Learning Factors and Determining Document-level Satisfaction In Search-as-Learning

by

Mustafa Abualsaud

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2017

© Mustafa Abualsaud 2017

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

An important use of search engines is as a tool for learning. Search engines help users find learning material and increases their knowledge in various topics. The underlying process of learning while web searching and which documents a search engine should return to enhance the learner's comprehension and learning is a new area of research. In order to build better search engines to supplement the learning process and overall satisfaction, documents the learner searches for should be investigated.

In this thesis, we propose six different factors that may be associated with learning and show which are significant in determining document-level satisfaction. We describe a lab-based user study in which each participant was assigned to a learning task with a pre and post quiz to measure their increase in knowledge after reading the selected documents. Using data collected at different stages of the study, our results indicate that documents with *broadness* of content, as well as *novelty* of information, are significant in determining satisfaction. We also show qualitative results that indicate a broader to more specific ordering of documents content is preferred for easier processing and retention of information.

Our study provides insight into the characteristics of documents learners prefer to read and the order these documents should be presented to the learner, and provides us a better understanding of the learning process that occurs during search-as-learning related tasks.

Acknowledgements

I would like to thank my supervisor Mark D. Smucker for his guidance and support throughout my research, and allowing me to research the topics of my interest.

I would also like to thank my readers, Maura R. Grossman and Gordon V. Cormack for reading my thesis and providing valuable feedback.

I also want to thank my father, Shaker Abualsaud, who passed away during my undergraduate studies. I will always be grateful for his encouragement. His memory will always be with me. My appreciation extends to my mother, Nawal Albayyat, for her patience and prayers, and my brothers and sisters for being supportive and encouraging. Words cannot express my gratitude and appreciation.

Finally, I would like to thank Valbona Bajrami, Malek Naouach, Hicham El-Zain, Abdelhadi Azzouni, Hemant Surale, Aiman Alharbi, Haotian Zhang and all members of the IR lab, for their support during my graduate studies and in my research.

Dedication

This thesis is dedicated to the memory of my late father, to my mother and to my siblings. This journey would not have been possible without you.

Table of Contents

List of Tables ix

List of Figures x

1 Introduction 1

1.1 Organization of the Thesis . . . 3

2 Related work 5

2.1 Study of Learning in Psychology . . . 5

2.2 Information Science . . . 7

2.3 Study of Learning and Technology . . . 9

2.3.1 E-Learning . . . 9

2.3.2 Massive Open Online Courses (MOOCs) . . . 10

2.4 Information Retrieval . . . 11

2.5 Assessment of Learning . . . 11

3 Learning Factors 14

3.1 Methodology . . . 14

3.2 Factors related to learning . . . 15

3.2.1 Readability . . . 16

3.2.2 Understandability . . . 16

3.2.3 Novelty . . . 17
3.2.4 Broadness . . . 17
3.2.5 Detailedness . . . 17
3.2.6 Reliability . . . 18

4 Study Design **19**
4.1 Topics used in the learning task . . . 19
4.2 Study Participants . . . 19
4.3 Search System and User Interface . . . 20
4.4 Tasks and procedure . . . 20
4.5 Quiz Questions . . . 40
4.5.1 Topic 1 Quiz Questions . . . 40
4.5.2 Topic 2 Quiz Questions . . . 41
4.5.3 Evaluating Quiz Answers . . . 41
4.6 Logging of Search Behavior . . . 42

5 Result **43**
5.1 Demographic data . . . 43
5.2 Pre-task result . . . 45
5.3 Learning outcome and learning factors . . . 47
5.4 Satisfaction and learning factors . . . 49
5.5 Understandability and learning factors . . . 58
5.6 User factor analysis . . . 59
5.6.1 Relationships between learning factors . . . 59
5.6.2 Relationship between search and topic learning variables . . . 60
5.7 Qualitative Data Result . . . 62
5.7.1 Saving Documents . . . 63
5.7.2 Document order . . . 65
5.7.3 Explaining the documents ordering . . . 67
5.7.4 Effect of ordering to learning process . . . 69

6 Conclusion **71**

6.1 Future work 72

References **73**

APPENDICES **77**

A Collected Data **78**

B Ethics **82**

B.0.1 Recruitment Poster 82

B.0.2 Ethics Form 84

B.0.3 Conset Form 94

List of Tables

4.1 Demographic questions. 21
4.2 Pre-task questions. 23
4.3 Ordering post-task questions. 31
4.4 Post-task questions and respond type. 36

5.1 Spearman correlation of learning factors with Satisfaction (S1). 50
5.2 Distribution of document satisfaction labels; N=350. 50
5.3 Distribution of learning factors labels on user-saved documents; N=350. . . 51
5.4 Spearman correlation of learning factors with Understandability (LF2). . . 58
5.5 Components loading table for learning factors, ranks, and document satisfaction. 60
5.6 Components loading table for pre-task and post-task questions. 61
5.7 Spearman correlation of learning factors with rank in reading order (RO). . 67
5.8 Spearman correlation of learning factors with rank in final order (FO). . . 67

A.4 Inventory of questions asked to participants. 81

List of Figures

4.1 Demographic interface. 21
4.2 Pre-task interface. 23
4.3 Search interface. 24
4.4 Search interface - list of saved documents. 25
4.5 Search interface - Proceeding to next task. 26
4.6 Search interface - help information. 27
4.7 Ordering task simulated work situation. 28
4.8 Ordering task interface. 29
4.9 Ordering task confirmation. 30
4.10 Ordering questionnaire interface. 31
4.14 Document feedback form where participants label each read documents on each learning factor as well as document satisfaction. 32
4.11 Documents reading task. 33
4.12 Documents reading task help information. 34
4.13 Proceeding from the documents reading task is not allowed before reading the 10 documents. 35
4.15 Post-task questionnaire interface. 37
4.16 Quiz questions interface. 38
4.17 Optional final ordering interface. 39

5.1 Participants answers to demographic question DQ6. 44

5.2 Participants answers to demographic question DQ7. 44
5.3 Participants answers to demographic question DQ8. 44
5.4 Participants answers to demographic question DQ9. 45
5.5 Participants answers to pre-task question PR1 for Topic 1 (Oil Spills). . . . 45
5.6 Participants answers to pre-task question PR1 for Topic 2 (Open Data). . . 46
5.7 Participants answers to PR2 for Topic 1 (Oil Spills). 46
5.8 Participants answers to PR2 for Topic 2 (Open Data). 46
5.9 Participants answers to PR3 for Topic 1 (Oil Spills). 47
5.10 Participants answers to PR3 for Topic 2 (Open Data). 47
5.11 Coders coding interface. 48
5.12 Readability Level and Satisfaction Level Frequency. 52
5.13 Understandability Level and Satisfaction Level Frequency. 53
5.14 Novelty Level and Satisfaction Level Frequency. 54
5.15 Detailedness Level and Satisfaction Level Frequency. 55
5.16 Broadness Level and Satisfaction Level Frequency. 56
5.17 Reliability Level and Satisfaction Level Frequency. 57
5.18 Parallel Analysis Scree Plot. 59
5.19 User factor analysis biplot of pre-task and post-task variables showing clustering of users points (black) and correlations between our used variables as vectors (red). Two vectors with close lengths and small angle between indicate two highly correlated variables. 62
5.20 Participants placement of Wikipedia documents in their reading order (RO) and final order (FO). 64
5.21 Participants labeling of reliability level of Wikipedia documents 65
5.22 Participants labeling of satisfaction level of Wikipedia documents 65
5.23 Percent of change in document positions at each rank. A high percent at one rank indicates that many documents at that rank in RO are no longer in the same position in FO. 66
5.24 Participants answer to whether the ordering has affected their learning process (OQ3). 69

5.25 Participants answer to how effective was the ordering towards their learning process (OQ4). 69

Chapter 1

Introduction

Information retrieval (IR) systems are viewed today as tools to retrieve information and satisfy the users' information needs. Users may use these tools to interact with the information content and gain knowledge and learn more about a certain topic.

To build search engines focused on supporting deeper learning experiences requires a better understanding of user behavior during search-as-learning tasks. Collins-Thompson et al. (2016) provided insight on what query strategies support human learning and what search behaviors are correlated with learning outcomes. In order to determine what affects the learning process at a deeper level, it is necessary to understand how the content of a document is written, presented, and communicated to the learner. a document's content can be considered by the user as: difficult to read, hard to understand, too specific, or providing little or no new knowledge. These factors can affect the way information is processed and retained, and the overall user satisfaction of a document.

An important topic in learning theory is the sequence of instruction provided to the learner (Mayer, 1977). Providing the learner with the appropriate sequence of instructional messages can make hard to comprehend topics easier to learn and vice-versa. In a learning-related task, an ideal ordering of documents should not contain much redundant information to a user's prior knowledge and should reflect a clearly articulated sequence which can heighten and accelerate a user's learning progression.

Towards the goals of determining how the content of documents affect user satisfaction, and understanding what sequence of information is preferred for improving users' learning process in web search, we conducted a lab-based user study of a search system with which participants completed a learning task. We propose six learning factors that are at the document and user-level and may influence learnability. Factors regarding the content of

the document are: 1) broadness of the information 2) level of depth and details, and 3) reliability level judged by the user. User-based factors are: 4) readability level, 5) easiness to understand information, and 6) novelty of the information in relation to users' prior knowledge. We further explain each factor and why we choose them in Chapter 3.

In our user study, participants use our search engine interface to search and save ten documents regarding a topic they are assigned to learn. We inform participants to choose ten documents they are interested in further reading and they feel are useful for learning the topic. After saving the ten documents, we ask the participants to re-order the documents in the order they prefer to read the documents in, and will allow for a more effective and meaningful learning process. We ask the participants to write an explanation on the reasoning behind the order they have specified, and to justify how such an ordering would help improve their learning process. After participants read each of their ten documents, we collect their document-level satisfaction judgment, and their labeling of each of the proposed learning factors using a five-point Likert scale (1=Not at all, 5=Extremely). After reading the ten selected documents, in the order in which they have specified, participants grade the effectiveness of their ordering in terms of learning and provide feedback on how effective the ordering was in helping them learn the topic better.

Using quantitative and qualitative data collected at various stages of the study, we aim to address the following research questions:

RQ1 : What learning factors are significant in determining user satisfaction at the document-level?

RQ2 : How do learning factors correlate with user's understandability of the document?

RQ3 : What behaviors and reasons do learners exhibit when ordering a set of documents to prompt self-learning?

RQ4 : What documents do learners consider useful for learning when using a search engine?

The importance of search as learning has been indicated at SIGIR 2016 search-as-learning workshop (Gwizdka et al., 2016), IIiX 2014 search-as-learning workshop (Freund et al., 2014) and SWIRL 2012 workshop (Allan et al., 2012). However, only limited research has been done in the information retrieval community to further make advances in understanding learning as a search task, and build search systems with a focus of supporting human learning. If we are able to gain a deeper insight on the underlying variables

that affect the learning process in web search, we are further on our journey towards a better understanding of users during learning-related tasks.

Using our collected data from our user study and our analysis of this data, we have found that:

- Broadness of content and the novelty of information in a document in relation to users' prior knowledge can influence the learners' satisfaction judgment of the document.
- The number of documents that are judged as "*Moderately*" novel or higher correlate with the learners knowledge gain score.
- There is indeed a recognized ordering preference that participants believe is effective towards their learning experience. Many participant showed interest in starting with documents that are broad and provide a high level view of the topic before proceeding to more complex material.
- Many participants have mentioned that their ordering of documents have allowed a clear and effective progression of information. Starting with documents providing basic definitions and understanding of the topic before going into more detailed information allowed for more effective comprehension.
- Participants have indicated interest in Wikipedia articles as their source of high-level knowledge and for providing a general introduction to the topic before delving into harder learning material.

1.1 Organization of the Thesis

The remainder of this thesis is organized as follows:

In Chapter 2, we discuss related work to the concept of learning using the web or technology. We discuss learning in psychology, information science and the latest research in information retrieval addressing learning as the main goal of the search process. We also discuss assessment methods for coding different type of student assessment questions to numerical values and the benefits of each type.

In Chapter 3, we discuss the six proposed learning factors that may affect the learning process in web. We also discuss their definition and why we have considered these factors in our user study.

In Chapter 4, we discuss our study design in detail, the two topics used in the study, the quiz questions for both topics and the logging of search behavior. We also discuss who are the study participants, and the search system and user interface participant to complete their tasks.

In Chapter 5, we discuss the result of our analysis. We show the relationship between learning outcome and the proposed learning factors, the relationship between document satisfaction and learning factors, understandability and learning factors and our factor analysis result to investigate the relationship between the learning factors themselves. We also show qualitative data result collected at different stages of the study to better understand the behaviour of learners during a learning-related task, how learners rank documents and the effect of documents ordering to the learning process.

Finally, in Chapter 6, we conclude our work and discuss potential future work that can further investigated to better design and implement search systems with focus on improving human learning.

Chapter 2

Related work

In this chapter, we review related work in psychology, information science (IS), search-as-learning in information retrieval (IR), and learning using web technology. First, we start by reviewing early studies in learning and memory in the psychology field. We then summarize previous work in IS discipline regarding study approaches and learners personality traits, both within search settings and beyond. We also review some of the forms of learning that is currently being used through the use of web technology. We review search-as-learning and related work of learning in the IR field. Finally, we describe assessment methods common in education theory that can help code answers to numerical scores.

2.1 Study of Learning in Psychology

There has been a considerable amount of research focused on learning theory and how individuals process and absorb information. Since 1975, psychologists and educators have shown interest in how humans acquire new knowledge in different areas such as child learning and teaching.

Studies on learning date back to 1885 with Ebbinghaus (2013) pioneering research on memory, in which he describes an experiment on the process of learning and forgetting. One of Ebbinghaus's groundbreaking results is the forgetting curve that describes the information loss of what an individual has previously learned, over time. Ebbinghaus claim is that the process of forgetting learned information with time applies to humans, but the more the newly learned information is being reviewed or recalled, the more time it takes to forget such information. Ebbinghaus also described a learning curve in which

he shows learning can be improved with experience. His findings came as a results of an experiment he was the only subject of. In the experiment, he introduced *nonsense syllables*, also known as the CVC trigram. A nonsense syllables is a set of consonant-vowel-consonant combination of letters that supposedly have no meaning in English. Ebbinghaus made more than 2000 nonsense syllables that he attempted to recall. After memorizing the nonsense syllables, he gave himself retention intervals ranging from several minutes to several days, and tested his memory by trying to re-learn the syllables counting how many syllables he was able to recall for each interval.

Although Ebbinghaus' results are significant, there are still concerns about his research method. Particularly that he was the subject of the experiment and he may not be considered as a typical learner, and whether nonsense syllables learning can generalize to learning different materials. Nonetheless, his findings exerted learning theory research and academic motivations in addressing important challenges in learning such as how does learning occur and transfer, and the role of memory in learning.

Wittrock (1974) work was also one of the early work in learning in psychology. Wittrock presented his model of how knowledge is being actively constructed by the learner by integrating cognitive process, prior knowledge, knowledge transfer, and generating in human learning. Central to his model is the notion that learners are able to actively construct their own cognitive meanings to the new learned material as a result of their prior knowledge and experiences in the material. The learner's self-constructed meanings represent the learner's comprehension of the topic. Wittrock's model predicts that comprehension of text is easier when the text includes semantic retrieval cues that enhances the learner's recall of relevant information. To test the model's claimed predictions, the Wittrock suggested two methods and conducted user studies to validate each method. According to the model, high-frequency words are associated with the recall of relevant information or experiences that in result facilitate the process of construction of meaning of the text. An experiment by Marks et al. (1974) with sixth grade students tested this method and found that reading comprehension was statistically significant when replacing one or two low frequency words with their high frequency synonyms. The other method was using familiar stories as retrieval cues to help learners recall relevant information. Wittrock and Carter (1975) study with sixth grade students found that by providing similar stories to the students, it facilitate the students' process of learning the definitions of unfamiliar vocabulary and overall comprehension of text.

2.2 Information Science

Different students have different *studying approaches* for learning tasks Heinström (2005). Heinström defined study approaches as different individual differences in the way students approach a learning task and proposed three pervasive studying approaches: *deep*, *surface*, and *strategic* types. A *deep approach* involves more time and active interest with intention to analyze and understand material A *surface approach* is faster and more passive in learning where learners would accept ideas without further analyzing them. A *strategic approach* uses time management effectively and involves more intentions and motivations to excel in tasks.

Previous research by Heinström (2006) investigated different information seeking patterns on students in various disciplines, and whether they correlate with subjects personality traits, studying approach and their discipline. Using a survey test comprised of 18 questions handed to several students in different areas, a factor analysis was used to find underlying relationships between personality traits and studying approaches. Heinström found that subjects' inner traits are more influential than their discipline, and that in exploration search, a broader scanning style is more typical, whereas specific-information seeking involve fast-surfing and deep-diving studying approaches. The study however, is more focused in "*traditional*" information environment and not an Internet-based information seeking.

Ford et al. (2001) is one of the studies from Information Science that investigated individual user differences in Internet searching. In particular, students cognitive styles, level of prior experience, Internet perceptions, study approaches, age and gender. These individual difference were measured using different questionnaires commonly used in the information science discipline, but devised for their particular study. Subjects participating in the study were presented with a simulated scenario and were asked to find relevant information using a built system which interfaced with AltaVista search engine. Each participant was assigned a relevance score based on the results of their queries. The measure of relevancy used was a simple dichotomous classification, where the retrieved list of documents were classified as relevant if it included material that illustrated some solution to their given simulated solution. The relevance judgment was based on a rerun of the queries made by the participants, after they have completed their study. The collection of documents classified are subject to a rapid and constant change, thus the relevancy scored assigned to participants may have been under or over estimated. An interesting finding related to participants study approaches was that poor time management and fear of failure was linked to low relevance scores. The study focus was on relevancy of information and did not assess their effect on learning.

An interesting work by Freund et al. (2016) investigated digital reading behaviour and whether the type of reading interface can influence user comprehension. In particular, whether or not presenting web documents in their plain text form helps with comprehension more than presenting documents in their original form with all the accompanying graphics and design. To address this question, Freund et al. conducted a user study with 41 university students participants. To measure comprehension, Freund et al. used a set of *micro-structural* and *macro-structural* questions. Micro-structural questions consisted of factual recall true-or-false questions and questions measuring conceptual understanding at the sentence level using the Sentence Verification Technique (SVT) (Royer et al., 1987). Each SVT question consisted of 12 sentences two of which are correct and represent the semantic of the text. Participants were asked to choose all sentences that represented the text. Macro-structural questions asked participants to choose three summaries from a set of six given summaries that most represent the semantic themes of the text. Freund et al. found that participant who were given the plain-text of documents achieved higher comprehension outcome but resulted in a longer reading time than people who read documents in their original format. Freund et al. (2016) suggest that designers of comprehension-centered search systems should allow users the option to display content of documents in plain text format to allow for more comprehension and less distraction of non-related information.

One of the major and influential frameworks in information science is Kuhlthau's Information Search Process model (Kuhlthau, 2004). Kuhlthau's model differentiate between two information seeking aspects: providing the source of information and providing the guidance for the construction of meanings process. Intervention support in source-related information seeking helps the user access information, whereas process-related intervention helps with the learning process. Kuhlthau's model is comprised of six stages centered around the user's experience in the process of information seeking: *initiation*, *selection*, *exploration*, *formulation*, *collection*, and *presentation*. In the *initiation* task, users recognize a need for information and lack of understanding in the topic. Users in this task often contemplate the problem and discuss possible approaches or topics they should pursue to fill the gap in knowledge. *Selection* task is where the user identifies and select the topics that need to be pursued to fulfill the information need. In the *exploration* task, users explore different information and may feel uncertain and confused as a result of their inability to precisely express the correct terms to find the right information to explore. Kuhlthau considers this as the most difficult task in the model as multiple rounds of this task might occur and due to increase of the user uncertainty, confidence in finding the right information may decrease. *Formulation* task involve feelings of decrease uncertainty and an increase in confidence and focus. At this stage, construction of meanings take place and

information become more clear to the user. *Collection* task is to gather the information related to the focused topic, while feelings of confidence continue to increase as well as interest in the topic. Finally, the *presentation* task to conclude the search and present or use the findings to fulfill the information need. Feelings of satisfaction or disappointment are common in this task depending on the success or failure of the search process.

2.3 Study of Learning and Technology

Technology has integrated into the lives of many of people and in our modern education, and has been shown effective in facilitating learning through different technology-based applications such as E-Learning platforms. With search engines, students can now search and access scholarly articles, books, university lectures, or even enroll in Massive Open Online Courses (MOOCs) offered from top universities. In this section, we describe some of these technology-based systems and show their potential in fostering learning.

2.3.1 E-Learning

Clark and Mayer (2016) define E-learning as *instruction delivered on a digital device (such as a desktop computer, laptop computer, tablet or smartphone) that is intended to support learning.*

Computers allow for a very flexible representation of media options and support different media elements such as text, static and animated graphics, videos and audio. With these elements, computers can offer a highly immersive and interactive environment that can be tailored to provide learning opportunities that cant be achieved without a digital environment.

Zhang et al. (2004) investigated the recent advances in E-learning technology and whether E-learning environment is more effective in supplementing learning than the traditional classroom environment. Zhang et al. indicated several advantages of traditional learning environments over E-learning, such as immediate feedback from instructors. E-learning, on the other hand, can be available to a larger or even a global audience, is self-paced and centered around the learner. To understand which learning environment is more effective, Zhang et al. implemented an E-learning platform with many capabilities and followed a set of principles to ensure that the platform gives flexible control to the learner in terms of content and style of learning, allows multimedia integration such as video lectures or PowerPoint slides, and provides interactivity such as providing learners with the

ability to ask and receive answers to questions in real-time. Zhang et al. conducted their experiment using their implemented E-learning platform and recruited English-speaking university students from different majors. Zhang et al. found that the performance scores of subjects in the E-learning group were significantly higher than subjects in the traditional classroom group, and explained that due to the nature of the traditional classroom environment, lectures are often sequential and do not offer much interactiveness.

E-learning can be a promising alternative to traditional classroom settings due to its ability to cover a larger audience yet be personalized and be learner-centric. By enhancing the interactiveness and personalization, more learners are engaged which may improve their knowledge acquisition and overall satisfaction with the learning experience.

2.3.2 Massive Open Online Courses (MOOCs)

Massive Open Online Courses have attracted many students around the world and has been an important development of online education the past few years.

In 2011, as an experiment to extend knowledge and skill around the world, Stanford University in California offered a free online course of one of its artificial intelligence (AI) courses taught by leading experts in the field of AI (Waldrop, 2013). Although registered online users were informed that they won't receive any university grade or credits, the course has attracted thousands of students around the would who have registered and completed the course. Many companies such as Coursera [1], which was co-founded by Stanford University professor Daphne Koller, and EdX [2] have realized the impact of MOOCs and have then been competing to offer more courses and partnering with more universities.

Koller have incorporated insight of previous research (Craik and Lockhart, 1972) in the Coursera platform. Craik and Lockhart show that passive listening to lectures reduces learning effectiveness, thus Koller have broken down content of university lectures to short 8-10 minute videos where people can pause and answer related question to increase engagement between students and the platform. Online discussion forms have also been Incorporated to allow interactions and collaborations between other online users.

[1] http://www.coursera.org
[2] http://www.edx.org

2.4 Information Retrieval

Although the importance of search-as-learning have been indicated in different previous workshops (Allan et al., 2012; Freund et al., 2014; Gwizdka et al., 2016), there are still a little research done to fully understand learning as it occurs during the search process and can we build better learning-centered search systems.

Kim et al. (2012) used automatic text classifiers that uses reading level and topic metadata to profile the level of "expertness" of web documents, but did not explicitly measure gain in knowledge of users when the profiling is in effect. It is also unclear whether learners consider the reading difficulty of a document as an important role in influencing their overall learning.

Collins-Thompson et al. (2016) work was the first in the information retrieval field that assessed learning outcomes in web search. In a lab-based study, they investigated measures and indicators that demonstrate learning experience in web search and whether different query strategy methods can support human learning. They have compared learning outcomes in three different query conditions: single query, multiple queries, or multiple queries with *intrinsically diverse result*, which is based on the author's previous work (Raman et al., 2013) on providing multiple subtopics covering the user's query topic and providing query suggestions for the subtopic. In their study, they have introduced a pre-task and post-task knowledge quizzes to measure the gain in knowledge in different query conditions before and after the searching task. The questions asked in the quizzes were written type questions that aim to cover different taxonomies in learning: remembering, understanding, applying, evaluating and creating. Participants written answers were then coded using a carefully developed scheme based on Bloom's revised learning taxonomy, Bloom (1956). Collins-Thompson et al. found that their multiple queries with intrinsically diverse result gave a larger advantage over other query conditions in terms of users knowledge gain. Our work builds upon that of Collins-Thompson et al. (2016) and uses the same learning topics, quizzes and their coding scheme.

