet. In addition to authoritative health information websites maintained by healthcare organizations and government agencies, social media sites are increasingly being accessed to obtain and share patients' experiences of diseases and treatments. Various surveys have found that people are using Internet health discussion forums for both informational and emotional support. However, it is not clear what kinds of health information people find useful in health discussion forums, and what relevance criteria they use in making relevance decisions. Most of the previous studies of relevance criteria have focused on relevance judgments at the document or document surrogate level in the context of purposive information searching with an information retrieval system or search engine. Moreover, few studies have made use of an eye tracker system to investigate eye movement behaviors during information seeking and the kinds of information users focused on.

The objective of this study was to identify the detailed relevance criteria people use when browsing a health discussion forum, for three categories of information needs: when seeking for their own health issue, seeking for other people's health issue, and browsing with no particular health issue in mind. An eye tracker system was used to capture users' eye movement behavior, and the text they skim over and focus on during browsing. These data were analyzed to find associations between users' eye movement patterns and the relevance criteria they used.

In the analysis, browsing a discussion forum was divided into two stages: *scanning* a set of post surrogates (post titles) in the summary result screen, and *reading* a detailed post content (including replies and comments from other users). Within each stage, the user will skim over some of the text or examine a text passage closely (i.e. eye fixation). By analyzing the text that people's eyes fixated on, the types of health information used in the relevance judgment can be determined.

It was found that participants with different types of health information needs (i.e. seeking for self, seeking for others, and with no particular issue) made use of different

types of relevance criteria and had different eye movement patterns. Participants seeking for their own health issue used mostly case-based relevance criteria such as symptoms, history of disease in their judgments, which took a longer time to scan post surrogates and reading the detailed posts. Participants seeking for other people were found to use mostly general knowledge-based criteria such as medical terms, cause of disease and basic treatments in their relevance judgments, which took a moderate amount of time to scan post surrogates and read post content. In contrast, participants browsing with no particular health issue used mostly their personal and general interest topics in their judgments, which took a relatively long time to scan post surrogates and to skim detailed post content.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere appreciation to my supervisors, Dr. Christopher, Khoo Soo Guan, and Dr. Chang Yun-ke for their invaluable guidance, support and encouragement during every stage of this study.

I am also thankful to my parents and wife for their concern and endless encouragement. Every time I feel tired, I will get energized with the thought of their love and support.

The work presented in this report has been carried out at the School of Communication and Information, Nanyang Technological University (NTU). The financial support from NTU is gratefully acknowledged.

Finally, I extend my gratitude to my friends who have shared the laughter and tears with me. They have made the past year a wonderful and enriching experience.

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vi
LIST OF FIGURES	viii
CHAPTER 1 INTRODUCTION	1
1.1 Background	1
1.2 Research Framework.....	8
1.3 Research Contributions.....	10
CHAPTER 2 LITERATURE REVIEW	11
2.1 Health Information Seeking.....	11
2.2 Relevance Judgment & Relevance Criteria.....	18
2.3 Eye Tracking Studies.....	26
2.4 Conclusion.....	35
CHAPTER 3 RESEARCH FRAMEWORK & METHOD	36
3.1 Research Framework	36
3.2 Choice of Health Discussion Forum	39
3.3 Participant Sampling	41
3.4 Experiment Procedure.....	44
3.5 Data Analysis.....	46
3.6 Results Of Preliminary Study	53
3.7 Derivation of coding scheme.....	55
CHAPTER 4 RESULTS	62
4.1 Quantitative Data Analysis	62
4.2 Content Analysis	77
4.3 Post Experiment Interview Analysis.....	92
4.4 Analysis of Exceptions.....	94
4.5 Discussion.....	98

CHAPTER5 CONCLUSION 103

5.1 Main Findings 103

5.2 Contributions to Research Methodology, and Interface Design 109

5.3 Limitations 111

5.4 Future Research Directions 112

REFERENCES 114

APPENDIX 1 Summary of relevance criteria from previous studies..... 124

LIST OF TABLES

Table 1.1	Stages of information seeking behavior in a discussion forum	9
Table 3.1	Stages of health information seeking within a health discussion forum, and the two types of relevance judgment associated with each stage	36
Table 3.2	10 health discussion forums retrieved using Google search engine...	40
Table 3.3	Summary of demographic information of research participants	42
Table 3.4	Summary of demographics for participants with different types of use contexts.....	43
Table 3.5	Interview guide to obtain participant's comments on video recording of eye movement behavior.....	45
Table 3.6	Definition of examining, skimming and regression	50
Table 3.7	Coding scheme for different types of health information	61
Table 4.1	Summary of statistics of scanning & reading session	64
Table 4.2	Summary of ANOVA & Post Hoc results for scanning duration	65
Table 4.3	Summary of ANOVA and Post Hoc results for skimming and examining within the scanning stage	69
Table 4.4	Summary of ANOVA & post hoc results for reading duration.....	71
Table 4.5	Summary of ANOVA & post hoc results for skimming & examining within the reading stage.....	74
Table 4.6	Ratio of examining duration to skimming duration for posts judged as relevant and non-relevant	77
Table 4.7	Health topics specified by participants in the experiments.....	80
Table 4.8	Percentages of health information in the set of skimmed surrogates compared to the fixated surrogate.....	83
Table 4.9	Percentage of fixated surrogates that were selected, for each type of health information.....	86
Table 4.10	Summary of types of health information that are important in determining the relevance of the detailed posts	91

Table 4.11 Coding scheme for post experiment interview data 93

Table 4.12 Post hoc test results for participants that were identified using
particular broad category of relevance criteria..... 101

LIST OF FIGURES

Figure 3.1	Screenshot of detailed post page with people’s eye fixations	47
Figure 3.2	Data sheet of user eye fixations	52
Figure 4.1	Distribution of average scanning duration for the three groups of participants	67
Figure 4.2	Distribution of average reading duration for the three groups of participants	73
Figure 4.3	Screenshot of post surrogates summary page	78
Figure 4.4	Screenshot of detailed post page	79
Figure 4.5	Percentage of post surrogates containing each type of health information for the three groups of participants	81
Figure 4.6	Percentage of selected post surrogates containing each type of health information	85
Figure 4.7	Percentages of particular types of health information in relevant posts.....	88
Figure 4.8	Percentages of posts (read) containing each particular type of health information	89

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

This is a study of consumer health information seeking on a particular type of social media site—a health discussion forum. This study focused on the relevance criteria used by people when browsing for health information. People’s eye movement behavior of skimming over the text and focusing on particular passages during browsing was captured with an eye tracker machine. Content analysis of the data was performed to identify the kinds of information people skimmed over and focused on when making relevance judgments, that is, deciding whether a post contains relevant information. The study also investigated associations between people’s eye movement patterns and the types of relevance criteria they used.

Increasingly people are seeking and accessing health information on the Internet. A telephone survey of U.S. adults conducted in 2012 by the PEW Internet & American Life Project (Fox & Duggan, 2013) reported that 59% of U.S. adults had sought health information on the Internet in the past year. Internet users sought various kinds of health information online. The PEW survey found that among the people who looked for online health information, 53% had sought information about specific diseases or problems, 43% for particular medical treatments, 27% for weight loss and control information, 19% for food safety information and 15% for drug information. But it is not clear how people decide what health information is

relevant for particular purposes (e.g., diagnosis) and in different situations, and what relevance criteria they use.

Relevance, *relevance judgment* and *relevance criterion* are difficult concepts to define. As early as 1975, Saracevic noted that *relevance* is a vague but well understood term without clear explanation. In other words, researchers have difficulty defining the term, but users have no trouble deciding whether a document or piece of information is relevant or not. A review of the literature is given in Section 2.2. For the purpose of this study, I define *relevance* as a user's subjective assessment of the relation between a piece of information or document and the user's situation. A *relevant* piece of information has a positive relationship to the user's situation, in that it evokes a positive sentiment in the user. In this study, *relevance* is operationalized in two ways:

1. Evaluative relevance: a piece of information or document is deemed relevant if the user says it is relevant.
2. Predictive relevance: a document is deemed likely to be relevant, if the user selects or clicks on a document surrogate (e.g. document title) to view the full document content.

Relevance judgment is then defined as a user's decision or process of deciding whether a piece of information or a document is relevant (evaluative relevance) or likely to be relevant (predictive relevance). This two-stage framework of relevance judgment is discussed later. I define *relevance criterion* as a reason that contributes to the user's relevance judgment. However, the reason can be expressed at different levels of abstraction. High-level reasons

include usefulness, topicality and quality. This study looks at the detailed information in the text as information cues that affect relevance judgment. So, in the context of this study, relevance criteria are the pieces of information (expressed in the document) that the reader uses to decide whether the information or document is relevant. In this thesis, they were operationalized as the information cues that users observe and help them to make the relevance judgment.

Most of the previous studies on relevance judgment were in the context of information retrieval in online bibliographic or full-text databases (e.g., Barry, 1994; Schamber, 1991;). More recently, researchers studied relevance judgments on Web search engines (e.g., Balatsoukas & Ruthven, 2012; Tombros, Ruthven, & Jose, 2003; Kim & Oh, 2007), and identified various kinds of relevance criteria used during the information search process. With the rising popularity of social media sites among Internet users, it is important to know what relevance criteria people use when browsing social media sites, and whether the criteria are the same as for search engines and online databases.

There are different types of social media applications with different characteristics and for each type of social media applications, the user information behavior varies. This study focused on a health discussion forum—HealthBoards.com (<http://www.healthboards.com/boards/index.php>). On a discussion forum, users post health issues, their views and responses to other people's posts. The posts on discussion forums are organized by topics and sub-topics in a hierarchical structure, which is different from Facebook and Instagram where people post information sequentially like in a news feed,

without organizing their posts. Users of a discussion forum focus on browsing by topic the posts and responses to posts, rather than checking only the responses to their own posts or the posts of specific people, as in Facebook. This is different from traditional online information searching where researchers have focused on relevance judgment in the context of searching, rather than browsing. Users may exhibit different browsing and relevance judgment behaviors on social media sites. I have not come across research reports of relevance judgment on social media sites, especially in the context of browsing.

The structure of a discussion forum bears some similarity to information retrieval systems. The user can select a topic by entering a query keyword in a search box or by browsing a classified directory of topics and subtopics. Having selected a topic, the system displays a summary result screen displaying a list of post surrogates (usually the post header, including the title, author, and the number of user views and replies). The user has to scan the post surrogates (mainly the post titles) to select posts that are likely to contain relevant information. When a post surrogate is clicked on, the system displays a detailed post screen showing the content of the main post and response posts from other users. As a post often contains response posts from other users, and responses to the responses, it is sometimes called a discussion thread. In HealthBoards.com, the responses to a post are appended to the end of the post content page, so a post and responses are equivalent to a full-text document in an information retrieval system. The user reads the post and its responses to determine whether post thread is relevant or likely to be relevant. The user also consumes the information by learning something.

There have been many studies of health information seeking. Several surveys have focused on what types of health information people are looking for and what are the demographic differences between user groups (e.g., Fox & Duggan, 2013). Other studies have focused on different motivations that people have when doing online health information seeking. People sometimes intentionally seek health information for their own problems (Lambart & Loiselle, 2007), and at other times encounter useful or interesting information by chance or with no particular issue in mind (Palsdottir, 2009). Fox and Duggan (2013) also found that people seek health information online sometimes for themselves and sometimes for others. This study distinguished between three types of health information seeking context:

1. seeking information for the user's own health issue,
2. seeking information on behalf of other people (i.e. for someone else's health issue),
and
3. browsing with no particular health issue in mind (but with possible tacit interest in particular health issues, which cannot be determined ahead of time).

This user context dimension is applicable to any health topic.

The objectives of this study were:

1. To find out the relevance criteria that people use to make relevance judgments on a health discussion forum, in three types of use context: seeking information for their own health issue, seeking for other's health issue, and browsing without a particular health issue in mind.
2. To find patterns of eye movement behavior associated with the three types of use

context (i.e. seeking for self, seeking for others, and browsing with no issue in mind), as well as the relevance criteria used.

This study makes use of an eye tracker machine to capture users' eye movements. Text passages that the user's eyes focus on are assumed to reflect information that the user is considering in making relevance judgments. These pieces of information are possible relevance criteria. In addition, the pattern of eye movements (e.g., the relative proportion of skimming and examining) suggest relative importance of particular types of information. If particular eye movement patterns are associated with particular categories of relevance criteria, they can provide additional support for the results, and can be used to infer the use of these types of relevance criteria in the future.

This study focused on one health discussion forum—HealthBoards.com (<http://www.healthboards.com/boards/index.php>). This health discussion forum is one of the biggest health forums: it has 1 million registered users, and thousands of new posts and replies everyday based on the statistics shown on the bottom of the main page. I searched for health discussion forums using the Google search engine and found 10 candidates for this study. I chose Healthboards.com as the study platform for its comprehensiveness, up-to-date information, and popularity. Details of the selection process are given in the Research Method chapter.

Eye tracker machines have been used in research studies to provide a different dimension of evidence than those traditionally obtained from user interviews. Buscher, Dengel and Elst (2008) developed quantitative measures of user eye movement behavior to quantify users'

skimming (quick eye movement over text) and examining (eye focus) durations when browsing or reading a document. They used these two types of eye movements to investigate the relevance judgment people tend to make when browsing the documents. Miwa, Egusa, Saito, Takaku, Terai and Kando (2011) used eye movement information to analyze people's information encountering, and to help research participants recall their thinking process when their eyes were focused on particular text passages during information seeking. Balasoukas and Ruthven (2012) recorded eye fixations to find out what people focused on during their Google search and how they judged the relevance of search surrogate results. In this study, the durations of text skimming and text examining were analyzed, together with the content of the text.

The study population was limited to Singaporean Chinese, and the participants were relatively healthy with no critical or terminal disease. Most of the participants were also undergraduate and postgraduate students and administrative staff of a university in Singapore.

The sampling focuses on adults of 18 to 50 years, with at least undergraduate education.. Most of the participants were associated with the university. Other than these, no other demographic biases have been noticed. As the task does not require health expertise – the users only have laymen/patient-level health knowledge. I think that the study results will hold with other similar groups. The focus is on basic human cognitive processes that everyone uses.

1.2 RESEARCH FRAMEWORK

In traditional relevance judgment studies in the context of information retrieval, participants are usually asked to make relevance judgments of document surrogates or the full-text documents retrieved. When browsing a discussion forum, users also have to assess the post surrogates first before selecting and reading the detailed post content and replies from users. In this study, I adopted a two-stage framework of relevance judgment proposed by Tang and Solomon (2001), Rieh (2002) and Crystal and Greenburg (2006). This two-stage framework of relevance judgment divides relevance judgment made by users during their information seeking process into two stages:

- *Scanning the surrogates* displayed by the information retrieval system to make *predictive* judgments—assessing the likelihood of relevance of the information (i.e. document content) represented by the surrogate;
- *Reading the documents* linked to the surrogates and making *evaluative* judgments based on the content of the document—assessing the content or content chunks (e.g., paragraphs or passages) in the document.

The two-staged framework of relevance judgment is incorporated into a three-stage information seeking framework given in Table 1.1. Stage 1 (searching or browsing the classified directory) is beyond the scope of this study. Stage 2 (scanning the post surrogates) is the stage where the user scans the list of retrieved post surrogates to determine which post is likely to be relevant (predictive judgment), and to select a post to read in detail. Stage 3 (reading the post content) is the stage where the user reads the detailed content of a post and

replies from other users, to evaluate the relevance of the information (evaluative judgment).

During the two stages (scanning surrogates and reading post content), users exhibit both skimming and examining eye movement behaviors. Users' eyes *skim* over the text (in the post surrogates or post content) to locate interesting areas of text for close examination. Users *examine* text passages (in the surrogate titles or post content) to make relevance judgment. This study made the reasonable assumption that users' main purpose during skimming and examining is to use information represented in the text to make relevance judgment. If this assumption is correct, then people's skimming and examining behavior can shed light on their relevance judgment during the health information seeking process. A few previous studies have used people's eye movements (in particular, eye fixations) to make inferences regarding relevance criteria people use in information retrieval situations (e.g., Balatsoukas, & Ruthven, 2012; Gwizdka, 2014), but not in the context of browsing social media sites.

Table 1.1. Stages of information seeking behavior in a discussion forum

Stage 1. Searching or Browsing the classified directory
Stage 2. Scanning (the document/post surrogates in the summary result screen) a. Skimming (over the list of surrogates) b. Examining (individual surrogates)
Stage 3. Reading (the document/post content in the detailed result screen) a. Skimming (through document/post content quickly) b. Examining (and absorbing the information in the text or document content)

1.3 RESEARCH CONTRIBUTIONS

This study adopted a different approach to the study of relevance judgment from previous studies:

1. Most previous studies of relevance criteria used either self-reported relevance criteria (in open-ended questions), or a questionnaire to elicit relevance criteria. A questionnaire will constrain the participants to a limited number of broad categories, that require inferencing and generalization from the specific criteria that the participants actually use. In this study, content analysis of the text with eye fixation allowed me to identify more specific relevance criteria, as a supplement to the criteria that the participants report.
2. Previous studies have identified general categories of relevance criteria, such as topicality and usefulness. These abstract categories of criteria have different specific meanings in different domains and contexts. For instance, *usefulness* can refer to an appropriate treatment that the user can adopt to alleviate a health condition. Few researchers have identified domain-specific detailed criteria for deciding relevance or usefulness. In specific domains such as health, more detailed criteria may be more illuminating and useful. Furthermore, identifying and coding abstract relevance criteria involves inferencing and generalization (categorization) by the coders, and some contextual and domain information is lost.
3. Most of the previous studies focused on relevance judgments of traditional documents or files retrieved by search engines. Few studies have been done on relevance judgments of detailed information reflected in text passages, and hardly any in the context of browsing social media sites.

CHAPTER 2

LITERATURE REVIEW

This literature review covers three major areas relevant to this study: health information seeking, relevance judgment in general and in the context of health information seeking, and eye tracking studies of user behavior.

