From Facet-based to Visualised Digital Library Interfaces

An Empirical Study of User Experiences with Digital Libraries

Doctoral Dissertation submitted to the
Faculty of Informatics of the Università della Svizzera Italiana
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

presented by
Maram Hassan Barifah

under the supervision of
Fabio Crestani and Monica Landoni

Dissertation Committee

Prof. Matthias Hauswirth	Università della Svizzera italiana, Switzerland
Prof. Vittorio Limongelli	Università della Svizzera italiana, Switzerland
Prof. Pia Borlund	Oslo Metropolitan University, Norway
Prof. Ian Ruthven	University of Strathclyde, United Kingdom

Dissertation accepted on

Research Advisor

Fabio Crestani

Co-Advisor

Monica Landoni

PhD Program Director

Prof. Walter Binder and Prof. Silvia Santini

I certify that except where due acknowledgement has been given, the work presented in this thesis is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; and the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program.

Maram Hassan Barifah
Lugano,

In memory of my beloved nephew,
Abdualaziz Alrumaim

"I don’t think that architecture is only about shelter, is only about a very simple enclosure. It should be able to excite you, to calm you, to make you think."

Zaha Hadid

Abstract

Digital library (DL) systems in the last decade have provided an unprecedented access to library content that were confined to their physical spaces. The last twenty years of the DL system renovation have witnessed several generations of online catalogs from the Online Public Access Catalogue (OPAC) late 1980s, to the next generation catalog systems (NGC).

DL systems are extensively used to facilitate access to collections of digital resources. Thus, NGC offers more user-friendly interfaces by providing functionalities like faceted navigation and filters. The facets are based on the unified index of metadata of the library's content, and it is believed that facets assist users to achieve higher task accuracy and satisfaction.

However, many researchers criticize the architectural elements of the traditional representation of contents in DL systems for two main reasons: firstly, the knowledge organisation schemes adopted to build DL systems were created by experts to classify resources, this results in inadequate use of the search tools i.e. facets and other DL functions by the end-users who are not familiar with the purpose of such techniques. Secondly, users are not allowed to deploy techniques of berrypicking in the search process, as no other information seeking behaviour (ISB) than the goal-directed ISB are supported.

The basic assumption of this study is that there is a need to provide alternative modes of access to information than the traditional facet-based DL interfaces.

The context of this study includes three different libraries serving three different communities of users. These libraries have in common similar facet-based interfaces that built on the traditional DL information architecture.

Our investigation follows a user-centered approach where the users are a central part in the design process of the system. Therefore, this research employs ethnographic techniques, that aim to explore the interactions of the real users along with other user attributes e.g. experience level, attitude, and affective variables including emotion and mood. In particular, this research investigates how users interact with the DL interfaces, and how they perceive the quality of the DL interface through: log file analysis, screen video recording, user experi-

ence (UX) study, and a large attitudinal survey.

Results from such investigation lead to the second phase of the thesis which is the redesign of the facet-based interface by adopting the visualization approach.

The visualization technique has been selected as it is supposed to provide an intuitive means for users to interact with the information. An intuitive interface is characterized by being easy to use and learn, requiring little to no previous knowledge, and just feeling "natural" to use.

Thus, the main concern of this thesis is to examine the effectiveness of the visualized interface as an alternative to the traditional DL interfaces, by considering the users' perspective. The evaluation of the visualized interface indicates that visualization is a promising solution that can enhance positive gauging of the participants' satisfaction, together with experiential values that invoke low level of frustrations and high enjoyment.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my father Hassan Barifah and mother Maryam Al-Salmi for their unlimited supports.

I wish to express my gratitude and appreciation to my supervisors Prof. Fabio Crestani and Dr. Monica Landoni whose guidance, support and encouragement has been invaluable throughout this study. I am also grateful to my doctoral thesis committee, Prof. Pia Borlund, Prof. Ian Ruthven, Prof. Matthias Hauswirth, and Prof. Vittorio Limongelli for the thoughtful comments and recommendations on this dissertation.

Further, my special thanks to the Saudi Ministry of Education for funded this project.

I acknowledge the support of RERO Doc administrators for their collaboration, and sharing the log files data. We also thanks Mr.Alessio Tutino, from University Library Lugano for the enlightening conversations on DLs, users and librarians.

I would like to express our deep gratitude to Effat University, Jeddah, Saudi Arabia for their support to conduct the laboratory user study. Special thanks to Dr. Ayman Eddakrouri, the administrator of Effat Library and Cultural Museum, and to the Graduate Studies and Research Department.

I also thank Dr. Saad Azzahri, from King Saud University, Information Science department for helping us to distribute the survey used in this thesis.

In addition, special thanks to my neighbor in Lugano, Mrs. Helen AlHariri and her family for their great support and help. Surely, not forget to thank IR lab group at USI.

Finally, a very big thanks to my family members for always encouraging me to pursue this thesis.

Contents

Figures

Tables

Part I

Introduction and Background

Chapter 1

Introduction

1.1 Motivation

Information retrieval (IR) is defined as "finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)" Manning et al. [2008] (p. 1). Mainly, IR research has focused on how precisely and completely text representations match queries against a store of indexed texts Sanderson [2010].

In reality, when users interact with IR systems, they do not only create queries and judge the relevance of the results. Instead, various forms of user engagement in the search process are involved. Understanding the user aspects when interacting with the information system is under the umbrella of an emerged distinct research area called interactive IR (IIR). Comparing to the traditional IR studies which eliminate the users from the IR evaluation, IIR studies have widened the research objects to focus on users' behaviors and experiences, as well as interactions between users-systems and users-information Kelly [2009]. Consequently, IIR research is called to consider various contextual factors, e.g. task types, user characteristics, user knowledge, search stages, and system features, that affect the dynamic user-system interactions when designing effective IR systems Xie and Cool [2009].

IIR is still a growing research area in the field of information science Joo [2013]. The context of this research considers digital libraries platform as an IR system.

The last 20 years of DL systems architectural renovations witnessed several generations of online catalogs from the Online Public Access Catalogue (OPAC) late 1980s to the next generation catalog systems (NGC) Yang and Wagner [2010].

OPAC provided the same access points as the card catalog. On the other hand,

NGC which is variously referred to as catalog overlays and discovery layers offers access to a greater diversity of sources than the traditionally available through a library's catalog. NGC is web-based applications that search in a unified index of metadata. Thus, it is characterized by having a central index from which to return results through a single interface of extensive content including library catalog records, resources from institutional repositories, indexing and abstracting services, vendor databases, and other scholarly information resources. NGC uses open platforms, open source license software, and open content where possible. For better supporting users' expectations, NGC offers more user-friendly interfaces by providing functionalities like faceted navigation and filters Nagy [2011].

The facets are based on the unified index of metadata of the library's content, and it is believed that facets assist users to achieve higher task accuracy and satisfaction Yeh and Liu [2011].

As many libraries adopted the NGC techniques, there is a growing interest of research in assessing the effectiveness of the current DL, on user search activities and task performance. Researchers are divided into two groups: the supporters who found that the DL search tools e.g. facets are useful and valuable features, and the opponents who confirmed that the current DL interfaces are confusing and overwhelming, and they call for innovative approaches, capable of handling the inherent complexity of DL systems with more uncomplicated and intuitive interfaces without a high learning curve Harkema [2015].

Consequently, a growing number of information science and technology researchers argue that the mismatch between users' real interactions and the functionality of the DL is due to the posited disconnect between information behaviour and practises (IBP) research and the development of information technologies and information services Huvila et al. [2019]. The main critique is that the design of new technology and systems is not guided by the insights of IBP research, regardless of the enormous number of studies produced by IBP research.

Users interact with the DL through the information space which is presented by its interface. In a search episode, the information behavior may change many times as the user information needs and interests are developed or their behavior is triggered by the affordances encountered on their way through the space of information Marchionini [2006].

Harkema [2015] considered the tradition design of the DL interfaces as an obstacle that kept the DLs as monument comparing to other growing and dynamic web sources, and he called for creating more opened and multifaceted information spaces. The current DL information architecture, the facet-based interfaces, supports only one type of information behavior, the goal-directed one. Thus, dif-

ferent researchers called to develop interfaces that go beyond the standard search functions, and promote different information behavior Pérez-Montoro and Nualart [2015]. Lueg and Twidale [2018] confirmed the need for alternative interfaces and functionality that intentionally designed for humans by saying:

> "*We seem to have become stuck designing interfaces that are more suitable for patient, logical, rational robots (or Vulcans) than for mammals who get tired, bored, exited, irritated, intrigued, or distracted, and who even change their minds about what they want to do.*" Lueg and Twidale [2018] p.409

To address age-old problems reported in the DL evaluation studies, this research adopts user-centered approach where the users should be a central part in the design process of the system Alben [1996]. This dissertation starts the investigation implicitly by exploring the usage patterns (UP) recorded in the log files (LF) of the DL system. The analysis of the LF will reveal hidden information about the user interactions. This leads us to investigate deeply the real usage of the DL functions; i.e. to what extent the DL's search tools are used?

This will be done by conducting three different user studies; one will investigate why searches failed and it will be remotely run study, another one will investigate deeply the user experiences with the DL and it will be conducted in a laboratory setting, and the last-one will enable us to collect a large-scale attitudinal survey.

Users' search behavior and cognitive processes can be significantly influenced by the type of search tools and interfaces they interact with. The basic assumption of this study, is that there is a need to provide alternative modes of access to information than the traditional facet-based DL interfaces.

To address this gap, we are going to redesigned the current DL interface taking in our consideration that:

> "*The ultimate goal of any discovery service, bar none, is to place content in the hands of the user or, more specifically, to discover, present, and deliver relevant content in a convenient, intuitive manner to today's researcher.*" Vaughan [2011] p.48-49

Thus, this dissertation will adopt the visualization approach as an alternative to the traditional DL interface. Visualization is not a new phenomenon in the IR filed but it has rarely been used in the DL context. Through visualization, complex information converted to a more user-accessible form. The features of the visualizations help to rapidly process information and provide knowledge that can not be recognized without visual data processing.

The ultimate aim of this study is to examine to what extent the visualized DL interface can be adopted and used intuitively by users. Different instruments will be utilised for the evaluation, and the results will be discussed to see whether the interactive visualized interface has a positive impact on the overall user experience or not.

The following sections summarize the research problem, research questions, thesis goals, methodology of this research, thesis outline, and publication overview.

1.2 Research Problem

Motivated by Nielsen [2020] who claimed that the structure of the website should be driven by the tasks the users came to the site to perform. Back to the eighties, the current DL interface heirs of the classical interfaces that built on a particular architectural anatomy including the controlled vocabularies and the filtering function through certain criteria e.g. author, year of publication, subject..etc Nualart et al. [2014].

Since then and in contrast to the other information systems, the DL interfaces have barely evolved on the basis of the findings provided by user studies, nor have the advances developed in specific disciplines, such as information architecture, or those derived more generally from user experience (UX) Pérez-Montoro and Nualart [2015]. Most of the IIR studies still remain in measuring how search results are relevant, useful or satisfactory as an indicator for system performance Borlund [2003].

This shortage result in significant limitations includes preventing deploying techniques of berrypicking in the search process Bates [1989], and supporting other ISB divergent, explorative behaviour. In her model as in figure 1.1, and in contrast to the assumption of 'standard' information retrieval that the user's information need remains the same throughout the search process, Bates [1989] emphasized the iterative nature of the search process. As new information encountered, new queries may be formulated, thus a search may lead to a new and unanticipated directions. Accordingly, the information needs are not satisfied by a single document but rather by a series of bits and selections of information found along the way.

Pérez-Montoro and Nualart [2015] confirmed that the available tools of the DL include classification support users with only goal-directed information needs, while there is a need to design the DL interfaces to support the divergent information needs. Björneborn [2010] claimed that the traditional design of the DL supports only the convergent information behaviour, the goal-directed behaviour,

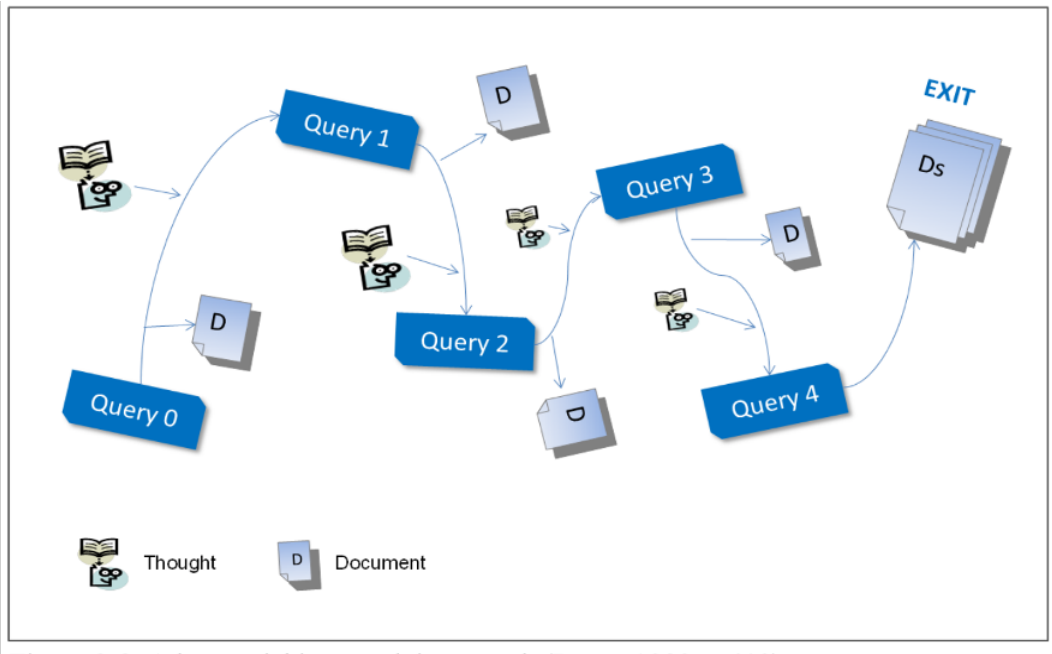


Figure 1.1. Bates [1989] Berrypicking ISB Model

and ignores the divergent or the explorative behaviour.

Therefore, this research attempts to fill this gap by firstly investigate the real usage of the DL tools by: (i) analysing the log files of a large-scale DL, (ii) conducting a series of user studies both online and offline. Secondly, the research examined the efficiency of the visualised interface as an alternative to the traditional DL interfaces.

1.3 Thesis Goals

Based on the shortcomings in the present literature, the aim of this study is threefold:

1. To provide empirical studies with the purpose of gain a comprehensive and in-depth analysis of users' interactions and perceptions in naturalistic settings,

2. To aid the researchers' understanding of user behaviour in DL domain by deploying user experience techniques which should be the ultimate criterion of the user-centered design, and leads to recommendations for improving DL functionalities, and

Questions	Method	Chapter
1- What are the potential UPs of a large-scale DL? *2- What are the main features that characterise the above UPs?*	Log files analysis	3
3- What are the reasons behind search failures from the searchers' perspective? *4- What are the primary emotions individual experienced in the failed searches?*	Online user study	4
5- To what extent the DL functions are usable by users with different expertise? *6- Will the advanced participants share similar perceptions of the DL compared to the non-advanced participants?* *7- What are the associated moods before and after the search experience in the DL?* *8 -What are the influential factors that might affect the moods before and after the search experience?*	Laboratory user study	5
9- How would user perceive Saudi Digital Library (SDL)? *10- What are the commonality between the studies?*	Survey	6
11- How do users perceive a visualised interface for a DL? *12- To what extent the new visualised interface can be adopted and understood by users?*	Online study	7

Table 1.1. Thesis's Research Questions

3. To examine the efficiency of the visualised interface as an alternative to the traditional DL interfaces

1.4 Research Questions

The research aims to answer one main question: *to what extent the DL's search tools i.e. facets and filtering functions are usable?* This question was divided into sub-questions. Different approaches were applied to answer those questions. We list the corresponding research questions in table 1.1 that have guided the research presented in this thesis.

1.5 Methodology

This research employs ethnographic techniques. Ethnography refers to "the use of in-depth observation, and often participation, of a human group, culture, or context, with the goal of developing a rich description of activities, interactions, beliefs, roles, and goals" Lazar et al. [2017]. Basically, ethnography is based on the understanding of complex human practices by an in-depth study.

In human-computer interaction (HCI), ethnographic research is particularly useful for understanding how stakeholders interact with a system to complete complex tasks. It can help in providing an understanding of the context in which specific interfaces or systems are developed and implemented. Therefore, ethnographic studies are usually conducted "in the wild" where the "action" of interest takes place e.g. in homes, workplaces, or educational settings" Lazar et al. [2017].

Ethnography methods including participant observation focuses on developing an understanding of the design problems by evaluating an existing design in order to support future design decisions. Therefore, in the early stages of a user-centered design project, ethnography methods are particularly useful Lazar et al. [2017].

This work aims to explore the interactions of the real users along with other user attributes e.g. experience level, attitude, and and affective variables including emotion and mood. Particularly, to explore how users interact with the interfaces, and how they perceive the interfaces quality through: log file analysis, screen video recording, UX study, and large attitudinal survey.

The context of this study includes three different libraries, namely: (1) RERO Doc [1]; acronym of REseau ROmand, designates the network of libraries in Western Switzerland. (2) Effat DL [2], academic DL of Effat University, and (3) SDL, Saudi Digital Library [3]. All the mentioned libraries have in common similar facet-based interfaces that built on the traditional DL information architecture.

This work consists of two main fundamental phases, each phase has sub-phases as it depicted in figure 1.2.

In phase (I): the evaluation, series of studies were completed, and thus their results lead to phase (II): the prototyping. Phase (I) started by exploring the usage patterns recorded in the log files of RERO Doc DL.

Then, as the problem identified, the author investigated deeply the problem by involving real users of three different DL systems. In order to guarantee more

[1] http://doc.rero.ch/

[2] https://library.effatuniversity.edu.sa/

[3] https://sdl.edu.sa

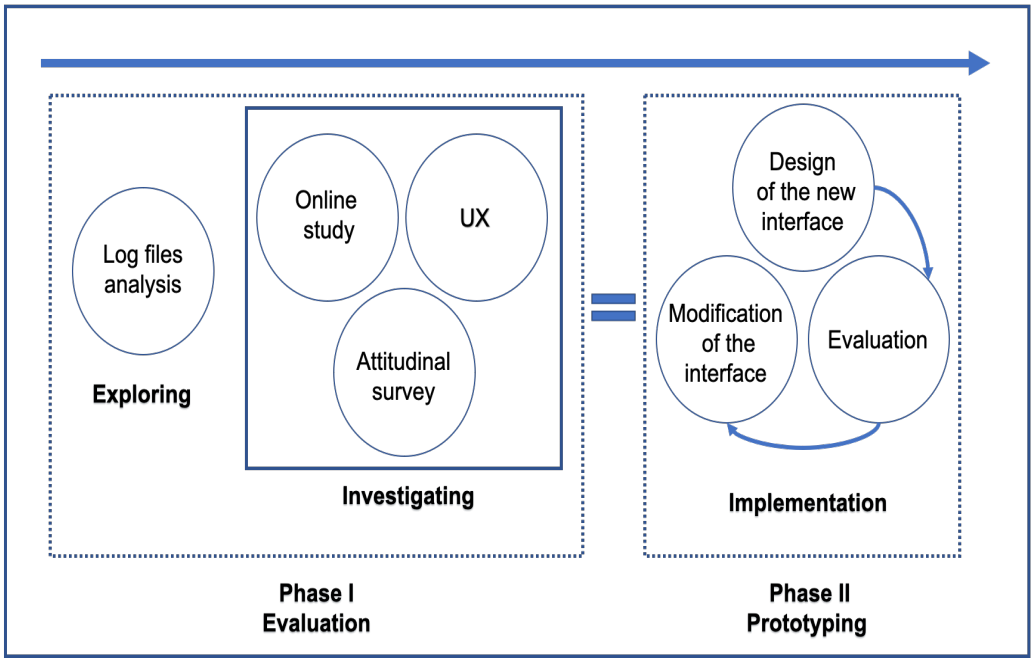


Figure 1.2. The Research Design

realistic results and eliminate the impact of the contextual factor, the experiments were carried out in different environments. The online study was designed to be run remotely in the users naturalistic settings, the UX study was conducted in the laboratory setting, and the attitudinal survey was distributed online. The results of the investigating sub-phase confirmed the findings of the LF analysis.

The outcomes of phase (I) inspired us to redesign the current DL of RERO Doc phase (II). The new design was evaluated by users, and accordingly some modifications were made to improve the new interface and enhance the user experiences.

1.6 Thesis Outline

The dissertation consists of eight chapters under five parts as follows.

Part 1 - Introduction and Background

- **Chapter 1 – Introduction**

  The first chapter started by the motivation to conduct this research, fol-

lowed by introducing the research problem, and then presented the research questions. Finally, the research methodology, and contributions of this dissertation were explained.

- **Chapter 2 – Background**

  The research problem was identified by reviewing the related works of different disciplines. The scope of the dissertation is wide, thus the background chapter covered three main areas: digital library literature, evaluation including the literature of the user aspects, and finally reviewed the information visualisation related works.

Part 2 - Exploring

- **Chapter 3 – Exploring Usage Patterns of a Large-scale Digital Library**

  LF analysis is known as an inexpensive method that enables researchers to reveal hidden information about the systems, and help to gain an overview of the user experiences with the system. Thus, this study started by analysing the LF of RERO Doc DL, the study and the results are presented in chapter chapter 3.

Part 3 - Investigation

The LF analysis assists us to identify aspects for further investigation, thus the this part consists of the following chapters:

- **Chapter 4 – Why Searches Fail?**

  Researchers believed that facets-based DL interfaces assist users to achieve higher task accuracy and satisfaction. To investigate this assumption, an online user study was conducted with the collaboration of RERO Doc aiming to understand why searches failed, and to what extent the facet-based DL interface functions are used. Chapter 4 described the study design and the results.

- **Chapter 5 – Evaluating User Experiences in a Digital Library**

  The result of the study in chapter 4 motivated us to redesign the current DL interface followed the user-centered design approach. The user-centered design approach requires a deep understanding of the user experience (UX) before the processing of the redesign.

Thus, the purpose of this chapter is to investigate the UX in the DL by considering threefold aspects: The *first* is to understand to what extent the DL functions are usable, the *second* is to investigate the user perceptions of the instrumental and non-instrumental quality of the DL, and the *third* is to examine the associated mood before and after the search experience. The design and results of this study is presented in chapter 5.

- **Chapter 6 – Large-scale Attitudinal Survey**

  So far, in this research the interactions with the DLs were examined implicitly by analysing the LFs, and explicitly by conducting online and laboratory user studies. In order to confirm or deny the commonality between the findings of the studies, a large-scale attitudinal survey was distributed to collect more data of the users' perspectives by considering a new DL with different population. Chapter 6 summarised the findings of those studies where the results inspired us to redesign the current DL interface by adopting the visualisation approach.

Part 4 - Prototyping

- **Chapter 7 – Can Visualisation be a Solution?**

  As a response to the problems reported in the user studies, we redesigned the facet-based interface, and the interactive visualised interface is presented. The goal of this chapter is to examine the effectiveness of the visualised interface as an alternative to the traditional DL interfaces.

Part 5 - Conclusion and Future Work

- **Chapter 8 – Conclusion and Future Work**

  Finally, the summary of this study is presented in this chapter, along with the future directions.

1.7 Contributions

This dissertation contributes to the filed of the DL in four main aspects as follows:

- **Framework for analysing DL log files**

Regardless of the importance of the LF analysis as a valuable source of the hidden information of the user interactions, the number of the LF analysis studies is low compere to other web platforms.

Discovering information from LFs is not a trivial process; it requires a strategic plan to gain a better understanding of the hidden information.

In this dissertation, we defined a framework for analysing DL log files to help and encourage DL administrators, information science researchers and students to conduct LF analysis.