2.5 Assessment of Learning

Methods involving assessment of learning have been widely used by instructors to assess students knowledge level. However, there has not been much research on evaluating learning in the context of search (Freund et al., 2016), until recently after a series of "Searching as Learning" workshops and publications. Freund et al. described several models and theories that are commonly used in work on searching as learning, such as assessment methods

and learning measuring methods. Schunk (1996) also described different ways that are currently being used by researchers and practitioners to assess and measure learning. In the remaining of this section, we describe summaries of these methods and how they are being used.

Direct observation Direct observation methods employ direct observations where an instructor observes his students behavior and procedure for completing a learning task to assess their knowledge. These methods work best in environments where the observations and behaviors can be specified and matched to a standard metric for evaluation. Two problems that can occur with direct observation methods is that it only focuses on what can be observed by the learner, thus disregarding the cognitive processes underlying the actions, and its inability to assess actual learning in the absence of certain actions the observers is waiting for.

Written responses Written response are widely used in tests, quizzes, homeworks and reports. These methods assess the level of knowledge in the topic to be learned, in which the instructor usually follows an instructional rubric to assess learners written performances on their learning tasks. Deciding whether learners have acquired some knowledge is done by measuring the difference in scores of a *pretest*, a test given prior to the learning task where the instructor assumes learners have little or no knowledge in the topic, and the follow-up test that proceeds the instructional unit. In our paper, we use written response as our method of assessing learning, and Bloom's Taxonomy of Education Objectives (Bloom, 1956) as the framework for developing questions that identifies a set of progressively complex learning objectives (Remembering, Understanding, Applying, Analyzing, Evaluating and Creating), and to design coding schemes to code written response to numeric scores.

Oral responses Oral response are very common learning assessments in schools, in which the instructor would call on students to answer their question and assess their oral responses. Instructors would then decide if learners show lack of understanding based on their answers. There may be problems with this method as anxiety about speaking or language difficulty often arises, and thus may not allow the the learners to completely reflect their thoughts.

Rating by others Another way to assess learning and knowledge that is commonly used in peer-reviewed academic work is to rate the subject's quality of work by different individuals or raters who are knowledgeable in the topic. An advantage of this method is

that it can be more objective and it reduces biases in the final assessment. This method can be expensive and time consuming for all raters.

Self-reports Self-reports are the learners assessments about themselves in the topic they are learning. This method can take different forms: self-score, questionnaires, interviews and think-alouds. Each of these forms have their own advantages and disadvantages. The choice of which form of self-report should be match with the type of learning task the learner should complete and the purpose of the assessment.

Chapter 3

Learning Factors

In the earlier chapter, we have discussed how web search engines provides us with convenience and efficiency in retrieving documents. Nonetheless, the collection of documents accessible through Internet search engines are large and is subject to rabid and sometimes unreliable changes. Providing incorrect information to users may mislead them into believing that the information is correct or arises confusion. At the same time, documents having redundant information and contributes no further knowledge to the user may slow down their learning process and affect their overall satisfaction with the documents and/or the search system. Similarly, providing documents with many unfamiliar words or advance vocabulary to learners who are unfamiliar with the topic is inconvenient for the learner. This instigates the necessity of further research to understand the fundamental factors that can influence learning and user satisfaction while web searching for learning purposes.

In our study, we propose different learning factors that are associated with learning and the content of the retrieved documents and user characteristics.

3.1 Methodology

The field of educational psychology has provided us with good understanding of the nature of the learner and of learning from both a cognitive and behavioral perspectives. As a result of much research done in this field, a large number of variables related to learning have been identified (Wang et al., 1990); From variables associated with the learning environment, such as the school size or its culture, to more specific individual students characteristics, such as their motivation level or cognitive ability.

Because our concern is with finding factors that may influence the learning process during web-search, we focus on investigating variables that are associated with the results returned by search systems, and those that are specific to the system's users. For this reason, we propose six learning factors and analyze which factor affects user satisfaction at a document-level, and investigate the importance of these factors in a rank order.

3.2 Factors related to learning

Commercial search engines can be seen as a tool that helps learners find documents related to a topic of interest. With the result returned by these systems, learners can select documents to read and process their content for comprehension and acquiring knowledge. The underlying process that starts from reading a document's content to perceiving and comprehending its information is the main focus of our study. If we are able to understand this process and how it can have an impact on the learners' experience, we will have a better understanding of what documents search systems should return when the aim is to improve human learning.

Jac J. W. Andrews (2015) address various differences between students including their cognitive and meta-cognitive abilities, and stresses the importance of students' diverse cognitive abilities and challenges to education planners. We suggest three learning factors that are user-specific: Readability, Understandability, and Novelty of documents.

We define document-specific learning factors to be associated with a document content and how it relates to the document's main general topic. Content writers who write educational or informative web material may choose to write their content in a way such that it covers many subtopics, provides little or extensive details in a subtopic, or gives accurate information supported by evidence or reliable sources. We suggest three document-specific learning factors: Broadness, Detailedness, and Reliability of information presented in documents.

We briefly describe each factor (labeled **LF1** to **LF6**) below:

- **Readability (LF1)**: The document language difficulty.
- **Understandability (LF2)**: The difficulty level of understanding to the reader.
- **Novelty (LF3)**: The amount of new information presented in the document relative to the learner's prior knowledge.

- **Broadness (LF4)**: The breadth of the information in relation to the document's main topic.

- **Detailedness (LF5)**: Level of details presented in the document.

- **Reliability (LF6)**: The level of reliability and accuracy of the information presented in the document.

In the next subsections, we explain our reasoning behind choosing these learning factors and why they may influence the process of acquiring new knowledge in web search.

3.2.1 Readability

A natural and critical aspect in supplementing learning during web-search sessions should be the language readability of documents returned by search systems. Readability of text may affect the learner's cognitive processes and therefore compromise the learner's ability to benefit from the text. The content readability level can be judged differently by people because not everyone has the same reading ability: content can be judged as highly readable or not readable at all. Experts in the field of their search may be more satisfied with documents showing advanced reading level that matches with their level of knowledge. Collins-Thompson et al. (2011) used reading-difficulty as a feature to a ranking algorithm and showed that reading level can be used to improve the relevance of web search results, but readability has not yet been assessed for its effect on learning. In this work, we focus on readability of documents content as judged by the learners themselves.

3.2.2 Understandability

Understanding instructional content to form a mental model is an important step in learning and acquiring new knowledge. Although understandability may be seen as a strong effect of readability level, there could be other factors that influence a learner's understandability of a document. For example, its content may be spread out across the document such that it is hard to extract useful information, or contains many distracting material leading the user to be unable to process its information effectively. The layout of the page can also be considered an affecting component. Studies have shown that web pages with many out links make it difficult to extract relevant information (DeStefano and LeFevre,

2007), and a non-linear presentation of text can increase cognitive load and decrease knowledge acquisition (Zumbach and Mohraz, 2008). We propose understandability as a learning factor and study its correlation with other factors.

3.2.3 Novelty

Methods of search diversification have been studied by many researchers in the information retrieval community (Clarke et al., 2011, 2008). In the case of learning a fairly broad topic, the task can be decomposed into multiple subtopic learning-tasks that as a whole cover intrinsically diverse aspects of the main topic. With the amount of relevant documents easily accessible using commercial search engines, it is common to find documents that share the same information or have very similar content. Because documents may contain partial or fully duplicate information that has already been learned, learners may search for information they are not familiar with to provide them with further knowledge in the topic. We consider novelty as a factor and investigate if it correlates with with learning outcome.

3.2.4 Broadness

A common teaching strategy instructors perform when teaching in a class-room setting is to start with a generalized overview of the topic before moving towards detailed material. This technique allows the learners to understand the material by creating and identifying relationships amongst the general concepts and the detailed to-be-learned concepts. Learners with little or no prior knowledge in a topic may search for documents covering a range of subtopics before exploring more detailed material. We investigate whether learners try to save broad documents and where they place such documents in their preferred reading list.

3.2.5 Detailedness

Learners with extensive prior knowledge or experts in the area of their search may experience the need to not only find relevant documents, but documents offering a higher level of detail that matches with their current level of knowledge. Similarly, learners that are slightly familiar with the topic and want to further expand their knowledge may search

for content offering more details than the typical overview-type documents. We investigate if detailedness of a document content is significant in determining document-level satisfaction.

3.2.6 Reliability

A large amount of information is easily accessible in the web. With the number of increasing tasks where content needs to be reliable and accurate, ranging from the academic, legal, medical, and other fields, trustworthiness of the information and how it effects learning is worthy of exploring. We include reliability as a learning factor and investigate its importance in learners' preferred ordering of documents.

Chapter 4

Study Design

In the following section, we explain how we designed our study to be aligned with our research questions, the collected data from participants, and at what stages of the study our data was collected.

4.1 Topics used in the learning task

We use the same two topics used by Collins-Thompson et al. (2016). Participants were assigned and asked to learn one of the two topics. Both topics are commonly discussed in major newspapers and do not require extensive prior knowledge in the area to start learning about. Participants were assigned a single topic in an alternating manner.

Topic 1: Oil Spills A scientific topic on environmental issues, which is unlikely for participants to be domain experts in but involves more concrete facts and descriptions.

Topic 2: Open Data A general topic on governments and public information, which participants may have little prior knowledge about but is more general and have broader descriptions.

4.2 Study Participants

Recruitment posters were posted at different areas and departments in our university (B.0.1). 36 subjects who are affiliated with our university participated in the study, however one subject was removed due to incomplete data collection, yielding a total of 35

participants that we use for our data analysis. The study was performed in five sessions, with a maximum number of ten participants per session. Both graduate and undergraduate students participated in the study. Since language proficiency can affect learning, we required all participants to be fluent in English. We also required participants to be familiar with using search engines and basic computer usage. The participants' age varied from 18 to 30, with average of 21.5 (SD = 2.85). Out of all participants, 17 were male and 18 were female, with 13 graduate and 22 undergraduate students.

The majority of the participants indicated that they often search the Internet to learn about a new topic and that search-engines were their primary destination to search and learn a new topic.

The study was tested with six pilot subjects. The purpose of the pilot study was to ensure the web interface is collecting the required data and for appropriate risk mitigation. The study took from an hour to an hour and a half to complete and participants were compensated $20 as remuneration for participating. The final count of participants assigned to topic 1 (Oil Spills) is 18, and 17 for topic 2 (Open Data).

4.3 Search System and User Interface

The search system was developed using Python Django web framework and was hosted on a web server that participants had access to from a private computer lab at our university. In order to limit distractions and noise, the computer lab was reserved and only participating subjects and the study coordinator were allowed in the room during the study. The retrieved documents returned by the system are provided by the Bing[1] API. Figure 4.3 shows a screen-shot of the search result interface used for our study.

4.4 Tasks and procedure

Before participants can proceed with their assigned tasks, we ask them to sign a consent form and complete a basic demographic questionnaire. Figure 4.1 shows the interface used to fill in the participants demographic information. Table 4.1 shows all the questions collected in the demographic questionnaire.

[1] http://www.bing.com

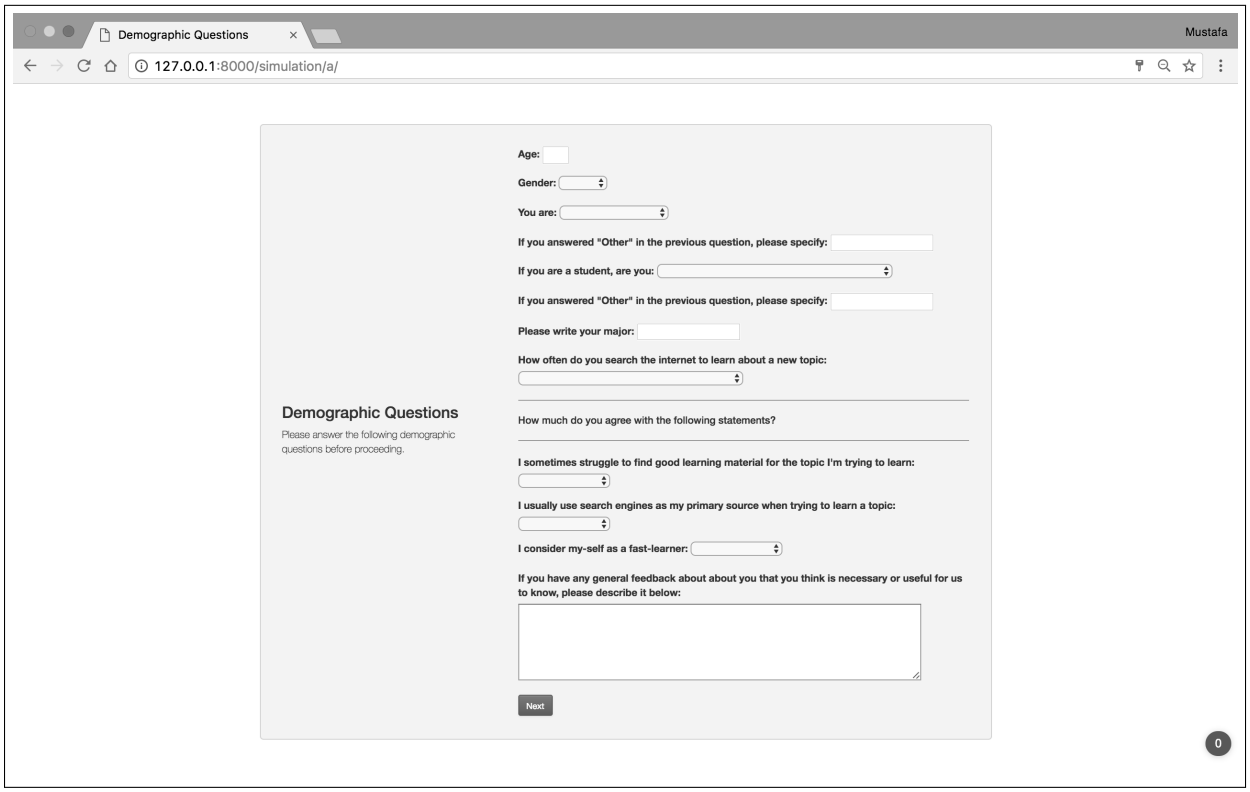


Figure 4.1: Demographic interface.

Question
DQ1: Age
DQ2: Gender
DQ3: Student Degree Level
DQ4: Faculty
DQ5: Major
DQ6: How often do you search the Internet to learn about a new topic?
DQ7: I sometimes struggle to find good learning material for the topic I'm trying to learn
DQ8: I usually use search engines as my primary source when trying to learn a topic
DQ9: I consider my-self as a fast-learner
DQ10: General feedback from the participant

Table 4.1: Demographic questions.

We then show a graphical tutorial on how to use the search system interface and what they are expected to perform at each task. Following the tutorial is a short quiz to test participants understanding of the study expectation and the correct use of the system. To ensure that participants have understood the tutorial and what they are asked to preform in the study, we do not allow participants to proceed until they answer all tutorial quiz questions correctly.

To make sure that the study depicts realistic situations and to develop interest in the participants point of view, we have constructed *simulated work task situation* (Borlund, 2000) at different parts of the study that align with the topic to be learned a participant.

Next, we describe each part of the study in the order participants undergo, along with our constructed simulated work situations. Participants were not allowed to return back to a task once it is completed and were only allowed to do each task consecutively. Throughout the thesis, we label each collected data item with an ID in bold font.

Pre-task: Participants were given a simulated work situation depending on which of the two topics (see Section 4.1) they were assigned to learn:

> "You are attending a university course on *Environmental Issues/Governments and Public Information*. One of the topics covered in the course is on *Oil spills/Open Data*. Your instructor told you to search and learn about the topic before attending the next lecture. Since this is the first time the instructor teaches the course, the instructor would like your help in providing ten good documents for learning the topic, and would appreciate your help in evaluating each document"

Figure 4.2 shows an example of the interface at the pre-task stage. Before completing the pre-task, we asked the participants list of question to measure their interest and familiarity of the topic. The complete list of questions asked at this task is in Table 4.2.



Figure 4.2: Pre-task interface.

Question	Response type
PR1: How much do you know about the topic?	1=Nothing 5=Very familiar
PR2: How difficult do you think it will be for you to learn the topic?	1=Very Easy 5=Very Hard
PR3: How interested are you in learning the topic?	1=Not at all 5=Very
PR4: Please write what you know about this topic with 3-5 sentences.	Written answer

Table 4.2: Pre-task questions.

Task 1: The interface used at this task was similar to a web-search engine interface (Figure 4.3, 4.4, 4.5 and 4.6). Each document in the search engine result page (SERP)

had a corresponding link next to its snippet to save the document to the participant's list of saved documents. Participants could enter as many queries as they want and save documents from different query results at different rank pages. We allowed participants to open any of the SERP documents to help with their decision in saving the document, but we asked the participants to only skim through the document's content at this stage of the study and save the document if they feel it is useful in learning the topic. We made these expectation clear during the tutorial and as a part of the tutorial quiz. Proceeding to the next task was allowed after the participant save their ten documents.

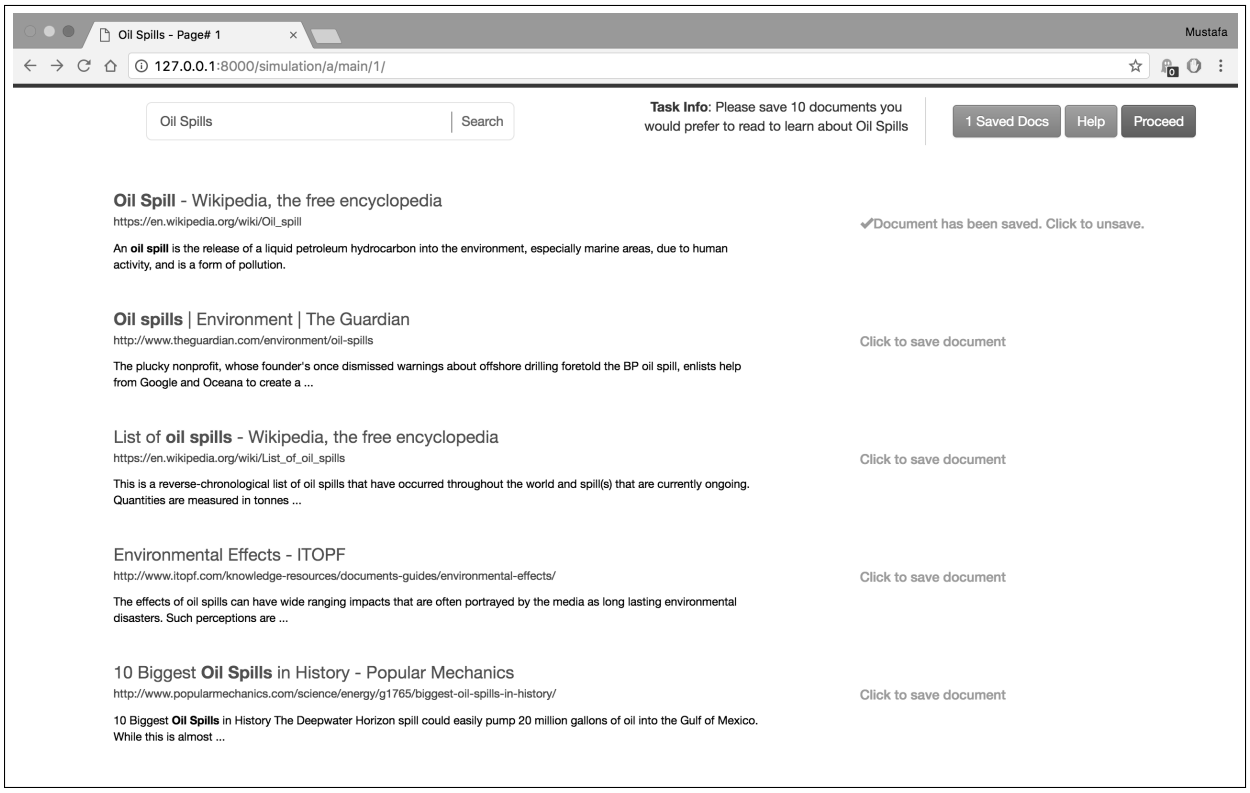


Figure 4.3: Search interface.

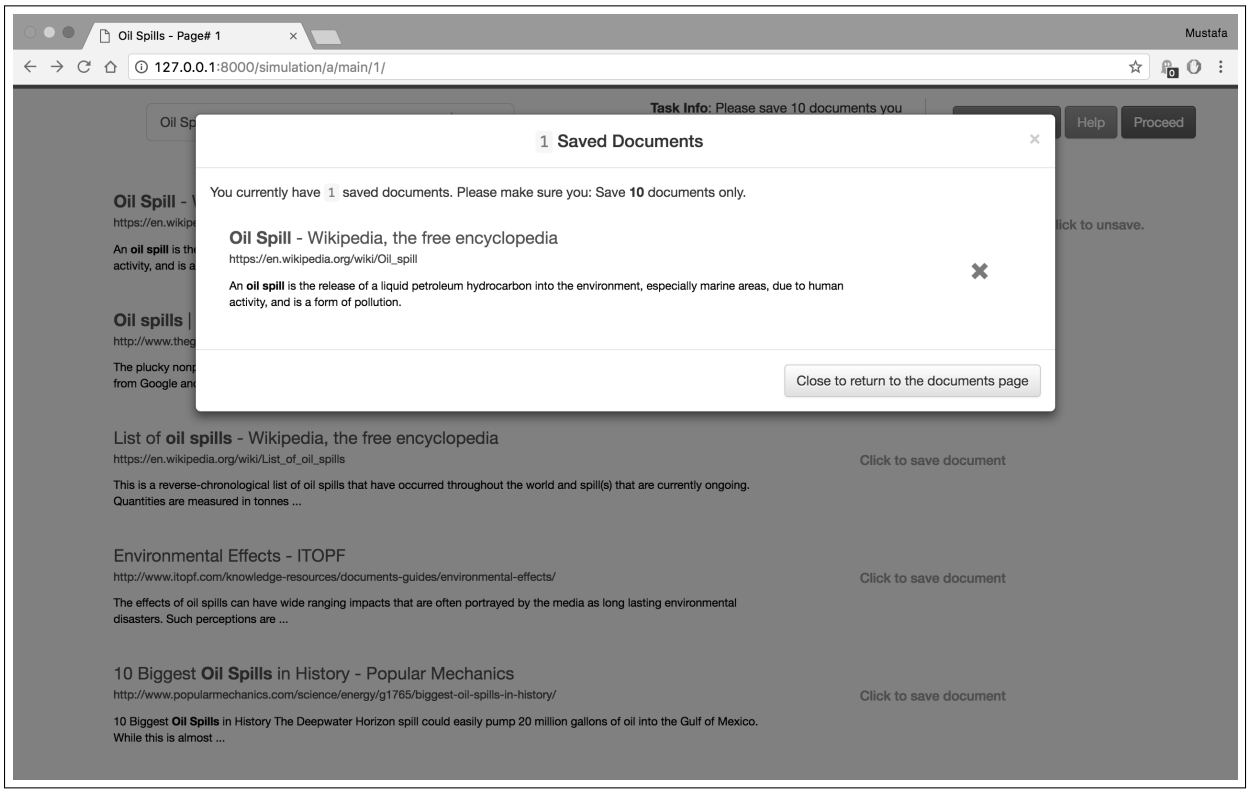


Figure 4.4: Search interface - list of saved documents.

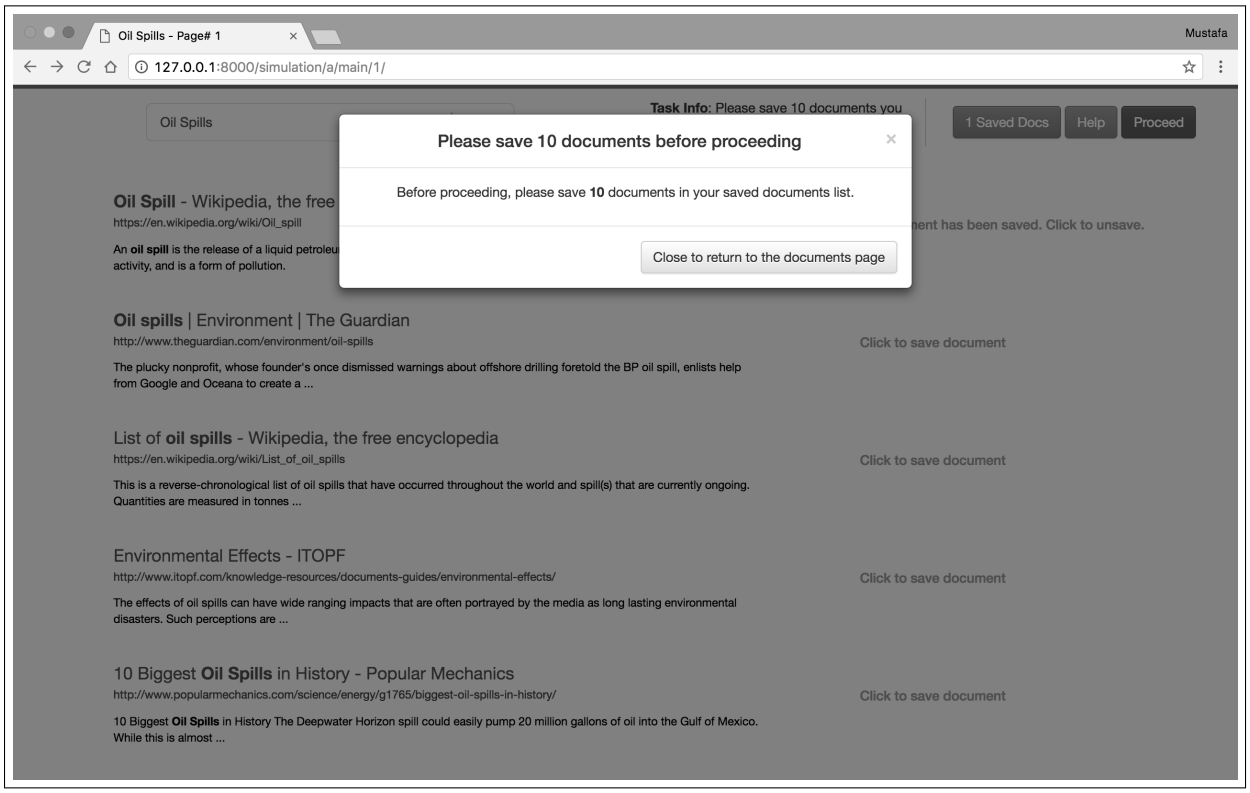


Figure 4.5: Search interface - Proceeding to next task.