2.1 HEALTH INFORMATION SEEKING

2.1.1 Health information seeking in general

The concept of health information seeking was proposed and discussed from the mid 1980s. Lenz (1984) described health information seeking behavior as a “series of interrelated behaviors related to health decisions that can vary along two dimensions: extension and method”. In this definition, *extension* can be understood as related to a topic and *method* can be understood as the way of dealing with a health problem. This is a general definition of health information seeking without providing more details. Later, Barsevick and Johnson (1990) characterized health information seeking as “actions used to obtain knowledge of a medical-related specific event or situation”. Conley (1998) added “verbal and non-verbal behavior” to the definition of Barsevick and Johnson (1990). Most authors adopted the general notion of information seeking behavior or information-seeking, and narrowed it to the specific domain of health-related information, rather than develop a new definition (e.g.,

Czeja, Manfredi, & Price, 2003; Meischke, Eisenberg, Rowe, & Cagle, 2005; & Rees, & Bath, 2001). Lambart (2007) noted that these authors didn't consider the context of health information seeking as an important factor when defining this concept.

Lambart (2007) suggested that an understanding of the concept of health information seeking behavior should take into account three aspects of context. The first aspect is to cope with a health-threatening situation. In some cases, people seek health information for coping with a medical situation. In other cases, people encounter relevant health information by just happening to come across interesting health information (Palsdottir, 2009). People can have different motivations or needs for seeking health information. The second aspect involves medical decision making. This medical decision making includes relevance judgment of medical information, which is the focus of this study. The third aspect is to change behavior or take preventive behavior. It refers to that people seek health information to make changes in lifestyle or some behaviors. However, behavior change is thought related to factors that are beyond the scope of this study. Hence the study did not include whether the users, in health information seeking, changed their behavior..

A number of studies have been conducted on the topic of health information seeking behavior on the Web, as more and more people are looking for health information on the Internet. A telephone survey of U.S. adults conducted in 2012 by the PEW Internet & American Life Project (Fox & Duggan, 2013) reported that 59% of U.S. adults had sought health information on the Internet in the past year. 35% of U.S. adults had accessed health information online to find out what medical condition they or someone else had. Among

these “online diagnosers”, 53% talked about what they found with their clinicians and 41% had their diagnosis confirmed by their clinician. Moreover, the survey found that among the online health information seekers, 16% had sought other people who shared the same concerns on the Internet, probably via social media sites.

The same trend was also found in European and Asian countries. A multi-country telephone survey conducted by WHO/eHealth Trends and Norwegian Center of Telemedicine between 2005 and 2007 found that among the 7000 people surveyed in 7 European countries—Denmark, Germany, Greece, Latvia, Norway, Poland and Portugal—71% of Internet users had used the Internet for health purposes. Among these, 29% had used online health information to decide whether they needed to see a doctor and 25% had used it in conjunction with the information from the doctor (Kummervold, et al., 2011). In Japan, a web-based survey conducted in late 2010 covering 10,685 participants over 16 years old found that 47% of the respondents had searched online for drug information in a variety of resources, such as search engine results, pharmaceutical company websites and social media sites (Kishimoto, & Fukushima, 2011).

2.1.2 Health information seeking on social media

There are different online sources available for people to search for health information, such as websites run by organizations, homepages owned by individual doctors, online support groups (e.g., discussion forums), and blogs and social media sites such as Facebook (Higgins, Sixsmith, Barry, & Domegan, 2011). The proliferation of various types of social media sites

(e.g. discussion forums, Facebook, blogs) has brought changes in people's health information seeking behaviors.

In a discussion forum, topics and subtopics are organized in a hierarchical structure, and user postings are organized into discussion threads (with a main post and responses/comments from other users). Users browse the posts related to one topic, rather than check one particular user's posts like in Facebook and Twitter. This is different from traditional online information searching where researchers have focused on relevance judgment in the context of searching an online database of records/documents, rather than browsing.

User-contributed health information content on social media sites is different from health information in traditional authoritative sources. User posts on social media sites provide both informational and emotional support to other users (Chuang & Yang, 2012). Information on social media sites is focused on users' experiences: health-related information shared is filtered through users' experiences and described in the patients' contexts. Health information provided is written in layman's terms, different from the content of authoritative health information websites containing medical jargon. An important aspect of social media postings is the emotional support provided by users. Users are found sometimes to join an online community to find or provide emotional support with others suffering from similar conditions (Wright & Bell, 2003). These two important aspects are considered in this study.

2.1.3 Consumer perspective of health information seeking

Consumer health information seeking refers to health information seeking by common people (non-health professionals) to address their health issues (or those of their family or friends), rather than for professional medical tasks or research. The characteristics of consumer health information seeking identified in previous research are as follows:

- Users frequently search for health information on the Internet, usually starting with a search engine.

Brodie, Flourney, Altman, Blendon, Benson, and Rosenbaum (2000) found that most American Whites, Africans, and Latinos relied on the Internet as their channel of health information seeking. Cline and Haynes (2001) found that individuals assigned more credibility to health information obtained via the Internet than any other media sources. Consumers often start searching by using a search engine rather than directly accessing medical portals or consulting librarians or references books. The search terms used were found not to match with medical terminologies (Zhang, 2012).

- They finish the information seeking session quickly.

Esyebach and Kohler (2012) found that the user on average took 6 minutes to obtain online health information. Although many health-related websites were created for the purpose of providing health information, users thought that too much time and effort were required to obtain health information in these websites (Arora, Hesse, Rimer, Viswanath, Clayman & Croyle, 2008). They want to quickly find relevant health

information and may switch to health social media sites instead. In the current study, the health information seeking/browsing sessions were in the range of 10-15 minutes.

- They tend to seek health information for themselves and others.

As mentioned in earlier section, based on the surveys on different continents, users were found to seek health information for their own health issue and for other's health issue (Fox, & Duggan, 2013), as well as browse for health information without a predefined aim. In this study, these three types of health information seeking were included.

- Most of the online health information seekers are below 60 years of age.

Surveys and other studies found that the majority of the online health information seekers are below 60. Fox and Duggan (2013) in their latest survey found that the majority are below 50. Younger people were found more likely to use the Internet to search for health information than old people (Cotten & Gupta, 2004). This might be due to their familiarity with the Internet and computer systems.

These characteristics of online consumer health information seeking were considered in this study:

- The health discussion forum used in the current study has an embedded search engine for use. The participant can also use the simple directory of topics displayed at the main page of the forum.

- Participants were asked to do health information seeking/browsing as long as they wanted, but none of them was found to spent more than 15 minutes.
- Different groups of users with different types of use context were recruited in this study.
- The recruited participants were below 45 years and were familiar with Internet searching, having conducted various kinds of information seeking on the Internet in the past.

2.1.4 Active versus passive health information seeking

Consumer health information seeking can be divided into active seeking and passive seeking or serendipitous seeking (Palstdottir, 2005). Wilson (2000) characterized active information seeking as a kind of information seeking when people experienced a lack of knowledge and purposefully seek information to satisfy the information need.

Another type of health information seeking is passive seeking, which refers to information seeking conducted by users without the expectation of finding particular information. They just come across some information they think useful or interesting (Wilson, 1997).

Williamson (1998) found this kind of accidental acquisition of information in her study of the Australian elderly.

A similar concept is the term *information encountering*, proposed by Erdelez (2004), which refers to discovering information unexpectedly. She pointed out the information found via this type of seeking are related with the general interests of people. She also found that this

kind of information encountering sometimes occurred not for people's own needs but for the needs of other people that they know and want to share information with. Another similar concept is *serendipity* studied by Foster and Ford (2003). They categorized different types of serendipitous encounters: not known to exist previously, in sources not expected to be included, and information of unexpected value. These kinds of encountered information are not thought of at the start of the information seeking session.

In sum, two major types of health information seeking behaviors are found in previous studies: active health information seeking and passive health information seeking. Among active health information seeking, users can seek health information for their health issue or seek for other's health issue. For the passive information seeking, users don't have a particular expectation in advance: they encounter and take note of health information based on their general interest. These users may also unexpectedly encounter information relevant to their health issue or that of family or friend. The current study distinguished between the three types of context: seeking for own health issue, seeking for others' issue and browsing without a predefined/particular issue.

2.2 RELEVANCE JUDGMENT & RELEVANCE CRITERIA

This section reviews the literature on relevance judgment, relevance criteria and studies of relevance judgments relating to social media sites.

2.2.1 Relevance judgment and relevance criteria

Relevance judgment has been an important area of research in information studies. Barry and Schamber (1998) pointed out that “throughout the history of information science, various writers have express a call for research that focuses on understanding of end users as the ultimate assessors of quality of information and of the systems that provide the information. The central part in these discussions has been relevance” (p. 220). They defined relevance judgment as the judgment of information quality between the user’s information problems and the information retrieved. As early as 1975, Saracevic (1975) had mentioned the concept of relevance and defined it as a vague but well understood term without clear explanation. Harter (1992) suggested a psychological definition of relevance, proposing that a given document serves as the role of causing reader’s cognitive changes in relevance judgment. Later, Taylor, Cool, Belkin and Amado (2007) defined relevance judgment as the process of evaluating documents or document representations in terms of relevance to people’s needs using a categorical scale. Other authors also provided their own definitions of relevance and relevance judgment. For instance, Swanson (1977) characterized relevance as new knowledge constructed by the requester with the context. This kind of definition focuses on the knowledge status change.

Regardless of the different definitions of relevance and relevance decisions, there have been two distinctive research approaches. In the 1960s and 1970s, researchers started with a system-oriented view of relevance research. This stemmed from research of information retrieval systems. This approach treated relevance as the association or similarity between the query terms and those with the retrieved documents (Lesk, & Salton, 1968). It focused on the

measures of recall and precision that took topicality as the main criterion of relevance (Talyor et al, 2007; Tang & Soloman, 1997). Most of the time, the searched keywords were used to match the retrieved documents.

However, this approach ignores the fact that relevance judgments are done by people rather than the retrieval system. Belkin and Kwasnik (1986) pointed out the user's role, situation and problems in the relevance judgment process. Schamber, Eisenburg and Nilan (1990) reviewed the different approaches taken in previous studies and found that there are various aspects of relevance, such as topicality, pertinence, and situation.

The other approach is the user-oriented approach, which focuses on the user's perception, estimation, and other factors that might influence their subjective assessment of relevance.

This approach gradually dominated the area of relevance research. This approach states that relevance is not a given, but "constructed" and "derived" by inference (Saracevic, 2007).

Saracevic pointed out that relevance was inferred by the information retrieval system by its algorithm as a kind of construction. However, it was interpreted by users not just focusing on the mere relevance created by the system (e.g., topical and term match). Users can find out more relevant information not just from the retrieved documents but also from other resources. Later, Vakkari and Hakala (2000) found that the relevance derivation was also related with people's expertise in the domain. All the findings support the idea that the user oriented view of relevance based on user derivation and inference rather than system construction should be investigated in more detail.

Since relevance is derived and inferred by people, it is important to find out what are the criteria and rationale people use when they make relevance decisions. Relevance criteria are thought to be the factors that contribute to user's relevance assessment/judgment (Taylor, Cool, Belkin, & Amadio, 2007). They also can be thought of as the clues that people look for to infer relevance (Saracevic, 2007). Several researchers at Syracuse University (Barry, 1994, 1998; Schamber, Eisenburg, & Nilan, 1990; Schamber, 1994) investigated users' subjective relevance judgments in various kinds of settings (e.g., students, working people, and different occupations). Based on their studies, several lists of relevance criteria were derived.

Schamber (1991) conducted an interview of 30 people who used weather information in their jobs about how they judged relevance of information from various kinds of media, such as TV and newspaper. He derived a list of 21 categories of relevance criteria from the interview data that can be grouped into 10 groups. They are: accuracy, currency, specificity, geographical proximity, reliability, accessibility, verifiability, clarity, dynamism and presentational quality.

Barry (1994) interviewed 18 students and faculty members to identify how they judged document information relevance in preparing assignments. He found that seven groups of evaluation criteria were used in the assignment preparation process. They are: information content of document, user's previous experience and background, user's belief and preference, other information and sources in the document, the source of the document, the document as physical entity and user's situation. Several user aspects were identified in this

study, which indicates that the users' own characteristics can also influence their relevance judgment, not just the objective features of the information itself.

Huang and Soergel (2006) carried out a study on how students make relevance assessments. History major graduate students were recruited and made 27000 relevance assessments on interview segments of Holocaust survivors and real user topics. They focused on the four types of relevance they found in their previous research (Huang & Soergel, 2004): direct relevance, indirect relevance, context relevance, and comparison relevance. They pointed out the comparison relevance was the similarity that was used to help to understand the topic. They further divided the comparison relevance into the following categories: external factors (such as time, and place), the participants, and act and experiences (such as attitude, feeling, treatment, experiences of activities). They provided an example of the judger to illustrate the importance of this comparison relevance. The judger made the critical connection between some cases by comparing different characteristics of these cases. This broad category of comparison relevance is applicable in the relevance judgment of health information during the seeking process on a health discussion forum since the user might compare the similarity between their own conditions and the descriptions in the post content. Especially the participants, feelings, treatment and experiences are the important criteria user might use for their relevance decisions.

There were still other studies conducted on relevance judgment in different environments and the criteria found in these studies tended to be mostly similar, such as clarity, accessibility, and accuracy (Maglaughlin & Sonnenwald, 2002). Some criteria are related to the working environment or task and problem situation, such as the emotional status of the user.

Later on, studies were not confined to document information. More types of information sources were used in the studies: audio (Kim, Oard & Soergel, 2003), images (Choi & Rasmussen, 2002), videos (Yang & Marchionini, 2004), and websites (Tombros, Ruthven, & Jose, 2003). They found that the types of criteria derived from document relevance judgment could be applied to these information sources as well.

With the development of Internet services, more people started to use the Internet to get information. The information retrieved from Internet can be of various types, not just limited to textual documents. Because of the difference between online information and document information, studies were conducted to find if the traditional relevance judgment criteria could be applied to online information. Kim and Oh (2007) conducted a study to investigate how online information users judge information relevance in social Q&A sites (Yahoo! Answers). They used a customized online crawler to extract online questions and their answers from Yahoo! Answers for various topics. They identified six groups of criteria by analyzing the content and attributes of the questions and answers: content, cognitive, utility, information sources, extrinsic and socio-emotional, most of which were found in previous studies. But they found that socio-emotional group criteria included the social aspects of the

environment, and the importance of each criterion used in the relevance judgment varied depending on the topics and environments.

A list of the main relevance criteria identified by previous researchers is given in Appendix 1.

These relevance criteria are rather broad high-level categories. They can have different meanings when they applied to different contexts. No specific health information relevance criteria were proposed in previous studies. The current study has derived a list of detailed relevance criteria used in assessing user-contributed health information.

2.2.2 Dynamics of relevance criteria

Spink, Greisdorf and Bateman (1998) conducted a study in order to find what criteria people used for relevance judgments and how they changed according to different judgment decision results (e.g., judged as relevant, partially relevant, and non relevant). In this study, 55 users were recruited to do an online search to resolve their problems. For the retrieved items, the participants were asked to judge their relevance to their problems and state the criteria they used during these judgments. The study found correlations between particular types of criteria and retrieved documents (relevant, partial relevant, and non-relevant). For the relevant items, topicality, specification and utility were frequently used, whereas for non-relevant ones, usefulness, wrong meaning and novelty were often used.

A later study conducted by Greisdorf (2003) further found that the types of relevance criteria and the number of relevance criteria used were not evenly distributed for relevant and non-relevant documents. He found that for the documents that were judged as non relevant mostly by one criterion but for relevant items 3 or more relevant criteria were used together. For non-relevant documents, “not on the topic” was the most frequent criterion used, while for the relevant documents, topicality, pertinence, motivation, usefulness were used more often.

Other scholars also found that relevance criteria were used differently depending on people’s search tasks or work tasks (Bruce, 1994; Tang, & Soloman, 1998; Vakkari, & Hakala, 2000).

These findings suggest that relevance judgment is not a simple one criterion decision making process, but a complex multi-dimensional process. However, these studies only sketched a broad outline of this complexity, rather than dig deeper into the details. In this study, the distribution of the different types of relevance criteria was investigated in relation to the three types of context (i.e. searching for self, for others or with no particular health issue) for relevant posts and non-relevant posts. In addition, the use of different types of relevance criteria at different stages in the information browsing session was analyzed.

2.3 EYE TRACKING STUDIES

2.3.1 Eye tracking studies on user behavior

Eye movements are thought to be directly linked with people's thinking and cognitive process (Rayner, 1998), such as reading and decision making. Scientists have studied how people read to infer their cognitive processes, and found that people's eyes exhibit different behaviors on the reading document. Eye fixation is one basic kind of eye movements analyzed by researchers in previous studies. Fixation refers to static or relatively static focus of the eye on a particular small area of the object. He found that people's eye movements were affected by cognitive processes occurring at the time when people fixated on something. He found that the fixation duration and number of fixations were related with people's sentence processing. Particular types of words, such as verbs received more attention than other types of words such as the sentence subject (Rayner, 1995). It indicated that people behaved differently when they process different texts, which were revealed by eye fixation

and saccade patterns. Hayhoe and Ballard (2005) pointed out that although no studies currently can reveal the internal cognitive process clearly and directly, a study of eye movements can serve as references to the internal process, and eye movement patterns can be interpreted by and related with particular mental tasks. They found that eye fixations were focused not on the points that were visually salient but the points that were spatial-temporally best for the related task/need. The authors demonstrated that users focused on task-related visual information rather than other information by showing an eye tracking snap picture where a user who wanted to make a sandwich focused on the table and the food components used to make the sandwich rather than other salient things such as refrigerator and doors. The duration of fixations were computed as evidence of mental task needs. It is reasonable to assume that during information seeking tasks, people might exhibit different eye movements and they can be used to infer their mental process such as relevance judgment on particular pieces of information.