The framework is explained in chapter 3, and the work is published as Barifah and Landoni [2019a].

- **Design an IIR naturalistic experimental platform**

  One of this research goals is to conduct evaluation experiments in a naturalistic setting as far as possible in terms of the task design and/or the experimental setting i.e. in the user everyday settings. Thus, a methodological contribution of this thesis is designing an experimental platform that enabled the participants to take part in a semi-naturalistic study as in chapter 4. We define it as semi-naturalistic given that users were working in their own environment while being guided by the simulated task scenarios. A simulated work task situation is "a short textual description that presents a realistic information requiring situation that motivates the test participant to search the IR system" Borlund [2016] (p.395). The results of this study was presented in ISIC 2020 conference.

  It is also worth noticing when considering our contribution to the design of naturalistic studies, the introducing of the *user-tailored* task that was generated from the study in chapter 4, and reused in chapter 7. The user-tailored task refers to self-generated tasks by the real users of a system, and it can be used later to design the simulated task scenarios to evaluate the system.

- **User experience (UX) study in DL**

  Researchers called to deploy UX techniques, which should be the ultimate criterion of the user-centered design, to address age-old problems reported in the DL evaluation studies. Evaluating a system should be moved beyond the usability and quantitative metrics for fully understanding the UX.

  To our best knowledge, DL research lacks to the UX studies, and lacks of empirical studies that combined different data collection tools for fully understanding of user interaction, and user experiences.

Thus we contribute to the filed by deploying a user-centered method for fully understanding the UX in the DL domain. The design and results of the study is explained in chapter 5, and this work published as Barifah et al. [2020].

- **Implementing visualised interface as an alternative to the traditional DL interface**

  DL interface design has barely evolved on the basis of the findings provided by user studies, nor have the advances developed in specific disciplines, such as ISB, or those derived more generally from user experience (UX) studies. To response to this research need, this study adopted the visualisation interface as an alternative to the facet-based DL interface.

  Although there were different attempts of implementing visualisation interfaces and tools in the DL systems, Pérez-Montoro and Nualart [2015] stated that visualisation has not yet been widely implemented in the DL for practical and methodological reasons.

  Examples of the practical reasons are: visualisation tools are not integrated in the standard search field-result list paradigm of DL which means users perceive them as being secondary tools Hienert et al. [2012]. Moreover, visualising results are not very intuitive to users due to the high level of abstraction and conceptualization of the visualisation architecture.

  Regarding the methodological reasons, along with other reasons, Pérez-Montoro and Nualart [2015] confirmed that most of visualisations prototypes used small collections of documents and their efficient use with large collections has not been demonstrated. Moreover, most of the proposed techniques were not tested with the involvement of end users, and this makes it difficult to draw any clear conclusions about their efficiency. Therefore, in response to such gaps, this research adopted the interactive visualise interface to a large-scale DL system. Followed by conducting a user-centred evaluation study to investigate the efficiency of the visualisation in DL platform. The interface design and the study result described in chapter 7.

1.8 Publication Overview

Most of the works presented in this thesis was published in different events as listed below:

- Barifah, M. and Landoni, M. [2019]. Exploring usage patterns of a large-scale digital library, 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL) Barifah and Landoni [2019a].
- Barifah, M. and Landoni, M.[2019c]. Obstacles to conducting log file studies in Saudi Arabia,With an Eye to the Future: HCI Research and Practice in the Arab World Barifah and Landoni [2019c].
- Barifah, M. and Landoni, M.[2019b]. Interactive search profiles as a design tool,International Conference on Human-Computer Interaction, Springer, pp. 18-30 Barifah and Landoni [2019b].
- Barifah, and M. Landoni, M., [2020]. Emotions associated with the failed searches in a digital library .In press (accepted at ISIC 2020).
- Barifah, M. Landoni, M., & Eddakrouri, A. [2020]. Evaluating the user experience in a digital library. Proceedings of the Association for Information Science and Technology, 57(1), e280 Barifah et al. [2020].
- Measuring the User Experience with a Visualised Digital Library, under review of CHIIR 2021.

Chapter 2

Background

Because the scope of this research is wide and covers different disciplines, various research domains have been investigated starting with an overview of the evaluation of the information retrieval (IR) and interactive IR (IIR) evaluation, moving to digital library systems as an IR system, then to the user aspects in relation to DL, and finally the literature of the information visualisation in DL was reviewed.

2.1 Evaluation

This section reviews the evaluation of IR, IIR, and DL systems.

2.1.1 Overview of IR's Evaluation

Information retrieval is defined as "finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers), Manning et al. [2008].

Evaluation is a fundamental aspect of both IR and Interactive Information Retrieval (IIR) research Kelly and Sugimoto [2013]; Zuva and Zuva [2012]. *System-oriented* evaluation approaches fail to capture the dynamic environment of IIR Tamine-Lechani et al. [2010]. For instance, Cranfield Paradigm, the standardised evaluation approach in IR research, evaluate retrieval systems based on a test collection and associated relevance judgements for extensive testing and comparison of retrieval algorithms Maxwell and Azzopardi [2016].

The system-centred method aims to measure the effectiveness and efficiency of the IR systems. The effectiveness is defined as "a measure of the ability of the system to retrieve relevant documents while at the same time holding back

non-relevant one" Rijsbergen [1979]. Therefore, IR models and algorithms are evaluated in terms of their ability to identify topical relevant documents Tamine-Lechani et al. [2010]. The efficiency of the system can be assessed including the coverage of the collection, the time lag, presentation of the output, and the effort involved by user in obtaining answers to a request Rijsbergen [1979].

Various metrics can be used to evaluate the IR systems. Precision and recall belong to the Set-based measures. Precision is the measurement of the fraction of retrieved documents that are relevant whereas recall is the fraction of relevant documents retrieved Sanderson [2010]. Rank-based measures are used to measure the position of the items in the ranked list e.g. mean average precision (MAP) and discounted cumulative gain (DCG). Other measures are used to evaluate different IRS problems e.g. measure the success of search tasks by using mean reciprocal rank (MRR). Or using S-recall to assess the variability or diversity of results Clough and Sanderson [2013]. The system-oriented approach is widely used due to the established parameters (metrics), thus the results are easy to compute and interpret and the process is fast compared to the user-oriented approach Akhigbe et al. [2011]. Also, the system-oriented tests are distinguished by being repeatable, inexpensive, and ideal for testing specific components e.g. learning a ranker Clough and Sanderson [2013]. Because the focus of the system-oriented approach is assessing the performance of the underlying search algorithms with little involvement of end users, the system-oriented approach has been criticised for lacking of insight into the user-system interaction and disregarding iterative and exploratory retrieval Petrelli [2008].

2.1.2 Interactive IR's Evaluation

IIR studies how users interact with IR systems and evaluates the users' satisfactions with the retrieved information Kelly [2009]. IIR systems are defined by Borlund [2000] as "those where the user dynamically conducts searching tasks and correspondingly reacts to systems responses over session time". Thus, users' behaviours, experiences and interactions with systems are the main focus of IIR's studies Kelly [2009]. Three ingredients are essential in IIR studies Borlund [2003]: the involvement of potential users as test participants, the use of dynamic and individual information needs, and the employment of relevance judgements. Different actions are embodied in the process of IIR including: "query formulation and reformulation, search result list interaction (scanning, assessing and clicking), document interaction (reading and judging relevance), and result list and session abandonment" Pääkkönen et al. [2017].

As the concern of the IIR evaluation is to study the ability of the users to en-

gage with a system in order to retrieve relevant documents, user-oriented evaluation approach was emerged. These are characterised by involving real users in the evaluation process and considering the context, human aspects, and interactions in relation to the retrieval processes Kelly [2009]; Borlund [2000]. When involving the users, a variety of evaluation metrics are used which can be divided into four standard classes Belkin [2010]; Catarci and Kimani [2012]; Kelly [2009]:

1. The *performance measures* include: interactive recall and precision, multi-level relevance and rank measures, time-based measures (efficiency), informativeness for evaluating search results, and cost and utility measures.

2. The *contextual measures* consider: user characteristics or individual differences, measures of information needs.

3. The *interaction measures* cover: number of queries, number of search results viewed, number of documents viewed, number of documents saved, and queries length.

4. The *usability measures* include effectiveness, efficiency, and satisfaction.

However the user-centred evaluation approach has been criticised for being costly and time consuming Azzopardi [2011], and producing data that are not reusable, and delivering results that are not reproducible Joho et al. [2007].

The context of this research considers digital libraries platform as an IR system. The following section covers the literature of the digital library.

2.1.3 Digital Libraries

DLs can be defined as a type of "distributed systems with the capability to store various electronic resources and provide convenient access for end users via networks" Zha et al. [2015]. Similarly Frias-Martinez et al. [2007a] defines DL as "collections of information that have associated services delivered to user communities using a variety of technologies".

Thus, the DLs are characterised by being information systems; (a) that provide information services, (b) where information should be organized and presented in usable ways, (c) that enable users to communicate with information, and (d) help users to satisfy their information needs Fox [2002]. DLs provide

authentic content where the collections or items are organised by knowledgeable professionals Agosti et al. [2012]. In most cases, their interfaces are structured hierarchically, providing different types of interactions: traditional keyword searching and browsing-related activities, including using multifaceted menu and inter-document linkages Han and Wolfram [2016].

2.1.4 Types of Digital Libraries

There are different types of digital libraries based on the data distribution and material localisation. Those are: the Stand-alone Digital Library (SDL), Federated Digital Library (FDL), and Harvested Digital Library (HDL). The SDL is self-contained where the holdings are digital scanned or digitized, and the material is localized and centralized in a fully computerized fashion with networked access. The ACM Digital Library IEEE Computer Society DL are examples of SDL. The second type is FDL that composes several autonomous SDLs to form a networked library based on common focus and topic with a single user interface. An example of the FDL is Networked Digital Library of Theses and Dissertation. FDL systems usually implemented discovery tools such as SCOPUS, Database Advisor, ELIN, Knowledge Cite Library, and OCLS' FirstSearch. Finally, the HDL a virtual library providing mainly metadata with pointers to the holdings that are "one click away" in Cyberspace. the Internet Public Library is an example of the HDL Kadury and Frank [2007].

2.1.5 Digital libraries' Information Architecture

Back to the eighties, the current DL interface heirs of the classical interfaces of the document databases on CD-ROM that built on a particular architectural anatomy including the controlled vocabularies and the filtering function through certain criteria e.g. author, year of publication, subject..etc Nualart et al. [2014]. Most of the DL utilized the discovery search interface (DSI). One of the design principles behind the DSI is the information architecture which is defined as the "combination of organization, labeling, and navigation schemes within an information system" Rosenfeld and Morville [2002]. Consequently, the DSI is a gate that enables access to the DL catalogue and to multiple databases through a single search box Gross and Sheridan [2011]. The DSI is recognized to possess an abundance of search tools: facets, advanced search, and sorting. Those search tools are designed to be dynamic filters, limits or refinements that assist users to modify a set of search results. Facets, in particular, are powered by the controlled vocabularies and detailed metadata that librarians have built and maintained, and they

are usually positioned vertically on the left side of the interface Hall [2016]. Accordingly, there are three main components in the DL interfaces namely: the data model describing the information available to the users (e.g. labeling and terminology), the navigational model including the procedures for gaining access to data model (i.e. search tools e.g. facets and advanced search), and the interface style referring to the surface presentation (e.g. style and color) Cushman and Rosenberg [1991].

2.1.6 Evaluation of Digital libraries

Digital libraries tend to satisfy user needs by providing authentic and useful content which will determine the extent of the usage of the DL Petrelli [2008].

Evaluation is defined as "the systematic process of determining the merit, value, and worth of something. Evaluation is broader than assessment and involves making a judgement as to the effectiveness of an assessment. Evaluation has a certain goal, methodology and devices or techniques. The goal suggests some evaluation criteria e.g. performance, which sometimes break down to evaluation parameters e.g. response time, availability" Fuhr et al. [2007].

Evaluations are classified into: *formative* evaluation which is conducted during the development phases of a DL system to minimise imperfections before release. In contrast, *summative* evaluation is conducted in the final stages of the development or after an initial release. *Comparative* evaluation where different systems or components are compared. And the *iterative* evaluation that are conducted regularly as new materials are added, metadata are updated and user characteristics changed or new users joined Xie and Cool [2009]; Petrelli [2008]. Due to the interdisciplinary research domain of the DL covering information retrieval, human-computer interaction, and information science, the literature is rich with various evaluation models, frameworks and criteria Tsakonas et al. [2004].

DL is not only a content-bearing tool, thus considering system retrieval performance solely is not sufficient to evaluate the overall system acceptability. Instead, the interactions between the three core components: *user*, *content*, and *system* should be considered. Such interactions are embodied in the usage of a system thus monitoring the usage provide useful information for evaluation and improvement processes Fuhr et al. [2007].

The three DL components: the *user*, *content*, and *system* are described by Tsakonas et al. [2004] and summarised as the following:

1. *Users*: IIR systems are designed to target users regardless of their numer-

ous and complicated attributes Belkin [2010]. Users tend to use digital libraries to fulfil their information needs. Their interactions are embodied in the ISB which are affected by different attributes including expertise in using the system, familiarity, frequency of using the system and tasks' knowledge. Lacking of expertise and knowledge result in poor usage of the system's functionalities and content identification. Therefore, it is important to enlarge the scope of IIR systems' evaluation by considering more metrics i.e user attributes to evaluate usage along with the system performance evaluation Behnert and Lewandowski [2017].

2. *Content*: it is the main motivation behind using a DL, and finding the relevance and useful content is always related to the users' information needs i.e. search tasks of the users. The availability and the usefulness of the content affect the users' seeking behaviour, query (re)formulation, and types of digital collections.

3. *System:* it is the mean of the interaction between the users and the content and it consists of IR mechanisms, interface structure, and information architecture.

The interaction among the three DL components is depicted in 2.1: the *user*, *content*, and *system* are classified by Fuhr et al. [2007] and Tsakonas et al. [2004] into: *content-system*, *user-system*, and *user-content* interactions, they summarised as follows:

Content-system interaction which can be evaluated by considering the performance of the system that depends on the structures and representations of the content. Such evaluation is done computationally where the user cannot evaluate it directly. Precision and recall are examples of the measures. To evaluate the interaction of the *user-system*, usability aspects are examined including (i) *effectiveness:* the accuracy and completeness with which users achieve specified goals i.e. user performance and error generation, (ii) *efficiency*: the resources expended in relation to the accuracy and completeness with which users achieve goals e.g. completion time. And (iii) *satisfaction*: the comfort and acceptability of use e.g. aesthetic comfort and readability Cabrerizo et al. [2010].

The motivation behind using a DL is embodied in the interaction between the *user-content* where the usefulness of the content in relation to the search tasks and needs is the criterion. Hügi and Schneider [2013] explained the usefulness features to include relevance, available format of the material i.e. text or multimedia, level of the available material e.g. meta data, abstract, full text, reliability, and coverage.

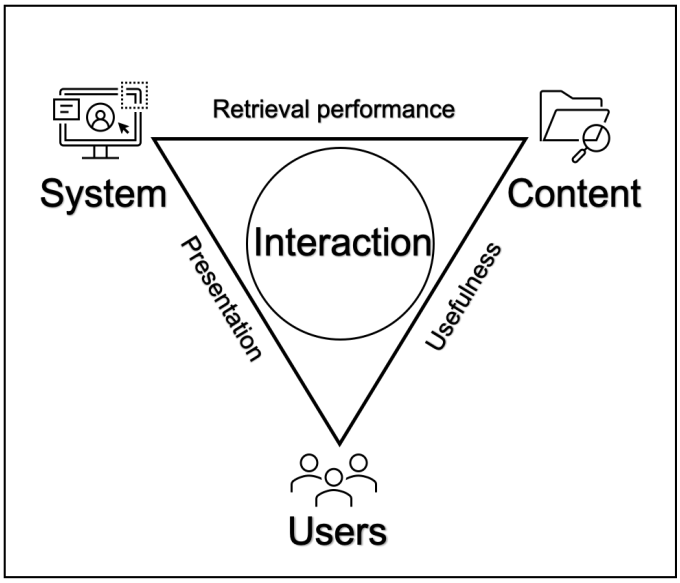


Figure 2.1. The Digital Library Interaction Components Tsakonas et al. [2004]

A holistic framework of the evaluation was introduced by Xie and Cool [2009] where they focused on the following features: interface usability, system performance, user satisfaction, collection quality and service quality.

It is important to mention that not all the evaluation criteria should be applied in one study, it depends on the aim of the study Tsakonas et al. [2004].

Depending on the aim and the focus of the evaluation, different approaches can be followed to conduct evaluation studies of IIR systems Kelly [2009]. The main approaches are: the naturalistic studies which can be conducted in the naturalistic setting and involving real users with real search tasks, or laboratory controlled experiments e.g. TREC-style Maxwell and Azzopardi [2016]. Undertaking IIR experiments with test subjects is laborious in terms of organising and running, time consuming, costly, and the experiments are difficult to reproduce Azzopardi [2011]; Kelly [2009]. However involving the end-users in the evaluation help system designers to design better systems and improve user experiences Kelly [2009].

In this thesis, different methods were followed. Firstly, we started by analysing the log files of the DL system to explore areas that might need further investigation. Secondly, a remotely user study was conducted to understand why searches failed and to what extent the DL functions are used. Thirdly, a laboratory user study run to gain a deeper insights into the user experiences in the DL context.

Accordingly, the literature of the log file analysis, search failed, and user ex-

periences were reviewed as follows

2.2 Log File Analysis

Providing IIR system users with better experiences requires a deep understanding of their information searching behaviour. Information searching can be defined as "users' purposive behaviours in finding relevant or useful information in their interactions with information retrieval (IR) systems" Xie [2012]. Users follow different search strategies or moves when interacting with systems. One method for understanding search strategies is investigation of users' digital footprints, or usage patterns (UP), found in log files (LFs) Ndumbaro [2018].

LF analysis refers to "the study of electronically recorded interactions between online information retrieval systems and the persons who search for the information found in those systems" Peters [1993]. LF analysis is an attractive approach due to the availability of the data considering a larger scale of interactions Hajek and Stejskal [2017]; Han and Wolfram [2016], and being less subject to bias and non-invasive, as it records interactions between systems and users that might not be clearly observed in laboratory experiments Bollen and Luce [2002].

The dynamic nature of the DL requires profound understanding of the user interactions. Thus, scholars utilised LF analysis to investigate two user-interaction levels: *clickstream* analysis and *fine-grained* actions analysis. The former concept refers to the higher-level analysis of interaction during a search session, and the latter refers to the investigation of sub-actions inside the UPs, with query-related analysis being an example Han and Wolfram [2016].

Previous scholars investigated UPs with the aim of assessing users' interaction and improving their search experiences. Chen and Cooper [2001] undertook a pioneering study of library UP recognition. They aimed to automatically detect UPs based on users' interactions extracted from the LFs of an online academic library catalog system. They employed clickstream analysis, and the session was the unit of analysis. Based on 257,000 sessions, UPs were characterised according to 47 operational implicit features of a session. The main features were session, search, display, relevance, error and help. Chen and Cooper used hybrid clustering and principal component analysis to identify, six distinct UPs: knowledgeable and sophisticated use, unsophisticated use, highly interactive use with good search performance, known-item searching, help intensive searching, and relatively unsuccessful use.

Similarly, Eason et al. [2000] investigated the UPs of an electronic journal collection by implementing K-means clustering analysis. Twenty-two months of

LF records were analysed along with surveys. The main variables analysed were frequency of use, breadth of use, depth of use, functions use, use of back and current issues, and read and print (e.g. PDF or HTML). Eight UPs were found: enthusiastic users, focused regular users, specialised occasional users, restricted users, lost users, exploratory users, tourists, and the searchers.

Nicholas et al. [2006] considered records from academic DL from June 2004 to December 2004 and investigated the potential UPs of accessed journals in a DL. They found four different UPs: accessing from search engines, subject lists, alphabetical or journal list, and a combination of the methods. The variables considered in this analysis were items viewed, session duration, time online, and site penetration. They concluded that the users who used more than three access methods were demanding users. Also, they established that the types of information needed, such as fact checking affected the patterns and access methods.

Frias-Martinez et al. [2007b] investigated the role of human factors, particularly cognitive styles, levels of expertise and gender differences, in stereotyping behaviour and perception of an academic library catalogue. They employed LF analysis in combination with questionnaires. Three clustering techniques were implemented: K-means, hierarchical and fuzzy clustering. Five different clusters emerged in relation to the system experts and functions used. They concluded that K-means clustering yielded positive results in terms of meaningful interpretation of results and ease of implementation.

Kovacevic et al. [2010] also implemented data mining techniques to analyse LF records of an academic DL, in particular K-means and Naive Bayes classification, with the aim of improving DL recommendation services. UPs were clustered based on the users' profiles and search history. They considered more explicit features including the users' institute, town, gender, age and services used. Consequently, services were recommended to new users based on previous similar user profiles and preferred services. Similarly, Renaud et al. [2015] analysed library UPs in relation to students' academic performance. They categorised the UPs according to explicit features such as academic positions, student class, and departments. Ndumbaro [2018] analysed the UPs of an academic OPAC system. The analysis of the LFs involved tracking preferred access points (e.g. search by title, author, subject or phrase, or multiple searches), and query-based behaviour to identify reasons for search failures. The dataset consisted of records for January to December 2015, with 532,113 search hits. Traditional descriptive statistics were used. Similarly, Papatheodorou et al. [2003] investigated the UPs of an academic DL to assess the preferred access points and construct user communities. Their unit of analysis was the query, and the Clique mining algo-

Study	Techniques applied	Dataset description
Chen and Cooper [2001]	hybrid clustering	257,000 sessions
Eason et al. [2000]	k-means cluster	22 months logs
Nicholas et al. [2006]	descriptive analysis	seven-month period
Frias-Martinez et al. [2007b]	k-means, hierarchical, fuzzy	48 records
Kovacevic et al. [2010]	k-means & Naive Bayes	230,000 records
Renaud et al. [2015]	statistical analysis & software	Weekly snapshot
Ndumbaro [2018]	statistical & content analysis	one year logs
Papatheodorou et al. [2003]	clique mining	32 months log files
Joo [2013]	statistics, kernel regression	sixty students entries

Table 2.1. Summary of the Existing Studies

rithm was implemented to cluster patterns. Two different UPs were identified: specialised users who frequently utilised "Author" as the access point, and a homogeneous user group whose behaviour was characterised by using all the main access points. They concluded that the variations in behaviour was due to the users' specialisation and the nature of the collections.

Joo [2013] analysed the LFs of a national DL to investigate search tactic patterns in relation to task types: known-item, specific information, and exploratory search. Transition analysis was implemented to trace common patterns of sequential tactics for each task. Hierarchical clustering and multi-dimensional scaling were employed for clustering search sessions by search tactic selection. Joo determined that the main features that differentiated the patterns were frequency of actions, spent time, change over time, and transition.

This brief review of related work indicates that LF analysis assists scholars and practitioners to gain deeper insights into users interactions and their searching experiences. The studies involving mining UP in the context of DLs can be classified according to: *type of the platform* (i.e OPAC system or academic DL), *level of analysis* (i.e. clickstream or fine-grained actions), *available features* (i.e. explicit or implicit), and *purpose of analysis* (e.g. identify user communities and interests, or enhance the system, services or interface). Classification enables selection of suitable analysis methods (e.g. descriptive statistics, clustering ,or classification) Hajek and Stejskal [2017]; Siguenza-Guzman et al. [2015].

We also noticed the lack of a standard dataset description among previous studies. Table (2.1) presents a summary of the studies, and shows considerable variation in dataset descriptions. Units used in the literature to describe datasets included data size, duration of data collection (e.g. week, month) and session or record numbers. A standard unit set is important to allow the findings to be comparable to others. Using a combination of data size by number of records

or measurement unit (e.g. GB) and session numbers is more informative than using one unit alone.