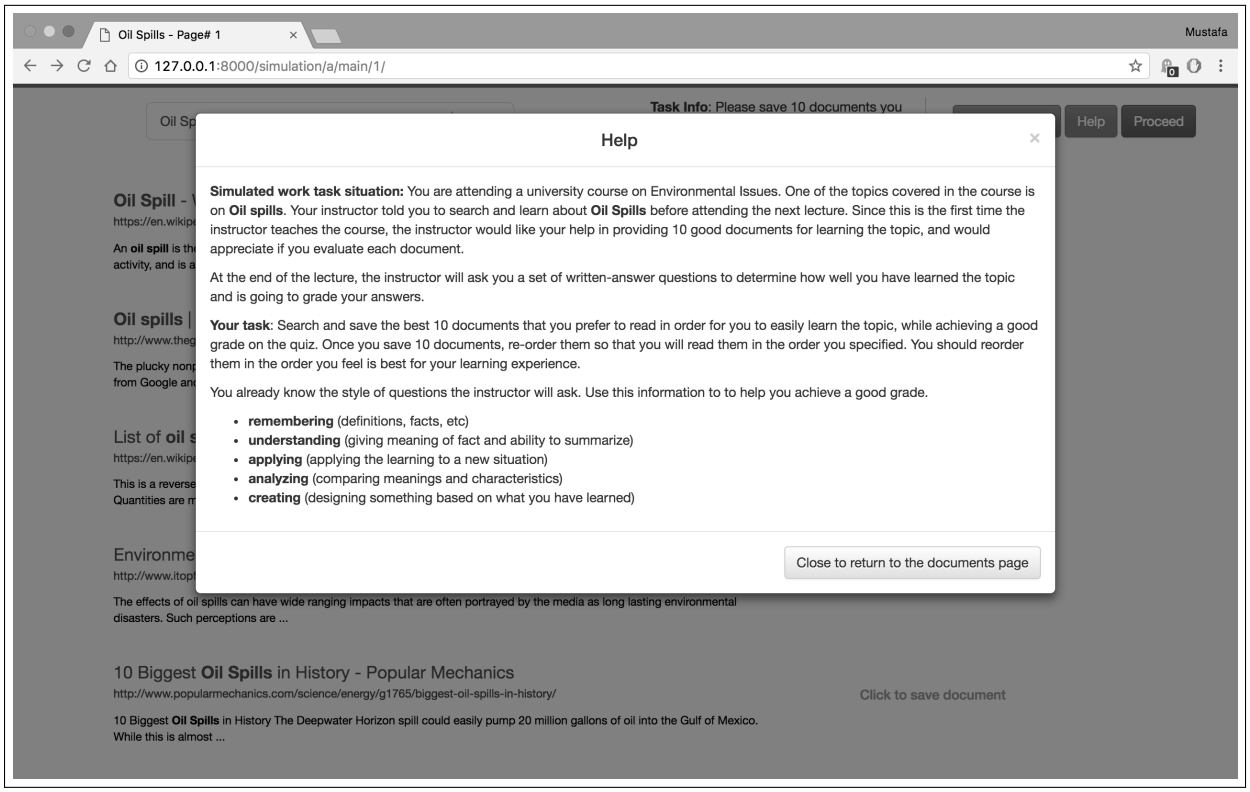


Figure 4.6: Search interface - help information.

Task 2: After the saving the ten documents, we asked participants to re-order the documents to a reading order that they feel will enhance their learning experience. Changing the position of a document could be done by holding the move icon and dragging the document vertically (Figure 4.8). The initial ordering of documents was set as the order in which they have saved the documents, but we note that all participants have indicated a specific reading order different from their saving order. We label the reading ordering user have specified as **RO**.

After completing task 2, we presented the following simulated work task (Figure 4.7):

> "You have saved 10 documents that you believe are good learning material. Since you will be quizzed at the end of the lecture, you decide to learn the topic by reading the documents you have saved in the order that will help you learn the topic better"

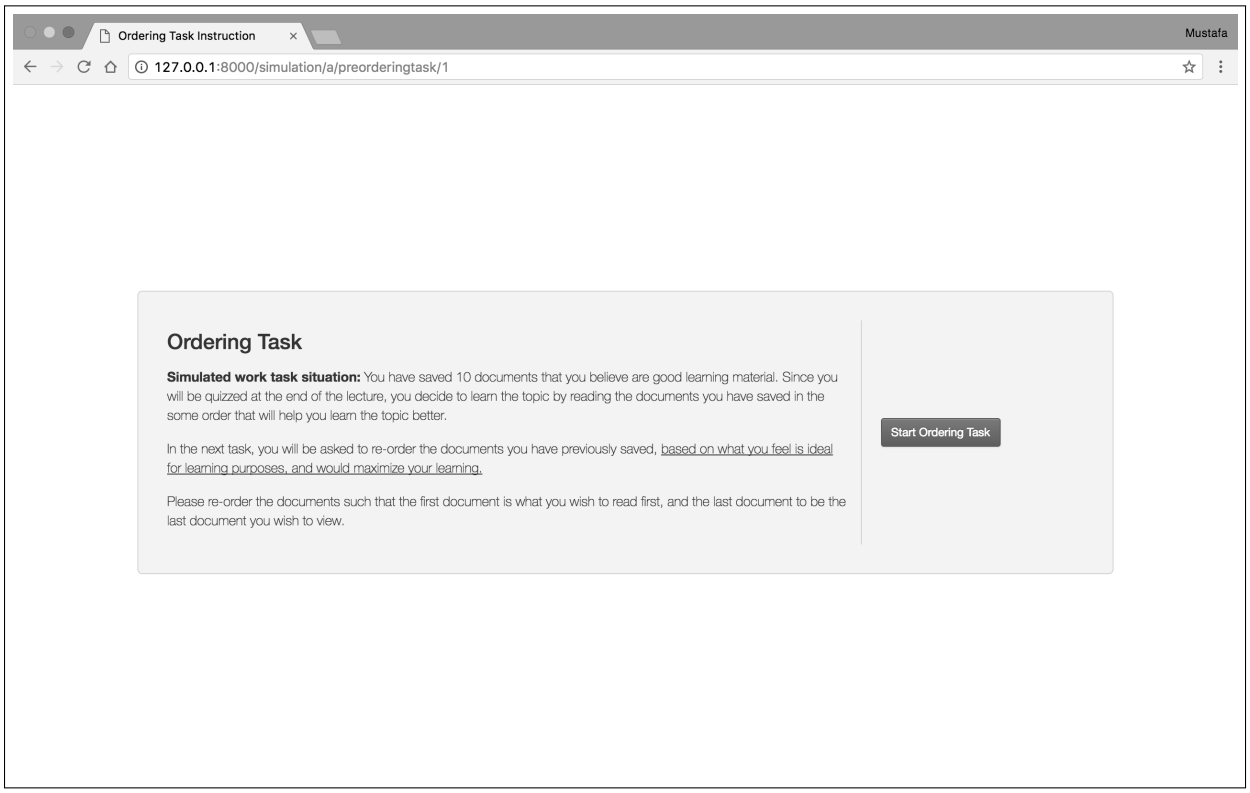


Figure 4.7: Ordering task simulated work situation.

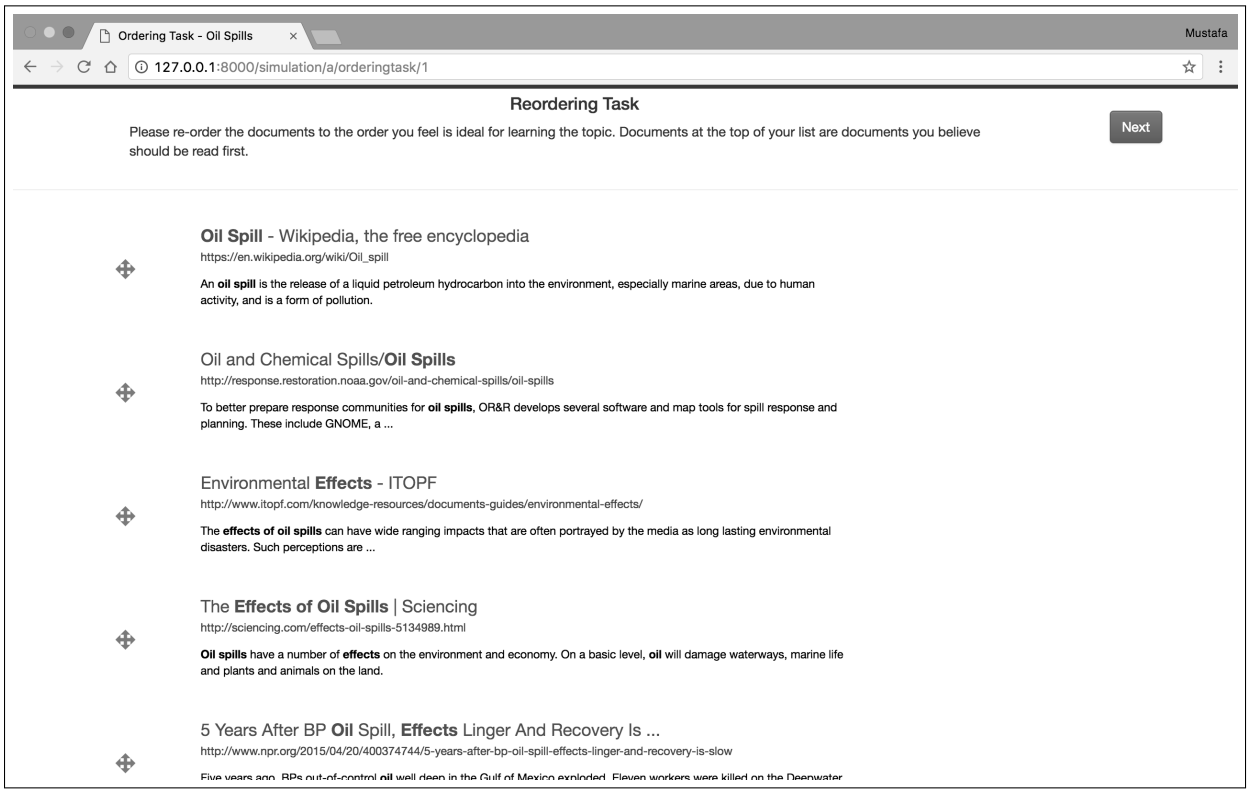


Figure 4.8: Ordering task interface.

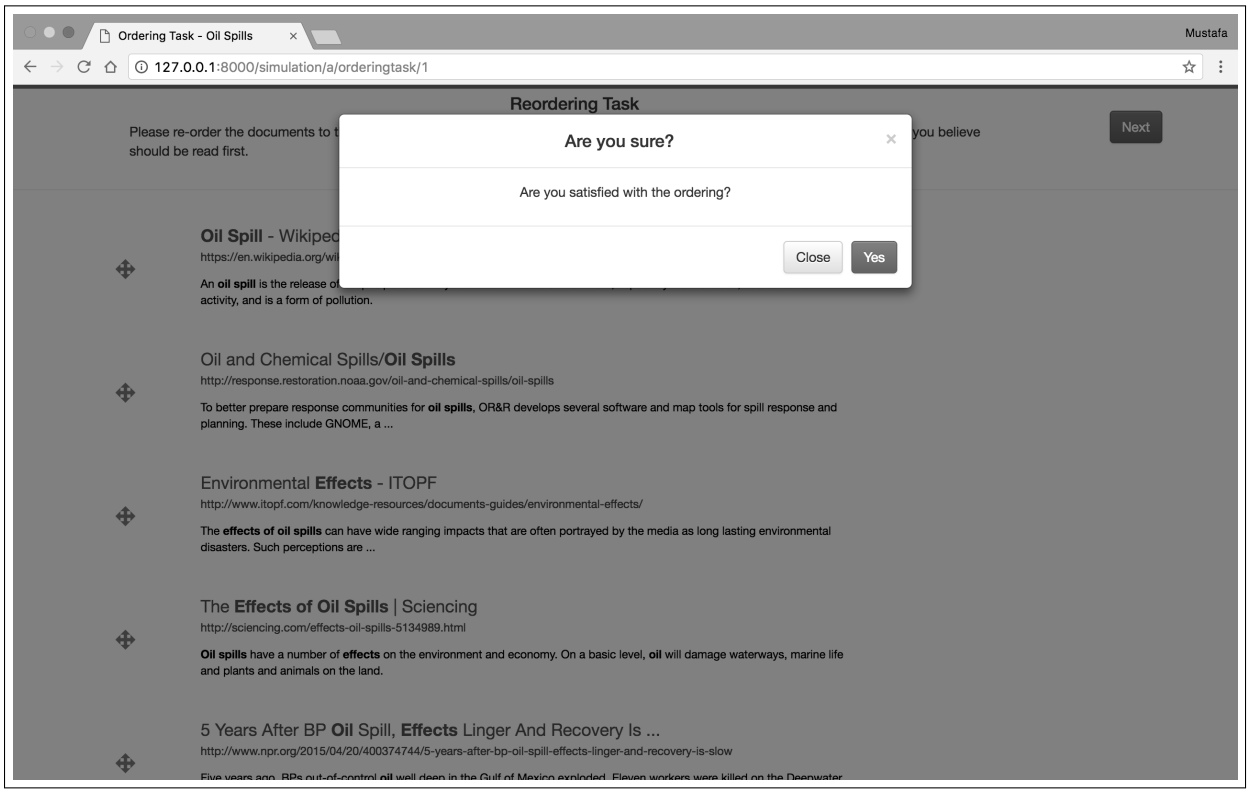


Figure 4.9: Ordering task confirmation.

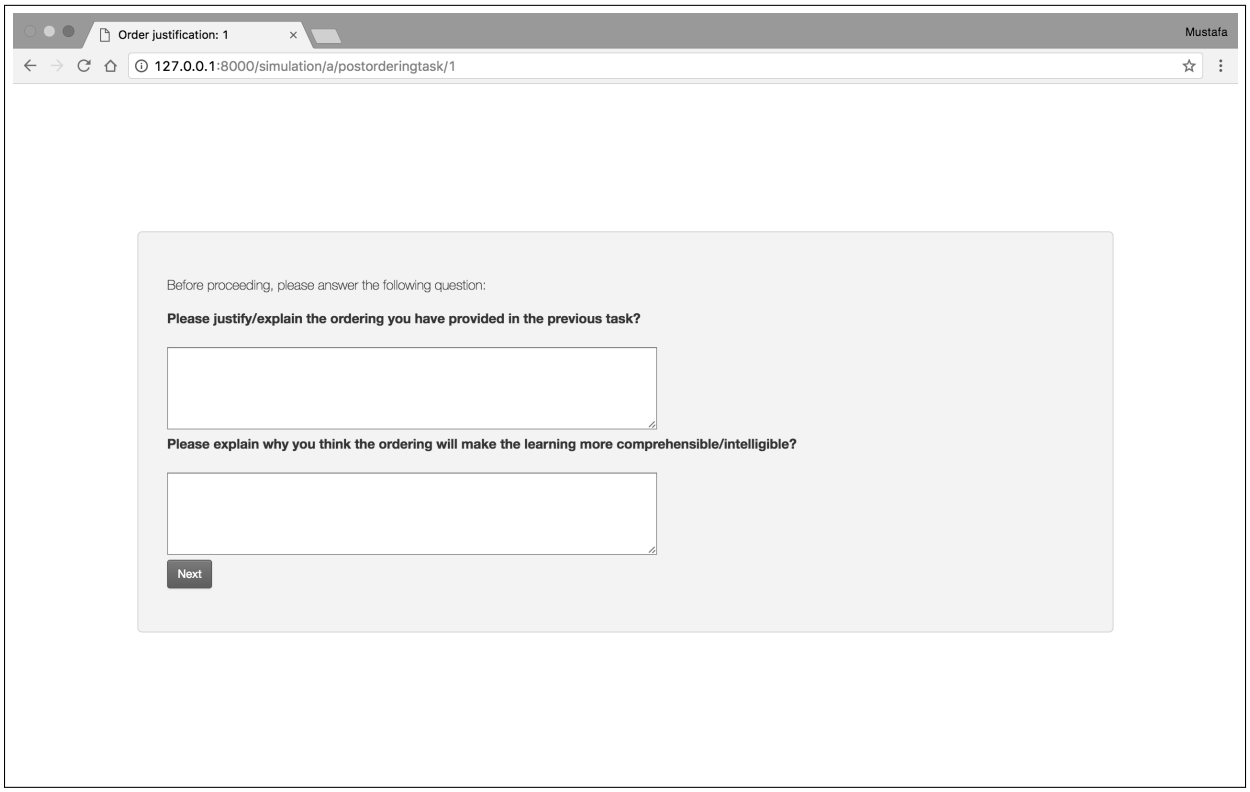


Figure 4.10: Ordering questionnaire interface.

Before starting Task 3, we asked participants questions in Table 4.3. These questions were asked to get a better understanding of why the learner have ordered their documents in their ordering and why do they thing it will help with their learning process.

Question	Response type
OQ1: Explain the ordering you have provided in the previous task	Written response
OQ2: Explain why the ordering will make the learning more comprehensible and intelligible	

Table 4.3: Ordering post-task questions.

Task 3 : After the re-ordering is complete, participants were asked to read through the documents they have saved, one-by-one, in the order they have specified in Task 2 (Figure 4.11). We presented the following simulated situation to the participant during this task:

"The instructor is grateful that you have provided him with ten documents that you believe are good learning material. Since you will be quizzed at the end of the lecture, you decide to learn the topic by reading the documents you have saved in the order you specified, and evaluate how useful they are after reading them."

Each document is hidden by default and will appear only if its previous document in the list has been read. After reading a document, participants were presented with a pop-up (Figure 4.14) with questions:

S1: to determine the participant satisfaction level of the document after reading it,

LF1 - LF6: to judge the document on each of six learning factors using a 5-level Likert scale (1=*Not at all*, 5=*Extremely*), and

DF1: to describe the reason for saving the document to their reading list.

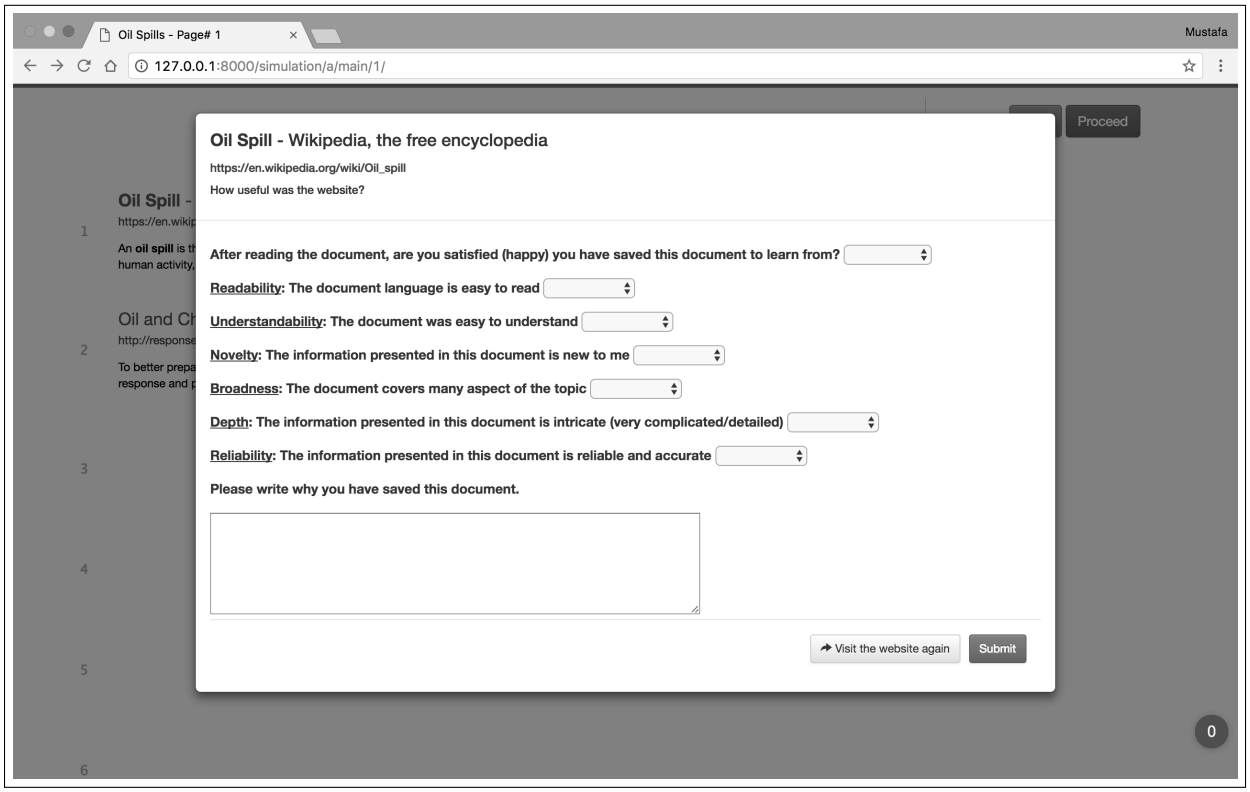


Figure 4.14: Document feedback form where participants label each read documents on each learning factor as well as document satisfaction.

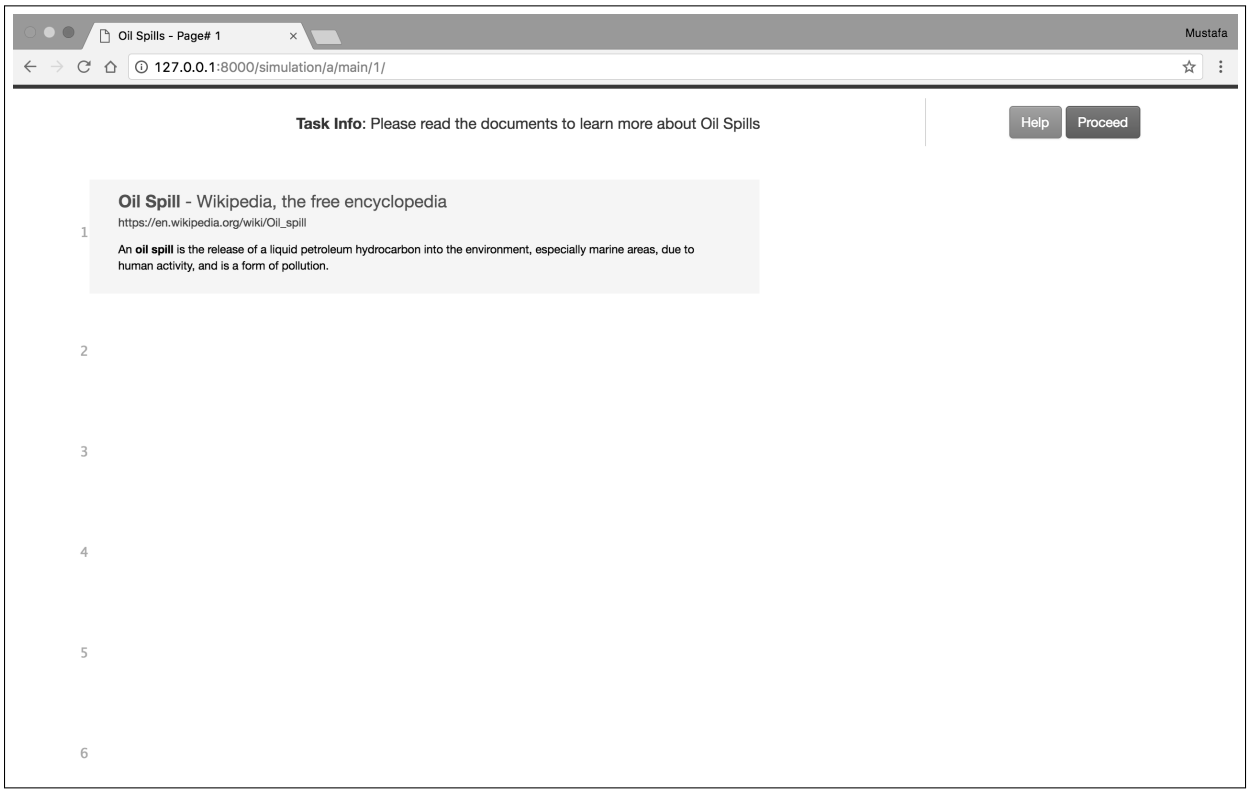


Figure 4.11: Documents reading task.