In information research, there are a small number of studies to investigate users' eye movements on information objects to understand people's mental processes, such as relevance judgment and quality judgment. The duration of fixation was often used as an indicator of user's mental information processing in relevance judgment. Cole, Gwizdka, Liu, Belkin and Zhang (2013) studied the relation between people's eye movements and their knowledge status and found that cognitive measures reflected/computed by eye movements and their patterns were associated with users' knowledge level. They recruited college students and used a questionnaire to determine their domain knowledge level in a medical

area and information search area. Then the participants' eye movements (duration of fixation, number of fixation, and reading speed) in searching related topics were recorded to compute cognitive measures, such as lexical access duration excess (the eye fixation duration required to have for cognitive processing before the lexical access) and perceptual span (the spacing of fixations in horizontal dimension during reading). They found these measures had high correlation with their domain knowledge level. In particular, perceptual span, median lexical access duration excess, and reading speed were found significantly correlated with user's domain knowledge level. Holsanova, Rahm and Holmqvist (2006) investigated eye movement patterns to understand how people made relevance judgments of newspaper threads. Other scholars also conducted studies on relevance judgment and quality judgment by tracking people's eye fixations and compute the fixation duration for online searching results (Google surrogates) (Balatsoukas & Ruthven, 2012). The detailed relevance criteria of Balatsoukas and Ruthven's study can be found in Chapter 3 (Table 3.6).

2.3.2 Measures used in eye tracking studies

Since eye tracker machines provide only raw measurements (e.g., fixation/saccade duration and fixation coordinates), researchers have developed different measures based on these basic types of data to investigate relevance judgment and other cognitive behaviors.

Buscher, Dengel and Elst (2008) carried out a study using eye tracker data to predict user's judgment of relevance on search engine results. They used a Tobii desk mounted eye tracker machine (Tobii X 60 series, the same model used in the current study) to capture user's eye

movements and used the ratio of the accumulation of reading to the accumulation of skimming, to predict user's relevance judgments. In this study, the original fixation data points on a particular document were merged into a bigger fixation region using a self-defined algorithm. This algorithm detected nearby eye fixations according to a distance threshold (30 pixels in their study) and merged them into a bigger fixation region/span. Then the transition/movement between two nearby fixation region in a particular line was calculated as the following categories of reading/browsing behaviors:

- Reading forward (distance between two nearby fixations: 0 to 11 letter spaces);
- Skimming forward (distance between two nearby fixations: 11 to 21 letter space);
- Long skimming jump (distance between two nearby fixations: 21 to 30 letter space);
- Short regression (distance: -6 to 0 letter spaces);
- Long regression (distance: -16 to -6 letter spaces);
- Reset jump (distance of less than -16 letter spaces) and;
- Unrelated movement.

Each type of eye movement behavior mentioned above was assigned a particular value and the cumulative value of each type within a particular document was analyzed and used to predict relevance. They used the ratio of the cumulative reading score and cumulative skimming score as the measure of relevance and associated with the user's feedbacks on the document relevance decision. They found that the ratio was positively correlated with the

relevance of the document. In particular, when the ratio increases, the document was more likely to be judged as relevant. The authors applied this fixation analysis method into a gaze filter method to improve the search results from a query (Buscher, Dengel & Elst, 2008b). They found that based on the gaze analysis, the method can contribute to 32% improvement of the search results for users with different tasks.

The ratio of reading to skimming reveals some association between people's eye movements and the relevance decisions they made. However, many measures were based on researchers' intuition, for example the threshold distances used to categorize different browsing behaviors. There are no theoretical justifications for these formulas. Moreover, researchers have not uncovered the underlying reasons of why users behave differently in their reading and skimming behavior. In this study, I adopted the ratio of reading to skimming but modified the formula based on my framework, as described in Chapter 3. In addition, I combined different quantitative and qualitative methods to identify both the association between eye movements and relevance judgments, and possible rationale behind users' different behaviors of relevance judgments.

Cole, Gwizdka, Liu, Belkin and Zhang (2013) conducted another study by using eye tracking data to predict user's knowledge acquisition process, one particular aspect of human cognitive process. Users were asked to finish a self-rating list of biomedical terms to reflect their knowledge in this domain. Then they were asked to perform a search and read the documents from the search results. They filtered out long reading sequences from all the text sentences with more than four fixations and developed four kinds of cognitive measures:

Lexical Access Duration Excess (the duration above 113 ms to indicate acquisition of word meaning), regression (fixations back to previous visited text), perceptual span (the distance that reflects the spacing of fixations and describes the length of texts that users take as a unit), and reading speeding. By simple linear regression model construction, they found three measures that were significantly correlated with users' domain knowledge level: perceptual span, median Lexical Access Duration Excess, and reading speed.

In their study, lexical access duration was similar to the common fixation duration, except it has a minimum value that indicates that longer than this value, the comprehension of word meaning occurs. This threshold setting was taken from a previous study by Pollatsek, Reichle and Rayne (2006). There is still no consensus on this minimum value and its indication of people's reading. Another measure—perceptual span—was calculated by a complicated formula based on the distance between fixation points. The authors pointed out that this measure was similar to that used by Buscher et al. (2008) which I adopted in this study.

Reading speed was also used in their study to infer the domain knowledge level. However, the reading speeding was affected by various factors, such as human characteristics, comprehension, and previous knowledge. Hence, this measure was not adopted in my study.

Saito, Terai, Egusa, Takaru, Miwa and Kando (2009) conducted a study on information seeking among university students. 16 students performed two seeking tasks: report writing task and trip planning task. Their eye movements were recorded by the eye tracker machine and analyzed. The researchers used several measures to compare between students with different tasks: search actions (e.g., search, link, back, jump, and bookmark), lookzones, the

number of fixation points in each search result, number of pages visited, and the duration of people staying at each particular page. They divided the entire web page into several lookzones and counted and compared the number of fixations within each zone. They found that the trip planning students had more fixations on the lookzone of toolbar while the report writing students had more fixations on the search bar. They also found that for the report writing task, undergraduates had longer browsing times on non-result pages than the result pages. In my study, I calculated the duration of staying at different types of pages within a health discussion forum: the duration of the scanning stage (staying at the post surrogate summary page) and the duration of the reading stage (staying at detailed post page). In addition, different types of eye movement behaviors (examining and skimming) within each stage were calculated and compared. Instead of counting the number of fixations in each particular lookzone, I counted the number of surrogates and sentences containing different types of health information that were fixated by participants, to find out whether these participants with different types of context focused on different types of health information and use them in their relevance judgments.

Other studies have been conducted focusing on content analysis of information focused on by people (e.g., Miwa, Egusa, Saito, Takaku, Terai & Kando, 2011; Balasoukas & Ruthven, 2012). Miwa, Egusa, Saito, Takaku, Terai and Kando (2011) combined the eye movement data with the post experiment interview data recording of users' feeling, thoughts and intentions about particular actions on the screen. They used the combined data to investigate information encountering during the information seeking process. They found that replay of

the eye movement data could help participants to recall the information encountering cases and lead to reconsideration of the definition of information encountering. They also found that eye fixation represents cognitive processing. In the current study, I assumed when people's eyes focus on a particular text passage, they are interpreting the information conveyed by the passage and making use of the information in the relevance decision making process.

Balasoukas and Ruthven (2012) investigated how people make relevance judgments on search results from the Google search engine. They developed a stepwise method to connect the eye movement data to users' actual relevance criteria used during the judgment process. The research participants were asked to think aloud during their browsing process. Then the audio recordings were transcribed into qualitative data. The content with the participants' eye fixations were also coded by duration and number of fixations and associated with the transcribed audio data. They found that eye movements, particularly the attributes of eye fixations (number of fixations, duration and frequency) reflected users' relevance dynamics. For example, the attention on topicality decreased along with the browsing process. People were found to put less effort on information related with topicality when they examine the search surrogate results since surrogates were assumed to be on the topic. When deciding non-relevant results, people tended to have long fixation duration when considering topicality and scope.

This approach of linking qualitative data of user's feedback with the quantitative data of eye movements can yield more insights into the relevance decision process, than relying solely on

quantitative eye-tracking measures. It focuses on the content of fixation instead of solely numeric values. Saito, Takusa, Egusa, Terai, Miwa and Kando (2010) combined the quantitative data of user eye movements with the qualitative data of their post experiment interview to analyze people's web searching process. However, Rayne (1998) pointed out that when people think aloud, they tended to think differently than in natural situations. People were found to have different reasoning processes than the normal reasoning process when they were required to think aloud. Hence, the thinking aloud process was replaced in this study with a post-experiment interview focusing on the commentary about what they were thinking during the searching process. This post-experiment interview was also used by Miwa et al. (2011) with the video replaying of user's eye movements on the screen. They found it useful for the participants to recall their behavior and comment on what they were thinking.

Another issue with previous studies is that they didn't pay enough attention to the use of skimming data. It is true that something related to skimming was considered in the study of Buscher et al. (2008). They defined user's skimming in terms of long distance jump between horizontally nearby fixations. But they didn't dig for the reason why people skim over some texts but fixate on other texts. In this study, the possible reason of skimming was provided and verified by user's comments from post experiment interview.

2.4 CONCLUSION

There have been very studies of health information seeking on social media sites, particularly of browsing behavior and relevance judgments on health discussion forums.

Previous studies have derived many categories of relevance criteria, but they are still rather broad high-level categories and they have not been test in the context of health information browsing on a health discussion forum. It is still unclear what detailed relevance criteria are used in assessing health information.

My study combined the quantitative analysis of eye movement data and content analysis of text with eye fixations to provide a more in-depth picture of relevance judgment and the associations between detailed relevance criteria for health information and people's eye movements.

CHAPTER 3

RESEARCH FRAMEWORK AND METHOD

3.1 RESEARCH FRAMEWORK

This section discusses the framework used in this study to distinguish between the different stages of health information seeking when interacting with a discussion forum, and the two types of relevance judgments associated with each stage. Table 3.1 outlines the framework.

Table 3.1. Stages of health information seeking within a health discussion forum, and the two types of relevance judgment associated with each stage

USER INFORMATION BEHAVIOUR	RELEVANCE JUDGMENT
Stage 1. Searching or Browsing	System retrieves and displays document/post surrogates matching the search query or selected directory category (relevance judgment by machine)
Stage 2. Scanning (the document/post surrogates in the summary result screen)	
a. Skimming (over the list of surrogates)	Select surrogates for attention/focus (unconscious relevance judgment based on keywords in the text that catches the eye)
b. Examining (individual surrogates)	Select surrogates to retrieve the associated (linked) document (predictive relevance judgment based on an estimation of the likelihood that the document contains relevant information)
Stage 3. Reading (the document/post content)	
a. Skimming (through the document/post content quickly)	Select text for attention/focus (unconscious relevance judgment)
b. Examining (and absorbing the information in the document/post content)	Deciding whether the information is relevant or likely to be useful in a use context (evaluative relevance judgment)

This framework of user searching and browsing in an information system was derived from two frameworks: the two-stage relevance judgment model (Tang & Solomon, 2001; Rieh, 2002; and Crystal & Greenburg, 2006), and the two-stage browsing model (Buscher, Dengal, & van Elst, 2008; and Zhang, 2012).

Tang and Solomon (2001) provided a two-stage relevance judgment model for online information searching, based on previous work of Saracevic (1996, 1997), Spink, Griesdorf and Batman (1998), Ingwersen (1996), and Schamber, Eisenberg and Nilan (1990). In this model, the users' overall relevance judgment of a document retrieved by an information system is divided into two stages, predictive relevance judgment and evaluative relevance judgment, as detailed below:

- *Predictive relevance judgment* of the document surrogate: the user scans the document surrogates retrieved by the information system to make judgments of the likelihood of relevance of the content of the documents. Based on the estimated likelihood or prediction of relevance, the user may retrieve the fulltext of selected documents to read.
- *Evaluative relevance judgment* of the document content: the user reads some or all the text in the document, and makes a relevance judgment based on the information in the text.

This two-stage relevance judgment model implies two major stages in user browsing of search results in an information system: scanning document surrogates and reading document content. These two stages reflect the typical design of information retrieval systems that

requires the user to first assess the list of document surrogates in the summary search result screen for potential relevance, and then select (click on) document surrogates that are potentially relevant to read the detailed document content. As its name suggests, predictive relevance judgment is based on an estimation of the likelihood of relevance. This kind of estimation is based on limited information provided by the surrogates, which do not provide enough information for a full relevance judgment. On the other hand, evaluative relevance judgment is based on full information in the document. The different nature of these two types of relevance judgment may involve different relevance criteria or different types of reasoning. Crystal and Greenberg (2006) pointed out that there are two different relevance judgments during different information seeking stages. In this study, I analyzed the relevance criteria used in these two stages of relevance judgment.

This two-stage relevance judgment model is suitable for the study of relevance judgments in health discussion forums because the structure of the forums typically have two kinds of pages: the surrogate summary page which lists post surrogates (with post title, number of user replies, number of views, and time of posting), and the detailed post page with full content of a particular post and replies from other users.

The other part of my research framework recognizes two types of eye-movements—skimming and examining—during information browsing. Buscher, Dengal and van Elst (2008) developed a skimming and reading model to investigate user's relevance judgments during the information seeking process. They categorized people's reading of retrieved documents into several types of skimming and reading behaviors, according to

whether the user's eye gaze moves quickly or is statically focused. They calculated the ratio of cumulative reading to cumulative skimming as a measure to predict the user's relevance decision on a particular document. They found that the ratio of examining to skimming is positively associated with the likelihood of judging information as relevant. My study makes use of the term *examining* instead of *reading* used in their study to better reflect that the user is closely focused on particular pieces of text information in contrast to *skimming*. I use the term *reading* for the entire stage of checking the detailed post content.

Skimming and examining are not found only in the reading stage (i.e. reading the document content), but also found in the stage of scanning search surrogates. Zhang (2012) conducted a study of health information searching by university students, and found that users exhibit skimming and examining not only during the stage of reading the retrieved fulltext documents but also during the stage of checking document surrogates to estimate likely relevance.

From these two studies, it is found that when the two stages of relevance judgment were made when scanning the surrogate or reading the full text document, the user will skim over some pieces of text information and examine other text information closely. Hence, I incorporated the two types of eye-movement behavior, skimming and examining, into the two-stage relevance judgment model as the research framework of my study, as outlined in Table 3.1.

3.2 CHOICE OF HEALTH DISCUSSION FORUM

HealthBoards.com (<http://www.healthboards.com/boards/index.php>) was the health

discussion forum selected for this study. A search on Google search engine with keywords “health forum”, “health discussion forum”, and “health information board” retrieved 10 candidate health forums, shown in Table 3.2.

Table 3.2. 10 health discussion forums retrieved using Google search engine

1.	Discussion forums on the Patient website (http://www.patient.co.uk/forums)
2.	Health Forum (http://www.healthforum.com/)
3.	HealthBoards.com (http://www.healthboards.com/boards/index.php)
4.	eHealth forum (http://ehealthforum.com/health/health_forums.html)
5.	netdoctor (http://www.netdoctor.co.uk/interactive/discussion/index.php)
6.	Consumers of Health Forum of Australia (https://www.chf.org.au/)
7.	Mental Health Forum (http://www.mentalhealthforum.net/forum/)
8.	PatientsLikeMe (http://www.patientslikeme.com/)
9.	Health Informatics Forum (http://www.healthinformaticsforum.com/forum/topics)
10.	Discuss Health Forums (http://www.discusshealth.org/)

I chose HealthBoards.com based on the user number, frequency of use, information update frequency, comprehensiveness and ranking on rating sites:

- Number of registered users: 1,079,219 registered users as at 2014.09.20. The second largest was PatientsLikeMe with 300,000 registered users;
- Number of posts: 879065 threads, and 4874692 posts and replies;
- Number of subsections on particular health conditions and problems: over 280 subsections;
- Number of daily online users: 3000+;
- Ranking: No. 1 health forum in Yahoo Health search

(https://search.yahoo.com/search?p=best+health+forum&fr=uh3_maghealth_web_gs).

3.3 PARTICIPANT SAMPLING

Fifty study participants were recruited from students and staff of Nanyang Technological University (a large government-funded university in Singapore). An additional 10 participants were recruited from my network of friends, who were working adults outside the university. The participants were invited by advertisements posted on the university campus notice boards, email invitation, telephone invitation and oral invitation. There were altogether 60 participants, comprising 30 full-time students, 14 part-time students, 4 staff members, and 10 working adults from outside the university. The following eligibility criteria were used in recruiting participants:

- Age: 18-50;
- Education level: Bachelor's degree or above;
- Health condition: no critical disease, including HIV, cancer, etc.;
- English fluency: competent in English reading and speaking.

The health condition criterion was used to ensure that the participants did not have a debilitating or critical disease, as such patients are expected to have rather different needs and information behavior. Two participants failed in the calibration of the eye tracker machine and were excluded from the study, leaving 58 participants who contributed data to the study.

The detailed demographic information is given in Table 3.3. Most of the participants were Singaporean Chinese (Singapore citizens) and Chinese nationals. The Singaporean

participants were all educated in English-medium schools. Chinese nationals were educated in Mandarin in their elementary school in China, but took English courses from secondary school to college. Before they came to Singapore, they had to pass one or more English tests such as IETLS and TOFEL.