The first phase of this research is to explore the UP of a Large-scale DL. The contributions of the chapter is to define a framework for analysing a large scale DL log files. In chapter 3 the study, the framework, and the results presented.

2.3 Why Search Fail?

The second phase of this research is to investigate the real use of the DL's functionalities. And this was achieved by conducting different user studies. Starting by understanding why search failed, and then a laboratory user study was conducted to measure the user experience when interacting with the DL. According to the literature, the researchers confirm that the DL functions were design with the aim of helping users to find resources, thus we focused on the failed searches with aim to identify to what extent the DL functions were utilised. This section reviewed the literature of the unsuccessful searches.

2.3.1 Reasons Behind the Failed Searches

The concept of success or failure of outcomes for information seeking was introduced by Wilson [1999] in his model of ISB. Mansourian [2008] defined search failure as "the situation in which users attempt to satisfy their information needs, but they fail to do so" (p. 29). Unsuccessful search experiences force users to refine their searches, ultimately resulting in low expectations, frustration, and less perseverance.

Generally, online search failures occur when systems fail to retrieve the desired information Drabenstott and Weller [1996]. Searchers might encounter failures when the system provides too many results, too few results, no results (zero hits), irrelevant or confusing results Lau and Goh [2006]; Xie and Cool [2009].

In the digital library domain, the failed searches have been investigated implicitly by analysing the query behaviour extracted from the log files, and explicitly by conducting user studies. The former category identified ill queries as the primer failure cause to include: typing and spelling errors Peters [1989], query formulation Antell and Huang [2008], irrelevant keywords, spelling mistakes, and incorrect use of Boolean logic Debowski [2001].

User studies helped to identify three main factors that could be behind failed searches, namely: task characteristics, system performance and design, and user

attributes. Examples of relevant task characteristics are: search types (e.g., unknown and known item searches) Slone [2000], characteristics of the information needed (e.g., vague and unfamiliar) Tang [2007]; Wildemuth et al. [2013].

The system effect includes the limited coverage of the digital library Behnert et al. [2017], and user interface Fast and Campbell [2004]; Blandford et al. [2001]; Capra et al. [2007].

The users are the centroid of the interaction process, they tend to search to fulfil information needs which defines as "the start state for information search, seeking and human information behavior to fill a gap in knowledge of the information user" Cole [2015] (p.4117). The information needs are addressed by contents that available on the IR systems. Thus, the effectiveness of such systems can be evaluated in terms of their support to achieve users' information needs or tasks. Ingwersen [2000] divides the information needs into: *verificative, conscious topical, and muddled topical* information needs where the first one is a fact-oriented task and the others are topical-oriented tasks. For each of these tasks it is possible to predict the associated search behaviour Borlund and Dreier [2014]. The association between INs and search behaviour is as the following:

1. *Verificative information need (VIN)*: it is characterised by a being well-defined and stable information need as the user wants to "verify information objects with known non-topical (structured) data, such as author names, journal name, and facts". The associated search behaviour is *querying and filtering* where searchers are more confident in formulating the queries and filter irrelevant information within a relatively short time.

2. *Conscious topical information need (CIN)*: it is a topical, well-defined, and variable information need, and the searcher aims to "clarify, review, or pursue information in a known subject matter and domain, where known subject matter signifies topical (unstructured) data about contents, such as terms, concepts, and image representation". The associated search behaviour is an *exploratory* search behaviour characterised by dynamic interaction including querying and navigating.

3. *Muddled topical information needs (MIN)*: the information need is topical and poorly defined where the searcher explores "new concepts and relations outside known subject matter or domain". Consequently, the search behaviour associated with MIN is *browsing, search loops, and trial and error* behaviour.

Task reveals the real information needs of searchers and their search intentions. Researchers distinguish between three levels of tasks: *work tasks* which is "an

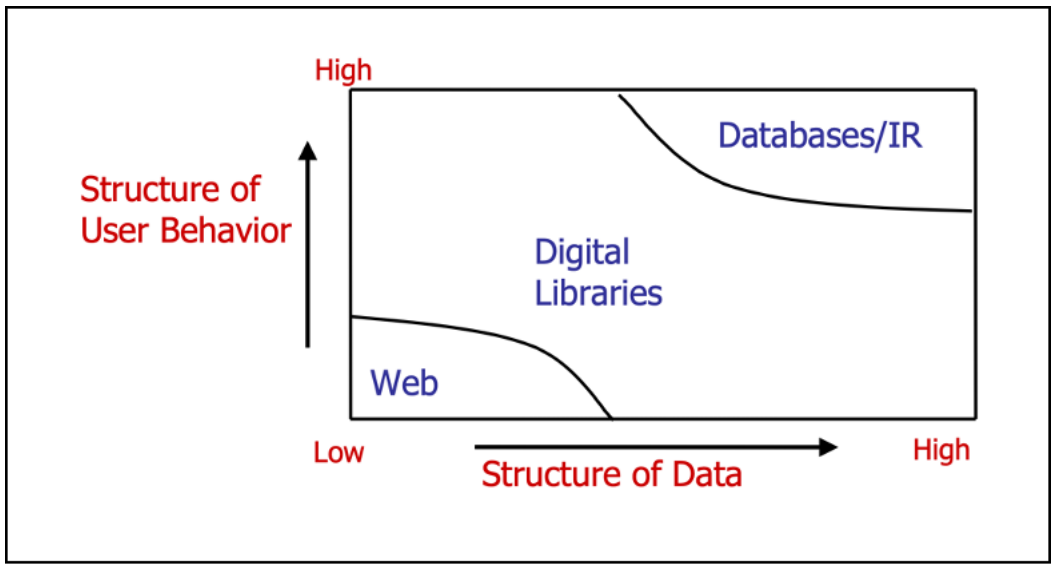


Figure 2.2. DLs in the Information Space Casarosa [2010]

activity people perform to fulfil their responsibility for their work" Li [2009]. *information seeking task* it is the subtask of the work task, and *information searching tasks* it refers to the micro-level of the information seeking task Byström and Hansen [2005]; Xie and Cool [2009].

The relationship between the information system design and the information seeking can be traced in figure 2.2. As the data structure is getting complex, users make more effort to locate the information Casarosa [2010]. Thus, it is vital to understand the information seeking process in order to build and design effective IR systems Belkin [2015].

Many studies have been conducted to investigate the most influential factors in ISB. According to the most cited and well-established models, the factors can be categorised into seven main categories: *information needs*; *roles and tasks* (frequency, predictability, importance, and complexity); *information sources and awareness* (familiarity, trustworthiness, packaging, timeliness,cost, quality, accessibility); *context* (cultural, organisational, social, sector's type); *socioeconomic* (wealth, contact networks, occupation); *personal* (education, attitude, experience, motivation, values physiological, affective or emotional needs, cognitive needs, demographics, environmental variable, personal style of seeking and personal relevance, person's degree of knowledge); and *situational* (situation specific need, available time, state of health) Wilson [1981, 1999]; Leckie et al. [1996]; Savolainen [1995]; Johnson et al. [1995]; Byström and Järvelin [1995]; Ingwersen and Järvelin [2005].

In their scoping review of individual differences in ISB and retrieval research, O'Brien et al. [2017] concluded that it is difficult to assert the influential individual differences to the information interaction due to the complexity of human experience.

In relation to the failed searches, Borgman [1996] identified the gap between the online catalogues' design and the search behaviour that could be the reason behind the search failed. She confirmed that interacting with the online catalogues requires a rich conceptual framework for information retrieval, where most of the end-users lack the conceptual knowledge for searching comparing to the expert librarian searchers. Recently, Xie and Cool [2009], identified user's domain knowledge and search experience as significant user attributes that cause the failed searches, besides the lack of awareness of the purpose of certain limiters, incorrect selection of the search index or misuse of the facets Trapido [2016]; Peters [1989].

2.3.2 The Affective Variable in Relation to the Failed Searches

Although the cognitive perspective has long dominated the area of information behaviour, affective behaviour includes emotion, mood, and preference needs to be considered as an integral part of developing a holistic understanding of an individual's approach to information seeking and use Julien et al. [2004]. Much of our daily experiences influence and are influenced by the emotions we feel, our experience with digital libraries has no exceptions. Emotions, as one of the affective variables, can be defined as "an integrated feeling state involving physiological changes, motor-preparedness, cognitions about action, and inner experiences that emerges from an appraisal of the self or situation" Mayer et al. [2008] (p. 508). Emotions not only regulate our social encounters, but also influence our cognition, perception and decision-making through a series of interactions with our intentions and motivations Scherer [2001]

A pioneer model that encompasses user affective experience as well as cognitive constructs within the information process is Kuhlthau's Information Search Process model Kuhlthau [1993]. Other research suggests that experiencing failure has marked emotional and psychological consequences across a range of individuals and settings Johnson et al. [2017]. As an example of the setting effect, Poddar and Ruthven [2010] found that the artificial task has higher uncertainty comparing to the genuine search tasks. They also identified the search task types, where complex search tasks have lower positive emotions and more uncertainty before and after searching. The study of Gwizdka and Spence [2007] found the time spent on searching (search duration) is associated with subjective

feelings of lostness in Web searches.

In the domain of the research in librarianship and information science, a seminal systematic review was conducted by Julien et al. [2004]. The review demonstrated the importance of the affective dimension e.g., emotion or confidence in human information behaviour. It also indicates that system-oriented research in librarianship and information science pays little attention to affective variables and encourages the research in the librarianship and information science community to move beyond a focus on system technicalities or cognitive aspects of searching behaviour to include affective variables.

Bilal [2000] investigated children's ISB to include the affective state, among other factors, on fact-based search tasks. The study concluded that children experienced negative feelings in a few cases including lack of matches and difficulty in finding the answer Bilal [2000]. The importance of considering the affective variables in information behaviour is confirmed by introducing the affective load theory (ALT) Nahl [2005]. Nahl [2005] reported how an investigation of students' affective loads revealed that individuals with higher affective coping skills (e.g., self-efficacy, optimism) were better able to function in information seeking situations, even when they had lower cognitive skills. Lopatovska [2014] found that there are direct relationships between primary emotions and search actions. Mckie and Narayan [2019] developed a Lib-Bot, a digital library chatbot, aiming to minimise the effects of library anxiety when searching in library databases and librarian services.

The positive affect of emotions has been investigated intensively in the research in librarianship and information science research, compared to the negative affect Fulton [2009]. Previous research confirmed the difficulties that searchers face when interacting with OPAC and digital library systems. Investigation of search failures from the searcher's perspective is scarce. Thus, understanding search failures, and the emotions associated with failure situations need further investigation. Such exploration might help the designers provide more usable and useful digital library systems.

Our brief review of studies on failed searches suggests that the affective variables have not yet been extensively investigated in the domain of the digital library. To this end, we inspect what are the emotions associated with failed searches. The description, results and discussion of this study presented in chapter 4.

The result of the study in chapter 4 motivated us to redesign the current DL interface followed the user-centered design approach. The user-centered design approach requires a deep understanding of the user experience (ux) before the processing of the redesign. The following section review the UX literature.

2.4 User Experience in DL

DL used to be evaluated explicitly by collecting data of attitudinal surveys e.g. Priestner and Borg [2016], or implicitly by generating reports of the usage from the LFs e.g. Debowski [2001]. The usability studies measure the performance characteristics; effectiveness, efficiency, and satisfaction Secretary [1998]. Massis [2018] and Appleton [2016] argued that the traditional quantitative metrics and measures are not sufficient to reveal the real value of the DLs. Such quantitative metrics failed to provide a full understanding of the quality of the experiences, instead the user experience (UX) should be the ultimate criterion of the user-centered design Alben [1996]. Cronin [2014] suggested to deploy UX techniques to address age-old problems reported in the DL evaluation studies including the lack of awareness of the search tools, incorrect selection of the search index or misuse of the facets Peters [1989]; Trapido [2016]. UX defines as "the overall effect created by the interactions and perceptions that someone has when using a product or service" Priestner and Borg [2016] (p.3). UX is mainly based on ethnographic methodologies that aims to explore the interactions of the real users along with other user attributes e.g. experience level, attitude, and affective variables including emotion and mood Priestner and Borg [2016]. The significance of the UX research embodied in increasing user satisfaction and improving the product quality Mahlke [2008].

The evaluation of the DLs is complicated due to the overlap of different interaction dimensions including the interaction with the task, interaction with the supportive technologies i.e. interfaces, and the interaction with the information resource Saracevic [2000]. Accordingly, DL evaluation studies could be classified into: interaction-design-based evaluation, task-based evaluation, and the overall performance evaluation studies. Our study focuses on the former one and it is based on the framework proposed by Mahlke [2008] as in figure 5.1. We contribute to the DL research by deploying a user-centered method for fully understanding the UX. The study described here was designed based on the framework defined by Mahlke [2008]. Basically, the framework identified three main factors that influence HCI; these are: system properties, user characteristics, and context parameters. Those three factors affect three central components of the UX, namely: the instrumental and non-instrumental quality perceptions, and emotional user reactions. The following sub-sections review the previous works in the DL sector which are related to Mahlke's framework.

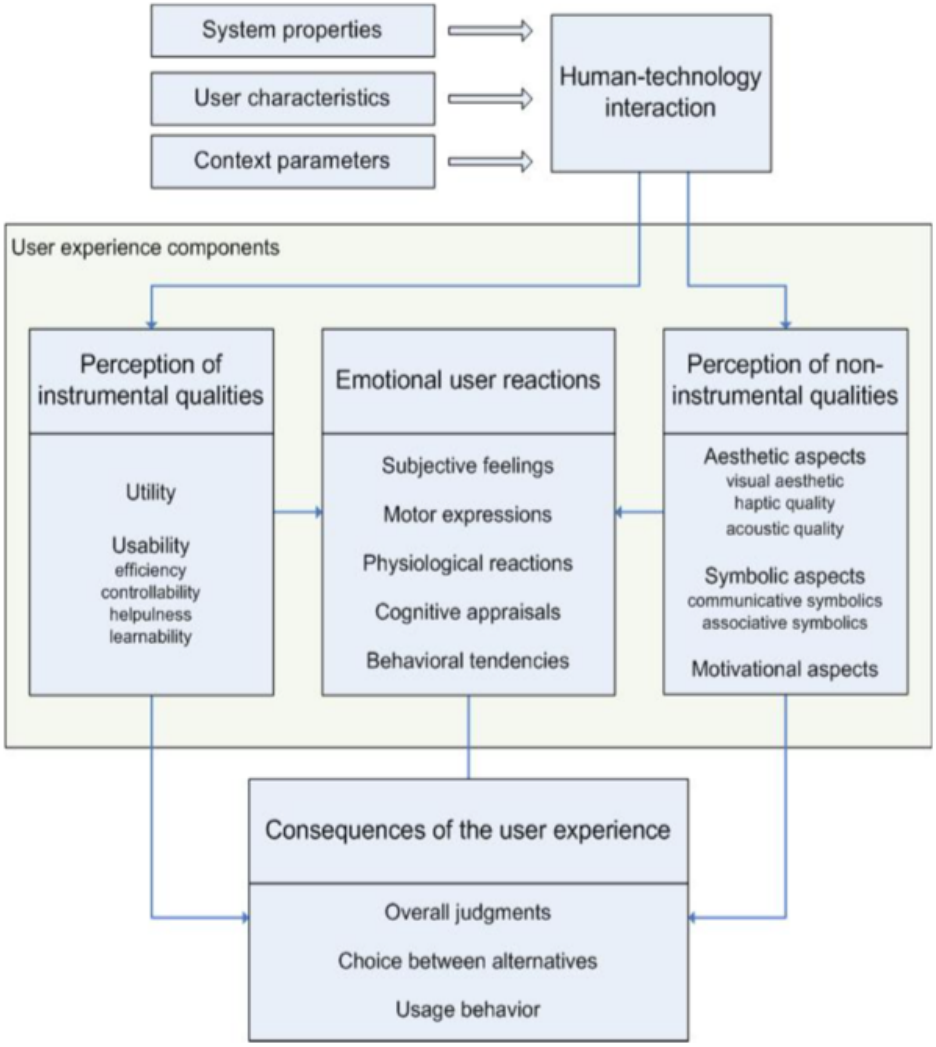


Figure 2.3. Interaction User Experience Framework Mahlke [2008]

2.4.1 System Properties:

Mahlke [2008] categorized the system properties into: the properties that affect the interaction indirectly (e.g. weight and the geometry of the product, and surface properties), and the properties that have direct effect on the interactions (i.e. the interface). DL interfaces, as the main interaction channel between the users and the system, is the focus of this study. Most of the DL utilized the discovery search interface (DSI) Asher et al. [2013]; Bossaller and Sandy [2017]. One of the design principles behind the DSI is the information architecture which defines as the "combination of organization, labeling, and navigation schemes within an information system" Rosenfeld and Morville [2002]. Consequently, the DSI is a gate that enables accessing to multiple databases through a single search box Gross and Sheridan [2011]. The DSI is recognized to possess an abundance of search tools: facets, advanced search, and sorting. Accordingly, three main components can be evaluated in the DL interfaces namely: the data model describing the information available to the users (e.g. labeling and terminology), the navigational model including the procedures for gaining access to data model (i.e. search tools e.g. facets and advanced search), and the interface style referring to the surface presentation (e.g. style and color) Cushman and Rosenberg [1991]. Evaluating the navigational model is the focus of this study. In the literature describing the effect of the DL interface including the discovery and OPAC interfaces, researchers are divided into two groups: the supporters who found that the search tools are useful and valuable features, because they can be used to reduce a large result set, explore unknown areas, and move through large information spaces e.g.Denton and Coysh [2011]; Niu and Hemminger [2015]. And the opponents who found that the current DL interface and search tools are confusing, and overwhelming e.g. Blumer et al. [2014]; Capra et al. [2007]. Students in particular do not always understand the mechanism of the search tools. Therefore, such tools are not used to their full potential and not often underutilized Dempsey and Valenti [2016]; Hamlett and Georgas [2019].

2.4.2 User Characteristics:

User knowledge, among other users' attributes, was identified as a significant factor that characters the interaction and determines the search outcome. Hölscher and Strube [2000] distinguished between two levels of knowledge: domain knowledge which shapes users' understanding of information needs and relevance assessment Liu and Zhang [2019], and searching knowledge refers to the expertise level of using the DL Deodato et al. [2016]. The effect of the domain knowl-

edge embodied in the query behavior and relevance judgement. The domain experts show more complex and relevant keywords compared to the novices who might have difficulties in generating appropriate keywords Zhang et al. [2005]. In terms of the searching knowledge, lack of appropriate knowledge of using the search tools negatively affect the speed and accuracy for locating information Rosman et al. [2016]. Often users lack the capability to choose the suitable advance functions that suits their information needs Fagan et al. [2012]. Besides, the misuse of the facets leads to failed searches Trapido [2016]. Too many elements in the interface could be counterproductive. Users are overwhelmed by the number of options available within the interface Johnson [2013].

2.4.3 Context Parameters:

A variety of context parameters influence UX. Examples of the context parameters are: the settings where the experiments are conducted i.e. controlled or naturalistic settings, observation time, user's degree of motivation, type of tasks or activity, and the nature of participating i.e. rewarded or volunteering Mahlke [2008]. As an example, Huurdeman et al. [2018] compared the UX of two different environments; the physical and DLs. Similarly, Zha et al. [2015] compared between the web and mobile libraries.

2.4.4 The Instrumental and Non-instrumental Perceptions:

Mahlke [2008] distinguished between the instrumental and non-instrumental qualities by referring the instrumental quality to the usability and utility values, whereas the non-instrumental qualities go beyond the functional objectives and include aesthetic and symbolic qualities. Nielsen [1993] considered the utility and usability as being the two important aspects which influence the system acceptance. Nielsen [1993] differentiated between the two terms as "..utility is the question of whether the functionality of the system in principle can do what is needed, and usability is the question of how well users can use that functionality" (p. 25). The instrumental quality is evaluated by conducting usability studies. In such studies, the users' perceptions are considered Johnson [2013]. Most of the usability studies based on predefined tasks that look at particular aspect of the interfaces. The non-instrumental qualities have gained little attention in the field of the DL. As Schmidt [2016] stated that "usability and utility seem more fundamental to a library's success. After all, it does not matter if a library has a beautiful building, a fancy website, and a professionally designed logo if it's difficult to use". Instead of measuring only the instrumental

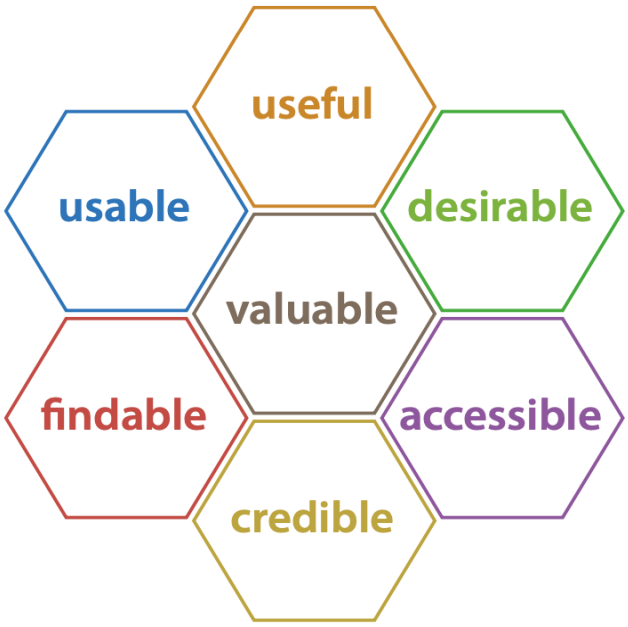


Figure 2.4. The 'Honeycomb' Model of User Experience of Morville [2004].

and non-instrumental quality perceptions, Morville's honeycomb model used in this study Morville [2004]. Morville [2004] claimed that evaluating a system should be moved beyond the usability and consider other facets that affect the UX. The honeycomb model, as in figure 2.4, consists of seven dimensions, those dimensions were explained in the results section.

2.4.5 Emotional Reactions:

Emotions, as one of the affective variables, can be defined as "an integrated feeling state involving physiological changes, motor-preparedness, cognition about action, and inner experiences that emerges from an appraisal of the self or situation" Lopatovska [2014]. A pioneer model that encompasses affective and cognitive constructs is Kuhlthau's Information Search Process model Kuhlthau [1991]. Julien et al. [2004] encouraged the researchers to move beyond the system technicalities or cognitive aspects, and consider the affective variables. The study of Bilal [2000] demonstrated that children experienced negative feelings when they experienced "lack of matches" and "difficulty in finding the answer". Zha et al. [2015] investigated the flow experience of the DLs and mobile libraries. Mood is an affective state that is closely related to emotion. Mood

guides our behavior, and have a broad influence on one's perceptions, judgments, and behaviour Thayer [1997]. The decision to include mood in our study was informed by previous research that emphasized the importance of the mood in HCI. For instance, unlike emotion, mood is a mild- or moderate-intensity feeling that usually lasts longer than does emotion, and it represents a summary of affective states Morris [2003]. Mood can be influenced by internal and external stimuli e.g. weather Watson et al. [1988]. Researchers investigated the effect of the mood prior to the search, and during the search. The study of Bilal and Bachir [2007] concluded that moods and attitudes prior to the search affected the search process. Lazar et al. [2006] found that the frustration levels during the search were negatively correlated with the mood after the session. Lopatovska [2014] suggested that search performance did not affect participants' moods.

To provide users with more pleasurable experiences, it is essential to understand the mood. Thus, we are more interested in exploring the factors that might affect the mood. Examples of techniques that measure the mood are psychological signals and self-report questionnaires Desmet et al. [2016]. In this work, a pictorial self-report scale called: 'pick-a-mood' was used. The scale, as in figure (3), was proposed by Desmet et al. [2016] as a solution for people who often have little time or motivation to report their moods.

The purpose of this chapter is to investigate the UX in the DL by considering threefold aspects. The first is to understand to what extent the DL functions are usable, the second is to investigate the user perceptions of the instrumental and non-instrumental quality, and the third is to examine the associated mood before and after the search experience. Thus the contribution of this chapter to the DL research is deploying a user-centered method for fully understanding the UX. The design and results of the study is explained in chapter 5.