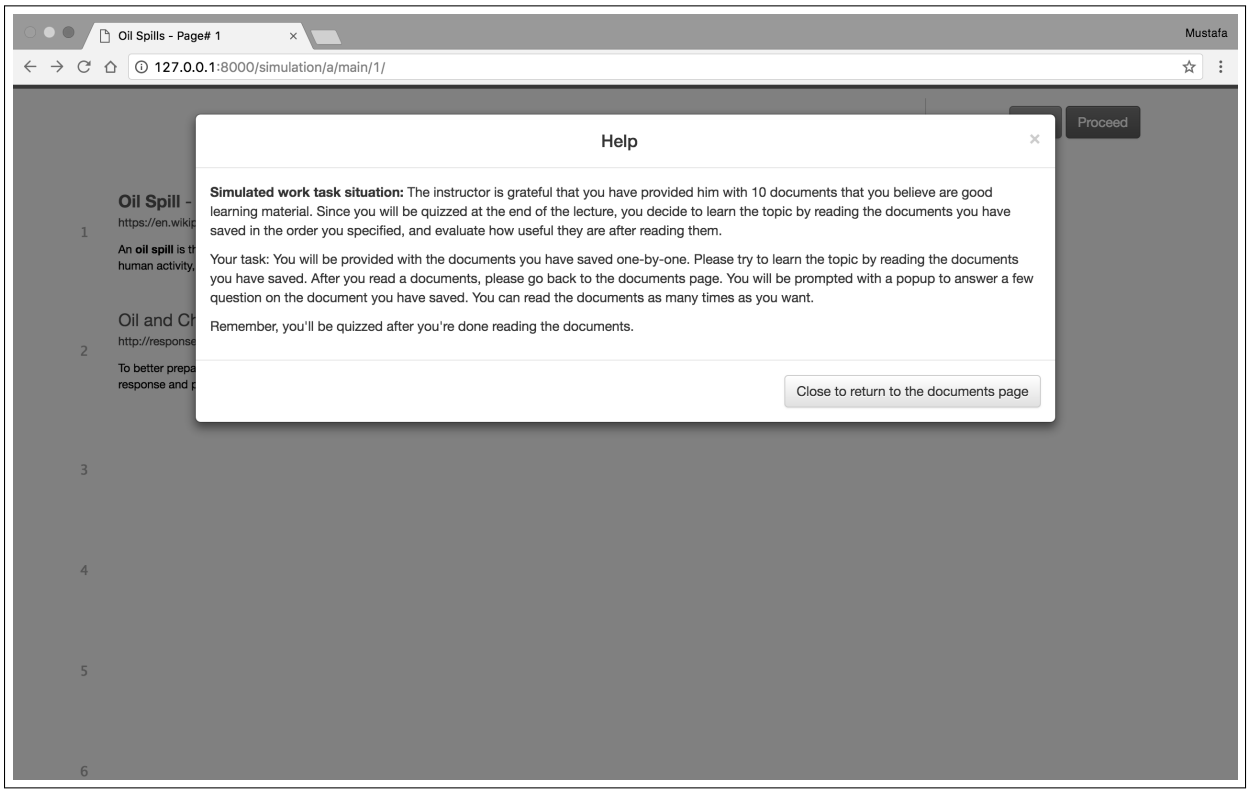


Figure 4.12: Documents reading task help information.

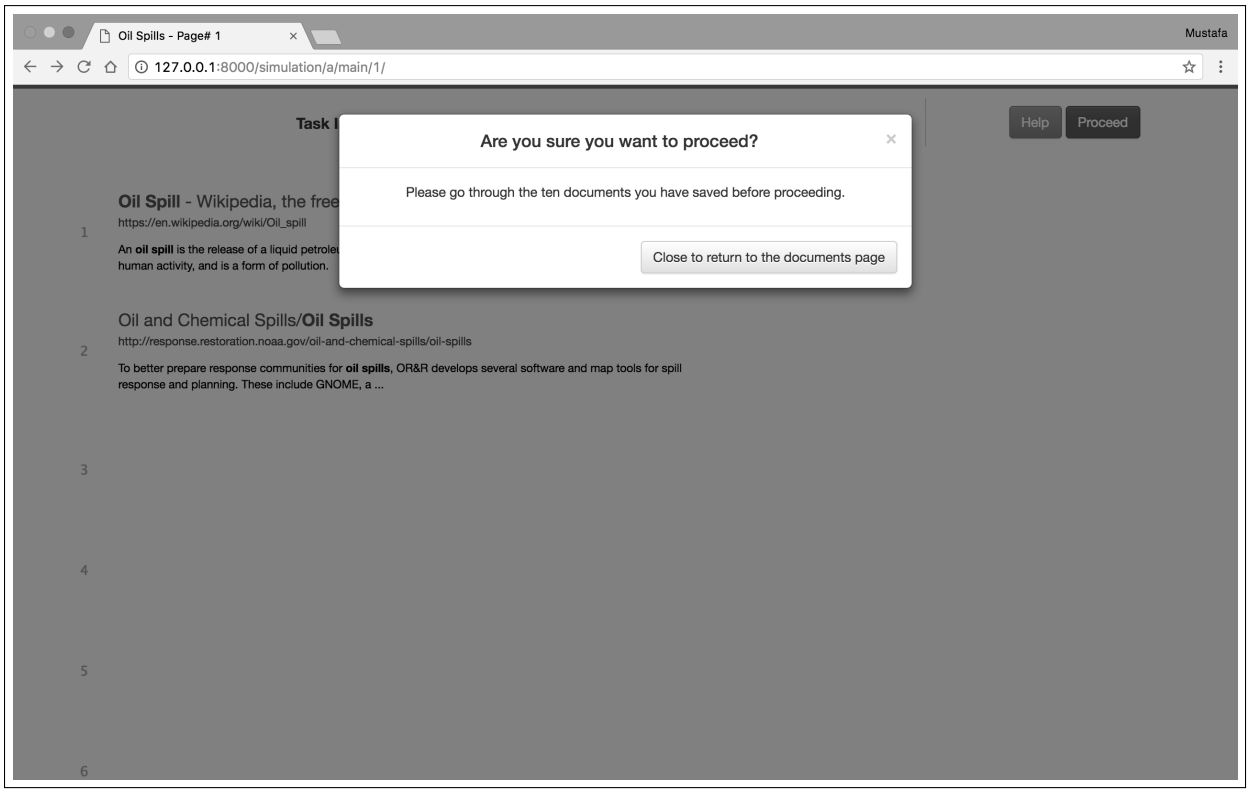


Figure 4.13: Proceeding from the documents reading task is not allowed before reading the 10 documents.

Post-task: After reading each saved document and completing all document related questions in Task 3, participants proceeded to the post-task (Figure 4.15). At this stage, we ask questions related to their interest, mood and difficulty of the topic, as well as questions on their ordering. Table 4.4 includes all the questions asked to participants during this task.

Question	Response type
PT1: I became interested in this topic.	1=Nothing 5=Very familiar
PT2: I would like to find more information about this topic.	
PT3: I would like to share what I learner with my friends.	
PT4: I learned useful information as a result of the documents I have read.	
PT5: I was able to develop new ideas or perspectives.	
PT6: The quality and readability of the documents content influenced my learning process.	1=Strongly disagree 5=Strongly agree
PT7: How difficult was it for you to learn the topic?	1=Very easy 5=Very hard
PT8: How difficult was it for you to find good learning material about the topic?	
PT9: How was your mood during the task?	1=Very enjoyable 5= Very not enjoyable
PT10: How would you grade your learning outcome?	0 to 100
OQ3: The order of documents affected my learning process.	1=Strongly disagree 5=Strongly agree
OQ4: How effective was the ordering of documents towards your learning process?	1=Not at all effective 5=Very effective
OQ5: If you feel like the ordering of documents affected your learning process in any way, please describe how.	Written response

Table 4.4: Post-task questions and respond type.

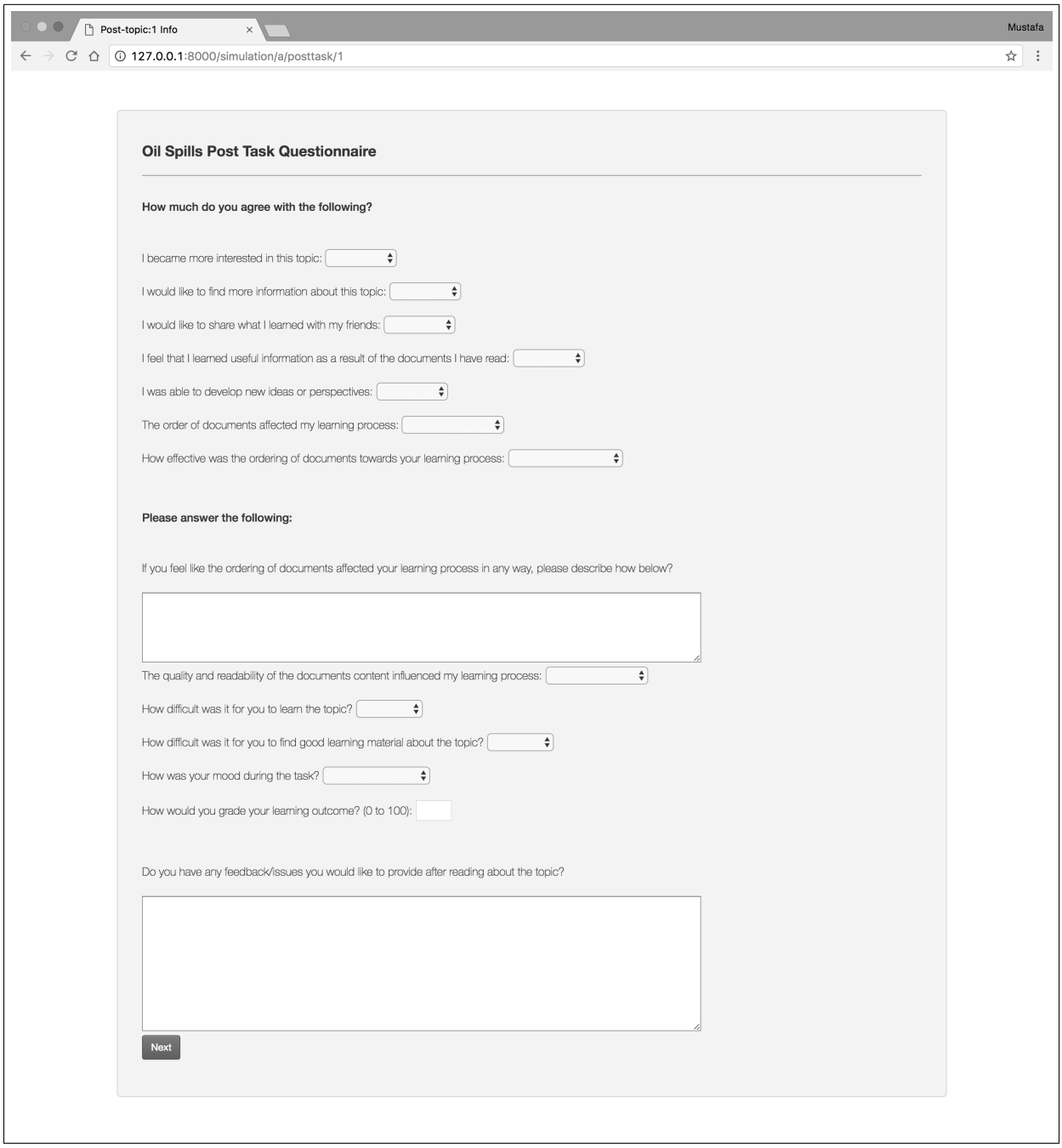


Figure 4.15: Post-task questionnaire interface.

Assessment quiz: To test participants knowledge after completing all tasks, we asked three low-cognitive assessment quiz questions **Q1**-**Q3** and three high-cognitive assessment quiz questions **Q4**-**Q6**. Each of the quiz questions addressed one of Blooms's six revised learning levels (Bloom, 1956) (We provide and discuss these questions in more details in Section 4.5). Each topic had different low-cognitive questions (**Q1** to **Q3**) but share the same high-cognitive questions (**Q4** to **Q6**). Figure 4.16 shows the interface where participants write their answer to **Q1**-**Q6**.

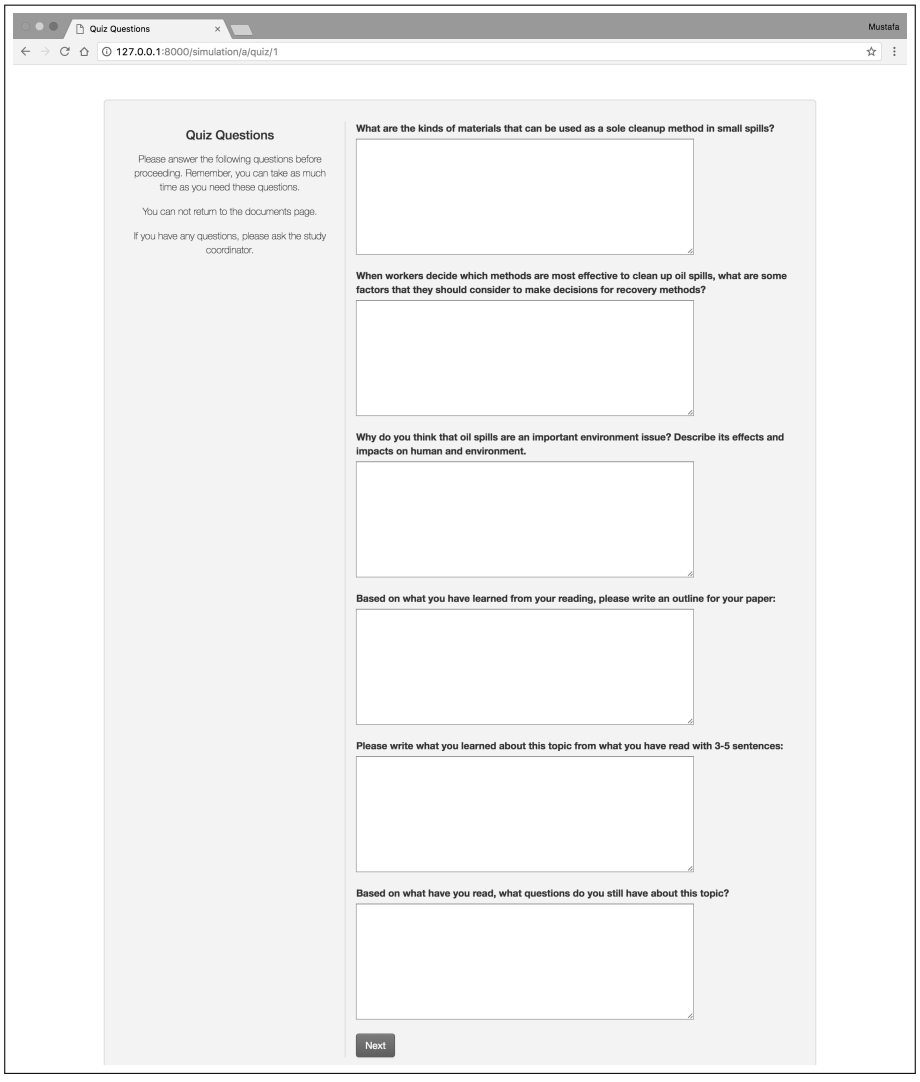


Figure 4.16: Quiz questions interface.

Optional Final Ordering : Finally, participants can optionally re-order their list after having read all documents carefully and completed the quiz (Figure 4.17). The reasoning behind this step is that participants may wish to provide a *better* ordering after having carefully read and judged all documents. We label the final ordering user have specified as **FO**. If no changes to the ordering was indicated, we keep **FO** the same as **RO**. We also asked participants **OQ6:** to explain the final ordering they have provided after completing the ordering.

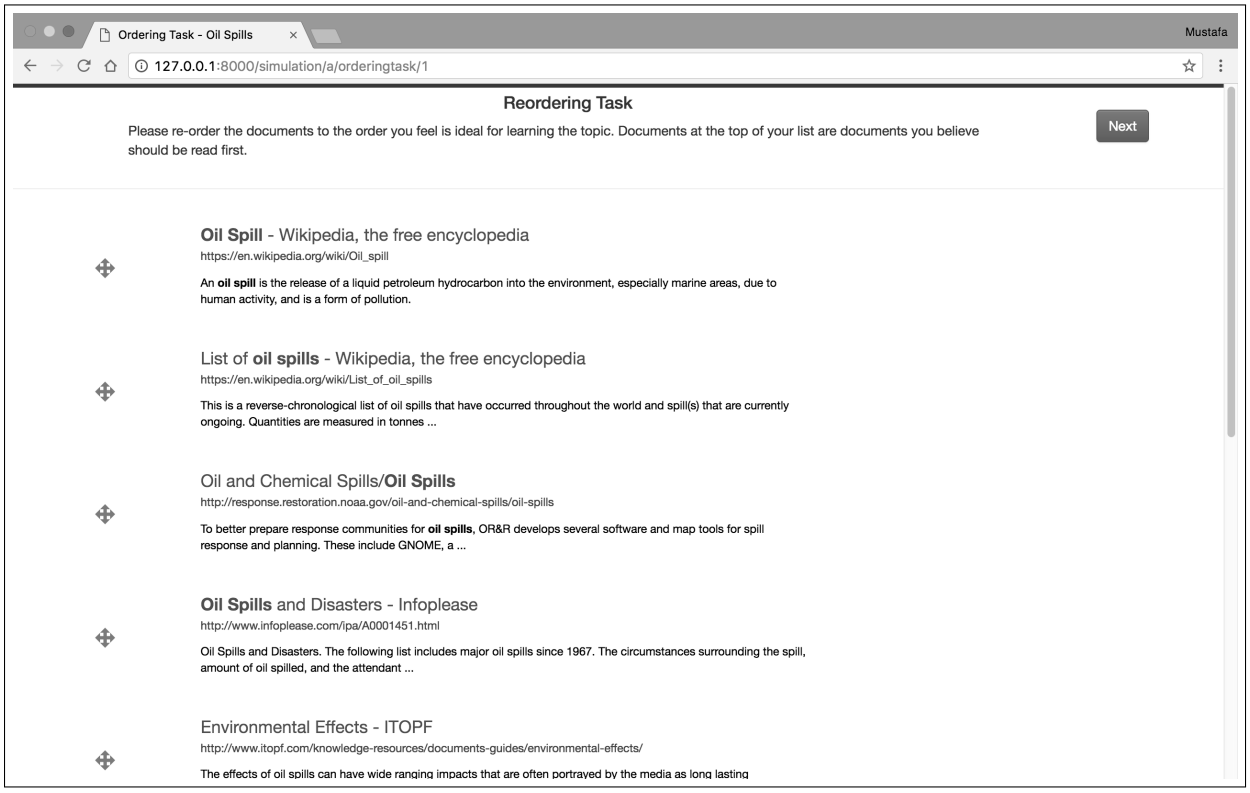


Figure 4.17: Optional final ordering interface.

An inventory of all questions asked during the study and the scale or type of answer provided by the participants is available in appendix (Table A.4).

4.5 Quiz Questions

Pre-task and a post-task assessment questions were introduced in our study to measure participants prior knowledge in the topic before completing the study tasks, and knowledge acquired after searching and reading more documents. In order to measure prior knowledge, we have asked participants question **PR4:** to summarize what they already know about the topic in few sentences.

In order to measure higher forms of learning that are more complex than remembering facts, such as processing and analyzing concepts, each question in the quiz task addressed a level of Bloom's revised learning taxonomy (Bloom, 1956). Each question and its intended assessment is briefly described below:

Lower Cognitive Questions:

Q1 Remembering: Recognizing facts and concepts

Q2 Understanding: Perceiving the instructional messages.

Q3 Applying: The ability to carry out procedures.

Higher Cognitive Questions:

Q4 Analyzing: Detecting multiple concepts.

Q5 Evaluating: Critiquing with current knowledge.

Q6 Creating: The ability to generate creative thoughts.

We explain how we code the written answers to each question in Section 4.5.3.

4.5.1 Topic 1 Quiz Questions

The following questions are asked to participants assigned to the oil spills topic.

Q1 Remembering: What are the kinds of materials that can be used as a sole cleanup method in small spills?

Q2 Understanding: When workers decide which methods are most effective to clean up oil spills, what are some factors that they should consider to make decisions for recovery methods?

Q3 Applying: Why do you think that oil spills are important environment issues? Describe its effects and impacts on human and environment

Q4 Analyzing: Based on what you have learned from your searching, please write an outline for your paper.

Q5 Evaluating: Please write what you learned about this topic from your searching with 3-5 sentences.

Q6 Creating: Based on your searching, what questions do you still have about this topic?

4.5.2 Topic 2 Quiz Questions

The following questions are asked to participants assigned to the open data topic.

Q1 Remembering: Is copyright protection available for works of the United State Government?

Q2 Understanding: In 2007, a number of open government advocates got together and claimed that government data shall be considered open if it is made public in a way that complies with some fundamental principles. Others added more principles since then. What are some examples of principles of open government data?

Q3 Applying: What kinds of individuals, communities, or organizations could be benefited as a result of accessing open data provided by government?

Q4 Analyzing: Based on what you have learned from your searching, please write an outline for your paper.

Q5 Evaluating: Please write what you learned about this topic from your searching with 3-5 sentences.

Q6 Creating: Based on your searching, what questions do you still have about this topic?

4.5.3 Evaluating Quiz Answers

Using the same coding metric used by Collins-Thompson et al. (2016), the criteria for scoring participants written answers to quiz question (**Q1** to **Q6**) and prior knowledge question **PR4**, addresses the cognitive process identified by Anderson et al. (2001) and adheres with the learning levels suggested in Bloom's Revised Taxonomy (Bloom, 1956).

The criteria consisted of seven points and tested participants' factual knowledge, such as recalling, recognizing and defining of facts and concepts and conceptual knowledge such describing, identifying and differentiating multiple concepts.

We consider a participant to have gained in knowledge if the difference in their **PR4** and **Q5** scores is positive. We label the knowledge gain score as **KG**.

4.6 Logging of Search Behavior

We have developed our interface to record the following:

- **Saved Documents**: Documents participant wish to read more carefully in the Task 3. Each participant provided a reason on why it has been saved (**DF1**).
- **User-judged documents**: User judgments on each learning factor **LF1** to **LF6** on a 5-point Likert scale, for each saved document.
- **Documents order**: The preferred reading order (**RO**) of the 10 saved documents per user, and their final order after reading all documents (**FO**).
- **Documents ordering feedback**: Written summaries explaining the ordering and how it will make the learning process more effective (**OQ1** to **OQ6**).
- **Quiz answers**: Participants written responses to lower and higher cognitive-level quiz questions on the topic (**Q1** to **Q6**).

Chapter 5

Result

We collected 350 saved documents that include participants satisfaction judgment of documents (**S1**), written response on why the document has been saved (**DF1**) and level of learning factors **LF1** to **LF6** reflected from the document after reading its content. We also collected 35 answers for each of the other question asked in Table A.4.

5.1 Demographic data

We collected different demographic and user data before participants started the study. Table 4.1 contains all the questions collected in the demographic questionnaire.

Participants age varied from 18 to 30, with average of 21.5 (SD = 2.85) with 17 males and 18 females. 13 graduates and 22 undergraduate students participated.

Figure 5.1 shows participants answer to demographic question **DQ6:** on how often they use the Internet to learn about a new topic.

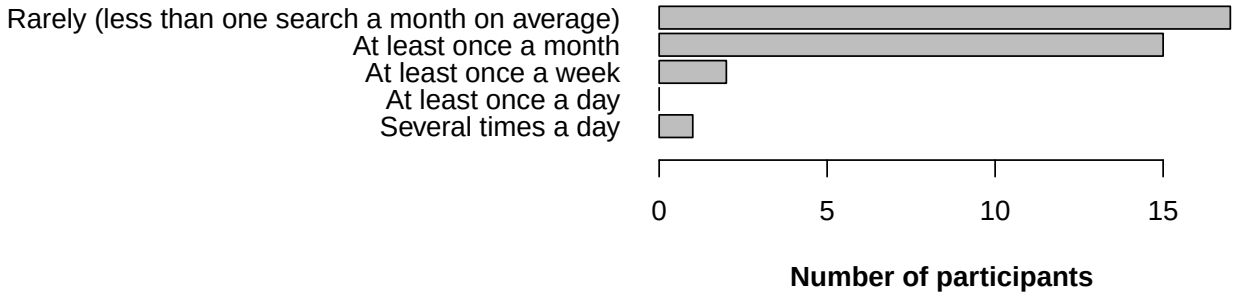


Figure 5.1: Participants answers to demographic question DQ6.