Table 3.3. Summary of demographic information of research participants (N=58)

Nationality	China Nationality: 31 Singapore Nationality: 24 Other Nationalities: 4
Education	Students: 40 Undergraduate: 12 Postgraduate: 16 PhD student: 12 Working (not full & part time student): 18 Bachelor's Degree: 5 Master's Degree: 9 Doctor's Degree: 4
Occupation	Full time student: 30 Part-time student: 14 NTU Staff: 4 Working adults (not NTU staff): 10
Age	18-20: 2 >21 to 25: 13 >25 to 30: 14 >30 to 35: 13 >35 to 40: 12 >40 to 45: 2 >45 to 50: 2

Table 3.4. Summary of demographics for participants with different types of use contexts

		Seeking for self (n=22)	Seeking for others (n=22)	Browsing with no (n=18)
Nationality	China Nationality	12	12	10
	Singapore Nationality	9	8	7
	Other Nationalities	1	2	1
Education				
	Students			
	Undergraduate	3	3	2
	Postgraduate	7	8	7
	PhD student	4	4	2
	Working (not full & part time student			
	Bachelor	2	2	3
	Master	4	3	4
	PhD	3	2	0
Occupation	Full time stude	11	12	11
	Part time stude	6	5	3
	NTU Staff	1	2	1
	Working adults (not NTU staff)	4	3	3
Age	18-20	1	1	0
	>21 to 25	5	4	4
	>25 to 30	5	5	4
	>31 to 35	4	5	4
	>35 to 40	5	4	3
	>40 to 45	1	2	0
	>45 to 50	1	0	0

3.4 EXPERIMENT PROCEDURE

The steps in the experiment procedure were adapted from previous studies using eye tracker machine (Miwa, Egusa, Saito, Takaku, Terai & Kando, 2011; Balasoukas & Ruthven, 2012; Miwa, Egusa, Saito, Takaku, Terai & Kando, 2011) as follows:

- Brief the participant on the general purpose of the study. Obtain the participant's agreement to participate in the study, and his or her signature on the consent form.
- Ask the participant whether he or she has some health issue to seek information for in the discussion forum. If no, ask whether the participant wants to seek information for the health issue of a relative or friend. If the participant cannot think of a health issue for self or others, then ask the participant to browse the discussion forum for health information. Ask the participant to describe the health issue in some detail, if the participant is seeking for own health issue or other's health issue.
- Give the participant an introduction to the eye tracker machine—what it does, how it works, and what the participant should do to obtain accurate results.
- Calibrate the eye tracker machine for the participant, following instructions in the eye tracker manual. The calibration is used to adjust the eye tracker system to the participant's gaze positions. The system indicates whether the calibration is successful or not.

- Ask the participant to look for information in the discussion forum that is relevant (or browse for interesting health-related information).
- Replay a video recording of the participant's search and browse session together with the indications of eye gaze positions and eye movements. Ask the participant to comment on the post surrogates with eye fixation as well as those selected, and the text passages in the post content with eye fixation. The purpose is to identify the participant's reasons why a post surrogate or detailed post is relevant or not. More details of the questions are given in Table 3.5.
- Pay the participant SGD 15 as incentive.

Table 3.5. Interview guide to obtain participant's comments on video recording of eye movement behavior

Summary result screen of post surrogates

1. For post surrogates with eye fixation, ask: *What attracted your eyes to these posts? Why is this interesting?*
2. For post surrogates clicked on, ask: *Why did you select this post for further reading?*
3. Look through the post surrogates without eye fixations, and point to a few titles that appear similar in topic to the post surrogates the participant had clicked on, and ask: *Why was this post not interesting?*

Detailed post screen

4. Point out text segments in the post content with eye fixation and ask: *Is this information useful or relevant?*
 5. Finally, ask: *Why do you think this post is relevant or not relevant?*
-

3.5 DATA ANALYSIS

In this study, qualitative and quantitative methods were used to address the two research objectives. For research objective 1—to find out the relevance criteria people use in 3 types of use context (i.e. searching for self, for others or with no particular issue in mind)—content analysis was performed on the discussion forum text with different eye movements as well as on the user commentary:

- *Analysis of post surrogates and post content with and without eye fixation.* The data were downloaded from the eye tracker machine and coded into different types of health information. The percentage of each type of health information found in the post surrogates and the detailed posts were computed. The purpose was to find out whether different types of health information were used in the different use contexts, and the importance of particular types of health information (i.e. relevance criteria) in making relevance judgments. A screenshot of a detailed post page showing text with eye fixations is given in Figure 3.1. The details of the coding scheme are discussed later in the chapter. The coding was performed by two coders. To check inter-coder reliability, both coders were asked to code about 20% of the data (the data for the first 12 participants). Cohen's kappa between the two coders was found to be 0.82, indicating good consistency. The conflicting codings were discussed with the coders to help them get a clearer conception of the categories. For the purpose of computation, I resolved the conflicting categories by selecting the category after reviewing with the two coders. The remaining 80% of the

coding data were divided evenly between the two coders.

- *Analysis of participant comments from the post-experiment interview.* The participants' comments on whether and why particular post surrogates or post content were relevant or not were analyzed. The purpose was to elicit the criteria and reasons for their relevance judgments, and find out what they thought about the health information on the screen during browsing. The participant comments were transcribed and coded based on keywords related to the different types of relevance criteria. I and another PhD student in the school did the coding, and Cohen's Kappa was 0.85 between the two coders. The detailed coding scheme is given in Chapter 4.

Figure 3.1 Screenshot of detailed post page with participant's eye fixations

For research objective 2—to find out patterns of eye movements associated with different

types of relevance criteria—different eye movement statistics collected by the eye tracker machine were analyzed to derive aggregate measures of user eye movement patterns.

Quantitative data from downloaded the eye tracker machine include the number of eye fixations on the screen, the duration of fixations, and the duration of skimming. In addition to these raw data collected by the machine, this study derived aggregate measures (e.g., total examining time for each session) to analyze the participants' behavior. In total, four types of derived aggregate measures were used: scanning duration, reading duration, examining duration, and skimming duration:

- *Scanning (post surrogates) duration* is the total duration of a scanning session, defined as the time the user spends on each post surrogate summary page. A scanning session starts when the system displays a post surrogate summary page, and ends when the user selects (clicks on) a post surrogate to display the content of the post thread. If the user moves to subsequent pages of the post surrogate summary, it is considered another scanning session. Thus, scanning sessions are separated by other scanning sessions and reading (the detailed post) sessions, described next. The whole user session typically has a few scanning (post surrogate) sessions and reading (detailed post content) sessions.
- *Reading (post content) duration* is the total duration of a reading session, defined as the time user spends on each detailed post page. A reading session starts when the system displays a detailed post page and ends when the user closed it or returned to the previous page. Thus, reading (post content) sessions are separated by scanning

(post surrogate) sessions.

- *Examining duration* is the total duration that people spend in focusing on the text, within either the scanning post surrogates or reading post content session. It is the sum of the individual fixation durations within a scanning session or a reading session.
- *Skimming duration* is the total duration of skimming (quick eye movements) within a scanning post surrogates session or reading post content session. It is the sum of the skimming periods between fixation points within a scanning or reading session.

The method for classifying *skimming* and *examining* behavior was adapted from Buscher, Dengal and van Elst (2008). Their operational definition of skimming and examining are shown in Table 3.5. This classification defines the nearby fixations as examining and distant fixations as skimming. They did not provide justifications for their classification method. The definitions of several kinds of eye movements are ambiguous. *Regression* in Table 3.6 is defined based on 0-16 letter spaces of moving the eyes back, whereas 0–11 letter spaces of moving the eyes forward were defined as *examining* and 11-30 letter spaces were defined as *skimming*. Hence, *regression* includes *examining* and part of *skimming* which makes it difficult to distinguish between them. In addition, skimming is the quick moving behavior between two eye fixations, including the duration of the eye fixations. In the current study, it is cleaner to separate skimming from examining (eye fixation), so their durations do not overlap.

Table 3.6. Definition of examining, skimming and regression (from Figure 2 in Buscher et al., 2008)

Type of behavior	Definition (in terms of distance between two nearby fixations in the same text line)
Examining	0 to 11 letter spaces
Skimming	11 to 30 letter spaces
Regression (moving back)	-16 to 0 letter spaces

A simplified method of classifying *skimming* and *examining* was used in this study: fixations (within one clause) above 100 ms each were considered as *examining*. The movement between these two fixations was considered *skimming* which included the eye saccades. This modified method is better in describing two different eye movements without confusion. Another advantage of this approach is that it is easier to process data exported from the eye tracker machine, since the fixation points are directly recorded by the system and *skimming* can be easily determined by finding the transition between two distant fixations on the same text line.

The detailed quantitative data processing steps are as follows:

- The quantitative data of eye fixations are exported from the system into Microsoft Excel tables (individual webpage) shown in Figure 3.2. The exported data include: fixation serial number (shown in Figure 3.2 as FixationIndex), type of eye movement (GazeEventType), duration of this fixation (GazePointDuration), and Fixation

coordinates (FixationPointX, FixationPointY) and the local machine time when the eye movement occurred (LocalTimeStamp), and other types of data not used in this study.

- Delete other types of eye movement data, leaving only the data of eye fixations (by reordering in the column of FixationIndex). The Figure 3.2 shows the results after reordering.
- Based on the coordinates of the fixations (FixationPointX, FixationPointY), delete fixations on the address bar, the ads of the webpage, leaving only the fixation data on the text.
- Sum the remaining fixation durations into the examining duration within each webpage.
- Use the LocalTimeStamp to calculate the skimming duration within each webpage. The difference between the durations of two consecutive fixations is the duration of the skimming between these two fixations. Then the individual skimming durations are summed up derive the skimming duration within a webpage.
- Calculate the scanning duration for each summary result page of post surrogates, and reading duration for each detailed post page by finding the recording duration for this particular page, represented in the Excel sheet as RecordingDuration.

real_experiment_1_health discussion forum_http_www.healthboards.com_boards_general-health_737317-pain-under-left-breast-upper-rib-area.html (CRC).xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Participan	Recording	MediaNar	SceneNan	SceneSegi	SceneSegi	SceneSegi	Recording	LocalTime	EyeTrackerTimestamp	FixationIn	SaccadeIn	GazeEven	GazeEven	FixationPt	FixationPt	GazePoint	GazePoint	GazePoint	GazePoint	GazePoint	StrictAver	StrictA
2	P01	813187	http://www					135819	18:51:08.4	1041573291611270		1	Saccade	100			1	306	125	296	70	101.07	229
3	P01	813187	http://www					135836	18:51:08.4	1041573291628020	1		Fixation	183	306	68	2	308	122	298	67	101.64	230
4	P01	813187	http://www					135852	18:51:08.4	1041573291644650	1		Fixation	183	306	68	3	321	126	311	71	106.04	228
5	P01	813187	http://www					135869	18:51:08.4	1041573291661270	1		Fixation	183	306	68	4	316	124	306	69	104.25	229
6	P01	813187	http://www					135886	18:51:08.4	1041573291678020	1		Fixation	183	306	68	5	321	128	311	73	106.07	228
7	P01	813187	http://www					135902	18:51:08.5	1041573291694640	1		Fixation	183	306	68	6	315	121	305	66	103.98	230
8	P01	813187	http://www					135919	18:51:08.5	1041573291711260	1		Fixation	183	306	68	7	319	119	309	64	105.22	231
9	P01	813187	http://www					135936	18:51:08.5	1041573291727890	1		Fixation	183	306	68	8	308	123	298	68	101.79	230
10	P01	813187	http://www					135952	18:51:08.5	1041573291744630	1		Fixation	183	306	68	9	311	123	301	68	102.72	230
11	P01	813187	http://www					135969	18:51:08.5	1041573291761260	1		Fixation	183	306	68	10	315	122	305	67	104.06	230
12	P01	813187	http://www					135986	18:51:08.5	1041573291777880	1		Fixation	183	306	68	11	325	119	315	64	107.32	231
13	P01	813187	http://www					136002	18:51:08.6	1041573291794500	1		Fixation	183	306	68	12	315	120	305	65	104.08	230
14	P01	813187	http://www					136019	18:51:08.6	1041573291811130		2	Saccade	33			13	279	52			92.25	255
15	P01	813187	http://www					136036	18:51:08.6	1041573291827890		2	Saccade	33			14	258	29			85.33	263
16	P01	813187	http://www					136052	18:51:08.6	1041573291844500			Unclassifi	33			15	260	38			85.85	260
17	P01	813187	http://www					136069	18:51:08.6	1041573291861120			Unclassifi	33			16	260	48			86.01	256
18	P01	813187	http://www					136085	18:51:08.6	1041573291877740		3	Saccade	17			17	260	30			85.97	263
19	P01	813187	http://www					136102	18:51:08.7	1041573291894480	2		Fixation	133			18	265	29			87.52	263
20	P01	813187	http://www					136119	18:51:08.7	1041573291911110	2		Fixation	133			19	262	28			86.63	264
21	P01	813187	http://www					136135	18:51:08.7	1041573291927730	2		Fixation	133			20	257	27			84.96	264
22	P01	813187	http://www					136152	18:51:08.7	1041573291944350	2		Fixation	133			21	263	33			86.85	262
23	P01	813187	http://www					136169	18:51:08.7	1041573291961100	2		Fixation	133			22	253	31			83.63	262
24	P01	813187	http://www					136185	18:51:08.7	1041573291977720	2		Fixation	133			23	268	36			88.71	261
25	P01	813187	http://www					136202	18:51:08.8	1041573291994350	2		Fixation	133			24	266	32			87.83	262
26	P01	813187	http://www					136219	18:51:08.8	1041573292010970	2		Fixation	133			25	267	38			88.24	260
27	P01	813187	http://www					136235	18:51:08.8	1041573292027590			Unclassifi	183			26						
28	P01	813187	http://www					136252	18:51:08.8	1041573292044340			Unclassifi	183			27						
29	P01	813187	http://www					136269	18:51:08.8	1041573292060960			Unclassifi	183			28						
30	P01	813187	http://www					136285	18:51:08.8	1041573292077580			Unclassifi	183			29						
31	P01	813187	http://www					136302	18:51:08.9	1041573292094200			Unclassifi	183			30						

Figure 3.2. Data sheet of user eye fixations

3.6. RESULTS OF PRELIMINARY STUDY

Before the formal study was conducted, a preliminary study was carried out to develop the coding scheme and to obtain preliminary results to be confirmed in the formal study. Fourteen graduate students were recruited for the preliminary study and they followed the same procedure as in the formal study. The preliminary results have been reported in Pian, Khoo and Chang (2014).

For research objective 1 (to find out the relevance criteria used by people in different use context), the preliminary results indicated that participants in the three use contexts focused on different types of information in the post surrogates and post content.

For participants seeking information for their own health issue, when scanning post surrogates, they focused mainly on the poster's symptoms and history of disease, terminology and description of disease. To a smaller extent they also focused on the poster's age, occupation, feelings about the symptom, and type and description of treatment. When reading the detailed post content, they focused on the same kinds of information with more emphasis on the details of the treatments. Data from post-experiment interviews indicated that this group of participants generally engaged in case-based reasoning to match post surrogates and post content with their own situation. This is supported by the results of our analysis of eye-fixation data, which indicated that the participants paid more attention to the posters' demographic information, health condition and subjective experience.

Participants seeking information for other people's issue focused on the terminology and description of the disease, and type of treatment when scanning post surrogates. When

reading post content, more focus was placed on the etiology and cause of disease, and description of the treatment. This is consistent with data from the post-experiment interviews, which indicated that this group of participants was looking for factual information/knowledge of the disease and treatments.

Participants browsing with no particular issue focused on topics of general interest such as smoking, pollution, sexual issues and weight loss, and rare or unusual issues.

For research objective 2 (to find out associations between eye movement patterns and relevance criteria used), the preliminary results again indicated different scanning, reading, skimming and examining durations for the three use contexts.

Participants searching for own health issue spent substantially more time scanning post surrogates and reading the selected post content than participants searching for others or with no particular issue in mind. Users searching for own issue also had the longest duration of examining (eye-fixation), followed by participants searching for others and participants with no issue. For posts that were judged to be non-relevant, users searching for own issue spent much less time reading them, but more of the reading time was devoted to skimming than examining. Users searching for others were particularly quick to decide non-relevance of post content. Users with no particular issue in mind spent more time skimming post surrogates and post content, than examining them in detail. Their browsing pattern was the same for relevant and non-relevant posts.

3.7 DERIVATION OF THE CODING SCHEME

The coding scheme used to code different types of health information was divided into three broad categories:

- Types of *case-based health information*,
- Types of *factual knowledge-based health information*,
- And types of *general knowledge-based health information*.

This scheme was derived partly from the literature, and partly from open coding I carried out in the preliminary study.

In the preliminary study involving 14 participants, the sentences with eye fixations were extracted and used as the basis for open coding. The sentences from seven participants were used as a training set to derive the coding scheme, and sentences from the remaining seven participants were used as a validation set to validate the coding scheme.

The three broad categories of health information as well as some detailed categories were adapted from the literature. These categories were used as a starting point for open coding. Each sentence within the training set was assigned to one detailed category under one of the three broad categories. However, when a sentence did not fit any of the existing categories comfortably, a new category was created under one of the broad categories.

The list of categories derived and the coding for the training set were reviewed by my thesis supervisor. The revised coding scheme was applied to the validation set of sentences, and it was found that all the sentences could be coded into the existing categories. The details of these coding categories are discussed below.

Case-based health information

Case-based health information refers to health information related to the patients' conditions, their own experiences, and their emotions. This broad category of health information was adopted from the literature. Fox and Duggan (2013) found that American Internet users searched for the experiences other users with similar health conditions. This broad category was divided into three subcategories: *patient demographics*, *patient symptom*, *emotional support* and *patient experiences of drug and treatment*.

Patient demographics include *age*, *job*, *nationality*, and *race*. These categories come from the literature. Fox and Duggan (2013) used age, race, educational level, and marital status in their survey report of online health information searching. Kishimoto and Fukushima (2011) used *age*, *education level*, *marriage status*, *gender* and *location* (residential area) in their online drug information studies. Cotton and Gupta (2004) also used *age*, *gender*, *race*, and *marital status* in their studies to investigate online and offline health information users.

I took *age*, *job* and *nationality* as the basic demographic categories in this study, while putting all other categories not mentioned in the literature in the *others* category. The *others* category comprise mostly infrequent types of information such as *race* and *location*. The anonymity of online discussion forums allow users to share and exchange information without knowing personal details. In the postings, *age* and *country* were sometimes mentioned to give readers a general profile and context for interpreting the shared information.