The results of this study inspired us to redesign the current DL interface by adopting the visualisation approach. The following section reviews the related work of visualisation.

2.5 Interface Visualisation

The gap between system design and research in HCI was reported in 1994 by Jeffrey and Chisnell [2008]. Back to the eighties, the current DL interface heirs of the classical interfaces of the document databases on CD-ROM that built on a particular architectural anatomy including the controlled vocabularies and the filtering function through certain criteria e.g. author, year of publication, subject ..etc Nualart et al. [2014]. Since then, and in contrast to the other information

Figure 2.5. Pick-A-Mood Scale Desmet et al. [2016]

systems, the DL interfaces have barely evolved on the basis of the findings provided by user studies, nor have the advances developed in specific disciplines, such as information architecture, ISB, or those derived more generally from user experience (UX) studies Pérez-Montoro and Nualart [2015]. This lack of attention to the quality of user interaction has prevented the deployment of advanced techniques of ISB e.g. berrypicking Bates [1989], and the exploration of ways to support different ISB i.e. divergent behaviour.

2.5.1 Challenges of the Traditional DL Interfaces:

According to the systematic literature review conducted by Gaona-García et al. [2017] to identify the main challenges in the field of the DLs and repositories, four main factors were identified as challenges in the DLs systems, they are:

Although *knowledge representation schemes* play a significant role in the classification, categorization, linking and management of digital resources which usually are carried out by domain experts, examples of strategies behind knowledge representation schemes are: hierarchical tree structure, and a faceted classification. These classification tools are not always used and exploited by the

end-users. This is due to the lack of awareness of the context of the classification scheme and how terms are interrelated, thus users might not be aware of possibilities of such functions. Consequently, the end-users ignore the benefits of using these types of interfaces.

The poor quality of the description of the resources through the use of metadata, this can be argued to be the most important factor that influences the location and access of digital resources. Examples of associated problems with describing the metadata are: ambiguity, inconsistency, redundancy, and lack of accuracy.

DLs offer different kinds of search and navigation mechanisms to facilitate access to relevant resources. *The misuse of the search and navigation methods* is another challenge that needs to be considered when designing and evaluating DLs. It encompasses the lack of a common and adequate terminology, problems in the accessibility of resources, and errors in the association of terms. Accordingly, users might face difficulties in finding resources, or returning to a previously accessed record.

All these listed challenges clearly affect the interfaces, therefore the final challenge relates to *the usability problems associate with the search interfaces*. Examples of the limitations associated with poor DL interfaces are: problems to locate and display resources, problems using the interface and interface's functionalities, combining search and navigation methods, higher learning effort and reliability of the search tools. In conclusion, the use of the DL is not always straightforward, because their interfaces and search mechanisms do not always offer and support adequate searching strategies Gaona-García et al. [2017]. Here, we present an alternative method to the faceted search interface, the visual search interface.

2.5.2 Information Visualisation

Information visualisation (*Infovis*) was adopted as an opportunity to support different ISB and enhance the overall user experience (UX). Infovis is a multidisciplinary research areas, including HCI, computer graphics, and cognitive psychology. It is an alternative approach to present textual data and lists to reinforce human cognition.

Infovis can be defined as "the use of computer-supported, interactive, visual representations of data to amplify cognition." Card [1999] (p. 6). It implies transforming and representing a wide variety of data, e.g. non-spatial and non-numerical data, in a visual form to help users to process them into understandable information Börner and Chen [2002]. The visualisation works as external cognition aids by reducing search time, facilitating the browsing and identifica-

tion of relevant metadata, and providing a quick overview of the coverage of a library system Börner and Chen [2002].

Compared with traditional non-visual DL interfaces, visualisation provides a more user-accessible form of interaction with information, offers more rapid search times, and more efficient formulation of queries as visualisation gives insights that could not be recognized without the visual processing of data Hienert et al. [2012]. Bauer [2014] stated that visualisation enhances interaction with the results, and facilitates the refinement of subsequent queries by showing different semantic relationships between the documents retrieved. Butcher et al. [2011] confirmed the important role of visualisations in supporting deeper cognitive processing during online information search and results evaluation.

Visualisation converts different types of data and expresses them visually to augment human cognition by leveraging human visual capabilities to make sense of the abstract information Gaona-García et al. [2017]. According to our best knowledge, Shen et al. [2019] conducted the most recent survey in the field of the infovis, and they summarised the most common visualisation methods as: *line* graphs which commonly used with one dimensional data, *plot* graphs e.g 2D/3D Scatterplot that project data into a two-dimensional or three-dimensional representation on screen, *map* graphs, a well known example is the heat map, *parallel* coordinates where a vertical line is used to project each dimension whose maximum/minimum values are scaled up/down to fit the upper/lower boundaries of the line, and finally the *radial* coordinates where multivariate data can be displayed in a circular or radial pattern.

In the field of DL, proposed visualisation solutions follow two directions: (i) improving the representation of the content of the retrieved documents or the entire database e.g. Nualart and Pérez-Montoro [2013]; Repke and Krestel [2020], or (ii) introducing interactive visualizations of the set of results; the set of documents returned by a search engine in response to some query e.g. Cugini et al. [2000]. The second stream is the focus of this research.

Pérez-Montoro and Nualart [2015] classified the studies that proposed different prototypes of the visual presentation of the results into: two-dimensional visualizations prototypes that visualise the relationships between the retrieved documents by using maps or clusters Andrews et al. [2002], or by using tables or grids Shneiderman et al. [2000], and the prototypes that used the three-dimensional visualizations of the retrieved results Cugini et al. [2000].

2.5.3 Visualization of Information in DLs

This section summaries different prototypes proposed in the DL sector. Different aspects of the DLs can be visualised including metadata, queries, co-authors and citations either on the document level or the result level Hienert et al. [2012].

A pioneer design in the DL sector was proposed by Robertson et al. [1991] who visualized hierarchical information structures on the keywords level with the Cone Tree technique. Later on, Kumar et al. [1998] proposed Interactive Timeline Editing and Review (ITER) to visualise DL metadata based on timeline technique. Andrews et al. [2001] introduced xFIND gatherer-broker architecture that helped in presenting traditional ranked list in an interactive scatterplot (Search Result Explorer) and using dynamic thematic clustering (VisIslands). Plaisant et al. [2002] presented a visualised interface based on the tree topology along with the integrated search and filter functions. Sheth and Cai [2003] implemented a radial tree layout method to visualize the National Library of Medicine's controlled vocabulary thesaurus. Keim et al. [2004] developed a technique based on the Information Seeking Mantra; overviews first, zoom and filter, and details on demand to visualise a dataset. Wei et al. [2010] combined interactive visualization techniques with advanced text analytics to help users explore and analyze large collections of text. For archival purposes, Xu et al. [2011] developed an interactive visual analytics application to help archivists analyze large-scale digital collections. Choo et al. [2013] proposed a visual analytics system, VisIRR, for document discovery based on tree technique.

Different visualisation tools were proposed too. For example, PaperLens was presented by Lee et al. [2005] to analyse and visualise papers to help users to discover research trends, patterns and relationships. At the same time, Marks et al. [2005] developed ActiveGraph by visualising data via a two- or three-dimensional scatter plot to provide users with a concise, customizable view of objects in a DL. Wong et al. [2011] introduced INVISQUE (INteractive Visual Search and Query Environment) to overcome the problem of list-based searches, where the information is represented by a two-dimensional spatial canvas, and search results are presented as index cards.

In conclusion, although there were different attempts of implementing visualisation interfaces and tools in the DL systems, Pérez-Montoro and Nualart [2015] stated that visualisation has not yet been widely implemented in the DL for practical and methodological reasons. Examples of the practical reasons are: visualisation tools are not integrated in the standard search field-result list paradigm of DL which means users perceive them as being secondary tools Hienert et al. [2012]. Moreover, visualising results are not very intuitive to users due to the

high level of abstraction and conceptualization of the visualisation architecture.

Regarding the methodological reasons, along with other reasons, Pérez-Montoro and Nualart [2015] confirmed that most of visualisations prototypes used small collections of documents and their efficient use with large collections has not been demonstrated. Moreover, most of the proposed techniques were not tested with the involvement of end users, and this makes it difficult to draw any clear conclusion about their efficiency. Therefore, in response to such gaps, in this research we aimed at investigating the efficiency of the visualisation in DL platform by designing and implementing a radial visualisation interface to a large-scale DL. Hence, we conducted a user-centred evaluation study to explore this area. The study and its result described in chapter 7.

Part II

Exploring

Chapter 3

Exploring Usage Patterns of a Large-scale Digital Library by Analyzing the Log Files

3.1 Introduction

Users follow different search strategies or moves when interacting with systems. One method for understanding search strategies is investigation of users' digital footprints, or usage patterns (UP), found in log files (LFs) Ndumbaro [2018]. LF analysis is an attractive approach due to the availability of the data considering a larger scale of interactions Hajek and Stejskal [2017]; Han and Wolfram [2016], and being less subject to bias and non-invasive, as it records interactions between systems and users that might not be clearly observed in laboratory experiments Bollen and Luce [2002]. Our explorations of the user interactions with the digital library started by analysing higher-level interactions of a heterogeneous population with the aim of looking into how people actually use a large-scale DL.

Thus, in this chapter, we aim to answer the following research questions:

1. *What are the potential UPs of a large-scale DL?*

2. *What are the main features that characterise the above UPs?*

To answer these questions, a log file analysis was conducted. Section 3.2 describes the materials including the study's platform (RERO Doc) DL and the dataset. Section 3.3 explains the experimental setup. Sections 3.4 and 3.5 explain the experiments and the results. Section 3.6 discusses the results.

3.2 Materials

The following sections describe the study platform (RERO Doc) DL, and the obtained dataset.

3.2.1 Challenges to Conduct Log File Analysis

Regardless of the importance of the LF as a valuable information source, we found it a challenge to find a DL who would agree to share the log files data. From our experience we contacted different DLs and they apologized for one or more of different reasons. Those reasons can be summarised as follows: firstly, most of the libraries apologised for privacy issues as they cannot share the IPs data, nor they could anonymize the IP addresses. This might result in inaccurate session identifications. Secondly, the librarians were afraid that extracting such data would be time consuming and required effort by system managers. Thirdly, some of the contacted libraries did not have their own log files either because the library did not implement a logger tool, or because it is a consortium library which has different subscriptions of different scholarly databases managed by third-party suppliers. Thus, general usage statistics (e.g. click through, search strings, and hourly search statistics) are generated by a third-party partner e.g. Ex Libris [1] and Primo. The libraries that we contacted were: Andersonian Library, University of Strathclyde [2], Library of KAUST, King Abdullah University of Science and Technology [3], RMIT University Library[4], SDL, Saudi Digital Library [5], and Zentralbibliothek Zürich, Zurich's central library [6]

The only DL that was willing to share its log files data was RERO Doc DL [7]. The next section presents RERO Doc DL.

3.2.2 RERO Doc

This study was conducted in collaboration with RERO Doc. RERO Doc is a DL connecting the libraries of Western Switzerland as a public service. It includes the majority of the university, patrimonial, public and specialised libraries from

[1] https://knowledge.exlibrisgroup.com
[2] https://www.strath.ac.uk/
[3] https://library.kaust.edu.sa
[4] https://www.rmit.edu.au/library
[5] https://sdl.edu.sa
[6] https://www.zb.uzh.ch/en/ueber-uns
[7] https://Doc.rero.ch

- IP0 - - [28/May/2017:06:25:44 +0200] "GET /record/288131/files/Faeh_Sira_2013_2016.pdf HTTP/1.1" 200 1143106 "https://scholar.google.com.au/" "Mozilla/5.0 (Windows NT 6.1; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/58.0.3029.110 Safari/537.36" 953196 0 NOSSL
- IP14 - - [28/May/2017:06:32:15 +0200] "GET /search?ifacets=author%3AColbois%2C%20Bruno%20%28Codir.%29&ifacets=keyword%3Ain%C3%A9galit%C3%A9s%20isop%C3%A9rim%C3%A9triques&jrec=1&keep_facet_filter=1&ln=it&so=d HTTP/1.1" 200 4404 "-" "Mozilla/5.0 (compatible; SemrushBot/1.2~bl; +http://www.semrush.com/bot.html)" 656233 0 NOSSL

Figure 3.1. Snippet of the Log Files

the cantons of Geneva, Fribourg, Jura, Neuchâtel and Valais, besides the law libraries of the Confederation.

The library offers free access to its contents and services to users located worldwide. It had approximately 6 million accesses from May 2017 to January 2018 from different countries (e.g. Switzerland, France, Germany, United States, Canada, and Algeria). The items retrieved cover many different domains (e.g. nursing, economics, health, language, computer science, and history). RERO Doc's catalogue lists 6 million items of various formats including dissertations, books, articles, periodicals, photographs, maps, digitised press, music scores, and sound recordings. RERO Doc supports four different interface languages: French, English, German, and Italian. The design of the system supports different information discovery activities; (searching, browsing, and navigating).

The library implemented a Google Analytics tool to gain regular statistical review of its usage. RERO Doc serves a diverse population coming from different countries, with languages and cultural variations. The analysis of LFs resulting from its users' interactions might provide a rich and realistic insight into different searching experiences. RERO Doc is an ideal investigation platform due to its varied population, rich content formats, and different functionalities.

3.2.3 Data Description

The study relies on the analysis of the LFs transactions. Examples of the information recorded in the LF texts are user identifications, date and time, and uniform resource locator (URL), which embodies the type of requests and content viewed. Figure 3.1 shows a snippet of the log files.

We collected data from the RERO Doc server for an eight-month period: May 2017—January 2018. This long period minimised the likelihood of atypical UPs

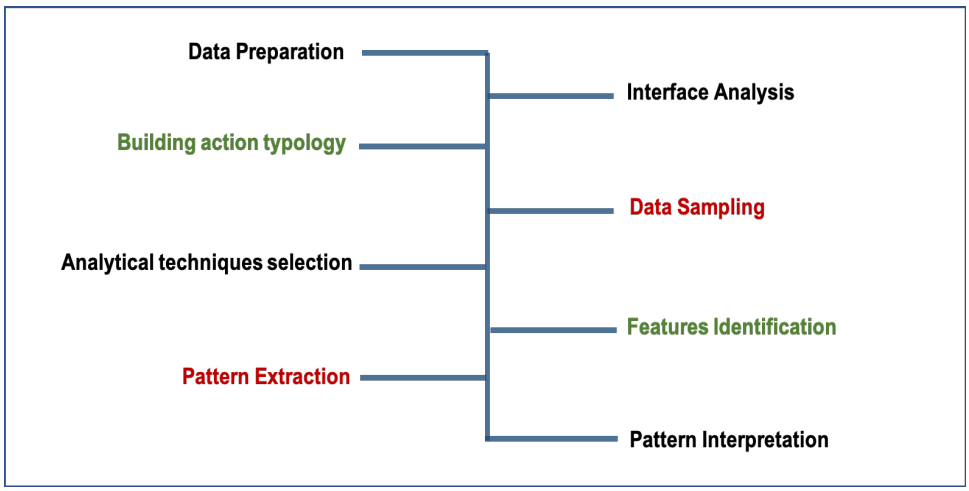


Figure 3.2. The Framework of Exploring UPs in LFs

that might exist within a shorter time. The dataset consists of 59 million records totaling around 20 GB.

3.3 The Experimental Setup

Discovering information from LFs is not a trivial process; it requires a strategic plan to gain a better understanding of the hidden information. Here we describe the experimental setup that we followed to discover the UPs. The phases are described in the proposed framework as in figure 3.2

3.3.1 Data Preparing and Processing

Because the obtained data was raw and could not be processed directly, the data was preprocessed through the following phases.

- ***Data Cleaning:*** the data was cleaned by eliminating erroneous records and requests for styles (e.g. "gif", "ico", "css").
- ***Data Parsing:*** the unstructured data in the LFs was parsed and split into meaningful parts, including user internet protocol (IP), time stamp, URL request, referrer and user agent.
- ***Sessions Determination:*** because not all search websites require login and not all the logs contain hypertext transfer protocol (HTTP) cookies, as in the case of RERO Doc, the sessions had to be determined.

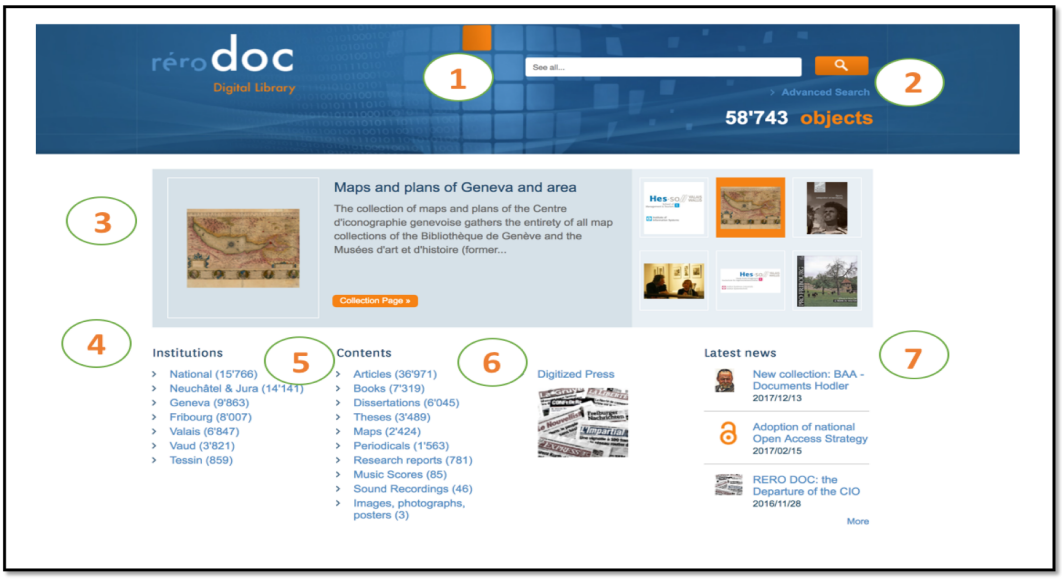


Figure 3.3. RERO Doc Interface

A session is a "common unit of interaction that is used in search log analysis" Russell-Rose et al. [2014]. Sessions determination is a crucial activity in LF studies. Sessions are used to contextualize interactions within a bounded sequence of actions. The sessions are used, as the units of analysis, to find search behaviour patterns Niu and Hemminger [2015], to investigate query modification behaviour Hollink et al. [2011], or to identify successful or failure interactions Jansen and Spink [2006].

Several approaches can be used to estimate the sessions' boundaries, including subject analysis of queries as in Guo et al. [2009], semantic analysis of topics, and time interval Islamaj Dogan et al. [2019]. Time interval is a commonly used session bounded as in Hollink et al. [2011]. The time interval is based on a threshold of user inactivity, and the 30 minutes timeout is considered a default method of breaking click-streams into sessions Niu and Hemminger [2015]; Srivastava et al. [2000].

Because this research excluded query analysis, the time interval approach was adopted. Thus, session boundaries were determined by the combination of user IPs+user-agents+ time (next log within 30 minutes).

As the sessions were determined, automatic session identifications (IDs) were generated. We concentrated on the session ID instead of the user IP because:

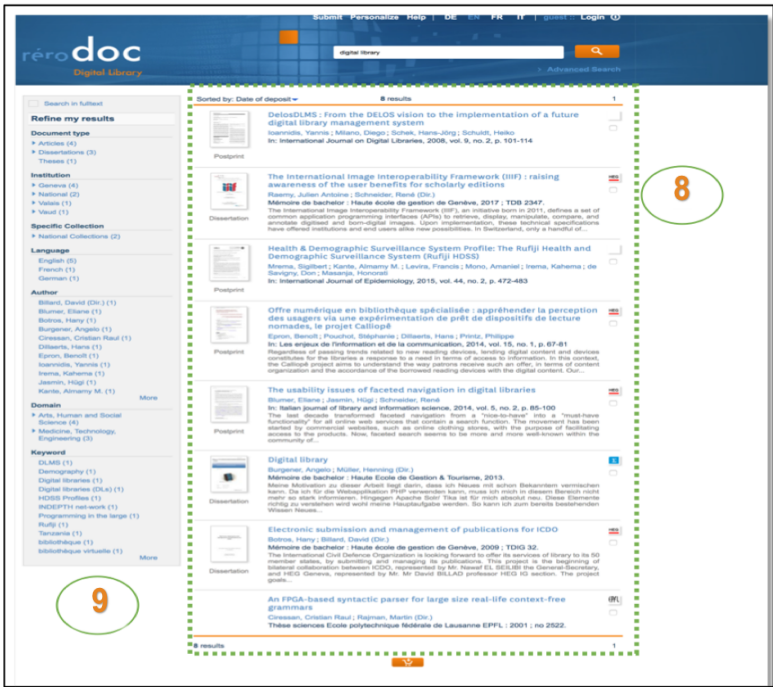


Figure 3.4. Result Page

- It is impossible to distinguish between different users as many users might share the same computers and the same IP addresses.
- The UPs of the same user are changeable based on his or her information needs or other contextual factors.

- ***Removing Non-Human Sessions:*** after the determination of the sessions, non-human requests were identified by considering the internet bot, crawl, and spider browser agents (e.g. SemanticScholarBot, Googlebot, Baiduspider, ZemlyaCrawl). Such requests were removed, and consequently the data was reduced to 9 GB with only user-generated requests totaled 28 million records.

3.3.2 Interface Analysis

For a better understanding of the potential UPs, we first familiarised ourselves with the interface design and its functionalities before starting to manipulate the data. The home page of RERO Doc is a multifaceted interface with functions that

facilitate and support information discovery (searching, browsing, and navigating). Nicholas et al Nicholas et al. [2006] differentiate between those activities as follows: searching, where the users submits queries through the search functions (i.e. simple or advanced); browsing, where users browse the content pages, lists, and menus; and navigation, where the users move around in the digital environment in pursuit of information. We adopted Nicholas et al. [2006] definitions to distinguish between the activities of RERO Doc users.

RERO Doc's interface, as shown in figure 3.3, has a hierarchical structure that gives priority to the search functions: the simple search (1), and the advanced search (2). Users may start their sessions by accessing the content from different landing pages (e.g. search engines, click links from emails, or directly through the home page).

Once the user starts a session from the home page, multiple functions are available. The users can navigate various collections (3) which are located on the middle of the home page. Also, the users may navigate RERO Doc's content by access via the institution (4), type of contents (5), or by navigating the digitised press (6), which is promoted on the home page to provide quick access to the digitised materials. The users also can view the latest news (7) about RERO Doc from the home page.

The header with the search function is permanent on the search results page. The results page as in figure 3.4 contains the result list (8) integrated with the browsing bar. The users can evaluate and browse the results page simultaneously with multiple facets (9). Facets (sets of "meaningful labels organised in such a way as to reflect the concepts relevant to a domain" Carevic et al. [2018]) enable the user to interact with the result lists, and help them to narrow down their search results based on certain features. RERO Doc offers the usual facets: author, keyword, language, keyword, and specific RERO Doc facets: institution, specific collection, domain, and type of Document.

The result page also offers extra functionalities such as filtering results by full text function, and sorting the results according to the date of the deposit, publication date, title and author. Once users click on a result, they may display a snippet, view a file with the Multivio viewer (a viewer application available in RERO Doc), or download the file. Otherwise, they may check similar records, change searching domain or collections, add a record to a personal list, or copy the link as a "permalink".

3.3.3 Building the Topology of Session Action

The URL requests embodied in the LFs reveal the users' actions. Analysing and understanding the structure of the website is a prerequisite for coding the URL requests. For example, Chen and Cooper Chen and Cooper [2002] built a hierarchical taxonomy of the website. Instead we built a code schema of the interactions extracted from the URL requests, as shown in table 3.1, consisting of 18 actions to be used in describing UPs.