Figure 5.2 shows participants answer to demographic question **DQ7:** on whether they struggle to find good learning material for a topic they are trying to learn.

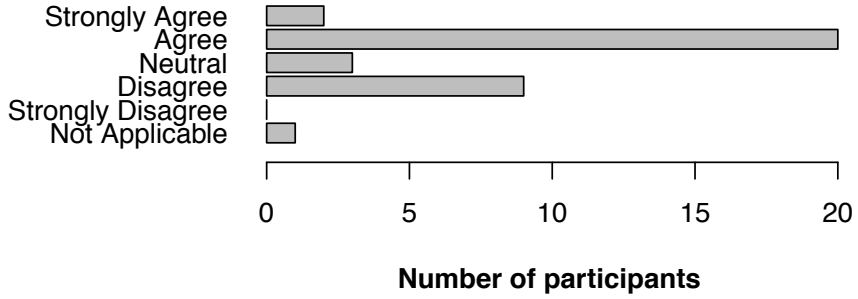


Figure 5.2: Participants answers to demographic question DQ7.

Figure 5.3 shows participants answer to demographic question **DQ8:** on whether they search engines as their primary source when trying to learn a new topic.

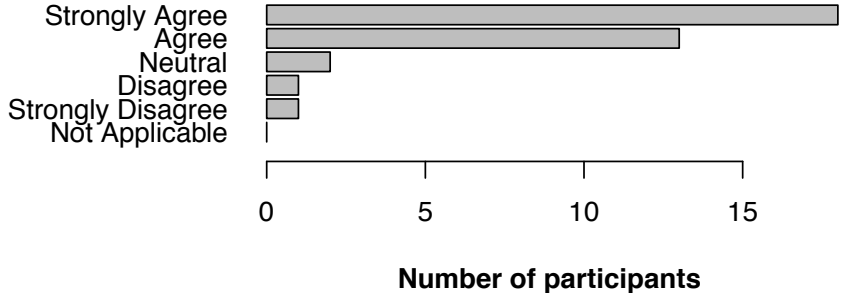


Figure 5.3: Participants answers to demographic question DQ8.

Figure 5.4 shows participants answer to demographic question **DQ9:** on whether they consider themselves as fast-learners.

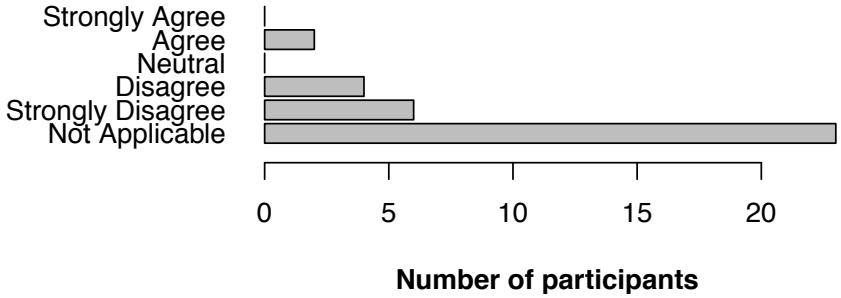


Figure 5.4: Participants answers to demographic question DQ9.

5.2 Pre-task result

Our pre-task questions (Table 4.2) were introduced to measure users' prior knowledge (**PR4**), familiarity level (**PR1**), perceived difficulty (**PR2**), and interest (**PR3**) for the topic participants were assigned to.

Figures 5.5 and 5.6 show participants answers to pre-task question **PR1** (Topic familiarity) for Topic 1 (Oil Spills) and Topic 2 (Open Data), respectively. In general, participant were more familiar with Topic 1 (Oil Spills) than Topic 2 (Open Data).

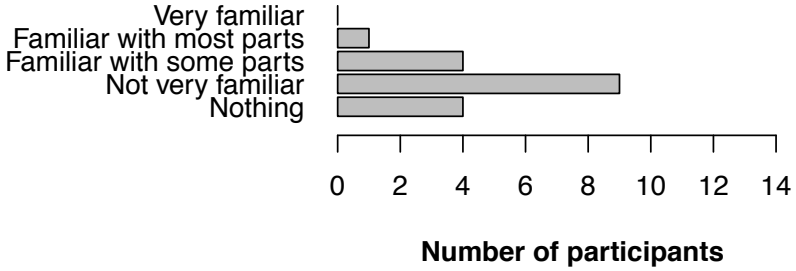


Figure 5.5: Participants answers to pre-task question PR1 for Topic 1 (Oil Spills).

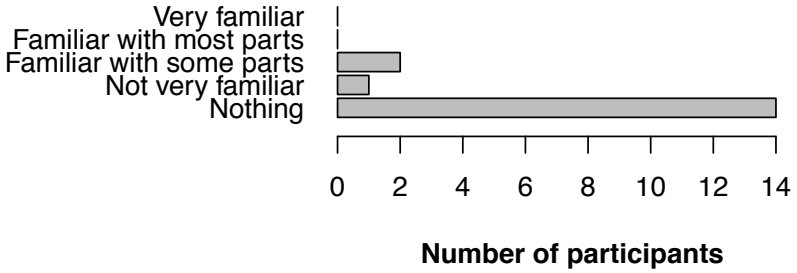


Figure 5.6: Participants answers to pre-task question PR1 for Topic 2 (Open Data).

Figures 5.7 and 5.8 show participants answers to pre-task question **PR2** (Perceived topic difficulty) for Topic 1 (Oil Spills) and Topic 2 (Open Data), respectively. Participants in both topics felt that it will *Moderately* difficult to learn the topic.

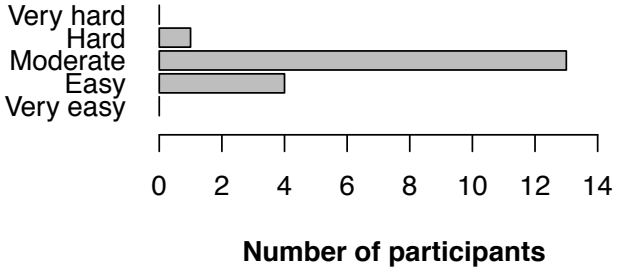


Figure 5.7: Participants answers to PR2 for Topic 1 (Oil Spills).

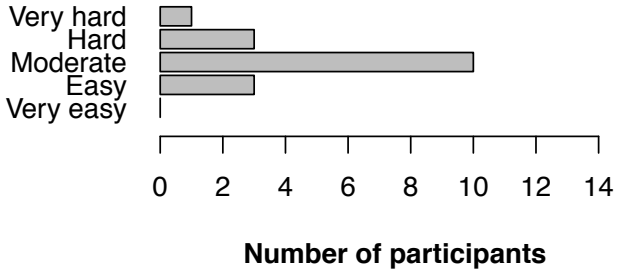


Figure 5.8: Participants answers to PR2 for Topic 2 (Open Data).

Figures 5.9 and 5.10 show participants answers to pre-task question **PR3** (Perceived

topic interest) for Topic 1 (Oil Spills) and Topic 2 (Open Data), respectively. More people were less interested in learning Topic 2 (Open Data) than Topic 1 (Oil Spills).

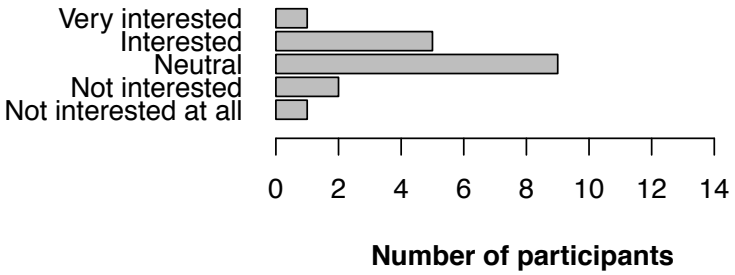


Figure 5.9: Participants answers to PR3 for Topic 1 (Oil Spills).

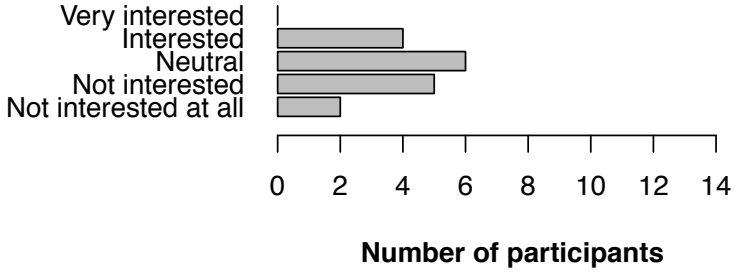


Figure 5.10: Participants answers to PR3 for Topic 2 (Open Data).

5.3 Learning outcome and learning factors

Two independent coders with no knowledge about the study details have applied the same 7-level coding scheme in Collins-Thompson et al. (2016) to score prior knowledge question (**PR4**) and the six quiz questions (**Q1** to **Q6**) for all 35 participants. Coders were provided with a 7-level criteria for each question, yielding a total of 84 different criteria that coders apply to assign a grade from 0 to 7 to each written response. If a criterion was evident in the written response, the score of that answer is increased by 1. Figure 5.11 shows the interface used by the coders to code participants answers. The interface design had participants answers in one side next to the 7-level criteria they need to complete for that given answer.

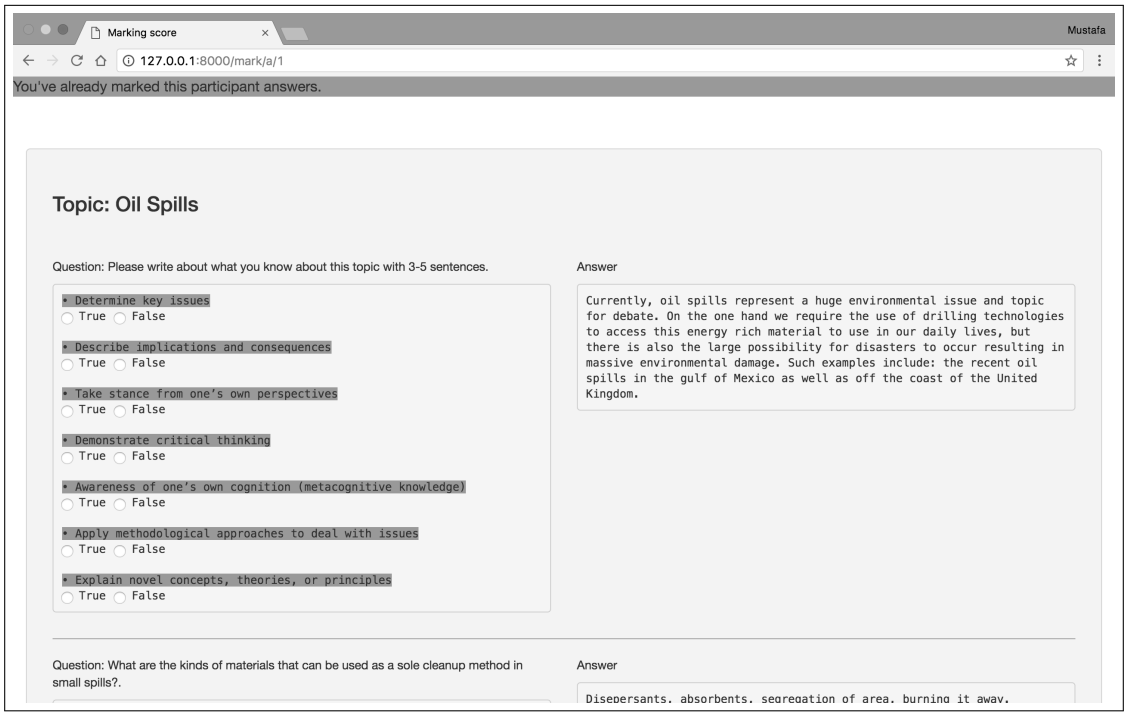


Figure 5.11: Coders coding interface.

As in Collins-Thompson et al. (2016), we used Holsti's coefficient (Holsti, 1969) as an index of the inter-rater agreement by calculating the ratio of coders agreement on the 7-level criteria to the number of coders judgments for each of the questions. A correlation coefficient close to 1 indicates higher agreement between the two raters. The mean inter-rater coefficient across all questions for Topic 1 (Oil Spills) and Topic 2 (Open Data) are 0.725, and 0.783, respectively. More specifically, inter-rater coefficient on prior knowledge (**PR4**) and current knowledge (**Q5**) responses are 0.756 and 0.655 for Topic 1 (Oil Spills), and 0.823 and 0.781 for Topic 2 (Open Data), respectively. We consider the learner to have gained in knowledge if their **Q5** score is higher than **PR4**.

We have conduct a Pearson correlation analysis to identify correlations between learners score and number of documents in each learning factors. We consider a document to fall under that learning factor if it was judged "*Moderate*" or higher. We use Pearson correlation because it evaluate the relationship between two continuous variables. Using Pearson correlation analysis, we found a significant correlation between the number of novel documents saved by the participant and their knowledge gain score **(KG)** (r=0.38, p=0.024 for the first coder, and r=0.37, p=0.03 for the second coder). We also found a

positive correlation between participants total score for all questions and their perceived learning score (**PT10**) that they answered after reading all documents (r=0.35, p=0.024 for the first coder, and r=0.35, p=0.045 for the second coder). We have found no correlation between the number of documents of other factors and the total score for lower or higher cognitive questions.

5.4 Satisfaction and learning factors

Previous research have investigated satisfaction as function of different factors such as relevance, readability, findability, and understandability (Verma et al., 2016). While searching for learning, however, the quality of document's content and users' prior knowledge in the topic can influence their satisfaction of the document. For example, users may be less satisfied with documents contributing little or no new knowledge, or simply with written quality that does not match with the user's cognitive ability. In this section, we show our attempt in answering our **RQ1**.

We examine whether the participants are satisfied with their ten saved documents, after reading each document in more details, and if one of the proposed learning factors can correlate with satisfaction judgment. Table 5.2 and 5.3 show the frequency distribution of learning factors labels and satisfaction labels on documents. If we were to assume that satisfaction is a direct function of a learning factor and conduct a Spearman correction analysis between the documents satisfaction judgment and the learning factor, the learning factors are significantly correlated with satisfaction ($p <= 0.003$) except for detailedness of documents (see Table 5.1). Because of the ordinal nature of our variables, we use Spearman correlation over Pearson as it determines the strength of monotonic relationships between ordinal variables (Haure and Kossoski, 2011).

Learning Factor	r	p-value
LF1: *Readability*$^+$	**0.24**	$\ll 0.001$
LF2: *Understandability*$^+$	**0.24**	$\ll 0.001$
LF3: *Novelty*$^+$	**0.17**	0.001
LF4: *Broadness*$^+$	**0.28**	$\ll 0.001$
LF5: *Detailedness*$^+$	0.08	0.15
LF6: *Reliability*$^+$	**0.16**	0.003

Table 5.1: Spearman correlation of learning factors with Satisfaction (S1).

Correlation analysis allows us to quantify the degree to which two variables are related. In order to identify relationships between more than two variables, we conducted an ordinal logistic regression[1] in the same manner as of Verma et al. (2016) work, to identify which of the learning factors is statistically significant in determining satisfaction of the document. Our analysis shows broadness ($p \ll 0.001$), and novelty ($p = 0.037$) are statistically significant with positive weights. Readability ($p = 0.060$) and understandability ($p = 0.057$) have positive weights and close to significance; we therefore believe that these two factors are worth further investigating and an increase in the number of subject should clarify their importance. Reliability and detailedness ($p = 0.480$, $p = 0.53$) are not statistically significant.

Figure 5.12, 5.13, 5.14, 5.15, 5.16, and 5.17 show the frequency of each learning factor level and the number of documents in each satisfaction level.

In order to improve search for learning purposes, we suggest that broadness of content and novelty of information in a document are considerable factors that should be incorporated into the retrieval algorithm if the goal is to increase use satisfaction in learning-related tasks.

	No	Somewhat	Yes
Satisfaction	34	69	247

Table 5.2: Distribution of document satisfaction labels; N=350.

[1]Using `rms::orm()` function in the `R` language.

	Not at all	Slightly	Moderately	Very	Extremely
Readability	6	24	53	161	106
Understandability	5	22	71	148	104
Novelty	29	69	110	101	41
Detailedness	75	76	101	76	22
Broadness	36	92	101	74	47
Reliability	5	26	107	148	64

Table 5.3: Distribution of learning factors labels on user-saved documents; N=350.

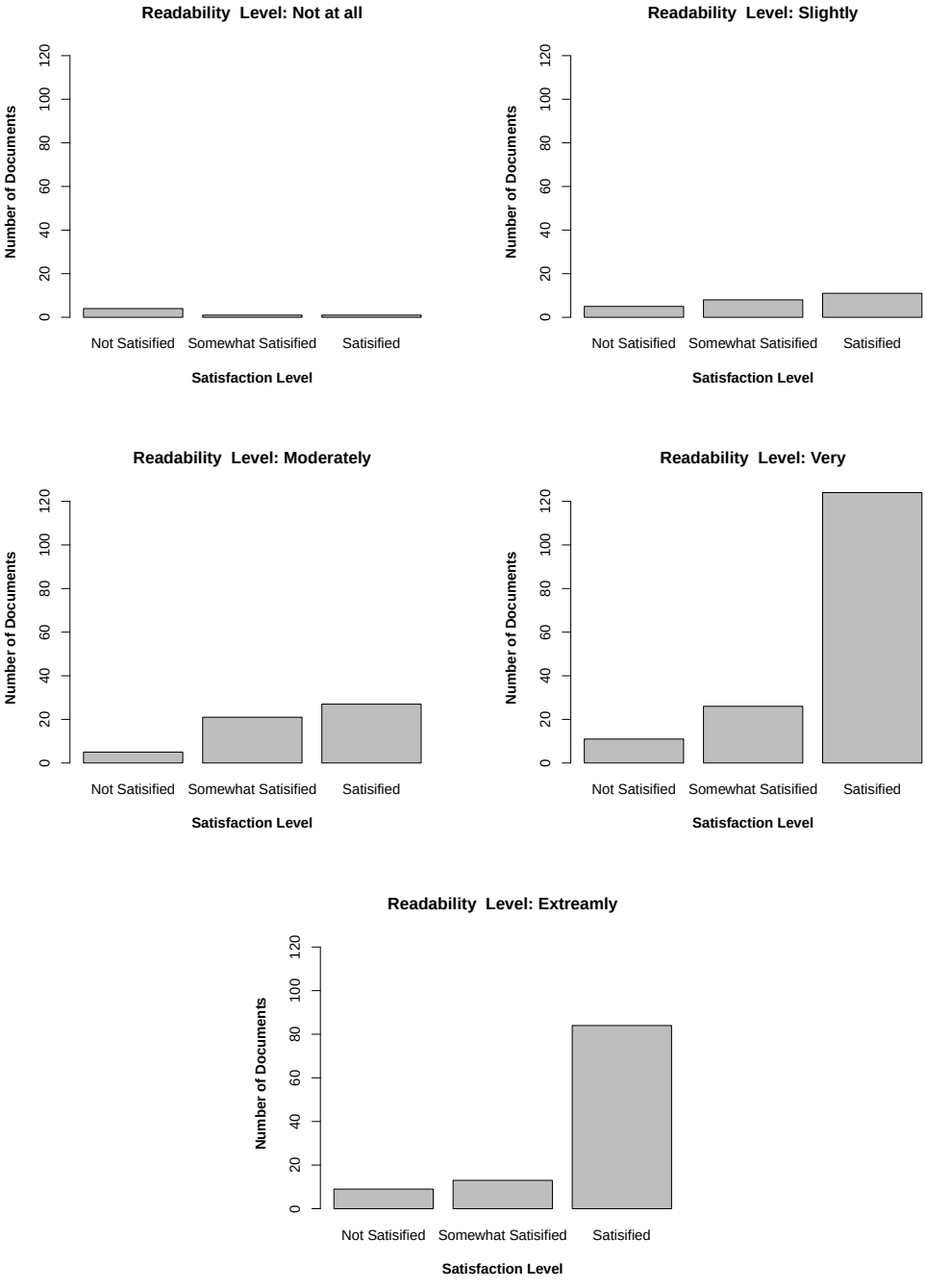


Figure 5.12: Readability Level and Satisfaction Level Frequency.

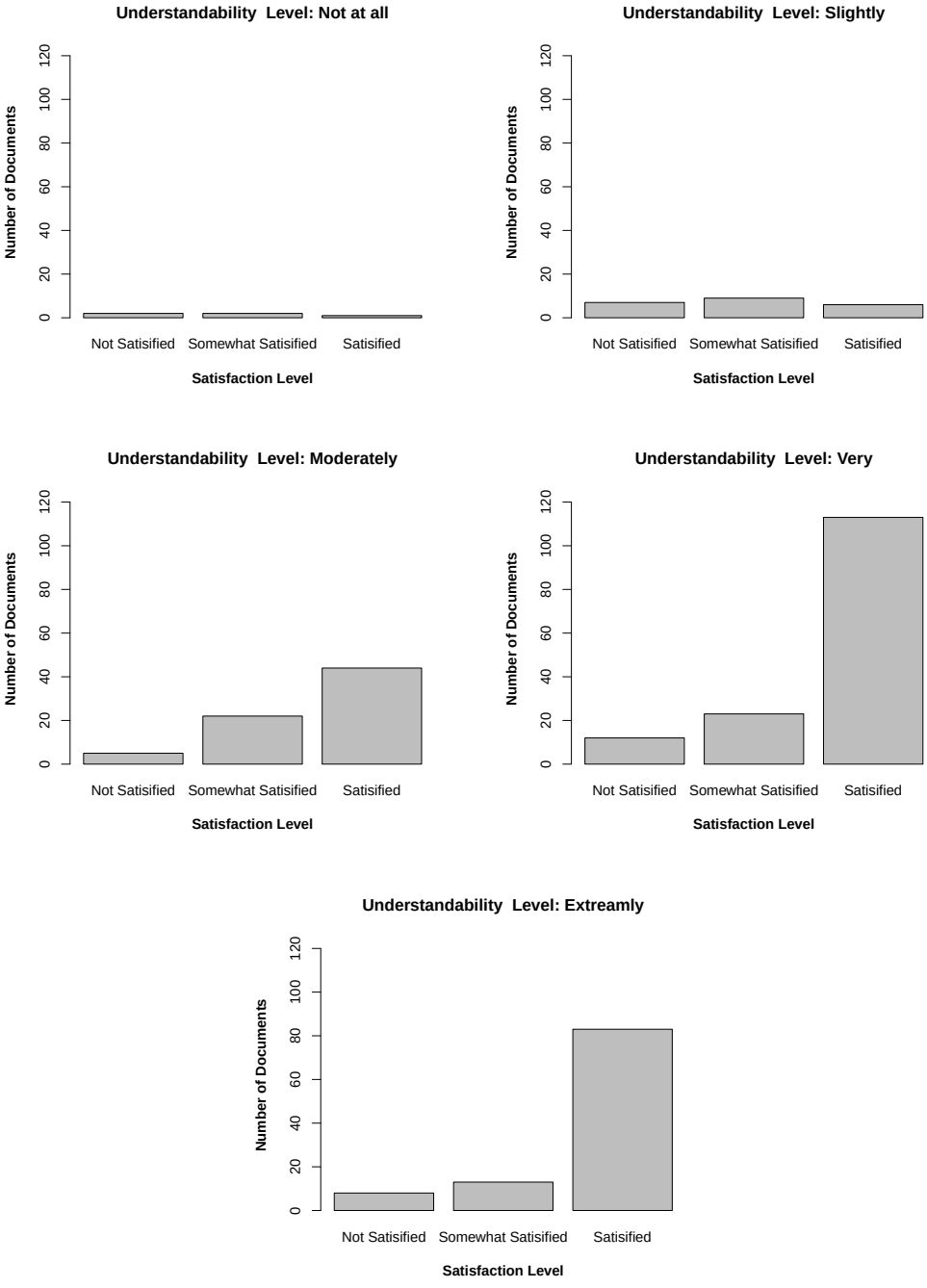


Figure 5.13: Understandability Level and Satisfaction Level Frequency.