Patient symptoms include *symptom description*, *subjective feeling of having the symptom*

(*condition*), and *patient history of disease*. The categories of *symptom description* and *patient history of disease* came from previous literature. Fox and Duggan (2013) found that American Internet users searched for specific disease conditions. Greene, Choudhy, Kilabuk and Shrank (2010) reported that users of an online Facebook group for diabetes posted their concerns of negative events to find out if their experience matched with those of other users. There is one category of health information not mentioned in previous literature: subjective feeling of (having) the problem/condition. During the open coding process, I found that some posts contained this category of health information—description of problem or condition based on the writer’s opinion. It sometimes starts with “I feel like...” or “I think it should be ...”. It is different from *symptom description* in that *symptom description* is the direct description of the patient’s problem or condition that has been confirmed by a health professional to be a symptom of the disease. The *subjective feeling* of having a symptom is a speculation of the user. Hence, this category was added in the sub-section of *patient symptom*.

Emotion support is another aspect of user related health information. It includes attitude to the problem, emotional status of having/knowing the problem, and other people’s attitude and support. These categories all come from previous literature, but were reorganized in this study. Chuang and Yang (2012) defined emotional support into the following types: *relationship, physical affection, confidentiality, sympathy, listening, understanding and empathy, encouragement and prayers*. However, these types are too detailed to be useful in this study. Moreover, *confidentiality, listening, understanding and empathy* can be

categorized into other's attitude since they represent other's viewpoints. Hence, I used *other's attitude and support* to represent all these detailed categories. However, these categories did not take into consideration the emotional aspect of the post writer. Hence, in this study, *attitude to the problem*, and *emotional status of having/knowing the problem* were added to represent the poster's own emotions.

Patient experience of drug and treatment were not frequently found in previous literature. In prior studies, drug and treatment information usually referred to formal or authoritative drug and treatment information. In this study, *patient experience of drug and treatment* focuses on the user's self report of taking the drug and treatment, including self-reported feelings and effect. The self-reported feelings and drug effects may not match the effects found in clinical trials and listed on the drug labels. The category of *patient experience of drug/treatment* includes the following subcategories: *perceived side effect*, *interaction with another health problem of patient*, *dosage used*, *description of used procedure*, and *caution or reminder from patient*.

Factual/general knowledge-based health information

Factual/general knowledge-based health information refers to health information that is based on medical or scientific facts, or proven and well-known factual knowledge. This type of health information comes from medical reports, surveys, medical journals and experiments, and has higher credibility than user self-reported health information. I subdivided it into two sub-categories: *etiology* and *drug information*.

Etiology is subdivided into *the cause of disease* and *the formal/professional description of the*

disease. These two categories come from the literature. Huges and Cohen (2011) found that there are professionally reviewed health information on some websites including formal descriptions of diseases and (formal) symptoms of diseases. They also found that the medical terms sometimes were used in the description of disease and drug effects. Hence, I added *terminology* in this category.

Drug information refers to the formal expression of drug names, effects, side effects, drug interaction, contra-indication, etc. These categories come from prior literature of online drug information research. Kishimoto and Fukushima (2011) found that Japanese health information users searched online for drug information covering the topics of side-effect, action mechanism, effect on body, effectiveness, generic drugs, drug interaction, drug dosage, and services of pharmacies. In this study, I reorganized them into the following subcategories: *efficacy*, *indication*, *contra-indication*, *interaction with other drugs*, and *pharmacology*. The *action mechanism* was replaced by *pharmacology*. The drug *dosage* has been included in the category of *patient experiences of drug*.

General-awareness based health information

There is another group of health information that is neither related to user's experience of health problem, nor to medical factual knowledge. I used *general-awareness based health information* to represent this broad category of health information. This is subdivided into: *general health issue*, and *curiosity based health issue*.

General health issue is related to public awareness and concern. During the open coding process, I found that 3 participants who had no particular health issue browsed for topics of

diet, weight control, and mental health issue. Fox and Duggan (2013) found that many US Internet users searched for health information about health and fitness. Hence, I used this term to represent these types of health information. It is subdivided into *common health issue*, *pollution*, *smoking*, and *hot topic of general interest*. A common health issue is not really a particular health condition, but a health area that the public is interested in, such as weight control and diet control. *Pollution* and *smoking* are well-known health-related issues. *Hot topic of general interest* refers to health issues that are prevalent or of major social concern, and not limited to particular diseases or topics publicized in the mass media. It includes drug safety and mental disorders such as depression.

Besides general health issues of public interest, I added *curiosity-based health information* in this category, subdivided into: *rare issue*, *interesting story*, *famous people* and *counter-intuitive information*.

The details of the coding scheme divided into three broad categories of health information are listed in Table 3.7.

Table 3.7. Coding scheme for different types of health information

User case-based relevance criteria
<u>A. Patient demographic</u>
A1. Age & gender
A2. Job & occupation
A3. Nationality
A4. Other categories
<u>B. Patient's symptom</u>
B1. Description of patient symptom
B2. Subjective feeling of having a problem
B3. Personal history of disease
<u>C. Emotional support</u>
C1. Attitude to the problem
C2. Emotional status of knowing the problem
C3. Other's attitude & support
<u>D. Patient's experience of drug/treatment</u>
D1. Perceived side effect
D2. Interaction with another health problem (experienced by patient)
D3. Dosage used
D4. Description of procedure used
D5. Caution or reminder (based on patient's experience)
Factual/general knowledge-based criteria
<u>E. Etiology</u>
E1. Cause of disease
E2. Description of disease
E3. Description of terms
<u>F. Drug information</u>
F1. Efficacy
F2. Indications
F3. Contra-indications
F4. Interaction with other drugs
F5. Pharmacology
General awareness-based criteria
<u>H. General health issue</u>
H1. Common health issue
H2. Pollution
H3. Smoking
H4. Hot topic of general interest
<u>I. Curiosity-based criterion</u>
I1. Rare issue
I2. Interesting story
I3. Famous people
I4. Counter-intuitive information

CHAPTER 4

RESULTS

Data analysis in this chapter is divided into three parts: quantitative data analysis, qualitative data analysis, and finally discussion of possible explanations for some of the findings.

4.1 QUANTITATIVE DATA ANALYSIS

In the quantitative data analysis, duration of scanning (post surrogates) and reading (detailed post content and replies), skimming and examining text, and the ratio of examining to skimming were calculated to investigate the relationships between these eye-movement parameters and the relevance judgments by participants in three types of use contexts (seeking for own health issue, seeking for others, and browsing with no particular health issue in mind). In addition, I recorded the number of scanning sessions and reading sessions for each participant, to investigate their associations with the three types of use contexts.

I also analyzed the average numbers of scanning and reading sessions. As each participant had different numbers of scanning and reading sessions, I used the average number of sessions per participant for each use context. There were about 20 participants for each of the three types of use contexts. The detailed results are given in Table 4.1. Recall that scanning refers to users checking the post surrogates to select which post to read in detail, the reading refers to users reading a detailed post content and replies/comments from other users. In this

table, it can be seen that participants seeking for own health issue had the lowest number of scanning and reading sessions (on average 5.6 scanning sessions and 5.5 reading sessions), followed by participants seeking for other people's health issue (on average 6.4 scanning sessions and 6.1 reading sessions). Participants with no particular health issue were found to have the highest number of scanning and reading sessions (on average 12.3 scanning sessions and 7.8 reading sessions). The high number of scanning sessions relative to reading sessions suggested that participants scan more pages of summary results of post surrogates (each new page is considered a new scanning session). This was because sometimes the participant did not select a post surrogate to read in detail within a page but went on to the next page of post surrogates (thereby starting a new scanning session). It was also noticed that the group of participants with no particular health issue had the highest difference between the number of scanning sessions and reading sessions—on average 4.5 more scanning sessions than reading sessions. In contrast, there was little difference between the number of scanning and reading sessions for the other two groups. This difference in scanning/reading behavior was an indicator that distinguished between users with no particular health information need and the other two groups. The reason for this difference could be due to people with no particular health issue not having a particular health topic in mind at the beginning of the session and hence they were less likely to find particular post surrogates that attract their attention enough to read the detailed post.

Table 4.1 Summary of statistics of scanning & reading session

Participant group	No. of participants	Average no. of scanning sessions	Average no. of Reading session
Seeking for self	18	5.6 (range: 4 to 8)	5.5 (range: 4 to 8)
Seeking for others	18	6.4 (range: 4 to 10)	6.1 (range: 4 to 9)
No health issue	22	12.3 (range: 8 to 16)	7.8(range: 7 to 14)

4.1.1 Scanning the post surrogates stage

In this section, the average scanning duration as well as examining and skimming durations with the scanning stage were calculated and compared. Average duration was calculated by taking the average for each participant (as a participant could have a few scanning/reading sessions) and then taking the average for the group of participants in each use context.

ANOVA (Analysis of Variance) was carried out to check for significant differences among the three groups. Post hoc Tukey's Honestly Significant Difference (HSD) test was used to identify significant differences between pairs of groups. The ANOVA and post hoc test results are shown in Table 4.2. The ANOVA test results indicated that the average duration of the scanning sessions was significantly different among the three groups of participants. It was found from the post hoc test that the average duration of the scanning surrogates stage for people seeking for self was higher than that for people seeking for others. Both of them

were lower than the average scanning duration for people with no issue in mind. All the differences were significant at level of 0.002.

Table 4.2 Summary of ANOVA & Post Hoc results for scanning duration

ANOVA test for the 3 groups

Scanning-duration

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1004.192	2	502.096	32.768	.000**
Within Groups	842.742	55	15.323		
Total	1846.934	57			

Group statistics

	participant_group	N	Mean	Std. Deviation	Std. Error Mean
scanning_duration	seeking for self	18	19.9333	1.97574	.46569
	seeking for others	18	15.2944	2.76522	.65177
	no health issue	22	25.3182	5.54803	1.18284

Tukey's HSD

(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
participant_group	participant_group				Lower Bound	Upper Bound
For self	For others	4.63889*	1.30480	.002	1.4959	7.7818
	No issue	-5.38485*	1.24408	.000	-8.3815	-2.3882
For others	For self	-4.63889*	1.30480	.002	-7.7818	-1.4959
	No issue	-10.02374*	1.24408	.000	-13.0204	-7.0271
No issue	For self	5.38485*	1.24408	.000	2.3882	8.3815
	For others	10.02374*	1.24408	.000	7.0271	13.0204

Participants seeking for their own health issue spent more time on scanning the post surrogates than those seeking for other's health issue, which was consistent with the finding of the preliminary study. The reason could be due to people seeking for their own issue need more time to check more details in the post surrogates to match with their own situation than people seeking for other's health issue. However, it was found that participants with no particular health issue spent even more time than the other two groups of participants, which was not consistent with the preliminary finding. The possible reason for this was that people with no health issue could not decide what was interesting at first and needed more time to scan more post surrogates.

A point to highlight was that the standard deviation of the scanning duration for the group with no health issue was also the highest among the three groups. It suggested that participants with no particular health issue was not clustered tightly, but exhibited wider range of scanning behavior, compared to the other two groups. Taking a closer look at the distribution of scanning duration (shown in Figure 4.1), it was observed that there were two clusters of participants with no health issue: one cluster of five participants had short scanning durations (of median 17.5 seconds), and a second cluster of 17 participants had a longer scanning duration (of median 28 seconds). This cluster of five participants had scanning durations that were similar to the other two groups. I conjecture that they actually had a tacit health issue in mind, which they did not specify to the researcher. Possible reasons

were:

1. They did not want to reveal their health issue to the researcher;
2. They had latent health information needs that were triggered by information on the screen;
3. While browsing, they developed an interest in a particular health issue.

Figure 4.1. Distribution of average scanning duration for the three groups of participants (group 1: seeking for self, group 2: seeking for others, group 3: no health issue)

As mentioned in the research framework, scanning was divided into skimming (quick eye movement over text) and examining (close focus on a particular text passage). I calculated and compared the average skimming and examining duration during the scanning stage. The detailed results are given in Table 4.3.

It was found that for the skimming duration, there was no significant difference between participants seeking for self versus participants seeking for others. But the skimming duration for these two groups was significantly less than for participants with no health issue ($\alpha=0.001$).

For the examining duration, there was no significant difference between participants seeking for own issue versus participants with no health issue. However, they were both significantly higher than for participants seeking for other's health issue ($\alpha=0.001$).

In addition, it was found again that the standard deviation of skimming duration for participants with no health issue was much higher than that of the other two groups, which provided more support for the two clusters observation.

**Table 4.3. Summary of ANOVA and Post Hoc results for skimming and examining
within the scanning stage**

Scanning-skimming

ANOVA test for the 3 groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	622.506	2	311.253	28.924	.000
Within Groups	591.854	55	10.761		
Total	1214.360	57			

Group statistics

	participant_group	N	Mean	Std. Deviation	Std. Error Mean
Scanning-skimming	Seeking for self	18	8.2167	2.35003	.55391
	Seeking for other	18	7.6722	1.66762	.39306
	No health issue	22	14.6818	4.63266	.98079

Tukey's HSD

(I) participant_group	(J) participant_group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
For self	For others	.54444	1.09346	.873	-2.0894	3.1783
	No issue	-6.46515	1.04258	.000	-8.9765	-3.9538
For others	For self	-.54444	1.09346	.873	-3.1783	2.0894
	No issue	-7.00960	1.04258	.000	-9.5209	-4.4983
No issue	For self	6.46515	1.04258	.000	3.9538	8.9765
	For others	7.00960	1.04258	.000	4.4983	9.5209

Table 4.3. Summary of ANOVA and Post Hoc results for skimming and examining within the scanning stage (continued)

Scanning-examining					
ANOVA test for the 3 groups					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	158.244	2	79.122	33.945	.000
Within Groups	128.199	55	2.331		
Total	286.443	57			
Group statistics					
	participant_group	N	Mean	Std. Deviation	Std. Error Mean
Scanning-examining	Seeking for self	18	11.6056	1.10638	.26078
	Seeking for other	18	7.6222	1.67070	.39379
	No health issue	22	10.6773	1.68944	.36019
Tukey's HSD					
(I)	(J) participant_group	Mean	Std. Error	Sig.	95% Confidence Interval
participant_group		Difference (I-J)			Lower Bound Upper Bound
For self	For others	3.98333	.50891	.000	2.7575 5.2092
	No issue	.92828	.48523	.145	-.2405 2.0971
For others	For self	-3.98333	.50891	.000	-5.2092 -2.7575
	No issue	-3.05505	.48523	.000	-4.2238 -1.8863
No issue	For self	-.92828	.48523	.145	-2.0971 .2405
	For others	3.05505	.48523	.000	1.8863 4.2238

4.1.2 Reading stage

Table 4.4 gives the details of the reading duration for the three groups of participants. It was found that the participants seeking for self had the highest reading duration, followed by participants seeking for others and participants with no particular health issue. But there was no significant difference between participants seeking for others and with no particular issue. These results were consistent with the findings of the preliminary study.

Table 4.4. Summary of ANOVA & post hoc results for reading duration

Reading

ANOVA test for the 3 groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4173.132	2	2086.566	19.965	.000
Within Groups	5748.257	55	104.514		
Total	9921.389	57			

Group statistics

	participant_group	N	Mean	Std. Deviation	Std. Error Mean
Reading	Seeking for self	18	67.9389	8.04123	1.89534
	Seeking for other	18	52.0500	11.19802	2.13940
	No health issue	22	48.1955	10.94856	2.33424

Table 4.4. Summary of ANOVA & post hoc results for reading duration (continued)**Tukey's HSD**

(I) participant_group	(J) participant_group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
For self	For others	15.88889	3.40773	.000	7.6805	24.0973
	No issue	19.74343	3.24915	.000	11.9171	27.5698
For others	For self	-15.88889	3.40773	.000	-24.0973	-7.6805
	No issue	3.85455	3.24915	.466	-3.9718	11.6809
No issue	For self	-19.74343	3.24915	.000	-27.5698	-11.9171
	For others	-3.85455	3.24915	.466	-11.6809	3.9718

The results indicate that participants seeking for their own health issue spent the highest amount of time reading the detailed post and replies, but participants with no health issue spent the least. The reason could be due to people seeking with their own health issue were familiar with their own condition and health history, and they needed more time to check and match their condition with the post's description. In contrast, people seeking for other's health issue or without a particular issue did not require such a long time to read all the details of the post description since they were not intimately familiar with the other person's condition or they did not have a specific case to match with.

Again it was found that participants with no health issue had a higher standard deviation for

the reading duration, than for the participants seeking for self. The reason for the high standard deviation might be due to the existence of two clusters of participants, consistent with the observation in the previous section on scanning duration that there were two clusters of participants. The details of these three distributions are shown in Figure 4.2.

In terms of the skimming and examining duration within the reading stage, it was found that participants seeking for their own health issue had the longest examining duration, followed by participants seeking for other's health issue and participants with no particular health issue.

Figure 4.2. Distribution of average reading duration for the three groups of participants (participant group 1: seeking for self, participant group 2: seeking for others, participant group 3: no health issue)

For the skimming duration, it was found that the participants with no particular health issue had the longest duration, but the difference between participants seeking for their own issue and participants seeking for other's health issue was not significant. It suggested that people with no particular health issue tend to skim over the detailed post to locate the text that are interesting to them, while the other two groups of participants focused closely on particular pieces of content and had similar skimming behavior. Hence they had less skimming time at this stage compared to people with no particular health issue. The details are given in Table 4.5.