3.3.4 Analytical Techniques

Analysis of UPs can be conducted to: (i) gain a holistic understanding of the UPs (e.g. frequency of databases access). Such analysis can be achieved by basic descriptive statistics (e.g. Arshad and Ameen [2015]). Traffic analysis tools can be used for such a purpose (e.g. Google Analytics); (ii) explore the potential UPs exhibiting similar behaviour by implementing unsupervised machine learning techniques (e.g. cluster analysis Chen and Cooper [2001]); or (iii) classify the UPs by mapping data into predefined classes by supervised learning algorithms Bhargav and Bhargav [2014].

Because the aim of this work is to explore the potential UPs of a wide range of users, unsupervised machine learning techniques were implemented —in particular, clustering techniques. Clustering was deemed a suitable technique for our problem for the following reasons: (i) the lack of inherent knowledge of the data and the UPs that might be presented in the data; (ii) the lack of ground truth data; and (iii) not enough complete clean-labeled data could be guaranteed; (iv) our ultimate aim is to identify similar UPs exhibiting similar behaviour. Cluster analysis is an unsupervised method aiming to organise data into classes considering the minimal intra-cluster distance; the similarity between objects within the same cluster, and the maximal inter-cluster distance; the dissimilarity between the objects of different clusters Romero and Ventura [2007].

We use an unsupervised clustering algorithm; K-means. It is a non-hierarchical algorithm defined as "a process of partitioning n-dimensional data into k sets to minimize the mean distance within each set" Xu [2013]. K-means gained its popularity because of its ease of implementation Frias-Martinez et al. [2007b].

The K-means algorithm starts by randomly choosing a centroid value for each cluster. After that the algorithm iteratively performs three steps: (i) find the Euclidean distance between each data instance and centroids of all the clusters; (ii) assign the data instances to the cluster of the centroid with nearest distance; and (iii) calculate new centroid values based on the mean values of the coordinates

Action	Description
Homepage Actions	
Simple Search (SS)	Submit query by SS function
Advance Search (AS)	Search by AS functions
Navigate by collection (NC)	Discover items by collections
Navigate by institution (NI)	Discover items by provider
Navigate by content (ND)	Discover items by item type
Navigate by press (NP)	Discover the press
Search result Actions	
View results list (VRL)	Evaluate results
Filter results by full-text (FT)	Include only FT
Browse result by facets:	Doc type (FT), Institution (FI), Domain (FD), Collection (FC), Author (FA), Keyword (FK), and Language (FL)
Sort the results by:	Ascending (SA), Descending (SD) Date (Default), Title (ST), Author (SU)
Snippet page Actions	
Snippet view (SV)	Evaluate a snippet
View item (VI)	Click item to view
Display item (DI)	View by Multivio Viewer (MV)
Download item (DO)	Item saves
Click similar record (SR)	similar items
Add to personal list (PL)	item is added to the account
Extra Actions	
Export the item as:	Dublin (ED), MARCXML (EX)
Share the item by:	Email (SE) or Social media (SM)

Table 3.1. Actions Typology of RERO Doc

of all the data instances from the corresponding cluster.

The algorithm chooses K data items randomly from X as initial centroids; the data points are assigned to the cluster with the closest centroid, and the process is iterated until the convergence criteria are met Padmaja and Seshasayee [2016]; HAJEK and Stejskal [2012].

K-means clustering was utilised due to the data size and its predominantly used in detecting UPs in DL environment. Further work will be conducted to investigate the performance of the K-means comparing to other clustering techniques.

3.3.5 Features Identification:

Similarly to Chen and Cooper [2001], a feature set based on the features frequency was defined. 31 features were identified as in table 3.2. In summary, each session consists of:

- **Access point (AP)**: describes the session starting point. Users can reach RERO Doc contents from search engines, click a link in an email, and from the RERO Doc home page.
- **Action type (AT)**: describes the content discovery action of the RERO Doc home page (e.g. searching or navigating).
- **Function used (F)**: the type of functions utilised (if any) during the interaction (e.g. facets or filters).
- **Termination points (TP)**: describe how the users finish a session.
- **Session duration (SD)**: refers to the time spent by the user during a session.

3.4 Experiments

As we had a large-scale dataset of around 28 million records, analysing the whole dataset would be a time consuming task. The dataset was generated over eight months, including before and after exams, annual holidays, and semesters. Figure 3.5 shows the distribution of the sessions over the entire period.

Analysing the UPs based on a sample time-frame (e.g. a month) might have biased the outcome, meaning the found UPs could not be generalised over a

Access and content discovery actions
frequency of access item from email links
frequency of access from search engine
frequency of search by simple search
frequency of search by advance Search
frequency of navigate by collection
frequency of navigate by institution
frequency of navigate of content
frequency of navigate by press
Function used
frequency of full-text filter
frequency of Document type facet
frequency of institution facet
frequency of domain facet
frequency of collection facet
frequency of author facet
frequency of keyword facet
frequency of language facet
frequency of sort the results by ascending
frequency of sort the results by descending
frequency of sort the results by title
frequency of sort the results by author
Termination point
frequency of view results list
frequency of snippet view
frequency of view item
frequency of display item with Multivio
frequency of download item
frequency of click similar record
frequency of add to personal list
Session duration
session durations were computed, and categorised into:
short (10-60 seconds),
average (>60-300 seconds),
long (900-1800 seconds),
and longest (1800-2700 second)

Table 3.2. Session Features

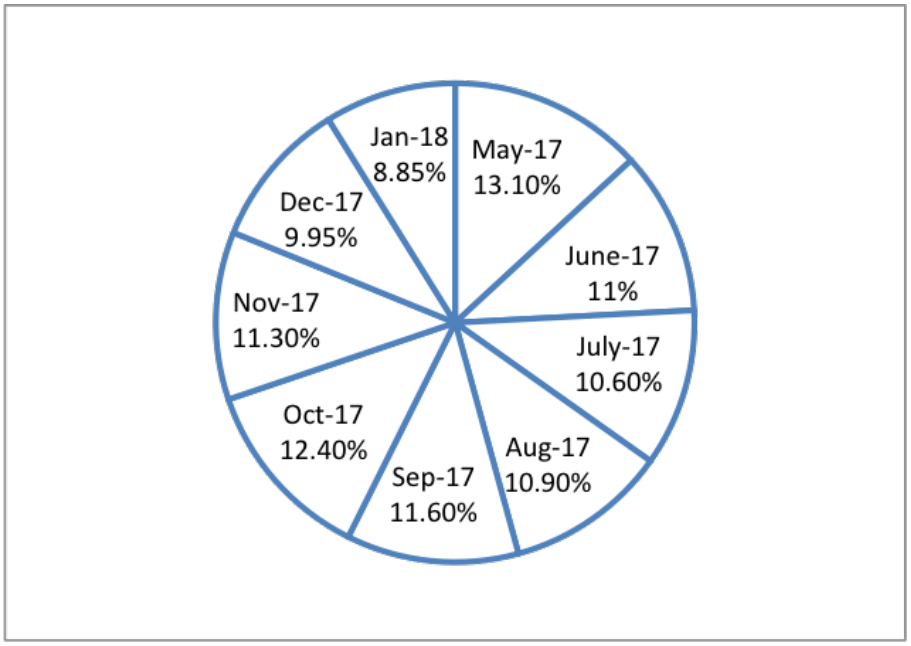


Figure 3.5. The Distribution of the Sessions over the Period

The dataset	Number of records	Number of sessions
Population	28M	6,004,005
Dataset 1 with 10%	2,151,194	600,400
Dataset 2 with 5%	1,066,808	300,320
Dataset 3 with 2%	408,019	120,080

Table 3.3. The Number of Records and Session of Each Dataset

period of time. Thus, three sub-datasets were created with the aim of measuring the consistency between clustering over the datasets. The experiments were run as follows:

- Create three datasets from the population with different sizes (10%, 5%, and 2%) without replacement; table 3.3 shows the total number of records along with the number of sessions for the population and the samples. It is important to reiterate here that the session is the unit of the analysis, not the records.

- The samples were built by conducting a random generation of the sessions across all the months;

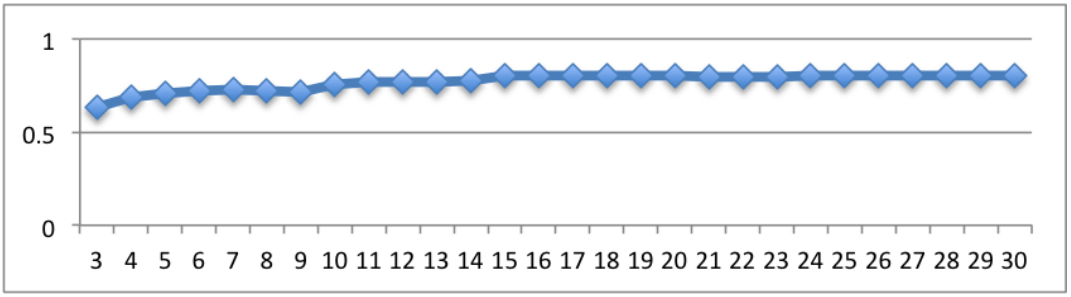


Figure 3.6. The Highest Silhouette Score for K

- Conduct clustering experiment with the 10% dataset with 600,400 sessions;

- For determining the optimal number of clusters, the clustering process went through different iterations to identify the optimal purity and dissimilarity of the data points. The clustering process started by making a random selection of 30 clusters.

  Based on the Silhouette coefficient score, a measure used to determine the neighbouring decision boundaries of the clusters (i.e. they have very few common features across clusters) Rousseeuw [1987], we could choose the optimal value for K clusters. The range of the Silhouette score is -1 to 1. Zero or negative scores indicate that the clusters are very close or overlapping and the clusters are not well formed. We found that 15 clusters recorded the highest Silhouette score of 0.80 as in figure 3.6 —indicating a better clustering;

- Identify distinct clusters based on the frequencies of the features as in table 3.2 of each cluster;

- To validate the found clusters, we considered the stability between the clusters' features across different datasets is a clustering quality measure. Thus;

- Conduct clustering experiments with the dataset 2 and dataset 3 by using the same *k* value of dataset 1;

- Identify distinct clusters of each dataset based on the frequencies of the features as in table 3.2;

- Check the frequency of the features of the datasets;

- Compare between the clusters across the datasets in relation to the clusters stability features similarity;
- Label the clusters and interpret their characteristics.

3.5 Results

Three main types of UP were found namely: item seeker, navigator and searcher. The item seekers are the users who reach RERO Doc from general search engines or links from emails, and are looking for authorised items. The navigators are the users who discover the content of the DL by navigating the content without submitting queries. The searchers use the simple and advanced search functions to interact with the DL.

Each of the main UPs has sub-patterns characterised by: access points, action type, function used, termination points, and session duration. Table 3.4 shows the patterns along with their characteristics. The description of the patterns is as follows:

3.5.1 Item Seekers

The item seekers represent the major users of RERO Doc across different datasets. These users arrived at RERO Doc either from general search engines (e.g. google.com) or arrived by clicking links in emails. Our finding is in line with the analysis of Google Analytics implemented by RERO Doc where most of the users were coming from google.com. This pattern seeks authorised information resources.

Their UPs were characterised by conducting one action i.e. viewing or downloading items without any further interactions with the system. View or download items are the traditional usage counts of the relevance items Chen [2018]. Their sessions duration range from short with 60 seconds, average with 60—300 seconds, and long 900—1800 seconds. We categorised them as: Firstly, the *satisfied* item seekers who spent shortest session duration, download items and leave.

The second category is the *Multivio* item viewer who spent short time and viewed the item with Multivio, a viewer application available in RERO Doc. The third and fourth patterns are *average* and *advanced* item seekers who spent longer time viewing items (60—300, and 900—1800 seconds), respectively. Figure 3.7 shows the patterns of the item seekers in terms of the download and view items across the datasets.

	Features				
Usage Patterns	access points	action type	function	termination point	session duration/sec
Item seekers (IS)					
Satisfied (SIS)	engine/email	view items	non	download item	60
Multivio (MIS)				view with Multivio	
Average (AIS)				view item	60-300
Advanced (DIS)					900-1800
Navigator (N)					
Light (LN)	RERO Doc	navigate	non	view result list	60
Average (AN)			domain/author/ Document/ & language	view/download	60-300
Advanced (DN)					900-1800
Press (PN)			non	non	60- 300
Searchers (S)					
Known item (KS)	RERO Doc	Search	author/ keyword	view result list	60
Light (SS)			non		
Average (AS)				view/download	300-600
Familiar average (FAS)			author/author/ Document/ & language		
Advanced			non		900-1800
Familiar advanced (FDS)			domain/author/ Document/ & language		
Sophisticated (PS)					1800-2700

Table 3.4. Characteristics of UPs

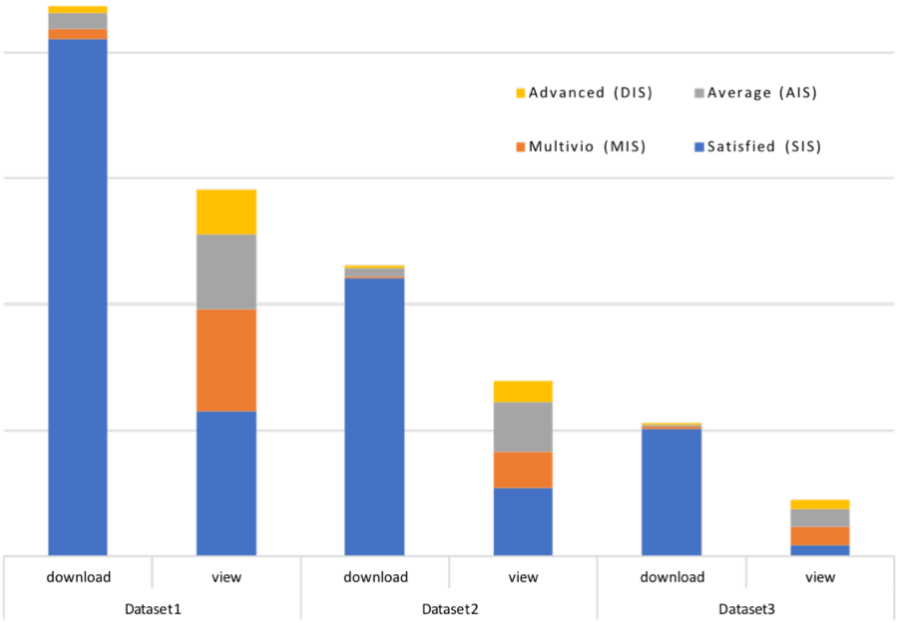


Figure 3.7. Patterns of the Item Seeker Across the Datasets

Datasets	Domain	Author	Document	Language
1	47%	28%	20%	5%
2	43%	32%	16%	9%
3	41%	29%	23%	7%

Table 3.5. The Dominant Facets of the Average Navigator

3.5.2 Navigators

The second main UP is the navigators. We found four behaviours belonging to this pattern:

1. *Light navigators*:

   This pattern started the navigation journey from the RERO Doc home page. They navigated the collection (NC). Their sessions were characterised by short duration (60 seconds), one single action (i.e. clicking on a collection), and ended by viewing the result list. This segment may represent (come-and-leave) visitors who navigate the library without viewing items. This might be because it was their first visit or they did not have specific information needs.

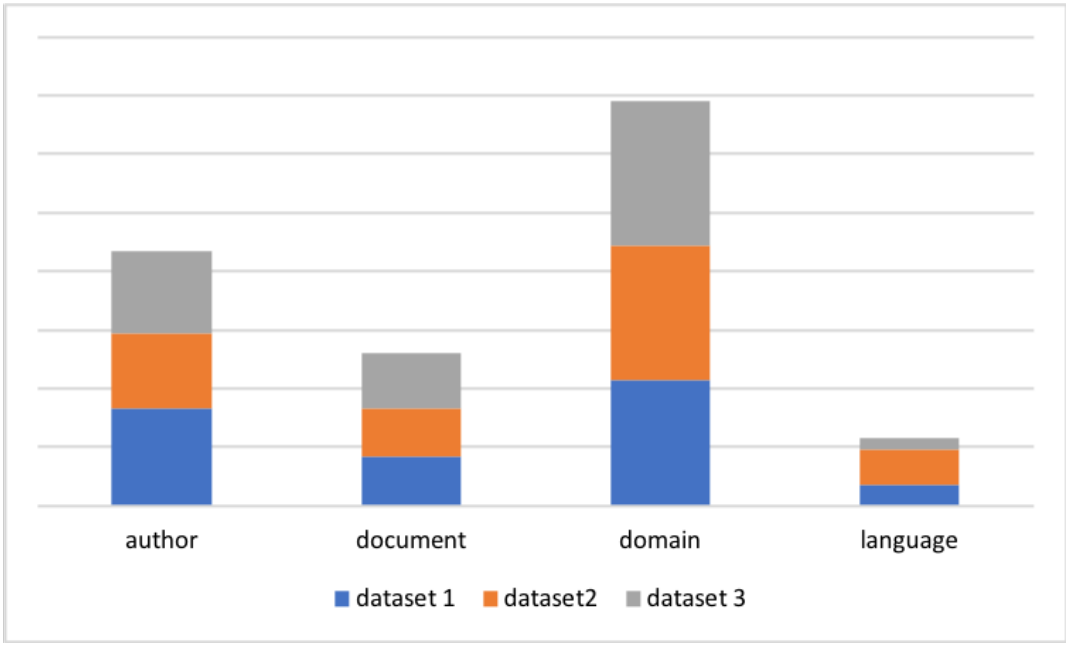


Figure 3.8. Facets Used by the Advanced Navigators

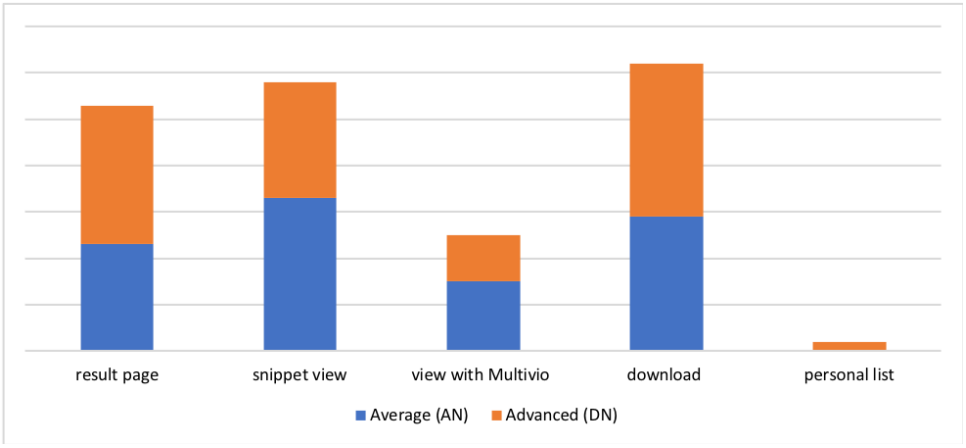


Figure 3.9. The Variations Among the Navigators

2. *Average navigators:*

   This is the second sub-pattern of the navigators whose sessions were of mid-range duration (60—300 seconds) and who navigated the content of RERO Doc by collection (NC) or content type (ND). Their sessions were characterised by being iterative: clicking on a collection, viewing a result page, evaluating the snippet, browsing the result pages by utilising domain, author names, Documents type, and language facets.

   Table 3.5 summarises the most dominant facets that were utilised by the average navigators across the datasets. Average navigators terminated their sessions by downloading and viewing items.

3. *Advanced navigators:*

   This pattern is similar to the previous one in terms of being dynamic and utilising multiple functions through many iterations. The difference is in their sessions duration, which were longer, ranging from 900 to 1800 seconds. We also found that domain and author names were the most dominant facets compared to the Document types and language facets as shown in figure 3.8.

   Another slight difference was found with the advanced navigators is that they were viewed the result pages more often than the snippet pages. They also downloaded and added items to their personal accounts. This might indicate the fulfilment of their information needs. In contrast, the average navigators viewed the snippet pages more frequent than the result pages. Such differences might suggest that the average navigators visit RERO Doc to explore the content rather than acquiring knowledge. The variations between the average and advanced navigators are depicted in figure 3.9.

4. *Press navigators:*

   We found a group of users who visited RERO Doc to navigate the available digitised press. Their sessions ranged in length from 60 to 300 seconds. The investigation of the press usage pattern with the specific press facets such as browsing by issue, name of media, and weekday, is beyond the scope of this work.

3.5.3 Searchers

The third main UP is the searchers. This pattern is distinguished by including more and varied sub-patterns. There are seven distinct patterns as follows:

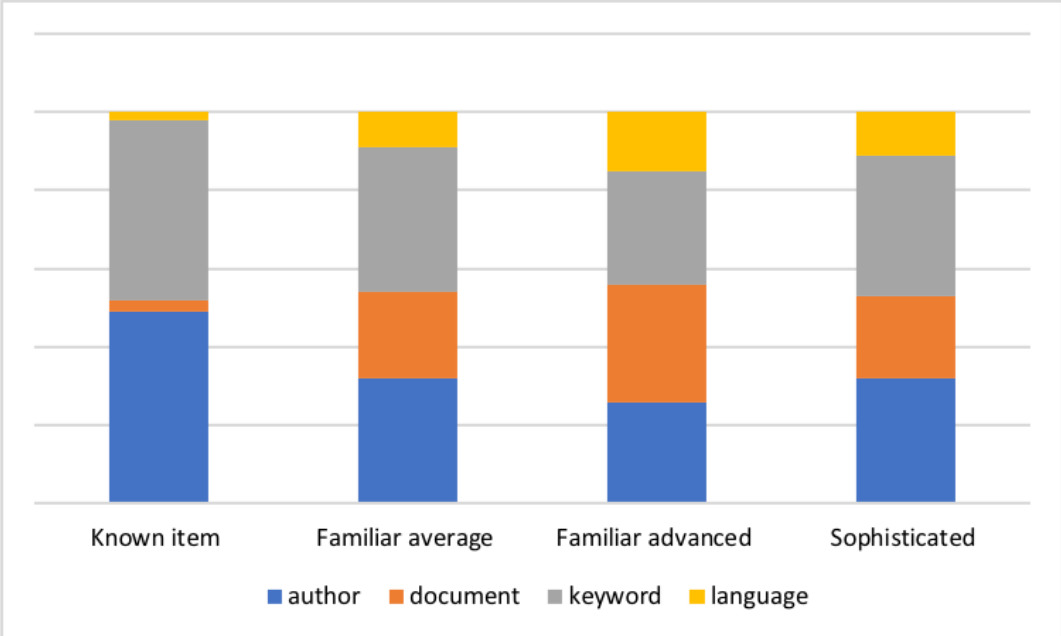


Figure 3.10. Facets Used by the Searchers

1. *Known item searchers:* This pattern was characterised by submitting queries through the simple search function and browsing the result page by utilising two main facets: author or keyword. Viewing result list pages was the termination point of their search.

2. *Light searchers:*

   This segment represents the simplest form of searching —submitting queries through the simple search and view result page without any further actions. Known item and light searchers involve the shortest sessions, lasting for 60 seconds.

   Their behaviour suggests that they seek well-defined and stable information; the user seeks known non-topical (structured) data, such as author names and book titles.

   The associated search behaviour is querying and filtering, where searchers are more confident in formulating the queries and filtering irrelevant information within a relatively short time Borlund and Dreier [2014].