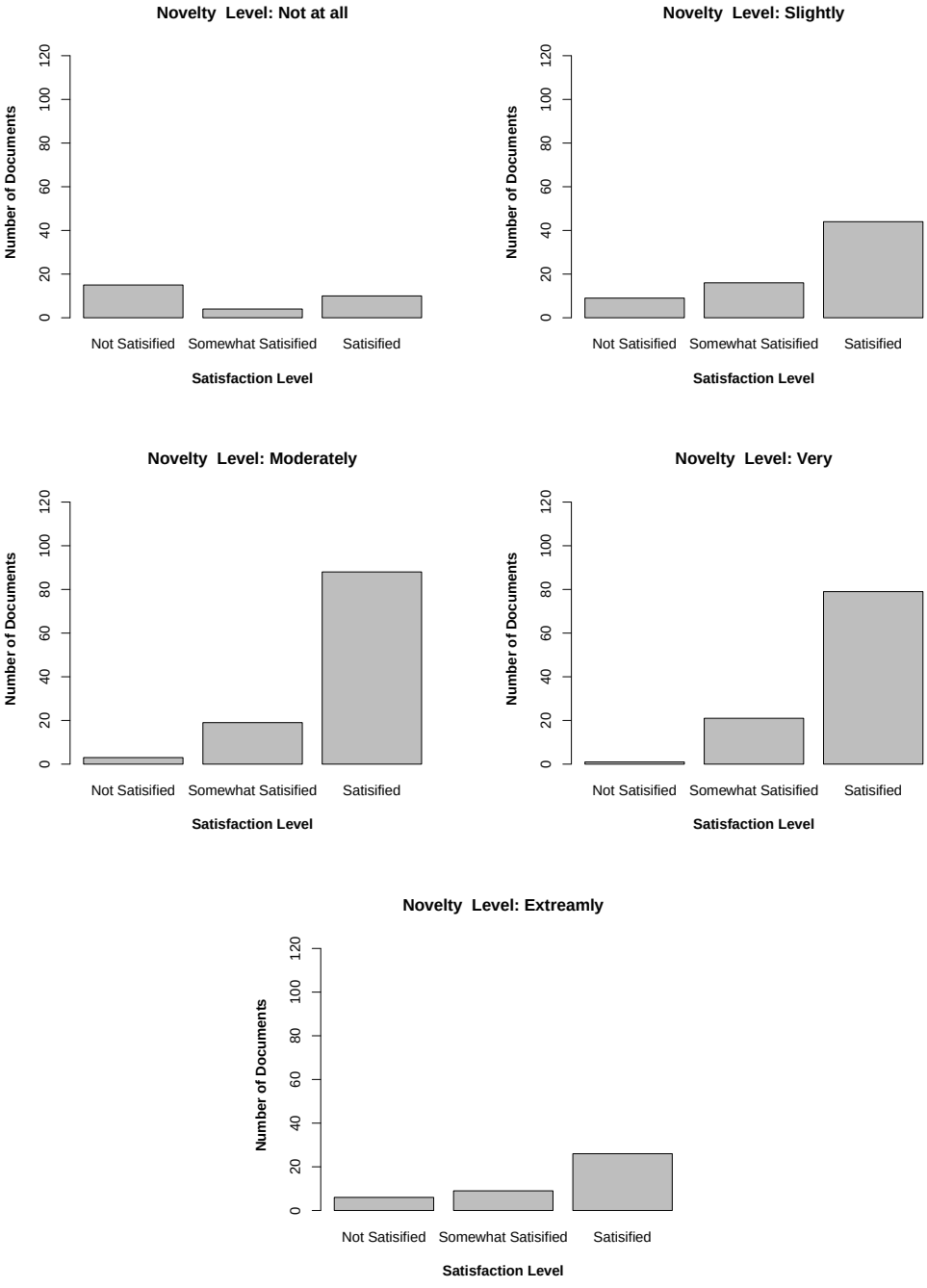


Figure 5.14: Novelty Level and Satisfaction Level Frequency.

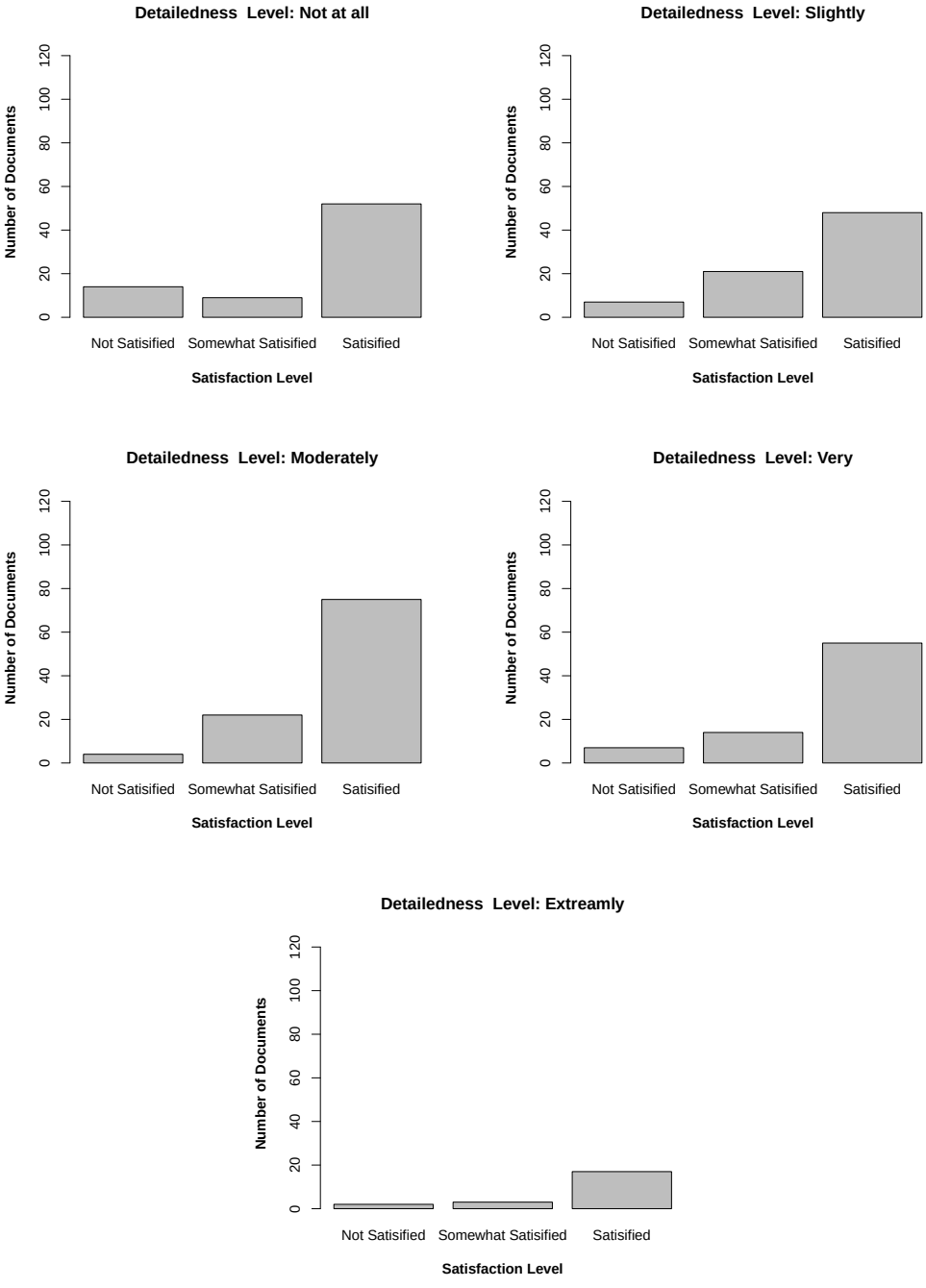


Figure 5.15: Detailedness Level and Satisfaction Level Frequency.

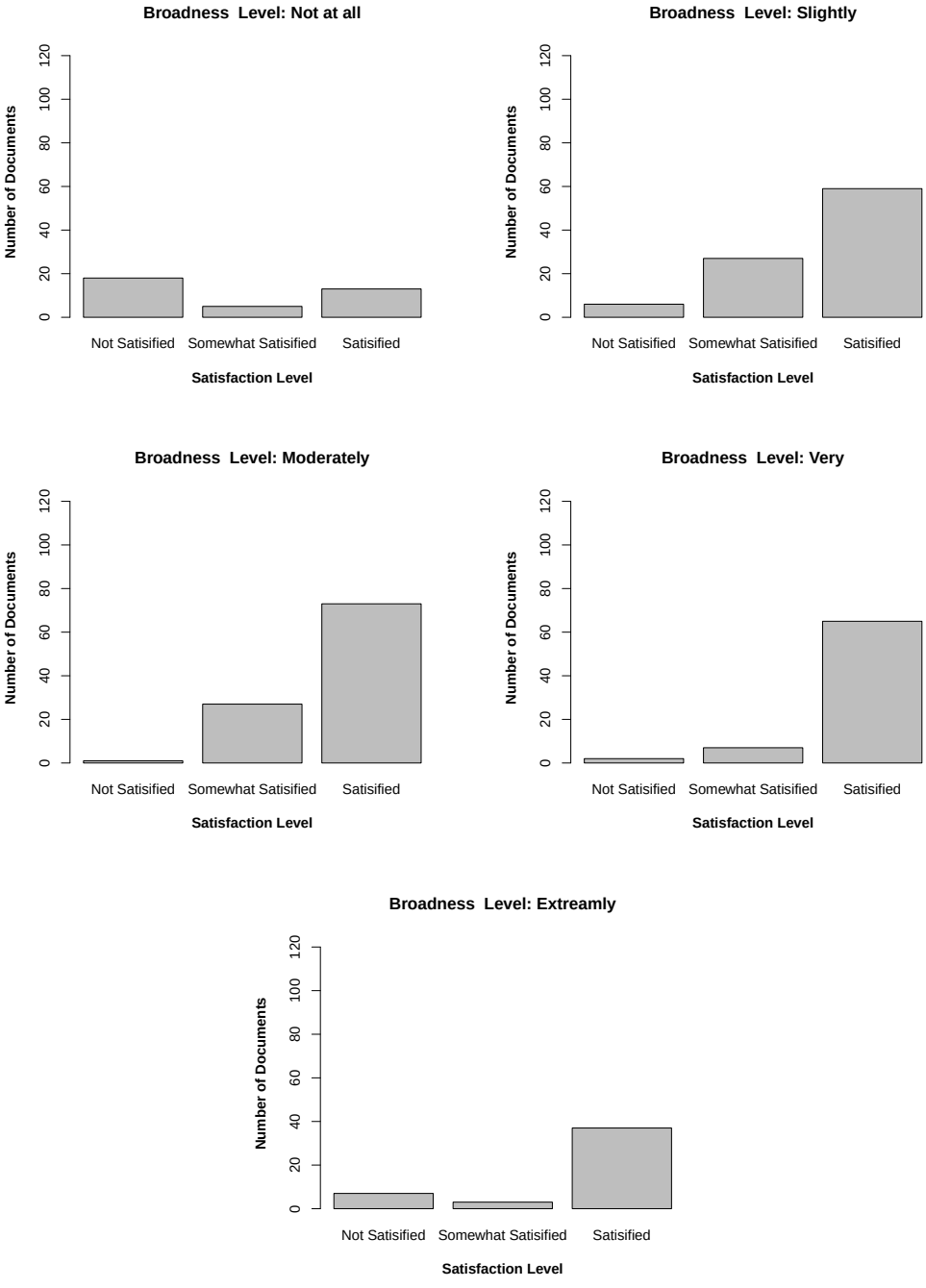


Figure 5.16: Broadness Level and Satisfaction Level Frequency.

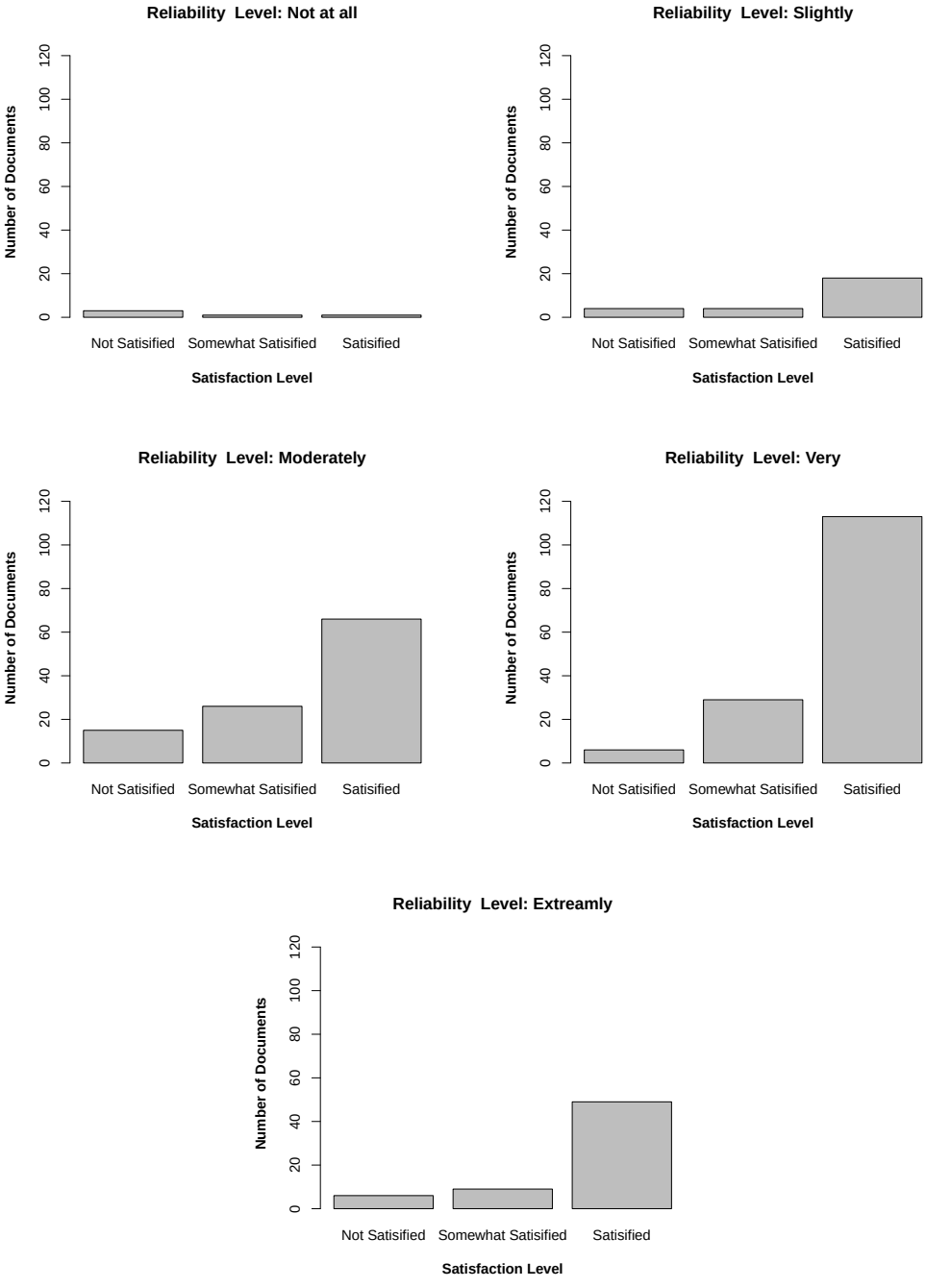


Figure 5.17: Reliability Level and Satisfaction Level Frequency.

5.5 Understandability and learning factors

In order to build better retrieval systems that improve learning, we need to identify what may promotes users' understandability of the topic (**LF2**). In an effort to achieve this, and to address our **RQ2**, we conducted a Spearman correlation analysis between our learning factors and user understandability factor **LF2**. Table [5.4] shows the result of the correlation.

Learning Factor	r	p-value
LF1: *Readability*$^{+}$	**0.80**	$\ll 0.001$
LF3: *Novelty*$^{-}$	-0.09	0.087
LF4: *Broadness*$^{-}$	**-0.14**	0.006
LF5: *Detailedness*$^{-}$	**-0.24**	$\ll 0.001$
LF6: *Reliability*$^{+}$	**0.17**	0.012

Table 5.4: Spearman correlation of learning factors with Understandability (LF2).

The readability level of text can have an effect on the success of communicating instructional messages effectively to the user, thus influencing their overall understandability. Our result indeed shows that readability of content (**LF1**) and users' understandability (**LF2**) are highly correlated.

Cognitive load theory assumes a limited capacity of working memory on the learner and show concern with cognitive load influencing the learning activity (Kirschner, 2002). Documents covering many subtopics (**LF4**), or providing intricate, detailed information (**LF5**) could increase cognitive load on the users due to the amount of information and thus negatively influencing users' understandability. Both **LF4** and **LF5** factors show a negative correlation with understandability.

Document judged as accurate and reliable (**LF6**) positively correlate with understandability. Novelty of information (**LF3**) has a negative weight but not significant.

5.6 User factor analysis

5.6.1 Relationships between learning factors

To understand the co-occurrence relation among the learning factors, we conduct an exploratory factor analysis (EFA) to determine the appropriate number of latent components (also called latent factors) we should use in our confirmatory factor analysis (CFA) (Larry, 2013).

We use *parallel analysis* (O'connor, 2000) to determine how many latent components we should retain. Figure 5.18 shows the parallel analysis scree plot. The number of components that lay above the red are the number of components we should retain. The result of the parallel analysis method on our data suggest that the number of factor analysis (FA) components to retain is 3, we therefore use k=3 components for our CFA. In our CFA, we use Varimax orthogonal rotation method which assumes components are not correlated (Fabrigar et al., 1999).

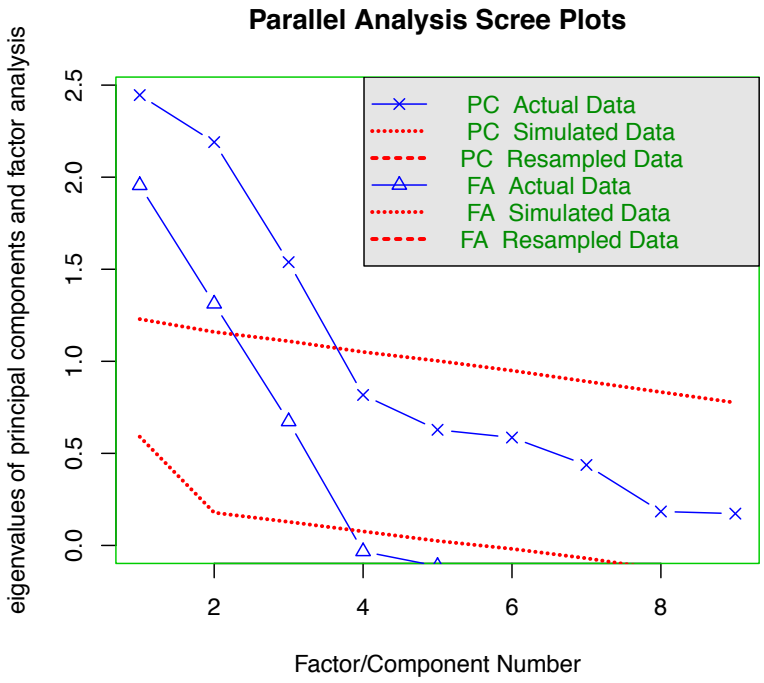


Figure 5.18: Parallel Analysis Scree Plot.

Table 5.5 shows the three components standardized loadings for the variables we kept. The value h^2 denotes the *communality estimate* (Worthington and Whittaker, 2006), which estimates the proportion of the variance of the variable that is shared with other variables. A value of $h^2 < 0.40$ indicates that the variable is less strongly correlated with its corresponding component (Worthington and Whittaker, 2006).

From Table 5.5, ranking related variables **RO** and **FO** are loaded together into component C2, while user-based (**LF1** and **LF2**) and document-based (**LF4** and **LF5**) factors are loaded separately into two different components, C1 and C2, respectively. **LF3** and **LF6** have $h^2 < 0.40$ thus less strongly to be correlated to its loaded component.

	C1	C2	C3	h^2
RO: *Rank*		**0.90**		0.83
FO: *Final Rank*	-0.13	**0.88**	-0.11	0.80
LF1: *Readability*	**0.92**	-0.11	-0.12	0.87
LF2: *Understandability*	**0.86**	-0.12	-0.15	0.77
LF3: *Novelty*			0.58	0.34
LF4: *Broadness*		-0.19	**0.68**	0.51
LF5: *Detailedness*	-0.15		**0.70**	0.51
LF6: *Reliability*	0.33		0.36	0.25
S1: *Satisfaction*	0.32	-0.15	0.32	0.23

Table 5.5: Components loading table for learning factors, ranks, and document satisfaction.

5.6.2 Relationship between search and topic learning variables

Our pre-task and post-tasks questionnaires were asked to measure the user interest and familiarity in the topic and their overall experience after searching and learning about the topic. In order to determine relationships between related variables before and after the learning process, we have conducted another factor analysis of pre-task and post-task variables. We choose number of components to retain to be $k = 2$ to discover whether there are two prominent groups of users were evident from the data. Table 5.6 shows the

two components standardized loadings for the pre-task and post-task questions. Figure 5.19 shows the result factor analysis biplot. The triangles points are users points plotted by their factor scores. Each variable from Table 5.6 is shown as a vector whose coordinates are the variable loadings. Vectors that are similar in length and direction indicate highly correlated variables.

	C1	C2	h^2
PR1: Familiarity with the topic.	0.24		0.05
PR2: Perceived difficulty of learning the topic.	-0.24	-0.31	0.15
PR3: Interest in learning the topic before searching and learning.	0.78	0.11	0.62
PT1: Interest in the topic after searching and learning.	0.62	0.40	0.55
PT2: Interest in exploring more new information.	0.87		0.76
PT3: Interest in sharing new information with friends.	0.44	0.34	0.31
PT4: User self judgment on learning useful information.	0.21	0.74	0.59
PT5: Ability to develop new ideas and perspectives.	0.36	0.37	0.26
PT6: Content quality affect on learning.		0.38	0.14
PT7: Difficulty of learning the topic.	-0.11	-0.63	0.40
PT8: Difficulty in finding good learning material.		-0.50	0.25
PT9: Experience during the learning task.	-0.51	-0.41	0.42
PT10: User self grading score (0 to 100).	0.39	0.59	0.50
OQ3: Agreement on order effecting learning.	0.21		0.07
OQ4: How effective the ordering was on learning.	0.50		0.27

Table 5.6: Components loading table for pre-task and post-task questions.

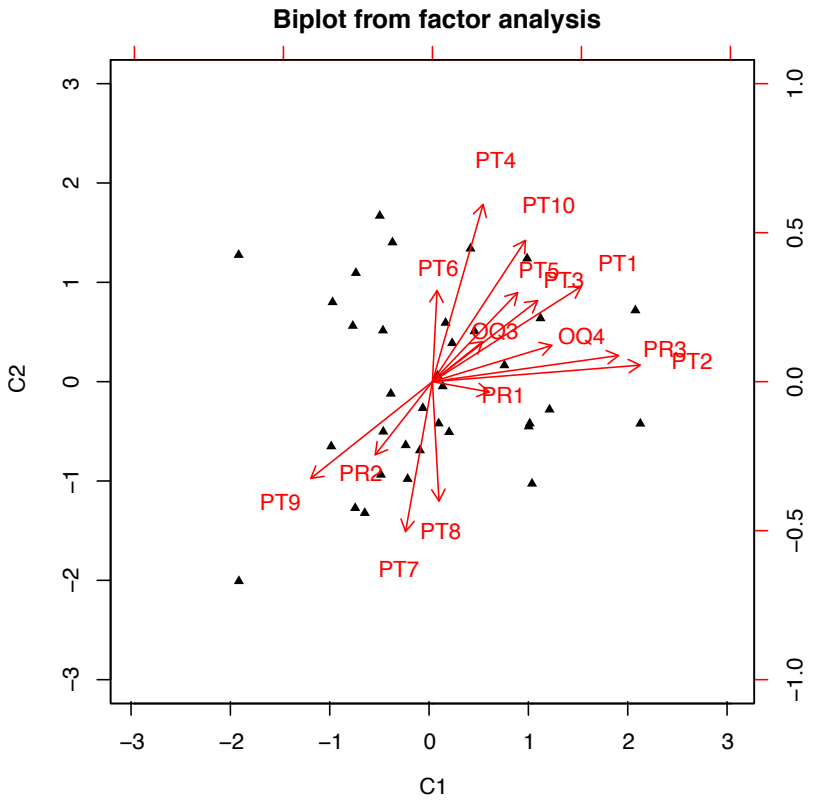


Figure 5.19: User factor analysis biplot of pre-task and post-task variables showing clustering of users points (black) and correlations between our used variables as vectors (red). Two vectors with close lengths and small angle between indicate two highly correlated variables.

From Table 5.6, variable measuring user interest in learning more about the topic (**PR3**) and variable measuring the user interest in exploring more new information about the topic (**PT2**) are clustered together into component C1.

5.7 Qualitative Data Result

Qualitative research methods can be an excellent method to gain rich and detailed information on why people act a certain way. We user our collected qualitative data to explain how learning through search is conducted and the justification on why the order of documents

can enhance subject's learning experience. To address our **RQ3** and **RQ4**, we explore the participants prospectives on why have saved their documents, how they re-ordered their saved documents, and how they felt the ordering will enhance their learning experience.