Table 4.5. Summary of ANOVA & post hoc results for skimming & examining within the reading stage

Reading-skimming

ANOVA test for the 3 groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7073.208	2	3536.604	19.052	.000
Within Groups	10209.528	55	185.628		
Total	17282.736	57			

Group statistics

	participant_group	N	Mean	Std. Deviation	Std. Error Mean
Reading-scanning	Seeking for self	18	19.5944	3.99967	1.88554
	Seeking for other	18	19.8611	4.35162	1.02569
	No health issue	22	26.5909	5.45614	1.16325

Table 4.5. Summary of ANOVA & post hoc results for skimming & examining within the reading stage (continued)

Tukey's HSD

(I) participant_group	(J) participant_group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
For self	For others	-.26667	2.02758	.991	-5.1506	4.6173
	No issue	-6.99646	1.93322	.002	-11.6531	-2.3398
For others	For self	.26667	2.02758	.991	-4.6173	5.1506
	No issue	-6.72980	1.93322	.003	-11.3865	-2.0731
No issue	For self	6.99646	1.93322	.002	2.3398	11.6531
	For others	6.72980	1.93322	.003	2.0731	11.3865

Reading-examining

ANOVA test for the 3 groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	643.834	2	321.917	8.700	.001
Within Groups	2034.990	55	37.000		
Total	2678.824	57			

Group statistics

	participant_group	N	Mean	Std. Deviation	Std. Error Mean
Reading-scanning	Seeking for self	18	18	47.7889	13.65499
	Seeking for other	18	18	31.1278	12.39316
	No health issue	22	22	21.1545	14.52206

Table 4.5. Summary of ANOVA & post hoc results for skimming & examining within the reading stage (continued)

Tukey's HSD

(I) participant_group	(J) participant_group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
For self	For others	16.66111	4.54151	.002	5.7217	27.6005
	No issue	26.63434	4.33016	.000	16.2041	37.0646
For others	For self	-16.66111	4.54151	.002	-27.6005	-5.7217
	No issue	9.97323	4.33016	.064	-.4570	20.4035
No issue	For self	-26.63434	4.33016	.000	-37.0646	-16.2041
	For others	-9.97323	4.33016	.064	-20.4035	.4570

The standard deviations of skimming and examining durations for participants with no health issue were again higher than for the other two groups.

I calculated the ratio of examining to skimming—to investigate its association with relevance judgment of detailed posts. It was found that the examining/skimming ratio for posts that were judged relevant was significantly higher than for posts judged not relevant, for all the three groups of participants. This was consistent with the finding of a study by Buscher, Dengal and Elst (2008). It was noticed that the standard deviation of the ratio for participants with no particular health issue was much higher than that for the other two groups. Again, there appears to be two clusters of participants with no particular health issue. The details are given in Table 4.6.

Table 4.6. Ratio of examining duration to skimming duration for posts judged as relevant and non-relevant

	Participant group	N	Post judged relevant	Post judged non relevant	Difference
Ratio of examining to skimming	Seeking for self	18	3.14	1.10	2.04
	Seeking for others	18	2.18	1.56	0.62
	No health issue	22	1.89	1.42	0.47
	No health issue*				
	cluster 1	17	1.67	1.69	-0.02
	cluster 2	5	2.76	1.40	1.34

4.2 CONTENT ANALYSIS

Content analysis of discussion forum texts with participants' eye fixation was carried out to find out what are the texts that participants focused on and what kinds of health information they paid attention to and possibly use in the relevance judgment process. In the content analysis, the sentences that participants fixated on were extracted from the screen shots and categorized into different types of health information by two coders. Different types of health information were related to different aspects of health issues, such as symptoms and history of disease related to the user's conditions (case), drug names and treatments which could be considered factual knowledge, and diet and exercise related information which were general

interest or lifestyle topics. The results of the analysis of these types of health information could provide possible explanations for eye movement patterns from the quantitative analyses. The example of screen shots that were used in the coding was shown in Figure 4.3 and Figure 4.4.

	Thread / Thread Starter	Last Post	Replies	Views	Board
Knee feels wierd Mortar790	06-20-2014 10:21 AM by Mortar790	0	121	Knee & Hip Problems
Why still excruciating pain in my knee? frustratedsoul	03-18-2014 04:58 PM by lacevgirl	4	339	Knee & Hip Problems
Knee pain Athlete Juschillin50	09-10-2013 05:15 AM by Juschillin50	0	232	Knee & Hip Problems
Weight loss suggestions for sedentary lifestyle. Terri D	04-06-2013 04:25 PM by kE893	4	1,500	Weight Loss
Plica Syndrome, PLEASE HELP! Rebeccajadeeva	03-18-2013 07:11 PM by CLB1988	1	382	Knee & Hip Problems
Left (Inner Knee) Pops out Hockey Chick	08-11-2012 07:21 PM by Hockey Chick	0	496	Knee & Hip Problems
Knee problem, please reply Jordansir	01-27-2012 06:00 PM by janewhite1	1	567	Knee & Hip Problems
Hard Fall = Severe Bruising & Painful right leg/knee mavis513	12-10-2011 11:25 PM by mscat40	2	5,278	Burns & Injuries
Right knee "locks" and "pops". Help! rdv0015	11-20-2011 06:14 PM by janewhite1	1	3,230	Knee & Hip Problems
Knee Pain and Odd Cold, Wet Sensation LordBreadsticks	11-02-2011 09:49 PM by LordBreadsticks	0	3,081	Knee & Hip Problems

Figure 4.3 Screenshots of post surrogates summary page (with eye fixations: the green spots)

Figure 4.4 Screenshots of detailed post page (with eye fixations: the green spots)

The results of the content analysis indicate that the three groups of participants focused on different types of health information: participants seeking for their own health issue focused mostly on symptoms, history of disease and treatment, which can be considered case-based relevance criteria that participants might use to match their own conditions. Participants seeking for others' health issue focused mostly on treatments, medical terms and cause of disease, which could be considered general knowledge-based relevance criteria. Finally, participants with no particular health issue in mind focused mostly on general health topics such as diet and exercise, public awareness topics such as smoking and air pollution, and interesting stories. These could be considered general interest-based relevance criteria. These are consistent with the preliminary findings summarized in the last chapter. The topics that participants sought for their own health issue and for others' issue are listed in Table 4.7.

Table 4.7. Health topics specified by participants in the experiments

The topics that participants sought for their own issue		The topics that participants sought for other's issue	
Joint pain	Decayed tooth	Calcium deficiency	Gastritis
Acne	Quinsy	Fever	Chinese medicine
Red spot on skin	Sprained ankle	Alternative medicine for bone	Child jaundice
Mosquito bite	Period disorder	Cancer	Pink eyes
Aging problem	Flu	Breastfeeding	Women's health
Back & foot	Sport injury	Diabetes	Lung problem
Knee problem	Gastritis	Obesity	Inflammation
Headache	Acid reflux	Vertigo	Fever
Block nose	Muscle pain	Rheumatism	Cough

4.2.1 Scanning post surrogate stage

In the scanning stage, I calculated the percentages of post surrogates containing particular types of health information. The percentages of post surrogates containing each type of health information are shown in Figure 4.5. Categories with percentages less than 5% were removed from the graph so that it can fit in the page.

Note: kinds of health information with percentage less than 5% of occurrences are left out

Figure 4.5. Percentage of post surrogates containing each type of health information for the three groups of participants

Abbreviations in the figure: AGE: age & gender; JOB: job & occupation; SYM: description of patient symptom; SES: subjective feeling of having a problem; HST: personal history of disease; DIS: description of disease; TRM: description of terms; TYP: type of procedure; PRO: description of procedure used; SMO: smoking issue; HOT: public hot topics; RAR: rare issue.

It could be seen in Figure 4.5 that the three groups of participants with different types of use contexts focused on different types of health information. For participants seeking for their own health issue, the following were the most common categories they fixated on:

- Description of patient symptom (B1 SYM: 58.3%)
- Personal history of disease (B3 HST: 57.6%)
- Description of disease (E2 DIS: 55.2%).

In comparison, for the participants seeking for other's health issue, the most common types of information fixated on were as follows:

- Description of disease (E2 DIS: 45.3%)
- Description of terms (E3 TRM: 47.5%)
- Description of procedure used (G2 PRO: 35.8%).

For participants browsing with no particular issue, the most common types of information fixated on were as follows:

- Smoking issue (H3 SMO: 32.4%)
- Hot health topic (H4 HOT: 28.4%) (e.g., sexual issue, weight control, exercise, diet)
- Rare health issue (I1 RAR: 23.5%).

A few other types of health information were also found important for each group of participants (with frequency above 20%):

- Participants seeking for their own health issue: self-estimation of disease, type of procedure, and terms of disease.
- Participants seeking for other's health issue: type of procedure.

There was a possibility that the percentages for the different types of health information with eye fixation merely reflected the overall distribution in the discussion forum. Table 4.8 compares, for each type of information, the percentage in the surrogates skimmed versus the percentage in the surrogates fixated on. The types of information with high percentage in the set of surrogates fixated on had low percentage in the set of surrogates skimmed. This

indicated that the participants intentionally focused on the particular types of health information in making relevance judgments.

Table 4.8. Percentage of health information in the set of skimmed surrogates compared to the fixated surrogates

Category of health information	% in surrogates skimmed	% in surrogates fixated on
For their own health issue		
B1 SYM symptom	29.3%	58.3%
B3 HST history of disease	18.4%	57.6%
E2 DIS description of disease	29.1%	55%
For other's health issue		
E3 TRM terms	25.1%	47.5%
E2 DIS description of disease	17.6%	45.3%
G1 TRT treatment	12.5%	28.4%
No particular issue		
H4 HOT hot topic	16.5%	32.4%
H3 SMO smoking	17.3%	28.4%
I1 RAR rare issue	13.2%	23.5%

I further calculated the percentages of different types of health information for the selected post surrogates (i.e. clicked on). The selected post surrogates were a subset of the post

surrogates with eye fixations. For the selected post surrogates, the most common types of health information found were:

For participants seeking for own health issue

- Description of patient symptom (B1 SYM: 88%)
- Personal history of disease (B3 HST: 75%)
- Description of disease (E2 DIS: 72%).

For those seeking for other's issue

- Description of terms (E3 TRM: 77%)
- Description of diseases (E2 DIS: 72%)
- Description of procedure (G2 PRO: 46.7%).

For those with no particular health issue

- Hot public health topics (H4 HOT: 68%)
- Smoking issue (H3 SMO: 56%)
- Rare health issues (I1 RAR: 28%).

The most common types of information were the same as those for fixated post surrogates, but with even higher percentages. The only differences were in the ordering of the top three types of health information for participants seeking for others and participants with no particular health issue. The details are shown in Figure 4.6.

Note: kinds of health information with percentage less than 5% of occurrences are left out

Figure 4.6. Percentage of selected post surrogates containing each type of health information

To further explore what are the more important types of health information in determining the relevance of post surrogates, I analyzed each type of health information separately. For each type of information, I calculated the percentage of post surrogates fixated on that were actually selected. If the participant clicked on the post surrogate, it suggested that the participant thought it likely to be relevant. I calculated the ratio of post surrogates selected over post surrogates fixated on, for each type of information. The formula used was:

$$\frac{\text{No. of post surrogates containing the type of information that were selected}}{\text{No. of post surrogates containing the type of information that were fixated on}}$$

For example, if there were 60 surrogates with fixations containing a description of a symptom, and of these 30 were selected, the percentage is 50%. The assumption is that if the participant

clicks on a post surrogate, the participant thought the post content likely to be relevant.

As a baseline for comparison, I also calculated the overall percentage of fixated surrogates that were selected:

$$\frac{\text{No. of post surrogates that were selected}}{\text{No. of post surrogates that were fixated on}}$$

The results are shown in Table 4.9.

It was found that for participants seeking for their own health issue, personal history of disease and description of patient symptom (B1 and B3) and the description of disease (E2) were important in determining the relevance of post surrogates. For participants seeking for other's issue, the treatment and disease factual information (E1, E2, E3 and G1) were important. In contrast, for participants browsing without particular issue, general interest (H2, H3, H4) and curiosity-based (I1, I2) information were important. This indicated that participants were using different kinds of relevance criteria in their relevance decisions.

Table 4.9. Percentage of fixated surrogates that were selected, for each type of health information

Type of health information	For own health issue	For other's health issue	No particular issue
A1 AGE age	32%*	0.23	0
A2 JOB job	29%	0	15%*
A3 NAT nationality	23%*	35%*	12%*
B1 SYM symptom	37%	21%*	0
B2 SES self estimation	31%	0	0
B3 HST history	32%	14%*	14%*
C1 ATT attitude	30*	0	0

Table 4.9. Percentage of fixated surrogates that were selected, for each type of health information

information (continued)			
C2 ES emotion	0	0	100%*
C3 OA other's attitude	13%*	0	0
D1 PSE felt effect	0	0	0
D2 INT felt interaction	0	0	0
D3 DOS dosage	21%*	0	21%*
D4 USE usedprocedure	0	25%*	0
D5 CAU caution	0	0	0
E1 RSN cause	0	45%	22%*
E2 DIS description	30%	31%	0
E3 TRM terms	21%	29%	0
F1 EFF efficacy	20%*	0	12%*
F2 IND indication	23%*	0	0
F3 CNT contra-indica	100%*	0	0
F4 INT interaction	0	0	0
F5 PHR pharma	0	17%	2%
G1 TRT treatment	16%	35%	21%*
G2 PRO procedure	22%	28%	100%*
H1 COM common	0	0	0
H2 PUL pollution	0	37%*	100%*
H3 SMO smoking	0	0	52%
H4 HOT hot	0	0	64%
I1 RAR rare	0	16%*	24%
I2 STO story	0	0	14%*
I3 FMP famous guy	0	0	0
I4 CONT contra	0	0	2%
Overall percentage	0.15	0.12	0.08

* Categories of health information with low percentage in fixated surrogates

4.2.2 Reading post content stage

Content analysis was carried out on the posts that were judged relevant by participants, i.e. that the participants indicated were relevant during the post-experiment interview. Figure 4.7 shows the percentages of the different types of information among all the relevant posts.

Figure 4.7. Percentages of particular types of health information in relevant posts

From Figure 4.7, it can be seen that the situation was similar with the scanning stage. For people seeking for their own health issue, the most common types of health information that participants fixated on were:

- Personal history of disease (91.7%);
- Description of disease (89.6%) and;
- Description of patient symptom (87.5%).

For participants seeking for other's health issue, the most common types of health

information were:

- Description of terms (92.3%);
- Description of procedure (90.5%) and;
- Description of treatment (88.7%).

For participants with no particular health issue, the most common types were:

- Smoking (66.7%);
- Hot topics (33.7%);
- Air and water pollution (33.7%).

Figure 4.8. Percentages of posts (read) containing each particular type of health information

For comparison, the percentages of posts that were read containing each type of health information are shown in Figure 4.8. It was found that for participants seeking for others, the

percentages of posts that were read containing personal or case related types of health information were higher compared to the percentages of posts judged as relevant. In fact, the percentages in posts judged as relevant were nearly zero, which indicated that participants seeking for others treat personal details as a sign of non-relevance. Other than this, the patterns were similar with those for posts judged as relevant.

To further investigate the importance of particular types of health information in determining a detailed post's relevance, I calculated the ratio of posts judged relevant over posts that were read, for each type of information:

$$\frac{\text{No. of posts read and judged relevant, containing this type of information}}{\text{No. of posts read, containing this type of information}}$$

It was found that description of symptoms, history of disease, and treatment procedure were the most important types of information for participants seeking for their own issue. For participants seeking for others, cause of disease, description of disease and treatment procedure were the most important types of information. In comparison, air pollution, smoking issue and rare cases were the most interesting types of health information for participants with no issue in mind. The details are shown in Table 4.10. These findings were similar to the findings for the scanning stage except that the treatment procedure was important in the reading stage for participants seeking for their own issue.

Table 4.10. Summary of types of health information that are important in

determining the relevance of the detailed posts

Type of health information	Of the posts containing this kind of information that were read, the percentage that were judged relevant
For their own health issue	
Description of symptoms (B1)	74.5%
Personal history of disease (B4)	72.3%
Treatment procedure (G2)	69.1%
For other's health issue	
Cause of disease (E1)	67.2%
Description of disease (E2)	60.5%
Treatment procedure (G2)	53.2%
No particular issue	
Air water pollution (H2)	55.9%
Smoking issue (H3)	45.2%
Rare issue (I1)	40.1%

From the results, it was concluded that the groups of participants with different types of use contexts (seeking for self, seeking for others, and no particular issue) focused on different types of health information and used them in their relevance judgment.

4.3 POST-EXPERIMENT INTERVIEW ANALYSIS

The participants were interviewed after the health information seeking session to obtain self-reported commentary on their relevance judgments. Questions were asked to find out what were the reasons that attracted participants' attention and the criteria/reasons why they judged particular post surrogate or detailed post relevant. Details were given in Chapter 3.

The participants' responses were coded to identify three broad types of relevance criteria: case-based criteria, basic/general knowledge-based criteria, and general awareness/curiosity-based criteria. To improve coding consistency, the coding was based on the presence of cue-words that reflect the three categories. For instance, if a participant's answer included the following terms or phrases "I compare", "similar to my situation", "the similarity between mine and the post", "familiar with the condition/symptom" and other similar phrases, then they were coded as case-based criteria. If the answer contained the phrase "I learned before", "I knew it before", "I searched for them in the past", "I heard" or other similar phrases, they were coded as basic/general knowledge-based criteria. The details of the coding scheme are given in Table 4. 11.

The coding was done by me and another PhD student in the school (not the same person as the coders for the content analysis of text with eye-fixation). The coders looked for cue-words or cue-phrases listed in Table 4.11, and used them to categorize participants into these three groups of criteria. If cue-words and cue-phrases belonging to more than one group were found, then they were simultaneously counted into different groups. The coding was quite straightforward and there was no difficulty distinguishing between the three categories of criteria during coding. A high Cohan's kappa of 0.91 was obtained for the two codings.

Disagreements between the two coders were easily resolved by checking more of the participant's answers.