   The two patterns; known item and light searchers differed in terms of the known item searchers were being more precise in utilising two specific facets. The light searchers might also represent the new visitors who were

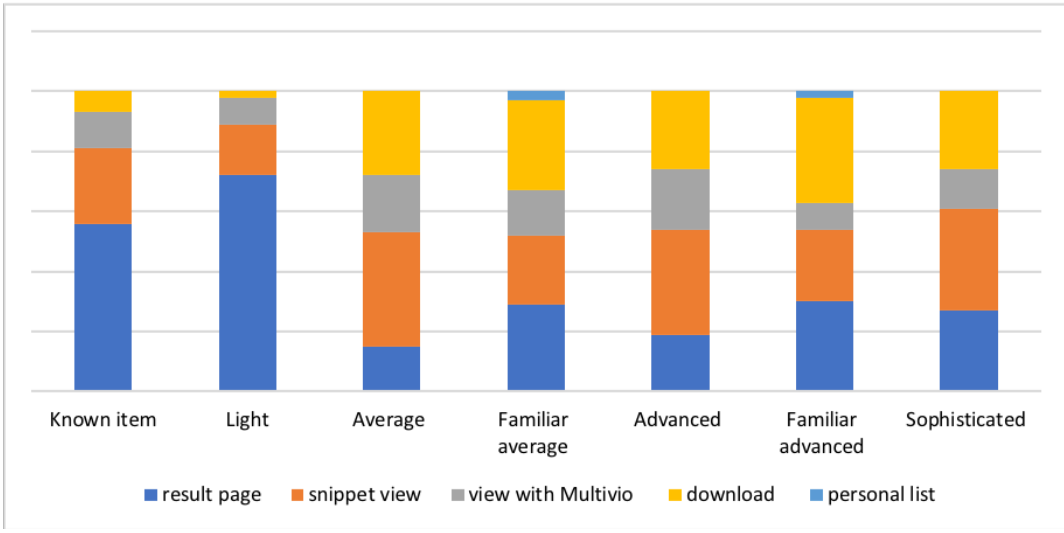


Figure 3.11. The Variations among the Searchers

not quite familiar with the system or the benefit of the facets.

3. *Average and familiar average searchers:*

   These UPs recorded mid-range session duration of 300—600 seconds, and their search behaviour was characterised by submitting queries, evaluating the results, browsing the results pages, and clicking items. We distinguished between the two average searchers by utilising the facets functions. We named the average searchers utilising such functions *familiar average* searchers.

   Their behaviour suggests familiarity with the system functions or that their information needs require precise criteria. Both patterns shared similar termination actions of downloading items more than viewing them. This might indicate that the information needs were fulfilled.

4. *Advanced and familiar advanced searchers:*

   This pattern is similar to the previous one in terms of heavy interactivity. The differences were that the advanced searchers recorded longer sessions (900—1800 seconds vs. 300—600 seconds) and utilised the advanced search function along with the simple search function.

   Again, the difference between the advanced and the familiar advanced searchers was in the utilisation of the facets functions during their searching. They terminated their sessions by downloading and viewing items.

5. *Sophisticated searchers:*

   This pattern involves the highest level of the interactions with the system. This segment represented the most sophisticated UP, with searching going through many iterations (i.e. formulating queries, browsing the results pages, and filtering the results). Also, their session duration were the longest (1800—2700 seconds). They viewed snippet pages more than result pages.

   The behaviour of the average, advanced and sophisticated searchers indicates that they seek to satisfy two types of information needs. The first is a conscious topical information need, which is a topical, well-defined information need, in a known subject and domain. The second is to fulfil a muddled topical information need, which is a topical and poorly defined need, exploring unknown subject matter or domains Borlund and Dreier [2014].

   In summary, the known items searchers utilised the author and keyword facets heavily as shown in figure 3.10. Compared to the familiar searchers, the average and advanced searchers viewed the snippet more frequent than the result pages as illustrated in figure 3.11. This might suggest the importance of the snippet page to the average searchers, and the result page to the familiar searchers.

   Another significant difference between the searcher patterns was the termination points. For example, the result page was the last termination points for the known item and the light searchers. In contrast, the highest percentage of the download items recorded by the familiar searchers. This might be due to their utilisation of the facets during their sessions. Adding items to the personal accounts and viewing them with Multivio viewer were the lowest termination points among searchers.

3.6 Discussion

The main aim of the research reported in this chapter was to discover the potential usage patterns mined from the LFs of a large-scale digital library. Considering the nature of the DL and the obtained dataset, LF analysis reveals valuable hidden information about the user interactions.

Three main UPs were identified: item seekers, navigators and searchers. Within those main patterns, sub-patterns were recognised. Comparing to previous UP studies, the item seekers and the navigators were the distinct UPs of

this study. Most of the findings of the previous studies considered searching as the main content discovery method. Thus, the searchers and their sub-patterns have been heavily investigated in the literature.

With the help of clustering technique, important insights into users' interactions were gained from the cluster analysis; this helped us to answer the second research question: *what are the main features that characterise UPs?*. We found that the UPs can be recognised by four main features: session starting points, content discovering actions, types of functions used (if any), termination actions, and session duration. Session duration is a significant indicator that allows patterns to be distinguished.

Key implications arising from the emerged UPs is that the implicit variations of the users can be leveraged to provide better user experiences in DLs content Han and Wolfram [2016]; Gooding [2016]. For instance, such knowledge can be used to:

3.6.1 Inform the design and development of interactive systems

DL requires iterative evaluation by involving real users. Such evaluation is time consuming and costly. The UPs can be used to inform system developers and designers when redesigning and reorienting DL. For example: different interaction activities can be promoted by:

- Visualising relevant part in the interface e.g. search tools, search results page, or documents. The visualisation might enhance the user experiences by introducing techniques based on external thesauri and suggested terms, or by implementing machine learning techniques as in the case of Ruotsalo et al. [2013]. For example, in this research, we developed an interactive visualised interface as in chapter 7.

- Providing query suggestions or auto completion functions. RERO Doc enhances the searching experiences by utilising Auto Correct function, it is also suggested to add the Auto Complete or "Popular Terms" function. This might accelerate the searching time for the sophisticated and advanced searchers.

- Better utilisation of the facet functions. Not all users are aware of the benefit of the facets. As a suggestion, the DL designers might provide an alternative search tools to the facets bar. For example, drop-down list of the facets instead of the current long static facets bar. Or adopt one of the

visualised techniques i.e. tree relation, or radial techniques. In this thesis, the radial technique was implemented as in chapter 7.

- Considering the heterogeneous users of RERO Doc, offering navigational icons (e.g. backward and forward) might improve the navigation interactions.

3.6.2 Inform the design of user studies

LF analysis is an inexpensive method that enables to gain an overview of the user experiences with the system. It might assist librarians and scholars to identify aspects for further investigation or validation via user studies. For instance, in our case, exploring the UPs helps us to:

- Develop different tasks: the emerged UPs, draw our attention to the important of designing different types of tasks. Instead of designing only known item search tasks, exploratory search task need to be considered. Careful consideration must be given to the different content discovery methods applied by the users. This outcome helps us to consider two different search tasks in our user studies as in chapter 4, chapter 5, and chapter 7.

- Identify individual variables: RERO Doc serves heterogeneous population. Thus, identifying individual variables to be further investigated is not a trivial task. Wide range of individual differences affect the ISB e.g. demographics, cognitive, and personality variables O'Brien et al. [2017]. The results of the UPs show considerable variation in using the system functions (e.g. facets). This might indicate the effect of the level of familiarity and expertise with the system functionality. Thus, the exploration of the UPs helps us to identify aspects for further exploring as in chapter 4 and chapter 5.

- Determine the role of the system aspects: IR systems are developed to support users to accomplish their tasks. The insights gained from the UPs encouraged us to further investigate the relation of the system functionality to generate better interactions. For example, in this research, the interface, as part of the IR system was considered and redesigned as in chapter 7.

- Decide about the data collection tools: the results show variation of utilising interface functions among users. Such valuable insights can not be easily investigated by the qualitative collection tools (e.g. questionnaires,

think aloud and interviews with users). This suggests a combination of passive behaviour tracking tools (e.g. eye tracking, screen recordings, or mouse hovers) along with the qualitative one. That is why, a UX study was conducted as in chapter 5.

In summary, the hidden information embodied in the UPs may enable system developers, designers, and librarians to provide extra support to the users.

3.7 Study Limitations

The study had the following limitations: it excluded fine-grained analysis, including semantic or query analysis as in Zavalina and Vassilieva [2014]. Realising that clustering is a powerful technique for discovering distinct UPs, it fails to show the sequential behaviour of the patterns which can be achieved by Markov chain techniques. Thus, a further investigation is required. Although LF analysis helped us to reveal the potential UPs, a full understanding of the context and motivations of the behaviour is missing. Thus, in order to develop a more comprehensive understanding of the user interactions and contextual factors, a user study will be conducted.

3.8 Conclusion

Our objective was to emphasise the importance of LF analysis as a method to reveal the hidden user interactions and its value as an inexpensive evaluation tool. DL webmasters may use such information when redesigning and reorienting DL services and enable them to provide extra support to specific system features if needed. Similar to previous studies reported in literature, this set of UPs, is specific to the RERO Doc interface, the available dataset, the type of the DL and its content, and the target users. Nevertheless, we believe that our investigation is a crucial step towards gaining a better understanding of the interactions between users and DLs.

The main message we gained from the exploring of the UPs that the DL functions i.e. search tools were not fully utilized across the heterogeneous users. Thus, this leads us to investigate deeply the real usage of the DL functions; i.e. to what extent the DL's search tools are used? This was done by conducting three different user studies; one was to investigate why searches failed and it was a remotely study as in chapter 4, another one was to validate the result of the online study and to evaluate the user experiences with the DL. The study was conducted

in a laboratory setting as in chapter 5. The last study was to collect a large-scale attitudinal survey as in chapter 6. The explanation of those studies is presented in the next Part III.

Part III

Investigating

Chapter 4

Why Searches Fail?

4.1 Introduction

The concept of success and/or failure of outcomes for information seeking was introduced by Wilson [1999] in his model of ISB. Unsuccessful search experiences force users to refine their searches, ultimately resulting in low expectations, frustration, and less perseverance.

To develop a holistic understanding of an individual's approach to information seeking, this chapter aims to understand causes behind search failures from the users' perception along with the associated emotions. Ultimately, such understanding might help us to design better user experiences and provide more usable and useful digital library systems.

Thus, this chapter attempts to answer the following research questions:

- RQ1: *What are the reasons behind search failures from the searchers' perspective?*
- RQ2: *What are the primary emotions individual experienced in the failed searches?*

The following section 4.2 described the methodology. Next, the data analysis and results presented in section 4.3. Findings of this study are reported in the discussion section 4.4. This is followed by the study limitations in section 4.5.

4.2 Methodology

The study was conducted online by inviting real users to carry out the experiments in their own natural settings. RERO Doc digital library, the library intro-

Figure 4.1. The Experimental Website

duced in 3.1.2, was chosen as the experimental platform.

Figure 4.1 shows the experimental interface of the DL where the information form (1) presented a summary of the study, procedures, and expertise self-assessment of using the DL; and the two main tasks the participants were asked to perform were explained in (2) the specific task; and (3) the exploratory task.

The participants were encouraged to use the digital library to run two search tasks: one specific and one exploratory, and to focus on topics of their own choosing; examples of the information needs of the participants are presented in table 4.1.

We let participants select their own topics for three reasons. Firstly, the purpose and the nature of the study design required self-assessment of the search results in relation to the real information needs of the participants. Secondly, this avoided the bias that might result from the task narrative and query construction. Thirdly, the digital library collections cover various knowledge domains including nursing, economics, health, computer science and history. Therefore, it would be impossible to design topics that would cover all these different domains and guarantee the engagement of the online participants. The tasks were

The specific task	The exploratory task
"The first original research paper which introduced random forests" *"I'm looking for a chapter in a book with the title 'The World Is Flat' "* *"The author of Introduction to Information Retrieval book"*	*"I would like to know more about applications and results of applying topic models on patents, what kind of algorithms exist, what tools are available for the same, if there are any open source implementation of those models etc."* *"The Applications of Voice Assistants in Health Care sector, I need to find more about the privacy issues, credibility and their design"* *"I need to look for the current web-based therapies applications, I'm more interested in elderly habits when using such applications"*

Table 4.1. Examples of the Participants' Information Needs

set in scenarios typical of an academic search for information and adapted wording from Hoeber et al. [2019] were used as follows.

> Scenario 1:
> *Suppose you are writing a paper for your specialized major and you need to look-up or verify a specific fact (e.g., the date of an event, the correct spelling of someone's name, the details of a specific research paper, the name of a book or author).*
> Kindly use RERO Doc as the information source to accomplish this task.
>
> Scenario 2:
> *Suppose you are writing a paper on a topic of your interest and you wish to gather more information to enhance your existing knowledge or discover new knowledge.*
> Kindly use RERO Doc as the information source to accomplish this task.

As the participants started the experiment, an informative form containing a description of the study, an introduction to the research team, and the electronic informed consent form were presented. Participants' demographic data, search skills and system expertise were collected in this form.

The participants were asked to fill out a pre-questionnaire before searching, and therefore the description of their information needs, perceived knowledge, familiarity with the topic and complexity level were collected explicitly.

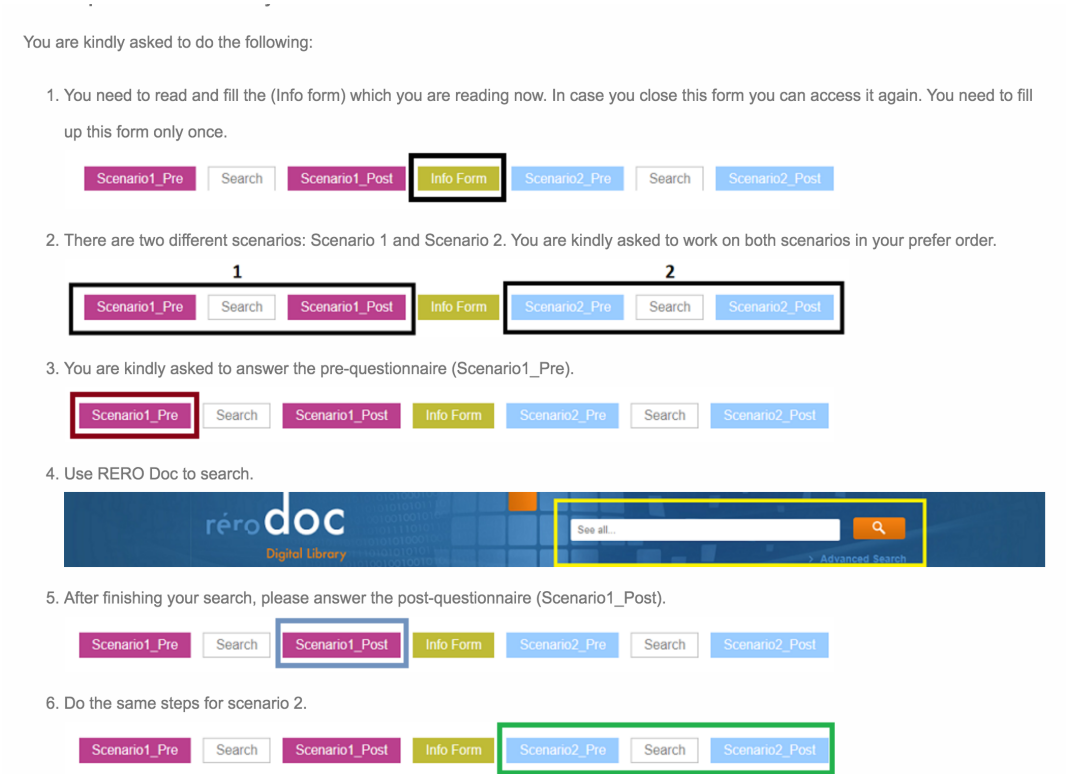


Figure 4.2. The Experimental Phases

Total # of participation	Uncompleted	Mixed	Successful	Failed
88	34	10	12	32

Table 4.2. Number of Submitted Searches

As the participants finished their search, by completing the post questionnaire, we collected their self-assessment of the overall search performance. Participants also answered open end questions regarding their experiences with the task and with the digital library in general.

4.3 Results

This section presents the results of the qualitative and quantitative data analysis. Only the unsuccessful searches were considered in the analysis. Table 4.2 shows the total number of submitted searches; eighty-eight is the number of the total participants, thirty-four of the participants were not completed and thus eliminated for one or more of the following reasons: one or both questionnaires were missing, participants did not conduct real searches for one or both scenarios, participants completed only one scenario, and participants failed to grasp the description of the scenarios e.g., they conducted both searches with the same task type instead of one specific and one exploratory.

The mixed participation refers to situations where both tasks made by the same participant where one task was successful and the other was failed.

Thirty-two of them assessed their search outcome of both tasks as failed and provided full questionnaire data. Those failure searches result in 64 search sessions and were all accounted for the analysis here.

4.3.1 Demographic Data:

The participants are representative of the general users of the library in terms of disciplines, academic status and other user characteristics.

As in table 4.3, of the 32 participants, 18 were female, 13 males, and one did not specify a gender. Around sixty per cent of the participants were 28-37 years old; 30% were aged 18-27, and three participants were aged 38-44. The participants came primarily from the Faculty of Computer Science (46%) and the Faculty of Communication Sciences and Business (38%). Less than 16% were from the Faculty of Medicine or the Faculty of Architecture and Engineering. Most

Gender	Age	Academic Status	Discipline
18 Female	28-37 19 participants 59%	Doctoral & master students 19 participants 59%	Computer science 15 participants 46%
13 Male	18-27 10 participants 31%	Bachelor students 7 participants 21%	Communication & business 12 participants 38%
1 Unspecified	38-44 3 participants 10%	Researchers & professors 6 participants 20%	Engineering & medicine 5 participants 16%

Table 4.3. Participants Characteristics (N=32)

of the participants (60%) were doctoral and master's degree students, followed by bachelor's degree students (20%), researchers (10%), and professors (10%).

4.3.2 System Expertise Level:

The level of the system expertise was considered as a significant variable according to the studies of Borgman [1996]; Xie and Cool [2009]; Trapido [2016] who confirmed that the current design of the digital library requires adequate knowledge of information retrieval mechanism.

In this study, the system expertise was self-assessed by the participants considering three aspects: *frequency of using the digital library*: (1) How often do you conduct searches in DLs?, *degree of familiarity* with the searching tools in digital libraries (e.g. facets, sort and advanced search): (2) Please indicate your degree of familiarity with searching tools in DL, and *rating of the digital library searching skills*: (3) How can you rate your searching skill on the DLs?

The participants showed various levels of the system's expertise that enables us to categorise them into non-expert (N=14) or experts (N=18). The non-expert participants were not familiar with digital library searching tools, they had rarely conducted searches within the digital library, and considered themselves as beginners or intermediate searchers. The experts, on the other hand, were quite familiar with the searching tools in the digital library, frequent users of the digital library, and regarded themselves as advanced searchers i.e. good at using advanced search functions e.g. Boolean operators and filter results by facets. Table 4.4 shows the expertise analysis results.

Questions	Frequency of searching DL				
1	Never	1-3 times per term	1-3 times per month	1-3 times per week	Daily use
	0	7	9	9	7
2	**Familiarity with the search tools**				
	Not familiar at all	Slightly familiar	Somewhat familiar	Familiar	Extremely familiar
	2	6	9	9	6
3	**DL searching skills**				
	Beginner	Intermediate	Advanced	Expert	
	7	8	14	3	

Table 4.4. System's Expertise Level

4.3.3 Domain Expertise Level

The topic characteristics, as an influence factor on the information seeking, were also considered in the analysis. Three main aspects of the topic characteristics were measured including: *level of the domain knowledge*: (1) How much do you know about the topic you are searching for?, *familiarity*: (2) How familiar are you with the topic you are searching for?, and *perceived difficulty*: (3) How easy do you think the topic is? The three aspects were self-assessed based on a five-point Likert scale.

For the specific task: most of the participants 90% reported that they had good or very good knowledge about the topic they searched for. Simultaneously, 92% of the participants reported that they had a easy or very easy topic. In terms of the familiarity with the topic, 80% of the participants rated their familiarity with the topic as familiar or very familiar.

For the exploratory task: while most participants agreed on level of familiarity, knowledge, and easiness of the specific tasks, they had more mixed reaction to the exploratory tasks and their characteristics.

For instance, in terms of the domain knowledge, fifty per cent of participants reported having good or very good knowledge, 23% had adequate knowledge and 27% had basic or no knowledge. Also, the participants reported their perceived level of difficulty: 41% as easy or very easy, 33% of participants rated their task as being of moderate difficulty and 26% reported that they had a very difficult or difficult topic. Familiarity with the topic also were reported as 43% of participants stated that they were familiar or very familiar with the topic, 23% were somewhat familiar and 34% were unfamiliar or not familiar at all with their topic. Table 4.5 shows the exploratory task characteristics. Accordingly, the participants were divided into knowledgeable searchers, who had good or very good

Questions	**Knowledge about the topic**				
1	No knowledge	Basic	Adequate	Good	Very good
	3	5	8	9	7
2	**Familiarity with the topic**				
	Not familiar at all	Slightly familiar	Somewhat familiar	Familiar	Very familiar
	2	9	8	10	3
3	**Easiness of the topic**				
	Very difficult	Difficult	Moderate	Easy	Very easy
	2	6	11	10	3

Table 4.5. Topic's Expertise for the Exploratory Task

knowledge about the topic, and so were very familiar or familiar with the topic (N=15), and non-knowledgeable searchers (N=17).

4.3.4 Search Performance:

The main objective of this work is to investigate the reasons behind search failures from the searchers' perspective. Searching for information is a process affected by *personal attributes* e.g. knowledge and experience, *task attributes* e.g. type and complexity, *system design* e.g. interface design and content coverage, and *search performance* or interaction outcome Xie and Cool [2009].

To measure the search performance, a user-oriented approach was considered here. Based on previous works, three main criteria were identified namely; the user-defined relevance of the information which depend on the users' knowledge and perceptions, and it is affected by factors including: search situations, the users' goals, knowledge level and beliefs, the nature of information being evaluated, constraints of time and effort and cost involved in obtaining information Savolainen and Kari [2006]. Considering relevance alone is not sufficient, the satisfaction with the found information needs to be considered Belkin et al. [2008].

Satisfaction defines as "the extent to which users believe the information system available to them meets their information requirements" Ives et al. [1983] (p.785). And, finally evaluate the overall outcome of the search process; once a person finished a search episode. The success of the search process is affected by subjective factors including search expertise, prior knowledge and interest in the topic, and flow experience Wirth et al. [2016]; Wang et al. [2017].

The three criteria were assessed by self-rating based on a five-point Likert

scale: relevance of the information found (highly relevant=1, not relevant=5), satisfaction with the information (completely satisfied=1, not at all satisfied=5), and search success (extremely successful=1, complete failure=5). Searches with only negative responses were considered in this study.

4.3.5 Behavioural Analysis:

Because search behaviour is recognised as an indicator of search quality Debowski [2001], behavioural signals which refers to the recorded interactions in the log files were quantified and treated as dependent variables. We compared the behavioural signals of experts and non-experts participants.

The main objective behind the behavioural analysis was to gain better understanding of the underlying causes of failed searches, in particular, to what extent the system functionality and the support tools e.g. advanced and filter functions were used across different expertise levels. Table 4.6 shows the metrics used to quantify the behaviour. The analysis covered more than 64 search sessions yielded the following observations.

Generally, the mean session duration for the specific task for both groups was two minutes which was shorter than the mean duration for the exploratory task; eight and five minutes for the expert and non-expert searchers respectively.

In the specific searches, the behavioural analysis showed that the advanced search functions and filters e.g. facets and sort functions, were never used by members of both groups. The results also indicate that the participants did not go beyond the first search results page.

The expert searchers recorded more complicated and dynamic interactions when they work on the exploratory task than the specific task. Their sessions lasted around eight minutes on average, going through different iterations, making light use of facets, and returning to the search results page very frequently.

Both experts and non-experts tended to click and view items very frequently rather than downloading them; it might be an indication of the unsuccessful searches. Like experts, non-experts spent more time on the exploratory tasks than the specific tasks i.e. five minutes on average, but less time than the experts. They also recorded fewer visits to the search result page, and rarely used advanced search functions or facets.

The result of the Mann-Whitney test, as shown in table 4.7, indicates that the significant difference between the expert and non-expert searchers is recorded only in two aspects: the session duration (Sig= $0.00 < 0.05$) and the number of visits to the result page (Sig= $0.00 < 0.05$) when searchers were dealing with exploratory tasks.