5.7.1 Saving Documents

Wikipedia as a source of knowledge

From the 350 saved documents, 39 were from Wikipedia. Articles from Wikipedia are generally considered good starting point to get a basic sense of the topic before moving into more in-depth content. We further look into participants answer to **DF1** to understand why they have saved Wikipedia documents. Participants showed interest in saving Wikipedia documents because:

— "Wikipedia is usually the first place I go to to learn about something new. From experience, the information has been reliable for learning purposes, even though I know that it can be changed by anyone. It usually has a good opening introduction, and touches on many different sub-topics related to the topic at hand. I usually just scroll through the headings to get a general idea of the possible sub-topics." *Participant 12*

— "Wikipedia is often my go to for learning about new things as it is often put into simpler terms and covers a range of information. I prefer to use Wikipedia so I have an idea of what information I should look for when I want to go deeper in my research." *Participant 14*

— "Wikipedia is a valuable source to read when looking to get a general understanding of a concept". *Participant 20*

— "Wikipedia has always been my go to website for any kind of study to start with. It provides me with information from basics to good detailing." *Participant 15*

— "It explains every topic briefly, starting with the definition of an oil spill, to human and environmental effects, occurrence rates and so on." *Participant 21*

— "Wikipedia is not a very accurate source and I would definitely not use this in my papers that I need to write for school as a reference. However, it is a good site to get a brief idea of the definition of the topic. It also gave a lot of information and topics relating to open data, which is very helpful." *Participant 26*

In Figure 5.20, we show where participants placed Wikipedia articles in their saved documents list in the order they want to read (**RO**) and their final ordering after they

have read all their saved documents (**FO**). Many participants have considered Wikipedia as their source for acquiring high-level knowledge in their assigned topic and have placed these documents higher in the list.

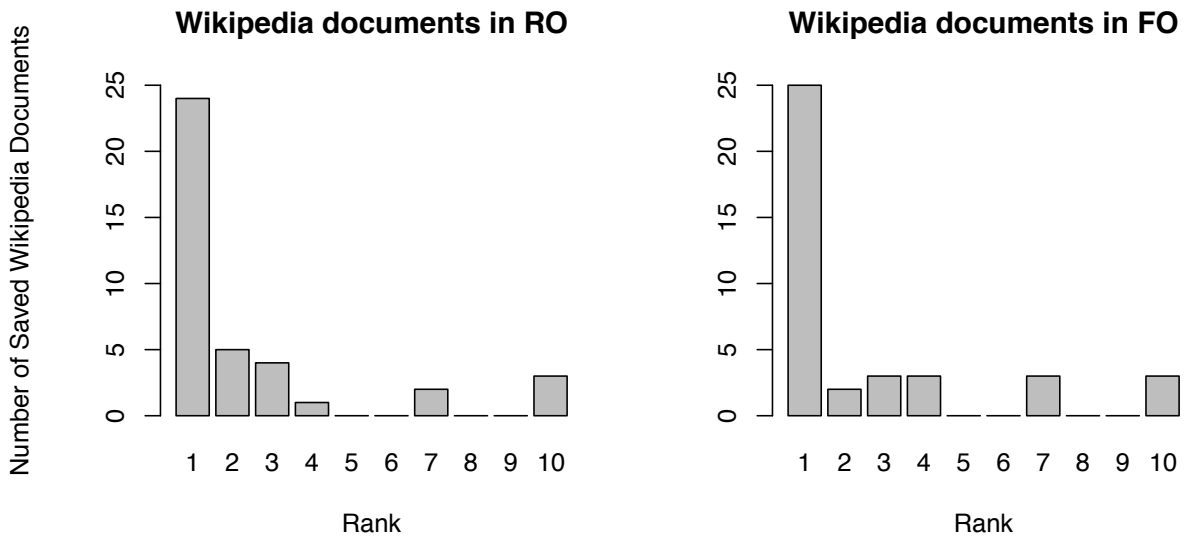


Figure 5.20: Participants placement of Wikipedia documents in their reading order (RO) and final order (FO).

Validating knowledge from broader documents

Figure 5.21 shows participants answers to how reliable they felt were the Wikipedia documents they have saved. Although many people have indicated *Moderately* and *Slightly* levels for Wikipedia documents, they were overall satisfied with the documents (Figure 5.22) with some participants indicating that:

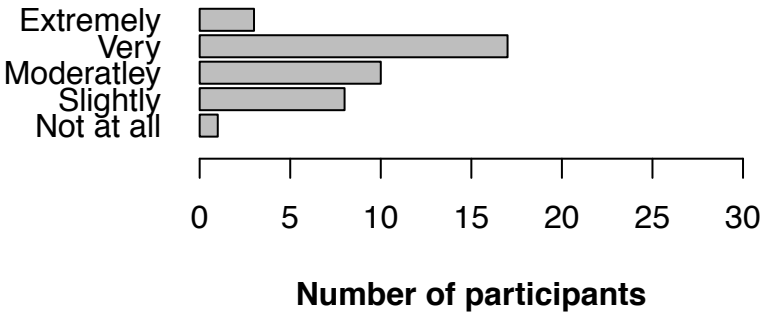


Figure 5.21: Participants labeling of reliability level of Wikipedia documents

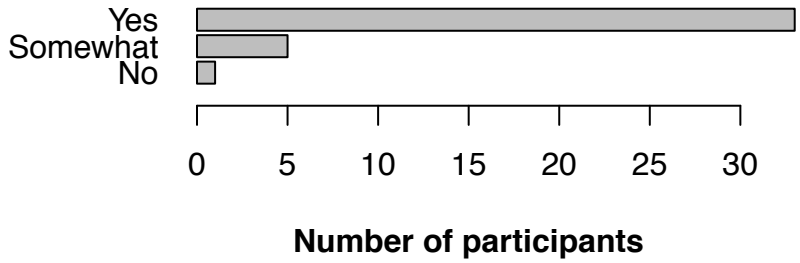


Figure 5.22: Participants labeling of satisfaction level of Wikipedia documents

— "I have highly ranked sources that provide definitions and clearly explain what Open Data is. Although Wikipedia is not a very reliable source, it gives me a brief summary of the concept so I am able to understand the rest of my research easier". *Participant 20*

— "Even if the information was wrong or biased, it could still point me in the right direction for new questions'.' *Participant 6*

— "... after, read several sites that are more legitimate to verify Wikipedia information'.' *Participant 12*

5.7.2 Document order

Participants may wish to change their reading order of documents after they have read all of the documents. Participants, after reading all documents, made more ordering changes

to lower-ranked or middle-ranked documents of their list than higher-ranked documents. In Figure 5.23, we show the percent of change from the reading order (**RO**) to the final order (**FO**) in documents at each rank. A zero percent at rank r indicates that all documents at rank r in the reading order were not modified and their rank was the same in the final order.

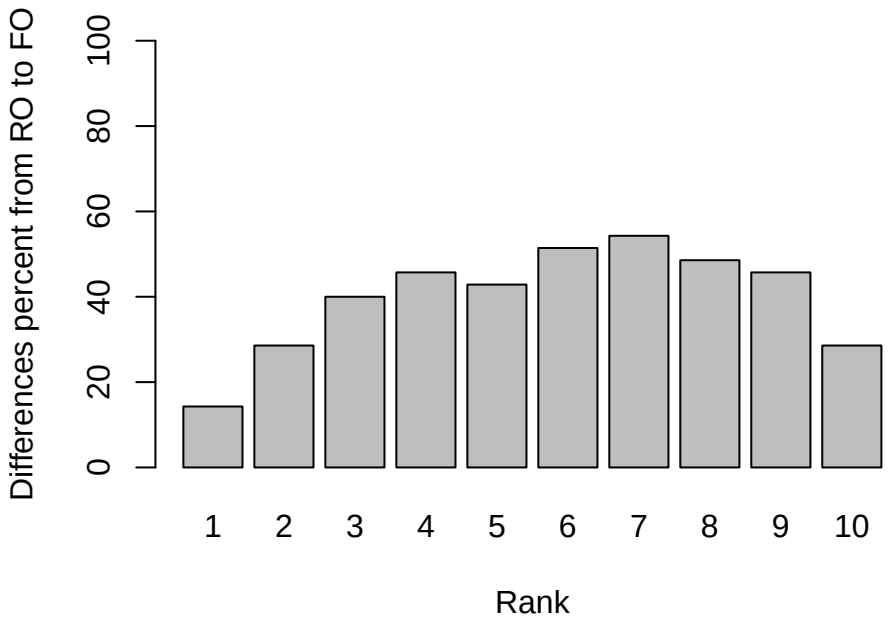


Figure 5.23: Percent of change in document positions at each rank. A high percent at one rank indicates that many documents at that rank in RO are no longer in the same position in FO.

Participants written feedback on the ordering tasks revealed useful information regarding our **RQ3**. We examine the final reading order by the participants and whether the order they would like to read the documents in can be correlated with the learning factors. Table 5.7 and 5.7 show the Spearman correlation coefficients and their p-values for correlation between rank of the document and the learning factors in the reading order (**RO**) and the final order (**FO**), respectively. Readability, understandability and broadness are

significant with negative weights. Negative weights indicates that the higher these factor are judged, their position in the ranked list is ranked higher.

Learning Factor	r	p-value
LF1 *Readability*$^-$	**-0.15**	0.005
LF2 *Understandability*$^-$	**-0.14**	0.011
LF3 *Novelty*$^-$	-0.07	0.168
LF4 *Broadness*$^-$	**-0.18**	$\ll 0.001$
LF5 *Detailedness*$^+$	0.06	0.264
LF6 *Reliability*$^+$	0.05	0.395

Table 5.7: Spearman correlation of learning factors with rank in reading order (RO).

Learning Factor	r	p-value
LF1 *Readability*$^-$	**-0.17**	0.0016
LF2 *Understandability*$^-$	**-0.17**	0.0014
LF3 *Novelty*$^-$	-0.07	0.197
LF4 *Broadness*$^-$	**-0.22**	$\ll 0.001$
LF5 *Detailedness*$^+$	-0.02	0.709
LF6 *Reliability*$^+$	0.00	0.963

Table 5.8: Spearman correlation of learning factors with rank in final order (FO).

Next, we show participants written responses on how they ordered the document and its effect on their learning experience.

5.7.3 Explaining the documents ordering

As participants save their ten documents, different useful information on their assigned topic can be scattered in different documents. As a result, a natural question to ask is the

order in which information should be presented to the learner.

We looked at **OQ1** and **OQ6** to further understand how documents were ordered. To enhance the learning process, and confirming what we have shown earlier, many participants felt that it is more effective to start with broader documents:

— "I want to read articles that are generalized and geared towards an audience that doesn't know much at all about the topic, and eventually browse links that have more specific information. At the end I saved a link to a webpage that has a long list of open data around the world, so that information from any one country may be found." *Participant 2*

— "I choose this order because I believe it portrays the most important issues regarding oil spills in order. For example, the Wikipedia article (first) gives a very general overview of oil spills, and gives the reader the majority of required knowledge to learn about oil spills. The second article, which talks about oils effects on animals, goes into more detail about the wildlife damage that an oil spill can create. This applies to the rest of the articles, but for different aspects of oil spills. As you move from article to article, you learn about one more aspect of oil spills. I also ordered the articles from most to least important, starting with environmental impacts and then moving on to economic and other side effects." *Participant 7*

— "I organized the articles by coverage of the topic. Since I don't know anything about it, I need something broad (like Wiki) to get me warmed up. Then some more thorough definitions" *Participant 8*

— "I think the most important thing about one topic is its definition and other related definitions, so I put those definition-related webpages first, and following with examples of this topic." *Participant 9*

— "The first document lays out the info well and briefly. I ordered the documents according to clarity, layout, how well it related to the topic, and the ability to answer my question of "what is open data" in a simple way while giving all the info I was looking for." *Participant 10*

— "I ordered each link based on how detailed the information will be. The more high-level and broad the information, the higher it was ordered. The links placed at the bottom would either give me more detailed information for curiosity purposes, or were presupposed to have similar information than the sources already chosen and placed at the top." *Participant 12*

— "I decided to order the list in a method where I would be able to obtain a brief summary of all the main points of the topic first and then afterwards I would be able to see the impact

of the topic." *Participant 27*

— "I ordered the first few documents in accordance with their being an overview of the issue of oil spills" *Participant 31*

5.7.4 Effect of ordering to learning process

We asked participant on whether the ordering has affected their learning process in **OQ3** and how effective was the ordering towards their learning process in **OQ4**. Figures 5.24, 5.25 show participants answers to **OQ3** and **OQ4**, respectively.

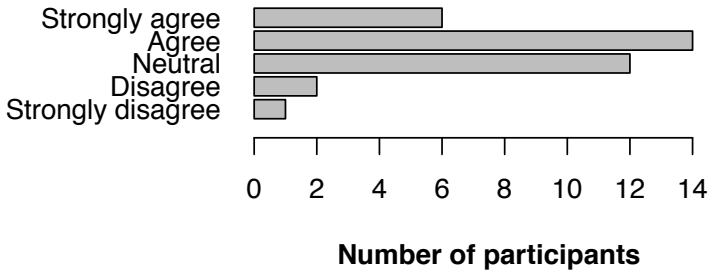


Figure 5.24: Participants answer to whether the ordering has affected their learning process (OQ3).

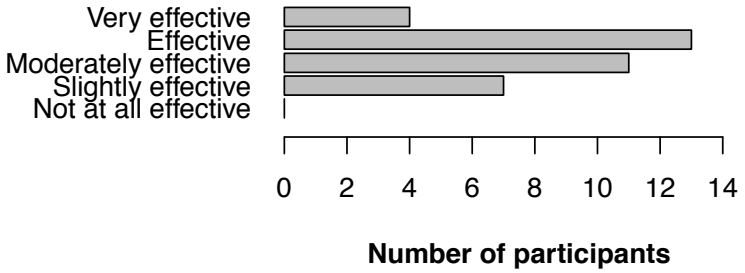


Figure 5.25: Participants answer to how effective was the ordering towards their learning process (OQ4).

Many participants agreed that the ordering of document has affected their learning with many participants indicating that it has been *moderately* effective or higher.

We also asked for their written responses in **OQ2** and **OQ5** to further understand the effect of ordering on their learning process. Some of the participants answers were:

—"I think that ordering the documents from most general to most specific allowed me to learn in a very linear and simple way. My thought process followed the order of the articles, and going to the next article was intuitive and allowed me to learn at my own pace." *Participant 7*

—"I was able to gain a broad, general understanding of the topic before delving into the more detailed aspects of it." *Participant 12*

—"It provided a clear and effective progression from basic understanding of the surface to a detailed analysis of some of the deeper issues." *Participant 30*

— "I as an individual was more keen during the start of learning when I had no knowledge about the topic. Studying relevant and important topics in the beginning helps in developing interest and base for the study. Structuring by the definitions followed by applications helped in better understanding." *Participant 22*

— "I think it was good that I didn't look at the more complicated ones first. It would have just made me confused and frustrated because they weren't giving me a straight, simple answer on what open data actually is". *Participant 18*

— "First by getting an overview of the meaning of open data will allow me to understand the topic. After that, the different opinions in the articles will help me formulate a better idea and understanding." *Participant 26*

— "I believe that ordering will make the learning more comprehensible/intelligible as it allows the reader to digest the information in a way and method that makes sense to them which will help them in understanding and processing the information" *Participant 27*

—"I have ordered the documents in a manner such that the first few documents help me understand the concept of Open data... This is because the data is ordered in a manner such that it provides the information in a clear systematic manner and helps me understand the topic logically, without any confusions." *Participant 22*

—"I am able to understand and learn a concept more thoroughly when I first look at clear and concise definitions and general examples. To maximize my understanding, it is then helpful to look at reliable government documents to further explore and understand examples of the concept" *Participant 20*

—"Forms a learning gradient. Always expanding on the previously obtained knowledge to make sure it sticks." *Participant 18*

Chapter 6

Conclusion

To improve the search process in supplementing learning or to design new tools that assist the learning process requires a new evaluation metric that takes into consideration the outcome of the search. After a learner completes his search, different forms of learning that are more complex than remembering, such as analyzing and applying and evaluating concepts, should be processed to allow for a more comprehensive learning experience.

To further improve the quality of information retrieval systems with focus in improving human learning, we clearly need more knowledge of what influences a learners cognitive process.

In this thesis, we described a lab-based study to investigate the influence of learning factors on determining user satisfaction at a document-level. We have proposed six learning factors associated with search systems results and its users.

We found that novelty of the information and broadness of the document content influence document satisfaction. To investigate whether participant have actually learned during the study, we measured their prior knowledge using a pre-test question and their acquired knowledge after the search in a post-quiz question. Our correlation analysis showed that the number of documents judged by users as moderately novel or higher correlate with their knowledge gain score. The qualitative results of the study show that there is indeed a reading-order preference among learners. Participants explained that a high-level to detailed ordering of content help with their thought process and is appropriate for learning progression. We also recognize a common theme among participant on why they save certain documents. Participants felt that Wikipedia articles are good starting point for learning a new topic due to the document's broadness in covering a range of subtopics and its ability to communicate content in simple terms.

Our study provides information retrieval system designers with insight on the documents a system should return if the aim is to improve user-learning and user satisfaction in the returned documents. We also show insight on the preferred document ordering that user prefer in learning-related tasks.

6.1 Future work

An interesting area of future work would be to investigate approaches in predicting user understandability of documents or identifying characteristics of documents in a systematic way. Level of broadness may be assessed by looking into documents text and investigating how many possible different subtopics can be retrieved.

Another area of future work that may supplement the learning process is through using a semantic-based question generating tools that generates questions from documents. Answers to these text-generated-questions are provided in the text the tool have used to generate the questions. Instead of allowing users to search through retrieved documents, read documents, and determine if they contain the answers to the questions users have in mind, users search through the list of questions and click on the questions they wants an answer to. Different meta-data of the question and its answers can be used for ranking questions. For example, readability level of the text passage the questions were extracted from can be used to determine the level of question difficulty and can be used for re-ranking questions. A large set of similar or exact questions generated from different documents of some topic may indicate that they are basic or popular questions, and are questions people should know the answer to when trying to learn the topic. Answers to questions in these sets can also have different writing style as they were extracted from different documents. Searching through questions instead of documents could allow for less time in the searching process and more time reading and understanding answers.

References

Allan, J., B. Croft, A. Moffat, and M. Sanderson (2012). Frontiers, challenges, and opportunities for information retrieval: Report from swirl 2012 the second strategic workshop on information retrieval in lorne. In *ACM SIGIR Forum*, Volume 46, pp. 2–32. ACM.

Anderson, L. W., D. R. Krathwohl, and B. S. Bloom (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Allyn & Bacon.

Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals. pp. 1–207.

Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of documentation 56*(1), 71–90.

Clark, R. C. and R. E. Mayer (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons.

Clarke, C. L., N. Craswell, I. Soboroff, and A. Ashkan (2011). A comparative analysis of cascade measures for novelty and diversity. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 75–84. ACM.

Clarke, C. L., M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon (2008). Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 659–666. ACM.

Collins-Thompson, K., P. N. Bennett, R. W. White, S. De La Chica, and D. Sontag (2011). Personalizing web search results by reading level. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 403–412. ACM.

Collins-Thompson, K., S. Y. Rieh, C. C. Haynes, and R. Syed (2016). Assessing learning outcomes in web search: A comparison of tasks and query strategies. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, pp. 163–172. ACM.

Craik, F. I. and R. S. Lockhart (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior 11*(6), 671–684.

DeStefano, D. and J.-A. LeFevre (2007). Cognitive load in hypertext reading: A review. *Computers in human behavior 23*(3), 1616–1641.

Ebbinghaus, H. (2013). Memory: a contribution to experimental psychology. *Annals of neurosciences 20*(4), 155–156.

Fabrigar, L. R., D. T. Wegener, R. C. MacCallum, and E. J. Strahan (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods 4*(3), 272.

Ford, N., D. Miller, and N. Moss (2001). The role of individual differences in internet searching: An empirical study. *Journal of the American Society for Information Science and technology 52*(12), 1049–1066.

Freund, L., S. Dodson, and R. Kopak (2016). On measuring learning in search: A position paper.

Freund, L., J. He, J. Gwizdka, N. Kando, P. Hansen, and S. Y. Rieh (2014). Searching as learning (sal) workshop 2014. In *Proceedings of the 5th Information Interaction in Context Symposium*, IIiX '14, New York, NY, USA, pp. 7–7. ACM.

Freund, L., R. Kopak, and H. OBrien (2016). The effects of textual environment on reading comprehension: Implications for searching as learning. *Journal of Information Science 42*(1), 79–93.

Gwizdka, J., P. Hansen, C. Hauff, J. He, and N. Kando (2016). Search as learning (sal) workshop 2016. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 1249–1250. ACM.

Haure, J. and T. Kossoski (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data.

Heinström, J. (2005). Fast surfing, broad scanning and deep diving: The influence of personality and study approach on students' information-seeking behavior. *Journal of documentation 61*(2), 228–247.

Heinström, J. (2006). Broad exploration or precise specificity: Two basic information seeking patterns among students. *Journal of the American Society for Information Science and Technology 57*(11), 1440–1450.

Holsti, O. R. (1969). Content analysis for the social sciences and humanities.

Jac J. W. Andrews, J. L. (2015). *Diversity education : understanding and addressing student diversity*. Nelson Education.

Kim, J. Y., K. Collins-Thompson, P. N. Bennett, and S. T. Dumais (2012). Characterizing web content, user interests, and search behavior by reading level and topic. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pp. 213–222. ACM.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning.

Kuhlthau, C. C. (2004). *Seeking meaning: A process approach to library and information services*. Libraries Unltd Incorporated.

Larry, H. (2013). Advanced statistics in research: Reading, understanding, and writing up data analysis results. *Publisher: Shadow Finch Media LLC (January 7, 2013). ISBN-13*, 978–0985867003.

Marks, C. B., M. J. Doctorow, and M. C. Wittrock (1974). Word frequency and reading comprehensiony1. *The Journal of Educational Research 67*(6), 259–262.

Mayer, R. E. (1977). The sequencing of instruction and the concept of assimilation-to-schema. *Instructional Science 6*(4), 369–388.

O'connor, B. P. (2000). Spss and sas programs for determining the number of components using parallel analysis and velicers map test. *Behavior research methods, instruments, & computers 32*(3), 396–402.

Raman, K., P. N. Bennett, and K. Collins-Thompson (2013). Toward whole-session relevance: exploring intrinsic diversity in web search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pp. 463–472. ACM.

Royer, J. M., B. A. Greene, and G. M. Sinatra (1987). The sentence verification technique: A practical procedure for testing comprehension. *Journal of Reading 30*(5), 414–422.

Schunk, D. H. (1996). Learning theories. *Printice Hall Inc., New Jersey*, 1–576.

Verma, M., E. Yilmaz, and N. Craswell (2016). On obtaining effort based judgements for information retrieval. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pp. 277–286. ACM.

Waldrop, M. M. (2013). Campus 2.0. *Nature 495*(7440), 160.

Wang, M. C., G. D. Haertel, and H. J. Walberg (1990). What influences learning? a content analysis of review literature. *The Journal of Educational Research 84*(1), 30–43.

Wittrock, M. and J. F. Carter (1975). Generative processing of hierarchically organized words. *The American Journal of Psychology*, 489–501.

Wittrock, M. C. (1974). Learning as a generative process 1. *Educational psychologist 11*(2), 87–95.

Worthington, R. L. and T. A. Whittaker (2006). Scale development research a content analysis and recommendations for best practices. *The Counseling Psychologist 34*(6), 806–838.

Zhang, D., J. L. Zhao, L. Zhou, and J. F. Nunamaker Jr (2004). Can e-learning replace classroom learning? *Communications of the ACM 47*(5), 75–79.

Zumbach, J. and M. Mohraz (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior 24*(3), 875–887.

APPENDICES

Appendix A

Collected Data

Task	Question	Response type
Demographic	**DQ1:** Age	0-99
	DQ2: Gender.	Male-Female
	DQ3: Student Degree Level.	Multiple choice degrees
	DQ4: Faculty.	Multiple choice faculties
	DQ5: Major.	Written response
	DQ6: How often do you search the Internet to learn about a new topic?.	Likert scale multiple choice
	DQ7: I sometimes struggle to find good learning material for the topic I'm trying to learn.	Likert scale multiple choice
	DQ8: I usually use search engines as my primary source when trying to learn a new topic	Likert scale multiple choice
	DQ9: I consider my-self as a fast learner	Likert scale multiple choice
	DQ10: General feedback from the participant.	Written response
Pre-task	**PR1:** How much do you know about the topic?	1=Nothing 5=Very familiar
	PR2: How difficult do you think it will be for you to learn the topic?	1=Very Easy 5=Very Hard
	PR3: How interested are you in learning the topic?	1=Not at all 5=Very
	PR4: Please write what you know about this topic with 3-5 sentences	Written response
Ordering	**OQ1:** Explain the ordering you have provided in the previous task	Written response
	OQ2: Explain why you think the ordering will make the learning more comprehensible and intelligible	Written response
	RO: Preferred reading order of the saved documents	Ranked list

Reading task	**S1:** Afer reading the document, are you satisfied (happy) you have saved this document to learn from?	1=No 3=Yes
	LF1: The document language is easy to read	1=Not at all 5=Extremely
	LF2: The document was easy to understand	
	LF3: The information presented in this document is new to me	
	LF4: The document covers many aspects of the topic	
	LF5: The information presented in this document is intricate (very complicated/detailed)	
	LF6: The information presented in this document is reliable and accurate	
	DF1: Please write why you have saved this document	Written response
Post-task	**PT1:** I became interested in this topic.	1=Nothing 5=Very familiar
	PT2: I would like to find more information about this topic.	
	PT3: I would like to share what I learner with my friends.	
	PT4: I learned useful information as a result of the documents I have read.	
	PT5: I was able to develop new ideas or perspectives.	
	PT6: The quality and readability of the documents content influenced my learning process.	1=Strongly disagree 5=Strongly agree
	PT7: How difficult was it for you to learn the topic?	1=Very easy 5=Very hard
	PT8: How difficult was it for you to find good learning material about the topic?	
	PT9: How was your mood during the task?	1=Very enjoyable 5=Very not enjoyable

	PT10: How would you grade your learning outcome?	0 to 100
	OQ3: The order of documents affected my learning process.	1=Strongly disagree 5=Strongly agree
	OQ4: How effective was the ordering of documents towards your learning process?	1=Not at all effective 5=Very effective
	OQ5: If you feel like the ordering of documents affected your learning process in any way, please describe how.	Written response
Quiz [a]	**Q1 Remembering:** Assessing the level of recognizing facts and concepts.	Written response
	Q2 Understanding: Assessing the level of perceiving and comprehending the instructional message.	
	Q3 Applying: Assessing the level of ability to carry out procedures in a given situation.	
	Q4 Analyzing: Assessing the level of recognizing multiple concepts and their relations.	
	Q5 Evaluating: Assessing the level of critiquing and knowledge a er reading and searching.	
	Q6 Creating: Assessing the level of ability to generate new ideas and write creative questions.	
Order	**FO:** Final order of documents after reading all documents	Ranked list
	OQ6: Please justify/explain the ordering you have provided.	Written response

Table A.4: Inventory of questions asked to participants.