Table 4.11. Coding scheme for post experiment interview data

Type of criteria/reason in relevance judgment	Keywords or phrases to look for
Case-based criteria	<p>Compare, match, comparison, similar, consistent</p> <p>Similar to my problem/situation/condition, same as mine/my problem/situation/condition/disease, like my problem/situation/condition</p>
Basic/general knowledge-based criteria	<p>Know, learn, familiar</p> <p>Heard/learn from, read this before from newspaper/book/magazine/Internet, somebody tell/inform me</p>
General awareness/curiosity-based criteria	<p>Interesting, funny, rare, weird, strange, curious</p> <p>Smoking, pollution, exercise, diet, fitness</p> <p>Everybody know, hot topic, public concern, environmental, healthy, well-being</p>

Some participants provided detailed explanations of why they judged one particular post

surrogate or detailed post as relevant. They mentioned in which ways they compared the post with their own conditions, such as age, location of problem, description of the pain, and diagnosis result. However, most of the participants provided only basic reasons. Based on the participant's initial answer, I followed up with more probing questions. As the interview was after the experiment session, the participants were somewhat tired, and not many gave very detailed explanations.

From the analysis of the answers, it was found that

- For participants seeking for their own health issue, over 70% (13/18) clearly used condition/symptom match in their judgment of post surrogates and detailed posts. Here is a quote from a participant:

"I found the guy in this post was quite similar to me in the position of headache. He also had the problem after a long time of sitting in front of the screen, the same as me. He did some massage but didn't work for a long time. I also tried this but found not very useful." – Participant seeking for his own health problem.

Three participants were ambiguous in their explanation, but mentioned some comparison between themselves and others:

"When I read the post I realized the boy didn't have a history of jogging as me. His pain was due to the injury of his leg. I don't think it is the same reason as me since my pain comes from the muscle." – Participant seeking for her own health problem.

The remaining two participants expressed other reasons to judge relevance. One wanted to find new information he did not know before. Another used the reference from

renowned medical journals or websites as the criterion of relevance.

- For participants seeking for other's health issue, nearly 85% (15/18) used medical and health knowledge they gained from school or previous searching to guide them in the relevance judgment. For example:

"I heard from my friend about acid reflux. So I want to know more about this and can share with him what I find here." – Participant seeking for other's health issue.

Three of them made the judgment with other reasons. Two of them mentioned novelty of the information as the reason they judged the post as relevant. The last participant did not trust information from normal users of this health discussion forum and wanted to find comments from health professionals.

- For participants with no particular issue, they were divided into two clusters. One cluster with 16 participants was found to judge relevance of post surrogates and detailed posts by their interest, curiosity, rarity and some well-known health issues (e.g., diet and exercise).

For example:

"I found it funny since I didn't expect a girl looking for losing some weight would believe that only magic pills can help them. She didn't do any exercise and that's impossible." – Participant with no particular health issue.

The other cluster of 6 participants judged relevance by a mixture of case-based matching and basic knowledge criteria. Moreover, it was found that these participants actually had latent health information needs. In some posts they used case-based criteria similar to the participants seeking for their own issue. In other posts, they used the prior medical knowledge they learned. I labeled these as "participants with latent health information

needs”.

Latent health information needs were detected when these participants were asked the reason for reading particular detailed posts. Since they were grouped as with no particular health issue, they were expected to respond with reasons related to personal interest and curiosity. Instead, their answers suggested that they had some health issue related with the topics of these detailed posts or had heard from their friends about a health issue. For example:

“When I come across the post, I recall I sprained my left foot three years ago. When I read the post, I find that the guy sprained the foot because of running, the same as me. So I continued to read and find it relevant. I can make use of his experience to avoid sprain in the future.”— Participant with no particular health issue.

“Actually my grandma had diabetes, so I want to know if any useful information in the post. I realized the man in the post used diet control and I heard before. Some reply suggests some herb medicine will help the patient to control the level of blood glucose. I think it useful and will tell my grandma.”— Participant with no particular health issue.

Their responses indicated that latent health information needs were activated at some point during the information seeking session when they came across familiar topics that triggered a memory.

Another possibility for the observed latent health information needs is that the participants were reluctant to share their real health information needs with the researcher at the beginning of the experiment. They became more comfortable later after reading posts of the

health issue. However, I did not detect clear evidence of this in the study.

4.4 ANALYSIS OF EXCEPTIONS

In this study, I found exceptional cases in each group of participants. In this section, exceptional cases of participants seeking for their self and others are discussed. The exceptional cases for participants with no health issue were identified as participants with latent health information needs, discussed in the previous section.

4.4.1 Participants seeking for their own health issue

It was found that participants within this group had the longest reading duration and examining duration within both the scanning (post surrogate) and reading (detailed post) stages since they needed to match their own conditions or symptoms with the description from the post surrogate or detailed post.

However, it was found that two participants had short reading duration and examining duration, because they only wanted to find new information or wanted to find a very strange problem. The first participant mentioned seeking for the topic of sprained ankle, having suffered from it for several years. He mentioned being quite familiar with this problem and knew a lot of details. Hence, during the browsing session, he just focused on the new information that he did not know before. Hence most of the time, he just skimmed over the detailed post and replies without focusing on the text information that he was already familiar with.

The second participant was seeking information for a strange problem. She suffered from a

headache that only occurred at the right side of the head. She found a few post surrogates similar to her case. But when she read further in the detailed post content, she found most of them non-relevant. Hence, she exited from the detailed post page at an early stage of reading, and didn't have long reading and examining durations.

4.4.2 Participants seeking for other's health issue

Overall, participants seeking for other's health issue have the second longest scanning and reading duration, as well as skimming and examining duration. However, three participants behaved quite differently. Two participants had very long reading duration and examining duration within the reading stage but very short skimming duration because of their reading habit. The third participant had long skimming duration but very short examining duration within the reading stage because he wanted to find out a particular drug name.

Of the two participants who had long reading and examining duration, one sought health information for his mother for her back problem, and another sought information for a son on a breastfeeding issue. They indicated that they usually read the detailed posts closely to find useful information. In addition, they said that close reading is their reading habit since childhood. Even for the posts they thought non-relevant, they still read carefully in case they missed something useful.

The participant who had long skimming but very short examining duration sought health information on diabetes for his family. He told me that he was looking for some particular drug names and terms that his family used. If he found the drug names in the context, he read carefully the sentence where he found it and neighboring sentences. Otherwise, he went

through the detailed posts without focusing on other types of health information.

4.5 DISCUSSION

The study results indicated that people made use of different types of information during relevance judgment and also exhibit different eye-movement patterns in three types of health information seeking contexts—seeking information for self, for others and with no particular issue in mind. The question now was whether there were associations between eye movement patterns and particular types of relevance criteria. Clearly it was difficult to relate eye-movement patterns to detailed relevance criteria. In the coding scheme for categorizing relevance criteria, I had grouped the criteria into three broad categories: case-based relevance criteria, factual/general knowledge criteria, and general awareness-based criteria. As case-based relevance criteria were associated with participants seeking information for own health issue, it suggested that this category of relevance criteria was associated with the eye movement patterns of this group of users. Factual/general knowledge criteria were associated with participants seeking health information for others, suggesting that this category of relevance criteria was associated with the eye movements patterns of this group of users.

To obtain stronger evidence for this association, I regrouped the participants into:

- Group 1: participants identified as using *case-based relevance criteria* in the post-experiment interview (N=16)
- Group 2: participants identified as using *factual/general knowledge relevance criteria* (N=15)
- Group 3: participants identified as using *general awareness based relevance criteria*

(N=17).

The eye movement patterns associated with these three groups were expected to be similar to those for the three use contexts. The results are given in Table 4.12.

Earlier, I had found that the group of participants seeking information for self had the following eye movement patterns:

- Scanning duration of post surrogates: higher than for participants seeking for others, but lower than for participants with no issue in mind.
- Reading detailed post duration: longest compared to the other two groups of participants
- Examining duration in the reading detailed post stage: longest

Table 4.12. Post hoc test results for participants that were identified using particular broad category of relevance criteria

(I)	(J)	Scanning-skimming		Scanning-examinin		Reading-skimming		Reading-examining	
participant_gro	participant_gro			g					
up	up								
		Mean	Sig.	Mean	Sig.	Mean	Sig.	Mean	Sig.
		Difference		Difference		Difference		Difference	
		(I-J)		(I-J)		(I-J)		(I-J)	
Case	General	.65773	.001	4.22786	.000	-.23135	.745	18.01723	.000
	knowledge								
General	General	-6.02458	.000	.90447	.167	-7.22176	.001	27.50034	.001
	awareness								
knowledge	Case	-.65773	.001	-4.22786	.000	.23135	.745	-18.01723	.000
	General								
awareness	General	-6.32115	.000	-3.22189	.000	-6.90476	.001	9.32445	.061
	knowledge								
General	Case	6.02458	.000	-.90447	.167	7.22176	.001	-27.50034	.001
	General								
knowledge	General	6.32115	.000	3.22189	.000	6.90476	.001	-9.32445	.061
	knowledge								

Table 4.12 indicates that these patterns also hold for the group of participants identified to use case-based criteria.

Participants seeking information for others were associated with the following eye movement patterns:

- Examining duration in the scanning surrogates stage: shortest
- Examining duration in the reading detailed post stage: second longest

These patterns also hold for the group of participants identified to use factual/general knowledge criteria.

Participants browsing with no particular issue were associated with the following eye movement patterns:

- Skimming in the scanning surrogates stage: longest
- Examining duration in the reading detailed post stage: shortest
- Skimming duration in the reading detailed post stage: longest

These patterns hold for the group of participants using general awareness criteria.

CHAPTER 5

CONCLUSION

This study sought to find out the information criteria people use for relevance judgments when browsing a health discussion forum for three broad types of use contexts—seeking information for own health issue, seeking information for other’s health issue and browsing with no particular health issue in mind. In addition, the study investigated the eye movement patterns associated with the three types of use contexts as well as categories of relevance criteria.

5.1 MAIN FINDINGS

It was found in this study that users browsing a health discussion forum made use of different types of information in relevance judgments and exhibited different eye movement patterns in the three types of use contexts –seeking information for self, seeking for others, and browsing with no particular issue in mind. Moreover, these eye movement patterns were also found significantly associated with three broad categories of relevance criteria. This has been discussed in the last section of Chapter 4. Users browsing for their own issue were found to use mostly cased-based relevance criteria such as symptoms, history of disease, and treatments in their judgments which took a longer time to scan post surrogates and to read the detailed posts. Participants seeking for others were found to use mostly general knowledge-based criteria such as medical terms, cause of disease and basic treatments in

their judgments which took them a moderate amount of time to scan post surrogates and read post content. In contrast, participants seeking with no particular health issue were found to be interested in general health topics, hot topics and rare health issues, which took a relatively long time to scan post surrogates and to skim detailed post content.

More details of the findings are as follows:

- In the scanning stage when the system displays a summary result page of post surrogates for the user to select, users with no particular health issue tend to spend the highest amount of time, followed by users seeking for their own issue and seeking for other's issue.
 - Within this scanning stage, users seeking for their own issue tend to have the longest examining duration, followed by users seeking for other's issue and users with no particular issue. This is probably because users seeking for their own health issue need a longer time to match their own conditions with the information in the post title.
 - Users with no particular health issue tend to have the longest skimming duration within the scanning stage, followed by users seeking for their own health issue and seeking for other's health issue. But there was no significant difference between users seeking for their own health issue and seeking for other's health issue. This is probably because users with no particular health issue need more time to skim over the post surrogates to decide what is interesting to read, as they do not have a pre-decided topic. On the other hand, the other two groups need less

time to decide relevance to a pre-decided health issue.

- In the reading stage, when the system displays the detailed post content for the user to read, users seeking for their own health issue tend to spend the highest amount of time in reading the detailed post and replies, followed by participants seeking for others and participants with no particular health issue.
 - Within the reading stage, users seeking for their own issue tend to have the highest examining duration, followed by users seeking for other's issue and users with no particular issue. This is probably because that they need to match their own conditions with the details of the post content.
 - There was no significant difference in skimming duration between users seeking for their own issue and participants with no particular issue, but both of their skimming durations were longer than that of participants seeking for other's issue. The long skimming duration for participants seeking for their own health issue is probably due the need to skim more text to find more details to match since there are many personal details in the post content. The reason for the long skimming duration for participants without a particular health issue is probably the need to cover more content to find interesting information.
- The ratio of examining duration to skimming duration in the reading stage (when reading detailed post content) was found higher for posts judged relevant than for post judged not relevant for all three groups of users. Thus, it is possible to determine whether a post is relevant from the examining and skimming duration.
- A small group of participants with latent health information needs was observed. They

did not specify a particular health issue at the beginning of the browsing session, but indicated in the post-experiment interview that they recalled a health issue relevant to themselves or someone they knew, triggered by information in posts. They were found to use a mixture of case-based and general knowledge related relevance criteria.

Looking at the results in more detail, participants seeking for own health issue focused mainly on the poster's symptoms and history of disease, and the terminology and description of disease, both when scanning post surrogates and when reading detailed post content. These case-based criteria could be considered more detailed categories of the *comparison relevance* category identified by Huang and Soergel (2006). They defined *comparison relevance* as the relevance derived from the similarity between two different cases. People who are seeking for their own health issue often compare their own situation (experiences, feelings, symptoms, and history of disease) with the description of the poster's situation.

Participants seeking for other people's issue focused on the terminology and description of the disease, and type of treatment available. The types of information that users focus on can be considered to be more detailed relevance criteria within the broader category of *content/information* in the framework of Cool, Belkin and Kantor (1993). The category of *content/information* could be interpreted as the factual knowledge of health issue and treatment. When people seek health information for others, they are often not familiar with the details of the patient and have to consider only generic medical information in their relevance judgment.

It could not be concluded that the relevance criteria identified in this study are the only ones

used when browsing a health discussion forum. People also use other relevance criteria implicitly, such as topicality, accuracy, presentation, and authority, but these cannot be determined just from content analysis of text with eye fixation. People seeking for their own issue and for other's issue must find the right topic before they can check and read other details. Participants sometimes check the poster's profile, which suggests that authority is also considered. I expected these implicit relevance criteria to become the focus of attention when they are violated. However, no instance of this was encountered in the study. These criteria might also be important considerations when users decide to actually use the information or adopt a recommendation in practice.

This study chose a particular health discussion forum—HealthBoards.com—as the study platform. It is representative of user contributed content in health discussion forums, as they are similar in structure and content. Hence, the results of this study are very likely to hold with other online health discussion forums. In addition, the results might also be applicable to other types of social media sites (e.g., Facebook groups for various diseases) with lots of user-contributed content. People seeking for their own health issue might look for the Facebook or blog pages of people with the most similar condition (i.e. identify the most similar person rather than the most similar post), as information on Facebook and in blogs is organized by person rather than topic.

In contrast, the results may not be applicable to authoritative health websites maintained by healthcare organizations and government agencies (e.g., Mayo Clinic and PubMed). People seeking for their own health issue on these websites may have difficulty matching their own

condition with the description on the websites. They have to use basic medical terms in searching and making relevance judgments.

The three groups of participants in this study represent participants with different levels of detailed medical knowledge. Participants seeking for their own issue normally have high level of detailed medical knowledge, including their own symptoms, history of disease, feelings, as well as the basic medical terms, causes and treatments. Participants seeking for others normally have low to moderate level of knowledge (depending on the closeness of the relationship to the patient), including medical terms, causes and treatments. In contrast, participants with no particular issue have low level of knowledge. The level of detail probably determines how people make relevance judgment. The people with the highest level of detail will make relevance judgment using detailed symptoms, history of disease, age, and occupation. The people with intermediate level of knowledge will decide relevance using basic medical knowledge. The people with the low level of knowledge will make relevance judgment based on interest since they don't have much to match. Hence, I conjecture that if people seeking for others are quite familiar with the conditions of their family members or friends, they might use case-based matching in their relevance judgments of health information.

The results of this study may be applicable to literature searching and browsing by researchers in academic disciplines. The expert researcher with detailed knowledge in a particular research topic might match detailed information in the published papers with the researcher's research. In contrast, a researcher who is new to the research area may be

looking for basic concepts and theories in the field.

In addition to the case-based and factual/general knowledge-based information criteria (i.e. symptoms, history of disease and treatment), participants in the study occasionally used other criteria, such as age, occupation, and nationality. These criteria depend to some extent on the characteristics of the health problem. For example, some female participants sought information on acne and skin issues. Therefore age and occupation become important relevance criteria in the detailed posts. In contrast, for the topic of mosquito bite, nationality criterion becomes important, as mosquitos and malaria is a major problem in some tropical countries. Hence, these less prevalent relevance criteria vary in importance depending on the type of health issue.

5.2 CONTRIBUTIONS TO RESEARCH METHODOLOGY, AND INTERFACE DESIGN

This study combines the analysis of quantitative data and qualitative data from the eye tracker system. In previous studies, most of the studies focused on one aspect of people's eye movements. Some calculated the quantitative measures, while others focused on the content of fixations.

This study also developed a classification scheme of health information types for content analysis coding. It identified fairly detailed categories of health information that might be found during health information seeking on a health discussion forum. It was derived mostly bottom-up from the types of information actually found on a health discussion forum. The coding scheme can be used with other types of social media.

The eye tracker system was found to be useful in recording eye movements of users browsing a discussion forum, to analyze patterns of skimming and examining behavior. It sheds light on users' mental process of relevance judgment, in addition to information from self-reported data.

This study also provides some implications for the design of more user-oriented health information systems. As it was found that people with different types of use context focus on different types of health information, the relevance ranking by the search function can assign different weights to different types of information, depending on whether the user is searching for self, others or with no particular issue in mind. The user can be prompted to select one of these use contexts when accessing the system. For users seeking for their own health issue, the summary page can display the post surrogates that best match the user's details, if available in the registered user profile. If the user is browsing with no particular topic in mind, the summary result page can display post surrogates that were clicked and viewed by the highest number of previous users, indicating topics of general interest.

A metadata field for posters to indicate the type of health information included in their post can help the system to filter and display posts that better match the use context and health profile of the user.

Finally, I theorize that the results of this study apply to non-health domains, as well as to other types of information sites and systems—that people in general use case-based criteria when assessing the relevance of information to their own issue or task, whereas people use factual-knowledge based criteria when assessing the relevance to other people's issue or task.