Metrics	Description
Session duration (SD)	Average duration of the session
Action variable	
Simple search (SS)	Number of search by SS
Advance search (AS)	Number of search by AS
Facet used (FU)	Number of FU
Result page (RP)	Number of visit RP
Query behaviour variable	
Number of query (Q)	# reformulating query per session
Query length (QL)	# of the term/s of each query
Search result actions	
Clicked result (CR)	Number of CR
View item (VI)	Number of VI
Downloaded item (DI)	Number of DI

Table 4.6. Metrics to Quantify the Behaviour

Expert searchers had longer session duration compared to the non-experts. At the same time, they return to the search result page very frequently while conducting exploratory searches. In terms of the specific task, the behaviour of both groups did not differ significantly in any behaviour signals.

4.3.6 Causes for Failure

To address RQ1, *what are the reasons behind search failures from the searchers' perspective?*, participants were asked to provide reasons why they thought they failed with their searches. A code schema was developed to analyse the answers, table 4.8 shows examples of the failed reasons.

The reasons were identified and classified into two main categories: digital library coverage e.g., zero hits, irrelevant results and too many or too few results, and usability problems, including accessibility and navigation issues. Figure 4.3 shows the frequency of the problems across both tasks-specific tasks and exploratory tasks.

4.3.7 Emotions Analysis:

To answer the second research question, *what are the primary emotions individual experienced in the failed searches?*, a textual analysis was conducted by using the

Task Types	**Simple task**			**Exploratory task**		
Statistical test	U	Z	**Sig**	U	Z	**Sig**
Session duration	83	1.6	.09	8	4.5	**.00**
Simple search	98	1.1	.23	81	1.5	.06
Advance search	119	.88	.37	105	1.5	.11
Facet used	114	.79	.42	112	1.2	.20
Number of query	92	1.54	.12	110	.66	.50
Query length	104	.92	.35	115	.42	.67
Result page	124	.08	.93	49	3.0	**.00**
Clicked result	124	.08	.93	99	1.0	.28
View item	122	.18	.85	120	.25	.79
Downloaded item	126	.00	1.0	99	1.2	.22

Table 4.7. Behavioral Analysis of both Groups for the Two Tasks

Linguistic Inquiry and Word Count (LIWC) software. LIWC is a psycholinguistic lexicon tool that is created by psychologists to help the non-specialists to detect psychological statistics in text. LIWC consists of different dictionaries where each word is associated with the emotions it evokes to capture word-emotion connotations. The software can detect different psychological categories embodied in individuals' verbal and written speech samples including emotional, cognitive and structural components Santos and Vieira [2017].

We analysed sixty-four excerpts reporting search failures extracted from the open-question in the post questionnaires and provided by the participants. We asked the participants to reflect on their failure experiences by answering the following opened questions: (1) *If you are not satisfied with your searching, can you please explain why?* and (2) *what do you think of your overall experience with the digital library?* Accordingly, the proportion of words that scored positively on LIWC categories was measured. Table 4.9 shows the average number of excerpts that contain at least one word associated with a particular emotion based on Plutchik's Wheel of Emotions.

The radar chart, as in figure 4.4, confirms that users show a positive attitude toward the digital library in general. This includes emotions such as trust, joy, and anticipation. While joy and anticipation dropped down when participants reflected on their search experiences, participants were still showing the same level of trustfulness in the digital library as an information source.

The top two emotions experienced in the failed searches and linked to negative feelings, were anger and sadness. Such unpleasant emotions might prevent

Category	Sub-categories	Examples
Coverage of the DL	Too many results	"I was expecting very **precise** responses" "I would prefer to find more **shorter** documents" "I wanted to know very basic things and the results returned were **too** many."
	Too few results	"The collection of books and research papers is very **limited**." "**Not** all the publications of the author were **found**" "I found **only** some articles related to my search topic"
	Zero hits	"I did not get **any** result back." "There were **no** such books" "I **couldn't** find any information about my search"
	Irrelevant results	"The papers found were not **exactly** what I was looking for." "The documents provided to the query were not **relevant**" "There were not major **key** sources"
Usability	Accessibility	"I did not find it easy to find how to **go** to the full text of the article" "I would prefer to be **able** to download the research papers rather than viewing it online with the viewer software." "I could not **access** the digitalized version of old research papers"
	Navigation	"It will be good if there is an icon to go back to search **results page**" "There is no date **facet**" "The **result page** include documents in other languages."

Table 4.8. Problems Reported by the Participants

Emotion	Search experience	Overall experience
Anger	2.1	2.2
Anticipation	0.9	1.9
Disgust	0.0	0.0
Fear	0.0	0.0
Joy	1	1.7
Sadness	2.2	2.2
Surprise	0.0	0.0
Trust	2	2

Table 4.9. The Distribution of the Emotions for the Search and Overall Experiences

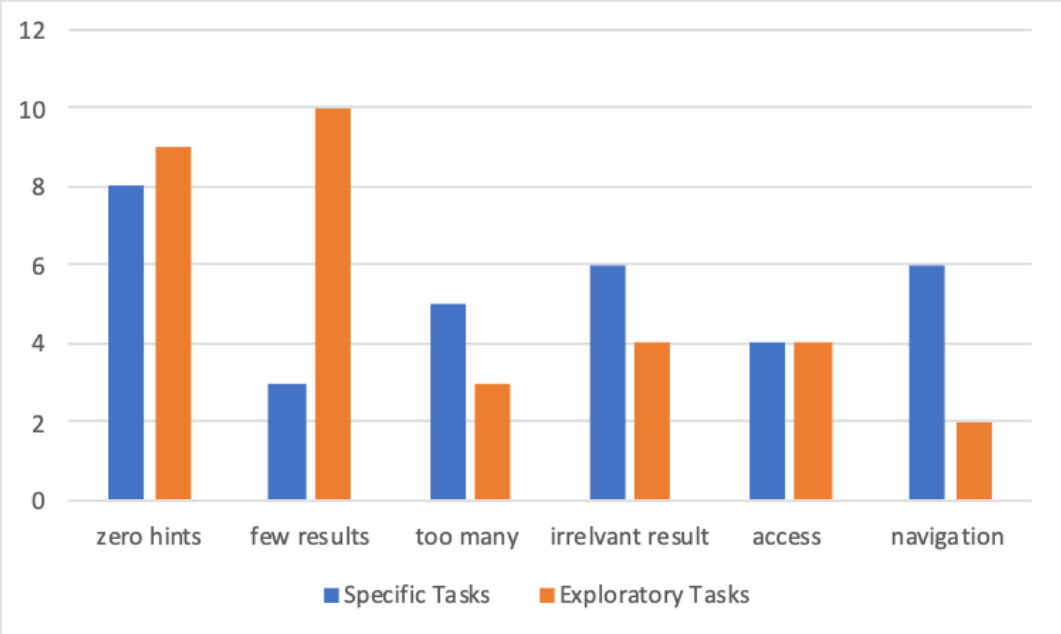


Figure 4.3. The Frequency of the Problems for both Tasks

user perseverance in achieving their information seeking goals. Being able to detect such emotions might help system designers to improve the design and user experiences with digital library.

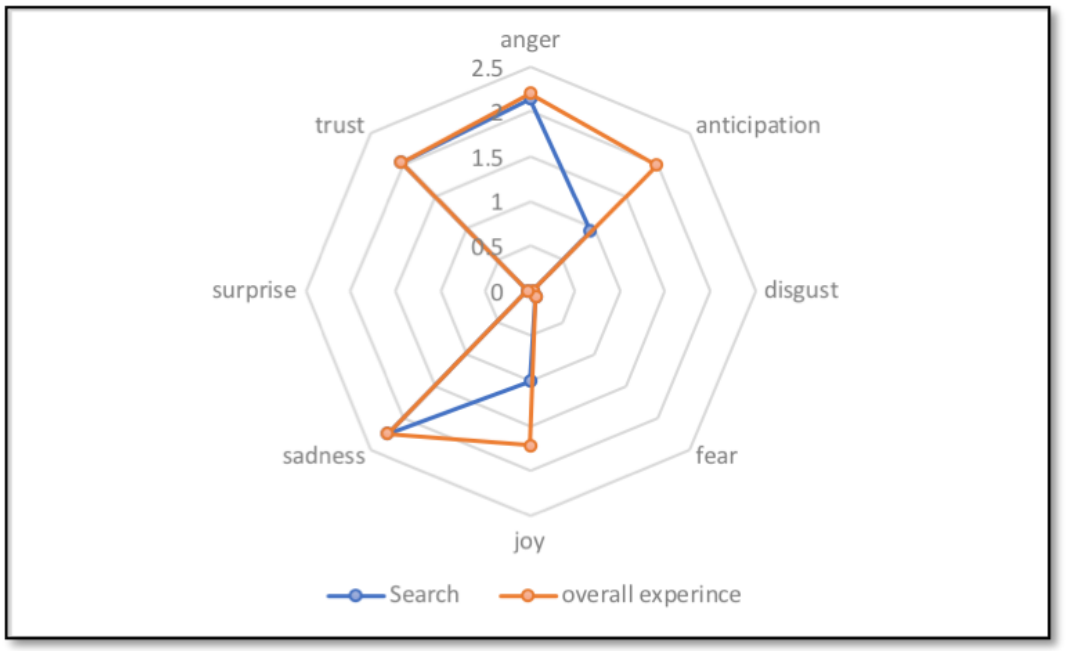


Figure 4.4. The Distribution of the Emotions

4.4 Discussion

Investigation search failures is crucial to enrich our understanding of the user experiences and help us figure out how to improve the system functionality. Log file analysis is the preferred sources to measure success and failure rates due to its availability and objectivity; however, studying the search failures along with the users' attributes provides a better understanding of the barriers to successful searching.

Identifying the unsuccessful search depends on the searchers' definition of an unsuccessful search Trapido [2016], and users' personal and cognitive characteristics, together with the nature of tasks Xie and Cool [2009]. Therefore, we can answer the first research question: what are the reasons behind the failure searches from the searcher's perspective?

Speculations from our study participants indicated two main reasons: the lack of coverage of the digital library and usability problems. This is in line with the previous work where Antelman et al. [2006], and Chan and O'Neill [2010] confirmed that the unclear coverage, vague representation of information objects, and ineffective search mechanisms result in unsuccessful searches. Similarly, Trapido [2016] discussed the effect of the size of the underlying database, systems search capabilities and interface design on the unsuccessful searches.

However, in the analysis of usage patterns extracted from the log files, the real use of support tools (e.g., advanced and filter functions) was examined across different expertise levels. The main objective behind such examination was to gain better understanding of the underlying causes of failed searches. We realised that the digital library functionalities were not used fully by both the experts and non-expert searchers.

The behavioural analysis demonstrated that several searches might have headed towards success but for various user actions that was precluded. Searchers might not understand how to craft an appropriate search strategy to meet their information needs. As an example, one participant complained about the inability to find old references quickly but did not use the sort function that would display the results chronologically. Another frequent comment was about the difficulty of retrieving English and non-English documents. RERO Doc offers a filtering results function with different facet types including language, but the usage patterns analysis showed little use of these discovery tools. Using facets could have helped to turn the failed searches to a success one, as Sadeh [2007] argued that facets could reduce a large result set to a manageable size.

Our observations are also in line with a number of previous works including Antelman et al. [2006] and Blumer et al. [2014] which confirmed the importance of facets as a discovering tool to reduce the frustrating experience with unhelpful result sets. However, Capra et al. [2007] reported the problems associated with facet implementation and highlighted the need to train users to better utilise facets, considering the fact that not all searchers find facets beneficial. The inadequate use of these tools could be due to a lack of awareness of their importance, or a poor conceptual understanding of the way they work. And both should be further explored as unreported causes for search failures.

We could conclude that there is a mismatch between the searchers' awareness, real interactions and the functionality of the digital library. Accordingly, there is still a call for bridging the gap between the digital library design, usability issues and user awareness. In conclusion, almost twenty years ago, Borgman [1996] called for bridging the gap between the search behaviour and the design of the online catalogue. Users are still experiencing search failures because of ill query formulation, and difficulty in identifying the appropriate access points.

As solutions, Borgman [1996] suggested to provide training to the users, and more importantly design better systems that feeds from the research results. Today, there persists a need to improve the digital library system and simplify its interface. The use of query suggestions or auto completion functions could better support the search experiences and speed up the time taken by struggling searchers as mentioned by Barifah and Landoni [2019a]. Designing simple, fast,

and easy to use search systems and interfaces could be the way forward. In particular, focusing on an interface that increases users' awareness of the facets and their accurate application remains a worthwhile goal Trapido [2016].

In terms of the emotional analysis, referring to the second research question, what are the primary emotions individuals experienced in the failed searches? researchers confirmed that investigating affective aspects along with other information searching factors facilitates our understanding of information behaviour more holistically as the emotional responses may influence the information search process and searchers' actions Kuhlthau [1993]; Fulton [2009]. It is important to not consider only how an individual navigates information, but also what emotions are experienced. The effect of the emotions plays a significant role in determining how successful the search sessions might be and to what extent users might persevere in accomplishing a task Gwizdka and Lopatovska [2009].

Prior works confirmed that emotion is an ubiquitous element in user interactions Brave et al. [2005], therefore system designers should consider it when designing usable and intelligent systems Karat [2002]. As literature points out that people's emotions, feelings, and attitudes can be conveyed through the words they use, in this work, we analysed 64 textual excerpts from the post-search questionnaire. Findings indicate that participants were experiencing sadness and anger as their primary emotions when they experienced search failure. This finding is in line with Johnson [2013] who confirmed that as failure experiences deepened, the feelings of sadness, defeat and frustration increased. This in turn has a detrimental impact upon cognitive functioning as it reduces the accuracy of memory recall Johnson et al. [2017].

Together with raising people's awareness of how to use the system, we need systems that are better tailored to the users' needs. The results of this study have implications for the system design as follows: from the behaviour analysis, the results showed that the search tools (e.g., facets and filters) were rarely used by the users.

One way to support users to achieve more successful searching is to replace the current digital library facet-based interface with a visualised interface as described and studied in chapter 7. Visualisation is not a new phenomenon but has rarely been used in the DL context. Ruotsalo et al. [2013] designed a scholarly system based on modeling interactive intent, which allows users to provide relevant feedback to the system using an interactive user interface. The interaction process is supported by a radar-based approach that enables users to easily find relevant results. As a future work, we are planning to design a visualised interface and evaluate the user experiences.

Another supporting solution is implementing a chatbot that might help users

to redirect their interaction with the digital library. Recently, Mckie and Narayan [2019] developed a Lib-Bot, a digital library chatbot, aiming to minimise the effects of anxiety related to using library databases and librarian services, and produce a sense of ease when using library resources Mckie and Narayan [2019]. Their ultimate goal when developing such a chatbot was to provide a successful and positive experience for the user.

4.5 Study Limitations

We do acknowledge that one of the limitations of this study is the relatively small size of the user sample of 32 participants, potentially threatening the validity of our results. However, the sample is still representative covering most relevant types of searchers in terms of demographic attributes (gender, age, academic status, and discipline), system expertise levels (expert vs novice) with different domain knowledge expertise. Another limitation is the dependence on self-reporting to elicit emotions although it is a popular method for emotions detection.

4.6 Conclusion

This chapter is an attempt to shed the light on causes behind failed searches and identify associated emotions. The nature of digital library search failures was examined from the searchers' perspective and compared to the real interactions recorded in the log files.

Based on the feedback from participants, we identified two main reasons behind unsuccessful searches: limited coverage of the digital library and usability issues. The behaviour analysis of the LF revealed that there was a mismatch between the searchers' awareness, real interactions, and the functionality of the digital library. Thus, the poor awareness of DL functionalities could be an unreported cause for search failures.

This study also investigated the emotions associated with failed searches and identified these as being anger and sadness. We conclude that digital library users are encountering the same search failure problems as reported in literature almost 20 years ago. Thus, there is a need to improve digital library systems and simplify their interfaces.

To overcome the limitations of this study as previously stated in 4.5, a controlled laboratory study was conducted to validate the results of the online study

here reported and explore the use of a different method to collect more affective data.

Chapter 5

Evaluating User Experiences in a Digital Library

5.1 Introduction

DL used to be evaluated explicitly by collecting data of attitudinal surveys e.g. Priestner and Borg [2016], or implicitly by generating reports of the usage from the LFs e.g. Debowski [2001]. The usability studies measure the performance characteristics; effectiveness, efficiency, and satisfaction Secretary [1998]. Massis [2018] and Appleton [2016] argued that the traditional quantitative metrics and measures are not sufficient to reveal the real value of the DLs. Such quantitative metrics failed to provide a full understanding of the quality of the experiences, instead the user experience (UX) should be the ultimate criterion of the user-centered design Alben [1996].

Cronin [2014] suggested to deploy UX techniques to address age-old problems reported in the DL evaluation studies including the lack of awareness of the search tools, incorrect selection of the search index or misuse of the facets Peters [1989]; Trapido [2016].

The evaluation of the DLs is complicated due to the overlap of different interaction dimensions including the interaction with the task, interaction with the supportive technologies i.e. interfaces, and the interaction with the information resource Saracevic [2000]. Accordingly, DL evaluation studies could be classified into: interaction-design-based evaluation, task-based evaluation, and the overall performance evaluation studies. Our study focuses on the former one and it is based on the framework proposed by Mahlke [2008] as in figure 5.1.

We contribute to the DL research by deploying a user-centered method for fully understanding the UX. The rest of this chapter is organized with the method-

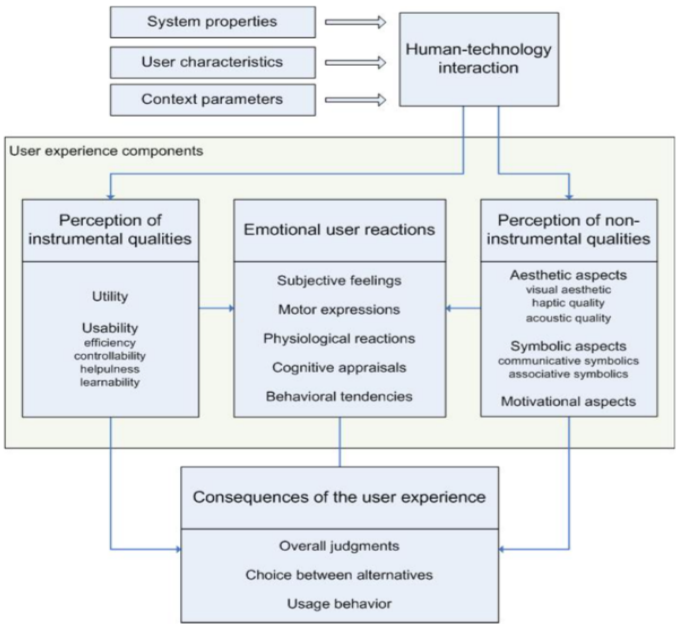


Figure 5.1. Interaction User Experience Framework Mahlke [2008]

ology presented in 5.2. The results and discussion can be found in sections 5.3 and 5.4.

The purpose of this chapter is to investigate the UX in the DL by considering threefold aspects. The first is to understand to what extent the DL functions are usable, the second is to investigate the user perceptions of the instrumental and non-instrumental quality, and the third is to examine the associated mood before and after the search experience. Thus, the research questions are:

- RQ1: *To what extent the DL functions are usable by users with different expertise levels?*
- RQ2: *Will the advanced participants share similar perceptions of the DL compared to the non-advanced participants?*
- RQ3: *What are the associated moods before and after the search experience in the DL?*, and
- RQ4: *What are the influential factors that might affect the moods before and after the search experience?*

Figure 5.2. The Digital Library Homepage Interface

5.2 Methodology

An academic DL, Effat library [1],was chosen. Real users of the DL (undergraduate students) were recruited by sending invitation emails. The experiment was conducted in a laboratory setting. To encourage participants, two course credits were offered, such credits have positive effect as experiments rewards Sharp et al. [2006]. seventy-two participants signed a consent form before the experiment. Only sixty-five participants managed to complete the experiment.

5.2.1 The Library

The library adapted KOHA as the integrated library system that manages the Library's Online Public Access Catalog (OPAC) and Summon as the discovery system. Over 350,000 digital scholarly materials relevant to the programs offered by the university are available. Examples of the academic databases subscriptions are: ACM, IEEE, ProQuest, EBSCOhost, ScienceDirect; Safari Book etc. Figures 5.2 and 5.3 show the DL interfaces.

[1]https://library.effatuniversity.edu.sa/

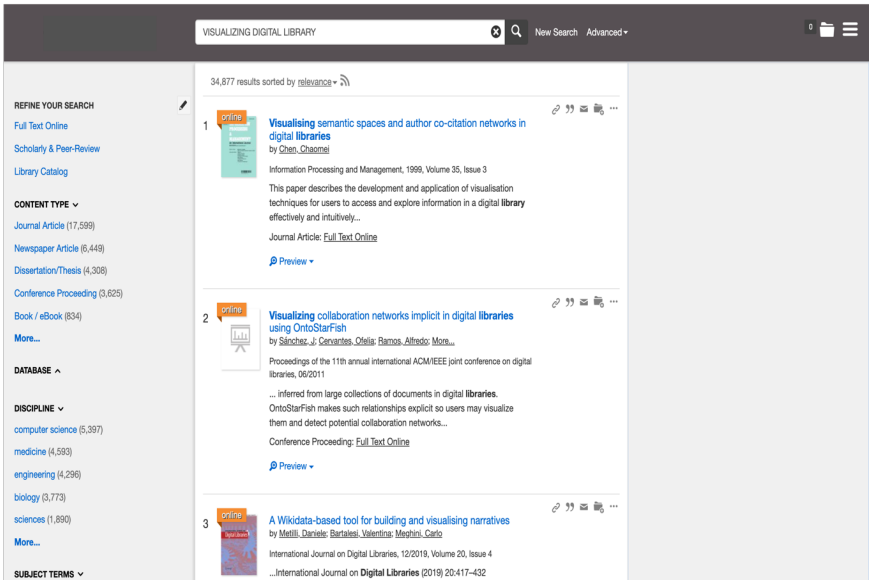


Figure 5.3. The Digital Library Search Result Interface

5.2.2 Experiment Phases

The participants were asked to fill-up a form that gathered basic information; major, and expertise level of using the DL. The participants were directed to conduct an exploratory search task. To make the experiment more realistic, a combination of simulated work task scenario and the participants' personal information needs was applied. Borlund [2003] developed the notion of the simulated task to eliminate the effect of the artificial setting. As the participants search for relevant and interesting information, realistic search interactions are ensured Borlund and Schneider [2010]. Thus, the real information needs of the participants were considered here. And the task was set in the form of a typical academic search and worded as:

> "*Suppose you are writing a paper on a topic of your interest and you wish to gather more information to enhance your existing knowledge or discover new knowledge.*" Kindly use Effat digital library as the information source and try to find at least three relevant sources.

We let participants to select their own topics, for three reasons. Firstly, the nature of the study design required self-assessment of the search results. Secondly, this avoided the bias that might result from the task narrative. Thirdly, the target sample came from different academic disciplines. Therefore, it would be a challenge to design tasks that would motivate all the participants. The experiment

Disciplines	Examples of the information needs
Architecture	*1-"deconstruction movement, and what is the impact of it, and some examples.."* *2-"vernacular architecture, and psychological science in its architecture"* *3-"history of Islamic architecture, particularly, the geometric designs of the Mihrab architecture"*
Business	*1-"how have the plans of vision 2030 impacted the overall economy and the GDP of Saudi Arabia?* *specific numbers and figures such as interest rates, employment level, gender gap, gender pay gap."* *2-"the impact of cryptocurrency on the stock market"* *3-"under employee well-being I would like to search for indicators that are dependent variable of employee well-being."*
Engineering	*1-"cellphone antenna types for Wi-Fi and Bluetooth"* *2-"solar panels, in particular what are the different sources of solar energy, and how does each of them work, and how it affects the voltage"* *3- "Engineering sustainability, the impact of robotics and automation in manufacturing"*
Humanities	*1-"emotional Intelligence and Personality Traits as predictors of Happiness"* *2-" how other writers used psychoanalysis to their advantage in writing"* *3-"the effectiveness of cognitive behavioral therapy on schizophrenia mental disorder"*

Table 5.1. Examples of the Participants' Information Needs

phases are based on Byström and Hansen [2005] work as:

- *Task construction:* participants were asked to complete a pre-experimental questionnaire, including a description of the information they would search for. Table 5.1 shows examples of the participants' information needs.