[a] See 4.5 for the exact questions asked for each topic

Appendix B

Ethics

B.0.1 Recruitment Poster

UNIVERSITY OF WATERLOO
FACULTY OF MATHEMATICS
David R. Cheriton School
of Computer Science

Participants Needed for Research in Search for Learning

- We are looking for volunteers to take part in a study of **search for learning** (if you use Google, Yahoo! or Bing to search the web for learning material) who are fluent in English (reading/writing) and able to use a computer.
- As a participant in this study, you would be asked to complete demographic and task-related questionnaires, search/view 10 documents, reorder documents and answer a quiz.

$20 for participating in a study

- Your participation would involve 1 session taking approximately 1.5 hours

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE#21633).
If you have questions for the Committee contact the Chief Ethics Officer, Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca.

- For more information about this study, or to volunteer for this study, please contact:

Mustafa Abualsaud
Department of Computer Science
at
Email: m2abuals@uwaterloo.ca

Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca
Please contact: m2abuals@uwaterloo.ca

B.0.2 Ethics Form

ORE OFFICE USE ONLY

ORE #_____

APPLICATION FOR ETHICS REVIEW OF RESEARCH INVOLVING HUMAN PARTICIPANTS

Please remember to PRINT AND SIGN the form and forward with all attachments to the Office of Research Ethics, EC5, 3rd floor.

Question A5 is incomplete. Please indicate Level of Project (1-General Info).

A. GENERAL INFORMATION

1. Title of Project: Documents Content Effect While Searching

2. a) Principal and Co-Investigator(s)
NEW As of May 1, 2013, all UW faculty and staff listed as investigation must complete the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans Tutorial, 2nd Ed. (TCPS2) prior to submitting an ethics application. The tutorial takes at least three hours; it has start and stop features.

Name Department Ext: e-mail:

2. b) Collaborator(s)
NEW As of May 1, 2013, all UW faculty and staff listed as investigation must complete the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans Tutorial, 2nd Ed. (TCPS2) prior to submitting an ethics application. The tutorial takes at least three hours; it has start and stop features.

Name Department Ext: e-mail:

3. Faculty Supervisor(s)
NEW As of May 1, 2013, all UW faculty and staff listed as investigation must complete the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans Tutorial, 2nd Ed. (TCPS2) prior to submitting an ethics application. The tutorial takes at least three hours; it has start and stop features.

Name	Department	Ext:	e-mail:
Mark Smucker	Management Sciences		mark.smucker@uwaterloo.ca

4. Student Investigator(s)

Name	Department	Ext:	e-mail:	Local Phone #:
Mustafa Abualsaud	Computer Science, School of		m2abuals@uwaterloo.ca	

5. Level of Project: **Specify Course:**

Research Project/Course Status:

6. Funding Status (If Industry funded and a clinical trial involving a drug or natural product or is medical device testing, then Appendix B is to be completed):

Is this project currently funded? Yes

- If Yes, provide Name of Sponsor and include the title of the grant/contract: NSERC

- If No, is funding being sought OR if Yes, is additional funding being sought? No
- Period of Funding:

7. Does this research involve another institution or site? No
If Yes, what other institutions or sites are involved:

8. Has this proposal, or a version of it, been submitted to any other Research Ethics Board/Institutional Review Board? No

9. For Undergraduate and Graduate Research:

Has this proposal received approval of a Department Committee? Not Dept. Req.

10. a) Indicate the anticipated commencement date for this project: 7/1/2016

b) Indicate the anticipated completion date for this project: 7/1/2017

11. Conflict of interest: Appendix B is attached to the application if there are any potential, perceived, or actual financial or non-financial conflicts of interest by members of the research team in undertaking the proposed research.

B. SUMMARY OF PROPOSED RESEARCH

1. Purpose and Rationale for Proposed Research

a. Describe the purpose (objectives) and rationale of the proposed project and include any hypothesis(es)/research questions to be investigated. For a non-clinical study summarize the proposed research using the headings: Purpose, Aim or Hypothesis, and Justification for the Study. For a clinical trial/medical device testing summarize the research proposal using the following headings: Purpose, Hypothesis, Justification, and Objectives.
Where available, provide a copy of a research proposal. For a clinical trial/medical device testing a research proposal is required:

Purpose: This project is part of a research program aimed to investigating how people use search engines for learning purposed. In order to improve web search systems and how people learn new topics, we need to be able to better understand how people use these systems, and how do people learn new topics.

Hypothesis: We hypothesize that the web documents and the order in which users view them will influence their learning experience. Users that were shown documents that are more readable and less complex should perform better in the quiz than users that were given complex documents. We also hypothesize that people with less knowledge in the topic to-be-learned will struggle to find the right keywords to formulate their search quires.

Justification for the Study: A key part of this is that we know that people often use web search engines when trying to learn more about a topic, but the results from the these web search engines are not necessarily useful for learning the topic. Current web search engines are highly optimized in for topic relevance, but lack information on searching for the purposes or learning a new topic, how much knowledge gain do user have after using the system, and how should documents be presented to the user.

Objectives: This project will collect data on the behaviour of users when trying to search for a documents on a topic to be learned, as well as evaluating the documents they believe are good learning material. Participants will also undergo a quiz to determine how much knowledge they gained after they read the documents on the topic to be learned. With this data, we can measure the user performance, and how much does the documents a participant view may affect their quiz score.

b. In lay language, provide a one paragraph (approximately 100 words) summary of the project including purpose, the anticipated potential benefits, and basic procedures used.

In this study, we will ask participants to use our specifically designed web-search engine to view or search for 10 documents on the topic to be learned. Participants then will have to read the 10 documents in some order they specify. They will evaluate the documents after they read them. After collecting this data, we analyze how do users perform in the quiz and study how the documents they have read effects their quiz answers.

C. DETAILS OF STUDY

1. Methodology/Procedures

a. Indicate all of the procedures that will be used. Append to form 101 a copy of all materials to be used in this study.

Computer-administered task(s) or survey(s) None are standardized.
Unobtrusive observations
Logging of computer usage

b. Provide a detailed, sequential description of the procedures to be used in this study. For studies involving multiple procedures or sessions, provide a flow chart. Where applicable, this section also should give the research design (e.g., cross-over design, repeated measures design).

This study will have one phase with 13 different tasks. This protocol uses with

slight modifications the protocol given by Toms et al. "WiIRE: the Web interactive information retrieval experimentation system prototype," Information Processing and Management, 40, 2004, pp. 655675.

Protocol:

1. Introduction

2. Consent Form

3. Demographic and Background Questionnaire

4. Overview of Experiment

5. Tutorial

6. Study Quiz

8. Searching Task

9. Re-ordering documents task

10. Reading Task

11. Post Task Questionnaire

12. Quiz

13. Re-ordering documents task

14. Exit Questionnaire

15. Thank You

Phase 8 will involve the participants being presented with web pages formatted similar to popular web search engines such as Google, Yahoo, or Microsoft's Bing. Participants will be asked to use this interface to view document summaries, view the underlying documents and change search queries, evaluate the documents.

Phase 9, 13 will involve the participants to re-order the documents in the order they would like to learn.

Phase 10 will involve asking the participants to read the documents they have saved.

Phase 12 will involve asking the participants a set of written-answer questions on the topic to be learned.

We will collect timing information and associated computer usage data unobtrusively during both phases of the study.

c. Will this study involve the administration/use of any drug, medical device, biologic, or natural health product? No

d. Will you be using or processing any biological materials such as human blood, tissue, cells or bodily fluids in the proposed research?
No

2. Participants Involved in the Study

a. Indicate who will be recruited as potential participants in this study.
UW Participants:
Undergraduate students
Graduate students
Faculty and/or Staff

b. Describe the potential participants in this study including group affiliation, gender, age range and any other special characteristics. Describe distinct or common characteristics of the potential participants or a group (e.g., a group with a particular health condition) that are relevant to recruitment and/or procedures. Provide justification for exclusion based on culture, language, gender, race, ethnicity, age or disability. For example, if a gender or sub-group (i.e., pregnant and/or breastfeeding women) is to be excluded, provide a justification for the exclusion.

Adults fluent in English, familiar with web search (e.g. Google, Yahoo, Bing), and capable of unassisted use of a computer with keyboard, mouse, and LCD monitor.

c. How many participants are expected to be involved in this study? For a clinical trial, medical device testing, or study with procedures that pose greater than minimal risk, sample size determination information is to be provided.

30 to 60 plus a couple of participants during the pilot phase. The phase of the study will involve 2 topics. We know that human performance in text retrieval varies across both humans and the search topics. Using a 10x2 block design (10 participants and 2 topics forming a 10x10 Latin square), we will collect a minimum of 30 participants for each topic. The total minimum number of participants for the study is 60, plus a couple of pilot testing participants. This will be a convenience sample of students and other adults of the University of Waterloo community.

3. Recruitment Process and Study Location

a. From what source(s) will the potential participants be recruited?
Posters across campus and grad Studies Mailing List

b. Describe how and by whom the potential participants will be recruited. Provide a copy of any materials to be used for recruitment (e.g. posters(s), flyers, cards, advertisement(s), letter(s), telephone, email, and other verbal scripts).
The student invistigator will recruit the participants. We will post posters around the University of Waterloo campus, and send an email to the grad studies mailing list

c. Where will the study take place? On campus: CPH 4335

4. Remuneration for Participants
Will participants receive remuneration (financial, in-kind, or otherwise) for participation? Yes
If Yes, provide details:
Participants will be paid $20 for the study that they participate in. Should participants need to leave before completing all phases of the study, participants will be paid on pro-rated time-spent basis. It is expected that the participants will need to spend around 1 hour to complete the study.

5. Feedback to Participants

Describe the plans for provision of study feedback and attach a copy of the feedback letter to be used. Wherever possible, written feedback should be provided to study participants including a statement of appreciation, details about the purpose and predictions of the study, restatement of the provisions for confidentiality and security of data, an indication of when a study report will be available and how to obtain a copy, contact information for the researchers, and the ethics review and clearance statement.
Participants will be advised that if they are interested in the outcomes of the study, they may contact the principal investigator at a later time to learn about any resulting publications.

D. POTENTIAL BENEFITS FROM THE STUDY

1. Identify and describe any known or anticipated direct benefits to the participants from their involvement in the project.
There are no known direct benefits to the participants from their involvement in the project.

2.Identify and describe any known or anticipated benefits to the scientific community/society from the conduct of this study.
Information retrieval (text search) has become part of daily life for many Canadians, as well as people around the world. This study has the long term potential to allow researchers to better evaluate retrieval systems. With better evaluation tools that allow for faster and more accurate evaluations, the rate at which retrieval systems improve should increase. The study will also help design a new retrieval systems optimized for learning. With better retrieval systems, people are able to find good learning material previously hidden. With better learning material given to the user, the more affective their learning process will be.

E. POTENTIAL RISKS TO PARTICIPANTS FROM THE STUDY

1. For each procedure used in this study, describe any known or anticipated risks/stressors to the participants. Consider physiological, psychological, emotional, social, economic risks/stressors. A

study–specific current health status form must be included when physiological assessments are used and the associated risk(s) to participants is minimal or greater.

Minimal risks anticipated.
Minimal risks anticipated.

Participants will be asked to use a computer with keyboard, mouse, and LCD monitor to answer brief questionnaires as well as to read and answer questions according to given result lists. These activities are common to everyday life and pose no greater risk. The topics to be learned are "Oil spills" and "Open data" and both of them deal with matters people can encounter in normal life. All documents come from popular search engines.

2. Describe the procedures or safeguards in place to protect the physical and psychological health of the participants in light of the risks/stressors identified in E1.
As the study involves only minimal risk, no explicit procedures or safeguards will be in place other than to provide a safe, usable computer system in a university computing lab commonly used by students.

F. INFORMED CONSENT PROCESS

1. What process will be used to inform the potential participants about the study details and to obtain their consent for participation?
Information letter with written consent form

2. If written consent cannot be obtained from the potential participants, provide a justification for this.

3. Does this study involve persons who cannot give their own consent (e.g. minors)? No

G. ANONYMITY OF PARTICIPANTS AND CONFIDENTIALITY OF DATA

1. Provide a detailed explanation of the procedures to be used to ensure anonymity of participants and confidentiality of data both during the research and in the release of the findings.
All participants will be issued an anonymous identifier (ID). The mapping from a participant's name to the ID will be maintained for the length of the study in case the participant forgets the ID. This mapping will be kept in a locked cabinet in a secure location during the study and will be destroyed at the completion of the study. After the study concludes, there will be no way to identify a participant to the data. All computer usage will be with computers in a University of Waterloo computer lab and not with personally identifiable computers, i.e. participants will not use their own computer. All data collected will be retained indefinitely and will be used for research purposes. We may refer to individual participants when describing the results or the study, and in these cases, we will always refer to "participant 1" or some other similar anonymous name. Participants' names will never appear in any publication that results from this study.

2. Describe the procedures for securing written records, video/audio tapes, questionnaires and recordings. Identify (i) whether the data collected will be linked with any other dataset and identify the linking dataset and (ii) whether the data will be sent outside of the institution where it is collected or if data will be received from other sites. For the latter, are the data de-identified, anonymized, or anonymous?
Each of the topics that we provide to the participants are defined by previous researchers. The documents list for each the topics to be learned will be from the Bing search engine. The initial documents lists are manipulated and fixed due to the control of quality. We may choose to distribute the data collected to other researchers. All data will be anonymized at the conclusion of the study and prior to any distribution, but each participant's data will remain identifiable as coming from an individual, i.e. "participant 1", "participant 2", etc. We will not publicly share this data, i.e. the data would only be made available to other researchers for research purposes.

3. Indicate how long the data will be securely stored and the method to be used for final disposition of the data.

Paper Records

Confidential shredding after indefinitely year(s).

Electronic Data

Erasing of electronic data after indefinitely year(s).

Location: Principal investigator's office (paper) and on secure computers.

4. Are there conditions under which anonymity of participants or confidentiality of data cannot be guaranteed?

No

H. PARTIAL DISCLOSURE AND DECEPTION

1. Will this study involve the use of partial disclosure or deception? Partial disclosure involves withholding or omitting information about the specific purpose or objectives of the research study or other aspects of the research. Deception occurs when an investigator gives false information or intentionally misleads participants about one or more aspects of the research study. No

Researchers must ensure that all supporting materials/documentation for their applications are submitted with the signed, hard copies of the ORE form 101/101A. Note, materials shown below in bold are normally required as part of the ORE application package. The inclusion of other materials depends on the specific type of projects.

Protocol Involves a Drug, Medical Device, Biologic, or Natural Health Product

If the study procedures include administering or using a drug, medical device, biologic, or natural health product that has been or has not been approved for marketing in Canada then the researcher is to complete Appendix A. Appendix A is to be attached to each of the one copy of the application that are submitted to the ORE. Information concerning studies involving a drug, biologic, natural health product, or medical devices can be found on the ORE website.

Please **check** below all appendices that are attached as part of your application package:

- Recruitment Materials: A copy of any poster(s), flyer(s), advertisement(s), letter(s), telephone or other verbal script(s) used to recruit/gain access to participants.
- Information Letter and Consent Form(s)*. Used in studies involving interaction with participants (e.g. interviews, testing, etc.)
- Information/Cover Letter(s)*. Used in studies involving surveys or questionnaires.
- Data Collection Materials: A copy of all survey(s), questionnaire(s), interview questions, interview themes/sample questions for open-ended interviews, focus group questions, or any standardized tests.
- Feedback letter *

* Refer to sample letters.

NOTE: The submission of incomplete application packages will increase the duration of the ethics review process.

To avoid common errors/omissions, and to minimize the potential for required revisions, applicants should ensure that their application and attachments are consistent with the *Checklist For Ethics Review of Human Research Application*

Please note the submission of incomplete packages may result in delays in receiving full ethics clearance. We suggest reviewing your application with the Checklist For Ethics Review of Human Research Applications to minimize any required revisions and avoid common errors/omissions.

INVESTIGATORS' AGREEMENT

I have read the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, 2nd Edition (TCPS2) and agree to comply with the principles and articles outlined in the TCPS2. In the case of student research, as Faculty Supervisor, my signature indicates that I have read and approved this application and the thesis proposal, deem the project to be valid and worthwhile, and agree to provide the necessary supervision of the student.

NEW As of May 1, 2013, all UW faculty and staff listed as investigators must complete the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans Tutorial, 2nd Ed. (TCPS2) prior to submitting an ethics application. Each investigator is to indicate they have completed the TCPS2 tutorial. If there are more than two investigators, please attach a page with the names of each additional investigator along with their TCPS2 tutorial completion information.

_____ _____
Print and Signature of Principal Investigator/Supervisor **Date**

Completed TCPS2 tutorial:
___YES ___NO ___ In progress

_____ _____
Print and Signature of Principal Investigator/Supervisor **Date**

Completed TCPS2 tutorial:
___YES ___NO ___ In progress

Each student investigator is to indicate if they have completed the Tri-Council Policy Statement, 2nd Edition Tutorial (http://pre.ethics.gc.ca/eng/education/tutorial-didacticiel/). If there are more than two student investigators, please attach a page with the names of each additional student investigator along with their TCPS2 tutorial completion information.

_____ _____
Signature of Student Investigator **Date**

Completed TCPS2 tutorial:
___YES ___NO ___ In progress

_____ _____
Signature of Student Investigator **Date**

Completed TCPS2 tutorial:
___YES ___NO ___ In progress

FOR OFFICE OF RESEARCH ETHICS USE ONLY:

_____

Date

Jannet Ann Leggett, JD
Chief Ethics Officer
OR
Julie Joza, MPH
Senior Manager, Research Ethics
OR
Sacha Geer, PhD
Manager, Research Ethics
OR
Nick Caric, MDiv
Research Ethics Advisor
OR
Janet McKeown, PhD
Research Ethics Advisor

ORE 101
Revised September 2016

Copyright © 2001 University of Waterloo

B.0.3 Conset Form

Title of Project: Investigating search for learning

Investigators: Mustafa Abualsaud, m2abuals@uwaterloo.ca

This letter is an invitation to consider participating in a study I am conducting as part of my Master's degree in the Department of Computer Science at the University of Waterloo under the supervision of Mark S. Smucker. I would like to provide you with more information about this project and what your involvement would entail if you decide to take part.

Summary of the Project:

This project is part of a research program aimed to investigating how people use search engines for learning purposes. In order to improve web search systems and how people learn new topics, we need to be able to better understand how people use these systems, and how do people learn new topics. A key part of this is that we know that people often use web search engines when trying to learn more about a topic, but the results from the these web search engines are not necessarily useful towards their learning. The information collected in this study will be used to construct models of human searching that can be used to improve researchers' ability to evaluate and create better web search systems. With better evaluation tools, researchers should be able to speed the rate of improvement in text search systems.

Procedure:

Your participation in this study is voluntary. Participation involves searching or viewing, and evaluating 10 web documents and re-ordering them based on what the participant think is best for their learning experience, and answering a quiz on the topic to be learned.

You will be asked to complete several brief questionnaires and to search and evaluate 10 web documents regarding the topic to be learned. Your task is to learn about the topic from the set of 10 documents and complete a quiz to determine how well you have learned the topic. The questionnaires that you will be asked to complete consist of a demographic questionnaire, a questionnaire on the topic to be learned, and a questionnaire on the topic after you have completed the reading task. You also be given a sheet of paper for your own note-taking purposes.

To participate, you must be a fluent speaker of English and require no assistance with using a computer with a keyboard, mouse, and LCD monitor.

The study should take approximately 1.5 hour.

We will record both your answers and your interaction with the computer. We may also make note of and record anything we observe, including what you say, while you are participating in the study. We also collect any notes you take on a piece of paper during the task.

You may stop participating in the study at any point and withdraw your consent without penalty.

Confidentiality and Data Security:

You will be issued an anonymous identifier (ID) as a participant in this study. The mapping from your name to the ID will be maintained for the length of the study in case you forget the ID. This mapping will be kept in a locked cabinet in a secure location during the study and will be destroyed at the completion of the study. After the study concludes, there will be no way to identify you to the data. All computer usage will be with computers in a University of Waterloo computer lab and not with personally identifiable computers, i.e. you will not use your own computer. All data collected will be

retained indefinitely and will be used for research purposes. We may refer to individual participants when describing the results or the study, and in these cases, we will always refer to “participant 1” or some other similar anonymous name. Your name will never appear in any publication that results from this study.

The web documents list for that we use comes from the Bing Search API. This is a publicly available API. By our very use of this search API, we will "link" with it, but we will not be linking your information collected here to any other information that concerns you personally.

We may choose to distribute the data collected to other researchers. All data will be anonymized at the conclusion of the study and prior to any distribution, but each participant’s data will remain identifiable as coming from an individual, i.e. “participant 1”, “participant 2”, etc. We will not publicly share this data, i.e. the data would only be made available to other researchers for research purposes.

Remuneration for Your Participation:

In appreciation of your time to complete the study, you will be paid $20. The amount received is taxable. It is your responsibility to report this amount for income tax purposes. Should you stop before completing the study, you will be paid on a pro-rated time basis ($5 for every 25 minutes).

Risks and Benefits:

There is minimal risk to you from participation in this study. Computer use and searching for relevant documents are common everyday activities and pose no anticipated risk greater than that encountered in everyday activities. The topics to be learned are those that might be used by an analyst and none of them deal with matters outside of what is commonly found in major newspapers.

There are no direct benefits to you from participation besides gaining more knowledge on the topic to be learned. However, we hope the study will provide results that can lead to advances in the evaluation and development of advanced text retrieval systems that will benefit society at large.

When information is transmitted over the internet, privacy cannot be guaranteed. There is always a risk your responses may be intercepted by a third party (e.g., government agencies, hackers). University of Waterloo researchers will not collect or use internet protocol (IP) addresses or other information which could link your participation to yourself.

Research Ethics Clearance:

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE#21633). If you have questions for the Committee contact the Chief Ethics Officer, Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca.

Thank you for your assistance in this project.

Questions:

If you have any questions regarding this study, or would like additional information to assist you in reaching a decision about participation, please contact me at m2abuals@uwaterloo.ca. You can also contact my supervisor, Professor Mark S. Smucker +1 (519) 888-4567 ext. 38620 or email mark.smucker@uwaterloo.ca.

CONSENT FORM

By signing this consent form, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

I agree to participate in a study being conducted by Mustafa Abualsaud, a Master's student in the University of Waterloo's Department of Computer Science. I have made this decision based on the information I have received in the information letter. I have had the opportunity to ask questions and request any additional details I wanted about this study.

If I participate in this study, I will be asked to complete several brief questionnaires and to search/view, and evaluate 10 web documents regarding the topic to be learned, and answer a quiz to determine my knowledge on the topic.

As a participant in this study, I am aware that I may decline to answer any question that I prefer not to answer. I am also aware that I may stop participating in the study at any point and withdraw my consent. Should I stop before completing the study, I will be paid on a pro-rated timely bases.

I am aware that all information that I provide will be anonymous with no identifiers retained to connect it to me.

I am aware that this study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Committee (ORE#21633). If you have questions for the Committee contact the Chief Ethics Officer, Office of Research Ethics, at 1-519-888-4567 ext. 36005 or ore-ceo@uwaterloo.ca.

I agree to participate in this study.
[Self-report questionnaires and searching/viewing/evaluating 10 documents (approximately 60 minutes)]

YES NO (Please circle your choice)

Participant Name: _____ (Please print)

Participant Signature: _____

Witness Name: _____

Witness Signature: _____

Date: _____