People use general-interest based criteria when they have no particular issue in mind. The particular types of information that are considered case-based, factual-knowledge or general-interest will, of course, depend on the particular domain and task. There are several factors that may affect the relative importance of case-based, fact based, and general-interest based criteria:

- How knowledgeable the person is about the domain as well as about his or her own issue or task;
- How important or critical the issue or task is to the person;
- How “fluid” the available information is: whether there is a wide range of information to choose from, and whether the information is customizable or prepackaged.
- In the case of seeking information for another person, how close the relationship is between self and the other.

5.3 LIMITATIONS

This study has some limitations that need further exploration and investigation:

- The participants are residents of Singapore (most of them are Singaporeans and Chinese nationals). This study did not include participants from other ethnic groups and nationalities.
- The participants did not have a critical health problem at the time of conducting the study. People with severe problems may exhibit different relevance judgment behaviors.

- The demographic profile of participants did not cover all segments of society. In particular, the participants were either undergraduate students or had at least an undergraduate degree.
- This study did not consider the influence of human factors, such as personality and attitude to the Internet. As pointed out earlier, two participants were found to have long examining duration due to reading habit developed in childhood. These human factors may have influenced the results of the study.
- The health information used in this study was written in English. The information written by other languages might influence people's eye movements and corresponding relevance judgments. For example, Chinese characters are quite different from English text in size and ways of organization.
- This study used an eye tracker system to provide detailed information on people's eye movements. However, sometimes the system's accuracy is affected when the participants move their heads too much. In a few cases, participants felt a bit tired and moved their bodies to find a better place to sit. Hence the accuracy of people's eye fixations is affected.

5.4 FUTURE RESEARCH DIRECTIONS

This study is just a start of research on health information seeking and browsing on social media by using an eye tracker machine to capture users' scanning, reading, skimming and examining behavior. Future work can be done in the following areas.

The study can be extended to other social media sites, such as Facebook, Twitter, social question and answer sites, blogs and drug review sites. They share some common features with health discussion forums, especially in having user-contributed content, while being different in some aspects including the ways they are organized. Twitter and Facebook arrange posts according to particular users rather than particular topics. The presentation of the text information is also different. For example, Twitter has the limitation of the number of characters. All these features might influence health information seeking behavior and relevance judgment. The study can also be replicated on authoritative health information websites.

More studies can be conducted on health information seeking by proxy—seeking information on behalf of other people. They can be divided into different groups based on the level of detailed medical knowledge they have. Associations between different relevance criteria, and the level and type of health knowledge the user has can be investigated.

The studies can be conducted on different types of health information websites to find out how the structural features of different types of websites influence relevance judgment. The authoritative websites are more formally designed and contain more medical research findings. It may influence relevance judgment when people seek for their own health issue. Some company websites and drug review sites contain visual information such as images and video advertisements for particular drugs or treatments. Visual information may interact with textual information to influence relevance judgment and decision making.

REFERENCES

- Arora, N. K., Hesse, B. W., Rimer, B. K., Viswanath, K., Clayman, M. L., & Croyle, R. T. (2008). Frustrated and confused: the American public rates its cancer-related information-seeking experiences. *Journal of General Internal Medicine*, 23(3), 223-228.
- Balatsoukas, P., & Ruthven, I. (2010). The use of relevance criteria during predictive judgment: an eye tracking approach. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1-10.
- Balatsoukas, P., & Ruthven, I. (2012). An eye-tracking approach to the analysis of relevance judgments on the Web: The case of Google search engine. *Journal of the American Society for Information Science and Technology*, 63(9), 1728-1746.
- Barry, C. L. (1994). User-defined relevance criteria: an exploratory study. *JASIS*, 45(3), 149-159.
- Barry, C. L., & Schamber, L. (1998). Users' criteria for relevance evaluation: a cross-situational comparison. *Information processing & management*, 34(2), 219-236.
- Barsevick, A. M., & Johnson, J. E. (1990). Preference for information and involvement, information seeking and emotional responses of women undergoing colposcopy.

Research in nursing & health, 13(1), 1-7.

Belkin, N. J., & Kwaśnik, B. (1986). *Using structural representation of anomalous states of knowledge for choosing document retrieval strategies*. Paper presented at the Proceedings of the 9th annual international ACM SIGIR conference on Research and development in information retrieval.

Brodie, M., Flournoy, R. E., Altman, D. E., Blendon, R. J., Benson, J. M., & Rosenbaum, M. D. (2000). Health information, the Internet, and the digital divide. *Health affairs*, 19(6), 255-265.

Bruce, H. W. (1994). A cognitive view of the situational dynamism of user-centered relevance estimation. *Journal of the American Society for Information Science*, 45(3), 142-148.

Buscher, G., Dengel, A., & van Elst, L. (2008). *Eye movements as implicit relevance feedback*. Paper presented at the CHI'08 Extended Abstracts on Human Factors in Computing Systems.

Choi, Y., & Rasmussen, E. M. (2002). Users' relevance criteria in image retrieval in American history. *Information processing & management*, 38(5), 695-726.

Chuang, K. Y., & Yang, C. C. (2011). Interaction patterns of nurturant support exchanged in online health social networking. *Journal of medical Internet research*, 14(3), e54-e54.

Chuang, K. Y., & Yang, C. C. (2012). Interaction Patterns of Nurturant Support Exchanged in Online Health Social Networking. *Journal of medical Internet research*, 14(3),

e54-e54.

- Cline, R. J., & Haynes, K. M. (2001). Consumer health information seeking on the Internet: the state of the art. *Health education research*, 16(6), 671-692.
- Cole, M. J., Gwizdka, J., Liu, C., Belkin, N. J., & Zhang, X. (2013). Inferring user knowledge level from eye movement patterns. *Information processing & management*, 49(5), 1075-1091.
- Conley, V. M. (1998). Beyond Knowledge Deficit to a Proposal for Information-Seeking Behaviors. *International Journal of Nursing Terminologies and Classifications*, 9(s2), 129-135.
- Cotten, S. R., & Gupta, S. S. (2004). Characteristics of online and offline health information seekers and factors that discriminate between them. *Social science & medicine*, 59(9), 1795-1806.
- Crystal, A., & Greenberg, J. (2006). Relevance criteria identified by health information users during web searches. *Journal of the American Society for Information Science and Technology*, 57(10), 1368-1382.
- Czaja, R., Manfredi, C., & Price, J. (2003). The determinants and consequences of information seeking among cancer patients. *Journal of health communication*, 8(6), 529-562.
- Erdelez, S. (2004). Investigation of information encountering in the controlled research environment. *Information processing & management*, 40(6), 1013-1025.

- Eysenbach, G., & Köhler, C. (2002a). *How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews* (Vol. 324).
- Eysenbach, G., & Köhler, C. (2002b). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *Bmj*, 324(7337), 573-577.
- Fitzgerald, M. A., & Galloway, C. (2001). Relevance judging, evaluation, and decision making in virtual libraries: a descriptive study. *Journal of the American Society for Information Science and Technology*, 52(12), 989-1010.
- Foster, A., & Ford, N. (2003). Serendipity and information seeking: an empirical study. *Journal of Documentation*, 59(3), 321-340.
- Fox, S., & Duggan, M. (2013). Health online 2013. Pew Research Center, 2013.
- Greisdorf, H. (2003). Relevance thresholds: a multi-stage predictive model of how users evaluate information. *Information processing & management*, 39(3), 403-423.
- Gwizdka, J. (2014). *Characterizing relevance with eye-tracking measures*. Paper presented at the Proceedings of the 5th Information Interaction in Context Symposium, Regensburg, Germany.
- Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, 43(9), 602-615.

- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in cognitive sciences*, 9(4), 188-194.
- Higgins, O., Sixsmith, J., Barry, M. M., & Domegan, C. (2011). A literature review on health information seeking behaviour on the web: a health consumer and health professional perspective.
- Holsanova, J., Rahm, H., & Holmqvist, K. (2006). Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurements. *Visual communication*, 5(1), 65-93.
- Huang, X., & Soergel, D. (2004). Relevance judges' understanding of topical relevance types: An explication of an enriched concept of topical relevance. *Proceedings of the American Society for Information Science and Technology*, 41(1), 156-167.
- Huang, X., & Soergel, D. (2006). An evidence perspective on topical relevance types and its implications for exploratory and task-based retrieval. *Information Research*, 12(1), 12-11.
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52(1), 3-50.
- Kim, J., Oard, D. W., & Soergel, D. (2003). Searching large collections of recorded speech: A preliminary study. *Proceedings of the American Society for Information Science and Technology*, 40(1), 330-339.
- Kim, S., & Oh, S. (2009). 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.
- KISHMOTO, K., & FUKUSHIMA, N. (2011). Use of anonymous Web communities and websites by medical consumers in Japan to research drug information. *Yakugaku Zasshi*, 131(5), 685-695.
- Kummervold, P. E., Chronaki, C. E., Lausen, B., Prokosch, H.-U., Rasmussen, J., Santana, S., . . . Wangberg, S. C. (2008). eHealth trends in Europe 2005-2007: a population-based survey. *Journal of medical Internet research*, 10(4).
- Palsdottir Lenz, E. R. (1984). Information seeking: A component of client decisions and health behavior. *Advances in Nursing Science*, 6(3), 59-72.
- Lesk, M. E., & Salton, G. (1968). Relevance assessments and retrieval system evaluation. *Information storage and retrieval*, 4(4), 343-359.
- Maglaughlin, K. L., & Sonnenwald, D. H. (2002). User perspectives on relevance criteria: A comparison among relevant, partially relevant, and not-relevant judgments. *Journal of the American Society for Information Science and Technology*, 53(5), 327-342.
- Meischke, H., Eisenberg, M., Rowe, S., & Cagle, A. (2005). Do older adults use the Internet for information on heart attacks? Results from a survey of seniors in King County, Washington. *Heart & Lung: The Journal of Acute and Critical Care*, 34(1), 3-12.
- Miwa, M., Egusa, Y., Saito, H., Takaku, M., Terai, H., & Kando, N. (2011). A method to capture information encountering embedded in exploratory Web searches.

Information Research, 16(3).

Pálsdóttir, Á. (2009). Seeking information about health and lifestyle on the Internet.

Information Research, 14(1).

Park, T. K. (1993). The nature of relevance in information retrieval: An empirical study. *The*

Library Quarterly, 318-351.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the EZ Reader model: Exploring

the interface between cognition and eye-movement control. *Cognitive Psychology*,

52(1), 1-56.

Rayner, K. (1995). Eye movements and cognitive processes in reading, visual search, and

scene perception. *Studies in Visual Information Processing*, 6, 3-22.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of

research. *Psychological bulletin*, 124(3), 372.

Rees, C. E., & Bath, P. A. (2000). The psychometric properties of the Miller Behavioural

Style Scale with adult daughters of women with early breast cancer: a literature

review and empirical study. *Journal of Advanced Nursing*, 32(2), 366-374.

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web.

Journal of the American Society for Information Science and Technology, 53(2),

145-161.

Saito, H., Takaku, M., Egusa, Y., Terai, H., Miwa, M., & Kando, N. (2010). Connecting

- qualitative and quantitative analysis of Web search process: Analysis using Search Units *Information Retrieval Technology* (pp. 173-182): Springer.
- Saito, H., Terai, H., Egusa, Y., Takaku, M., Miwa, M., & Kando, N. (2009). How task types and user experiences affect information-seeking behavior on the Web: using eye-tracking and client-side search logs. *Understanding the User-Logging and Interpreting User Interactions in Information Search and Retrieval (UIIR-2009)*, 19.
- Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, 26(6), 321-343.
- Saracevic, T. (1996). *Relevance reconsidered*. Paper presented at the Proceedings of the second conference on conceptions of library and information science (CoLIS 2).
- Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 2126-2144.
- Savolainen, R., & Kari, J. (2006). User-defined relevance criteria in web searching. *Journal of Documentation*, 62(6), 685-707.
- Schamber, L. (1991). *Users' Criteria for Evaluation in a Multimedia Environment*. Paper presented at the Proceedings of the ASIS Annual Meeting.
- Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: toward a dynamic, situational definition*. *Information processing & management*,

26(6), 755-776.

Spink, A., Greisdorf, H., & Bateman, J. (1998). From highly relevant to not relevant: examining different regions of relevance. *Information processing & management*, 34(5), 599-621.

Swanson, D. R. (1977). Information retrieval as a trial-and-error process. *The Library Quarterly*, 47(2), 128-148.

Tang, R., & Solomon, P. (2001). Use of relevance criteria across stages of document evaluation: on the complementarity of experimental and naturalistic studies. *Journal of the American Society for Information Science and Technology*, 52(8), 676-685.

Taylor, A. R., Cool, C., Belkin, N. J., & Amadio, W. J. (2007). Relationships between categories of relevance criteria and stage in task completion. *Information processing & management*, 43(4), 1071-1084.

Tombros, A., Ruthven, I., & Jose, J. M. (2003). *Searchers' criteria for assessing web pages*. Paper presented at the Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval.

Toms, E. G., O'Brien, H. L., Kopak, R., & Freund, L. (2005). Searching for relevance in the relevance of search *Context: Nature, impact, and role* (pp. 59-78): Springer.

Vakkari, P., & Hakala, N. (2000). Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*, 56(5), 540-562.

- Williamson, K. (1998). Discovered by chance: The role of incidental information acquisition in an ecological model of information use. *Library & Information Science Research*, 20(1), 23-40.
- Wilson, T. D. (1997). Information behaviour: an interdisciplinary perspective. *Information processing & management*, 33(4), 551-572.
- Wilson, T. D. (2000). Human information behavior. *Informing science*, 3(2), 49-56.
- Wright, K. B., & Bell, S. B. (2003). Health-related support groups on the Internet: Linking empirical findings to social support and computer-mediated communication theory. *Journal of Health Psychology*, 8(1), 39-54.
- Yang, M., & Marchionini, G. (2004). Exploring users' video relevance criteria—A pilot study. *Proceedings of the American Society for Information Science and Technology*, 41(1), 229-238.
- Zhang, Y. (2012). Consumer health information searching process in real life settings. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1-10.

Appendix 1:

Summary of relevance criteria from previous studies

Schamber (1991) -- 3 categories

• Information (Accuracy, Currency, Specificity) • Presentation (Clarity, Presentation quality, Dynamism, Understandability)	• Source (Reliability, Accessibility, Verifiability)
-

Park (1993) -- 3 categories

• Internal context (criteria pertaining to user prior experience, such as expertise, educational level) • Problem context (criteria pertaining to the motivation of search and so on.)	• External context (criteria pertaining to the searching task and environment, such as purpose of search, stage of search)
-

Cool, Beklin, Frieder, & Kantor (1993) -- 6 categories

• TOPIC (How a document relates to a person's interest)
Defines the topic itself; On/not on the topic; Focus (directly on topic, or not); Part of topic; Treatment (deep/superficial); Important (+/-) • FORMAT (Formal characteristics of the document)
Lists; Diagrams; Statistics; Pictures; Class text; Book review; Title; Introduction; Division into topics • VALUES (Dimensions of judgment - these are modifiers of other facets)
Interest (+/-); Amount (lot/little); Specificity (specific/general); Goodness (+/-); Usefulness (+/-); | <ul style="list-style-type: none"> • CONTENT/INFORMATION (Characterization of what is 'in' the document itself)
Basic concepts; Facts/factual; Explanation; Examples; Definitions; Connections; Description; Reasons; Ideas; Tips; Guidelines; Technical knowledge; Interview; (About) people; Variety; Point of view; Survey; History; Level of detail • PRESENTATION (How a document is written/presented)
Organization; Matter-of-factness; Precision; Writing style; Understandability; Technicality; Scientificness; Simplicity/complexity • ONESELF (Relationship between person's situation and the other facets)
Need; Utility; Desire ('want'); Like; Teaches; Informs; |
|--|--|
-

Age (of document); Entertainment value (+/-); Precision (precise/vague); Bias (+/-); Authority (+/-)	Supports understanding; Use to which document will be put (e.g. references, quotes)
<hr/>	
Barry (1994) -- 5 categories	
Information content The content itself is relevant. <ul style="list-style-type: none"> • User belief and preference Information consistent with user's thought 	<ul style="list-style-type: none"> • User background experience User's experience/history is relevant. • Source of document Metadata is relevant
User situation The information is relevant to user's condition (working, environment).	
<hr/>	
Fitzgerald & Galloway (2001) -- 12 categories	
<ul style="list-style-type: none"> • Interesting • Useful & helpful • Banned idea • Specificity • More is better • Serendipity (relevant to other) 	<ul style="list-style-type: none"> • Specific idea • Specific use • Divergent • Background • Essential (important to this issue) • Prior knowledge (familiarity)
<hr/>	
Toms, O'Brien , Kopak & Freund (2005) -- 5 categories	
<ul style="list-style-type: none"> • System relevance (rank, on-list/not-on-list) • Cognitive relevance (certainty, modified questions) • Situational relevance/utility 	<ul style="list-style-type: none"> • Topicality (aboutness) • Motivational (satisfaction, ease of use, familiarity, perceived time cost) • Completeness
<hr/>	
Savolainen & Kari (2006)	
<ul style="list-style-type: none"> • Ability to understand • Affectiveness • Cost • Currency 	<ul style="list-style-type: none"> • Accessibility • Clarity • Curiosity • Familiarity • Novelty • Security • Time constraints • Usability • Variety
Language <ul style="list-style-type: none"> • Reliability 	
Specificity <ul style="list-style-type: none"> • Topicality • Validity 	
<hr/>	
Kim & Oh (2009) -- 5 categories	
<ul style="list-style-type: none"> • Content (accuracy, specificity, clarity, rationality, completeness, writing style) 	<ul style="list-style-type: none"> • Cognitive (novelty, understandability, effectiveness, solution, feasibility)

- **Information source** (preference, expertise)
- **Socio-emotional** (emotional support, attitude, effort, experiences, agreement, tast, humor)
- **Extrinsic**

Balasoukas & Ruthven (2010) -- 12 categories

• Topicality	• Affectiveness
• Scope	• Recency
• User background	• Serendipity
• Quality	• Format
• Tangibility	• Ranking
• Resource type	• Document characteristics
-