- *Task performance:* participants interacted with the DL.

- *Task completion:* once participants finished searching, they reported on their search performance using the post-experimental questionnaire.

Beside the pre-and post-questionnaires, the screens of the participants were recorded, with their permission. The purpose of the screens recording is to analyze the search behavior. Finally, 'pick-a-mood' scale was applied before and after the experiment.

Based on students' availability and preferences, the experiments were conducted within six sessions during a week at different time spots i.e. at the beginning of the day, the mid, and by the end of the day. Participants were informed that there were no time limits for completing the task or the whole experiment.

Figure 5.4. Experiment's Lab

Prior to the experiment, a structural video was introduced for each participant illustrating the processes of the experiment. Figure5.4 shows the lab where the experiment was conducted.

5.2.3 Pilot Study

A pilot study was conducted beforehand with two objectives: (1) to decide the number of the tasks, (2) to examine the questionnaires questions. Initially, it was planned to consider two tasks; one specific task (e.g. name of a book) and one exploratory task. Unfortunately, the five students involved in the pilot study were negatively affected by the duration of the experiment. Thus, in order to avoid uncompleted tasks and/or random answers to the questionnaire, we decided to focus on one task only: the exploratory task.

Questions	Frequency of searching DL				
1	Never	1-3 times per term	1-3 times per month	1-3 times per week	Daily use
	8	22	25	9	1
2	**Familiarity with the search tools**				
	Not familiar at all	Slightly familiar	Somewhat familiar	Familiar	Extremely familiar
	11	19	20	12	3
3	**DL searching skills**				
	Beginner	Intermediate	Advanced	Expert	
	30	19	16	0	

Table 5.2. System's Expertise Level

5.3 Results

The analysis of the data considered four aspects: identifying the expertise levels based on the self-assessment, analyzing the information behavioral extracted from the screen recording, understanding the participants' perceptions towards the system qualities, and comparing the pre-and post moods.

5.3.1 Participants Expertise Levels:

The total number of the participants who completed the study is 65 coming from different disciplines; Faculty of Architecture and Design (30 students), Faculty of Humanities (14 students), Faculty of Business (12 students), and Faculty of Engineering (9 students). Expertise level was measured by self-assessment of three aspects namely: *frequency of searching the DL*: (1) How often do you conduct searches in DLs?, *familiarity with the search tools*: (2) How can you rate your searching skill on the DLs?, and *searching skills*: (3) How can you rate your searching skill on the DLs?

As it is shown in table 5.2. The frequency of using the DL was: (1) daily base, (9) 1 to 3 times per week, (25) 1 to 3 times per month, (22) 1 to 3 times per semester, and (8) never. Whereas the degree of familiarity with the search tools was: (3) extremely familiar, (12) familiar, (20) somewhat familiar, (19) slightly familiar, and (11) not familiar at all.

Regarding the searching skills, the scale of Huang [2014] was adopted in this study, none of the participants considered themselves to be expert, defined as fluent in use advanced search functions, and with a sound understanding of back-end information retrieval mechanisms. 16 participants rated themselves as advanced searchers who are good at using advanced search functions e.g. Boolean

operators and filtering results by facets. 19 participants considered themselves to be intermediate; need some help when searching DL, and 30 participants were beginner; just learning how to search for information in the DL, need lots of help. Accordingly, we categorized the participants as non-advanced (N=30) or advanced searchers (N=35).

The non-advanced participants were not familiar with DL search tools, they had rarely conducted DL searches from one to three times per semester, and considered themselves as beginners searchers. The advanced searchers, on the other hand, were somewhat familiar with the search tools in the DL, frequent users of the DL i.e were using the DL once to three times per week and per month,and regarded themselves as intermediate and advanced searchers.

5.3.2 Behavioral Analysis:

In order to answer RQ1: *to what extent the DL functions are usable by users with different expertise levels?*, we compared the search behavior of the advanced and non-advanced searchers, and investigated if there was a significant difference by conducting Mann-Whitney statistical test. Because search behavior is recognized as an indicator of search quality Debowski [2001], behavioral metrics were extracted from the screen records.

The behavioral metrics, as in table 5.3, were quantified and treated as dependent variables. We compared the behavioral metrics of the advanced and non-advanced to investigate if there were significant differences between the behavior of the two groups.

Generally, the mean session duration for both groups were around 20 minutes. The descriptive analysis shows that the advanced searchers recorded more complicated and dynamic interactions compared to the non-advanced in terms of going through different iterations, returning to the search results page very frequently, scrolling down under the first 10 results, and using the search tools. Comparing to the non-advanced searchers who frequently selected the first search results, recorded fewer visits to the search result page, and rarely used advanced search functions or facets.

Those differences were confirmed by the result of the Mann-Whitney test except the later metric (using of the search tools), as in table 5.4. The results indicate that the significant difference between the advanced and non-advanced searchers is recorded only in two aspects: the number of revisits to the result page (Sig= $0.00 < 0.05$) and the scrolling down under the first 10 results (Sig= $0.00 < 0.05$).

Both groups shared two common behavior: (i) the starting point of the search

Metrics	Description
Session duration (SD)	Average duration of the session
Action variable	
search tools (ST)	# of ST used during a session
Result page (RP)	# of revisit RP during the session
Scroll down (SD)	# of search considered results under 10
View item (VI)	# of VI during the session
Query Behaviour	
Number of query (Q)	# reformulating query per session
Query length (QL)	# of the term/s of each query

Table 5.3. Metrics to Quantify the Behaviour

was the default simple search function, although the interface of the DL provides various access points to different contents and databases as in figure 5.2, and (ii) the importance of the search results page for both groups, as in figure 5.3, where both groups reformulated their own queries and restarted their own search from the result page.

5.3.3 The Instrumental and Non-instrumental Quality Perceptions:

Morville [2004] 'honeycomb' model as in figure 2.4 was used to understand the participants' perceptions of the DL and to answer the second research question: *will the advanced participants share similar perceptions of the DL compared to the non-advanced participants?* The 'honeycomb' model consists of seven dimensions namely; usefulness, usability, desirability, findability, accessibility, credibility, and value. The following section presents the participants' perceptions of each of those dimensions. Table 5.5 describes the measurement of each dimension. Figure 5.7 and table 6.4b depict the participants' perceptions.

The *usefulness* dimension investigates to what extent a system is useful to particular users. The weighted averages of both groups were (2.1) with SD 0.89, and (1.95) with SD 0.63, respectively, the weighted averages laid in the interval [1.80-2.59] that equal to "agree" on the 5-point Likert scale. The results indicate a positive attitude of both groups in terms of the usefulness of the DL as an information service provider, and the usefulness of the search tools as searching assistance techniques.

The *usability* dimension examines to what extent the system is easy to use.

Variable	U	Z	Sig
Session duration (SD)	426.500	1.30	.19
search tools (ST)	369.000	2.61	.09
Result page (RP)	75.000	6.53	**.00**
Scroll down (SD)	120.000	6.24	**.00**
View item (VI)	514.500	.14	.88
Number of query (Q)	392.000	1.84	.06
Query length (QL)	514.500	.14	.88

Table 5.4. Behavioral Analysis of the Advanced and Non-advanced Participants

Dimension	**Measurement**
Useful	1-The digital library is useful to find search-related materials comparing to other information sources 2- The search features are useful tools to enhance my searching in digital library
Usable	1- Interacting with the digital library interface is not easy 2- I feel frustration when I use the search tools 3- the search tools complicated the search process
Desirable	1-The interface is attractive 2-The interface is appealed to my senses 3-my experience with the digital library was rewarding
Findable	1-The search results list shows too many results 2-The search results list shows too few results 3-T he search results list shows non-relevant results
Accessible	1- I could easily access the item 2- I could easily download the item
Credible	1-The digital library provides authentic materials 2-The digital library is my primer source for the search-related materials
Valuable	1- Searching in digital library requires time comparing to other information sources 2-The digital library is too difficult, I usually search elsewhere

Table 5.5. Honeycomb Dimensions

Similar perceptions (positive)	Similar perceptions (intermediary)	Different perceptions
Usefulness	Findability	Usability
Desirability	Valuable	Accessibility
Credibility		

Table 5.6. Participants' Perceptions Towards the DL

The advanced participants showed intermediary attitude "neither agree-nor disagree" toward the usability of the DL with a weighted average (2.9) and SD (0.94) which placed in the interval [2.60-3.39]. Comparing to the non-advanced participants who had a negative attitude toward the usability of the DL and the search tools where the weighted average is 2.0 with SD 0.83 equal to "agree" with interval [1.80-2.59].

The *desirability* dimension measures the non-instrumental quality of the system concerning the aesthetic design elements. Both groups were recorded positive attitude towards the interface in terms of the aesthetic aspects. The weighted averages for the advanced and non-advanced participants were 2.43 with SD 0.91, and 2.37 with SD 0.78, respectively.

The *findability* dimension inspects whether the users can locate what they are looking for. It depends on the content coverage, and the interface design. Both groups revealed intermediary answer between disagreement and agreement "neither agree-nor disagree" that the DL content is findable in terms of the quality and quantity, where the weighted averages were 3.12 with SD 1.05 for the advanced participants, and 2.7 for the non-advanced with SD 0.85. This could reveal that the quality and quantity of the DL content might need to be presented in a better way. This is in line with the behavioral metrics were the advanced participants scrolled down under the ten first results looking for more results.

The *accessibility* dimension meant to measure the accessibility of the system to people with disabilities. Since our sample does not contain participants with disabilities, we evaluated the content accessibility by evaluating login restrictions, links validate, and easiness of finding the download icon. The advanced participants showed a positive attitude toward the content accessibility with a weighted average 2.37 and SD 1.13. Comparing to the non-advanced who had a negative attitude to the accessibility with a weighted average 3.5 and SD: 1 equal to "disagree" according to the interval [3.40-4:19]. The differences in the accessibility perceptions might explain the differences in the behavioral analysis

Likert-Scale	Interval
Strongly agree (1)	1.00 - 1.79
Agree (2)	1.80 - 2.59
Neither agree-nor disagree (3)	2.60 - 3.39
Disagree (4)	3.40 - 4.19
Strongly Disagree (5)	4.20 - 5.00

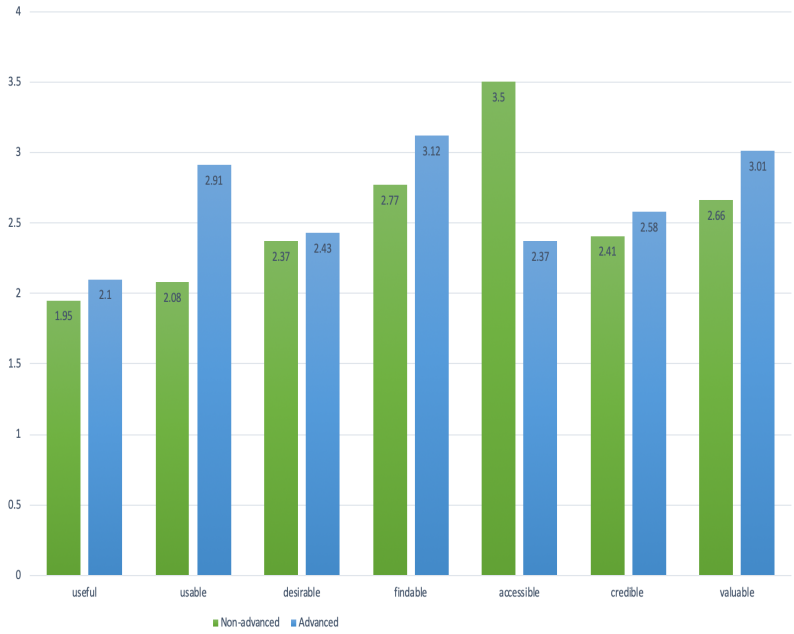


Figure 5.5. The Perceptions Towards the Instrumental and Non-instrumental Quality of the DL

where the non-advanced participants viewed less documents comparing to the advanced participants.

The *credibility* dimension investigates if the users believed that the DL is trustworthy. Both groups were adopted positive attitudes toward the DL as an authentic source of information with weighted averages of 2.58 and SD 1.03 for the advanced participants, and 2.41 with SD: 0.86 for the non-advanced participants which were placed in the interval [1.80-2.59] equal to "agree".

The *value* dimension examines the overall satisfaction levels of the users with a system. Both groups reported an intermediary answer between disagreement and agreement "neither agree-nor disagree" towards the value of the DL with weighted averages 3.01 with SD: 1.11 For the advanced participants and 2.6 with SD: 1.10 for non-advanced participants. This might indicate that the participants could use alternative system if they do not have to use the DL.

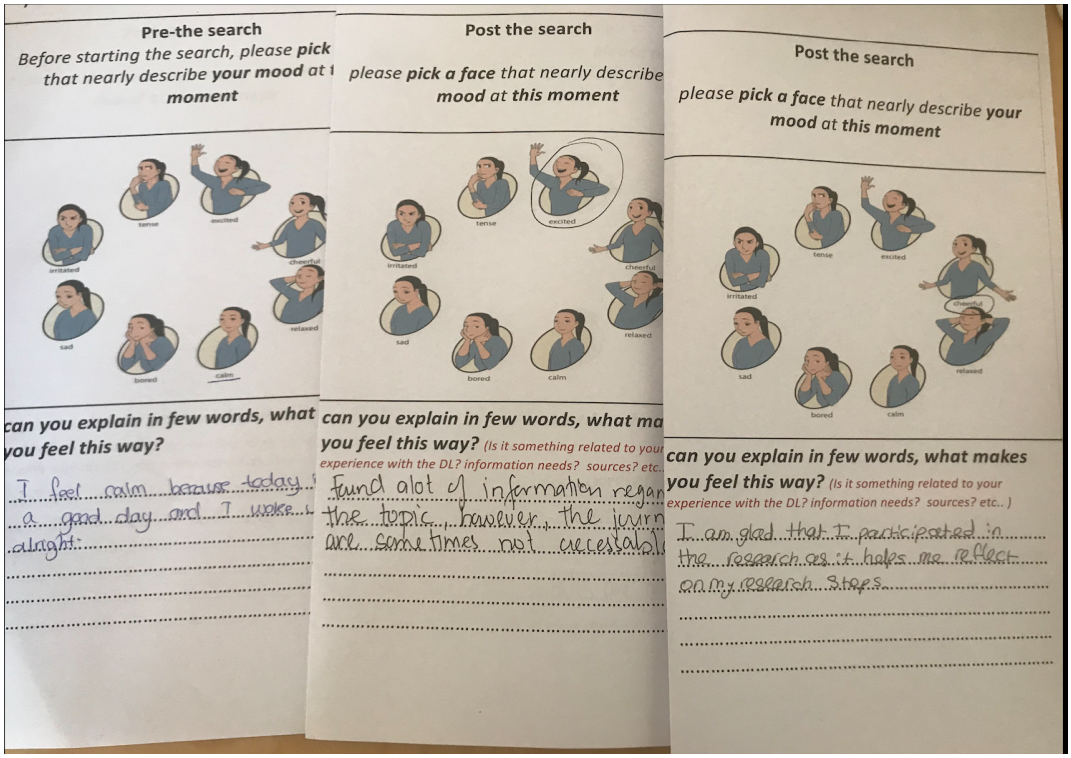


Figure 5.6. Examples of the Participants' Answers

5.3.4 The Associated Moods:

To answer RQ3: (3a)*what are the associated moods before and after the search experience in the DL?*, and RQ4: *what are the influential factors that might affect the moods before and after the search experience?*, 'pick-a-mood' scale as in figure 2.5 was selected as a mood tracker before and after the experiment.

The aim of tracking the mood is to investigate if the moods were affected by the nature of the interactions and the search outcome, if so what are the main factors behind such changes.

Overall, participants showed varied moods as in figure 5.7 due to various factors. Initially, we investigated the pre-experiment moods; more than half the sample, 36 participants, showed pleasant moods; 16 were calm, 10 were relaxed, 8 were excited, and 2 were cheerful. The rest of the participants 29 showed unpleasant moods; 12 of the participants were bored, 10 were tense, 7 were irritated, and no one showed sadness mood.

The participants were asked to explain the reasons behind the experienced moods, figure 5.6 presented examples of the participants' reflections. Accordingly, we categorized the reasons into: personal, contextual, and educational

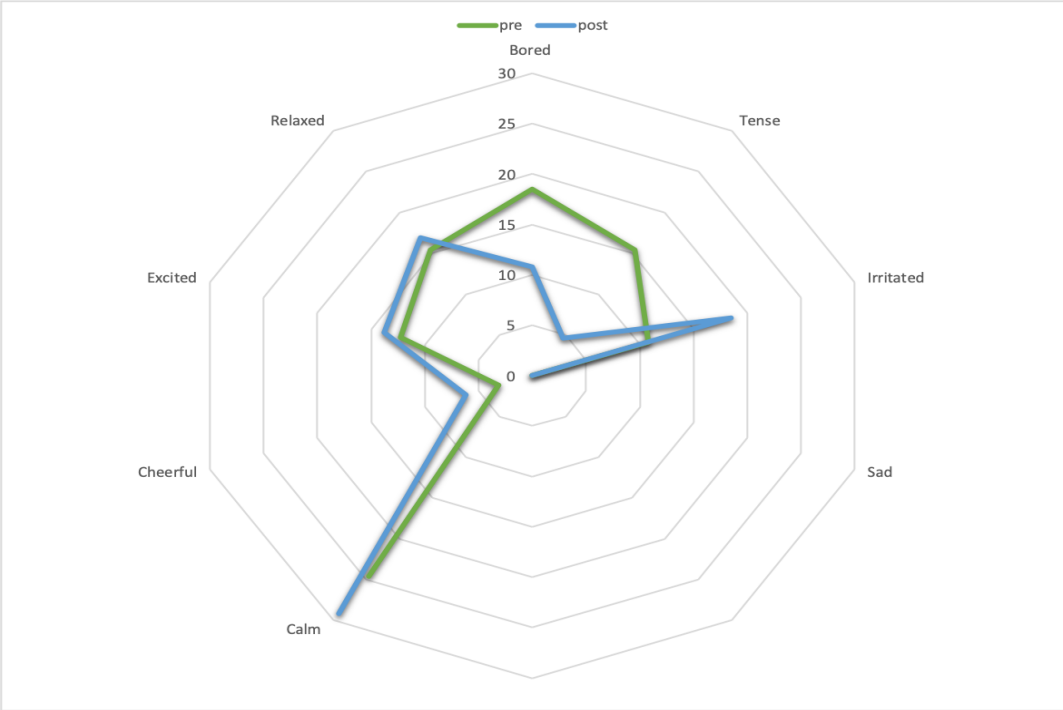


Figure 5.7. Participants' Moods Before and After the Experiment

factors. The calm mood is more associated with personal issues e.g. "*I'm calm as I slept well at night and had a productive day so far*". The relaxed mood is associated with contextual issues i.e. time of the experiment and the lab atmosphere, e.g. "*I finished my classes, so I'm going home early*", and "*I do feel relaxed in the lab, I like the quietness of the place and the temperature is just perfect*", and educational issues e.g. "*I just knew my midterm mark, I'm satisfied with the result*". The cheerful mood is related to educational reasons e.g."*I just submitted my midterm project*". The excitement mood is related to the personal issues, e.g."*it is so rare to be part of such experiment, thank you for sending the email*".

For the non-pleasant moods, participants, who were bored, mentioned personal reasons "*I'm not a fan of taking surveys thus it is not something I enjoy*", and contextual reasons; time of joining the experiment e.g."*I'm so bored because I have two mid-terms today*". The tense mood is often associated with personal issues; confidence level of using the DL, e.g. "*I'm not sure if I'm skillful enough to do search on the DL*", and educational issues e.g."*I'm preparing for my graduation project, its due is soon*". Finally, participants who reported they experienced irritated mood, they refereed the reasons to: personal issues e.g. family problems, contextual issue e.g. temperature of the lab, or educational issue e.g. bad

Mood	Pre-experiment	Post-experiment
Calm	6	19
Relaxed	10	11
Excited	8	9
Cheerful	2	4
# pleasant	**36**	**43**
Bored	12	7
Tense	10	12
Irritated	7	3
Sad	0	0
# unpleasant	**29**	**22**

Table 5.7. The Mood Before and After the Experiment

performance of the exam.

As the participants finished their searches, they were asked to pick a mood that reflected their moods at the current moment. Overall the number of the participants reported unpleasant mood was dropped, and the pleasant mood number increased as it is depicted in table 5.7. By the end of the experiment, 22 participants reported unpleasant moods; 12 were tense, 7 were bored, 3 were irritated. In contrast, 43 participants showed pleasant mood; 19 were calm, 11 were relaxed, 9 were excited, and 4 were cheerful.

Based on the explanations provided by the participants, we noticed that the reasons were shifted to be more search-oriented reasons. Thus, we classified the reasons into: personal and DL issues (content and usability).

Starting from the pleasant moods; the relaxed and calm participants rationalized their experienced mood to personal issues including previous knowledge, preference of the searching topics, expertise of using the DL, and cognitive state e.g. "*I enjoyed my search as I really love what I'm investigating*, *I'm calm because I was concentrating and reading different articles*. Participants who also expressed calm mood justified the reasons to the DL itself e.g. "*I feel relaxed as I found relevant content, also the topic is modern, so I could easily find enough information*", *I'm pleased with the useful search, the search went smoothly, and I hadn't faced any challenges*", and "*I had very fruitful search, may be because I'm not under pressure and it's not an urgent task*.

Those who reported cheerful and excited mood, justified the reasons to the quantity and quality of resources, e.g."*I found very informative materials, and a lot of resources that meet my needs*". The participants also were excited as they

discovered more or noticed a shift in their knowledge state, e.g. "*my search was successful as I discovered new trends in my study area*", and "*this experiment helped me to reflect on my searching skill*".

As the pleasant moods were more associated with the positive experiences, the unpleasant moods were resulted from (i) the insufficient of the quality and quantity of the sources. The insufficiency caused the irritated and tense moods e.g "*Every time I want to check an article, I find it challenge to access the article*", "*My search experience is under expectation. Although there were too many resources, I found it a challenge to choose between the results*", and "*I spent long time looking for relevant material but I couldn't find any!*". Or, from (ii) personal issues e.g. as in responses of bored and tense participants: "*I don't like to search in the DL, I need time to use the library*".

5.3.5 Challenges and Suggestions for Improvement

Two more multiple choice questions were asked: to evaluate the experience with the DL in general, and to gather user requirements to improve the DL interface in the future.

The first question was about the most difficult challenges that the participants usually face when they interact with the DL. The participants were encouraged to choose more than one option. Among six situations which identified from the previous works Blandford et al. [2001]; Rosman et al. [2016], the advanced participants reported that the most difficult situations were: "I found it a challenge to evaluate the relevant of the items to my information needs", and "I can't decide where I should start my search" as it depicts in figure 5.8. The later challenge was also reported by the non-advanced participants as challenge number 1. Followed by "I found it a challenge to download the item".

The second question: "If you have the chance to modify the current digital library interface, what would you like to do?" aims to collect opinions for improving the interface.

As shows in figure 5.9, the majority of the participants from both groups went for: "It would be better to add a graphical network shows other relevant resources related to my search", followed by "It would be better to add a search trail that shows my searching history", and finally to "It would be better to design a simpler interface". The suggestion "It would be better to remove the facet bar" received less support.

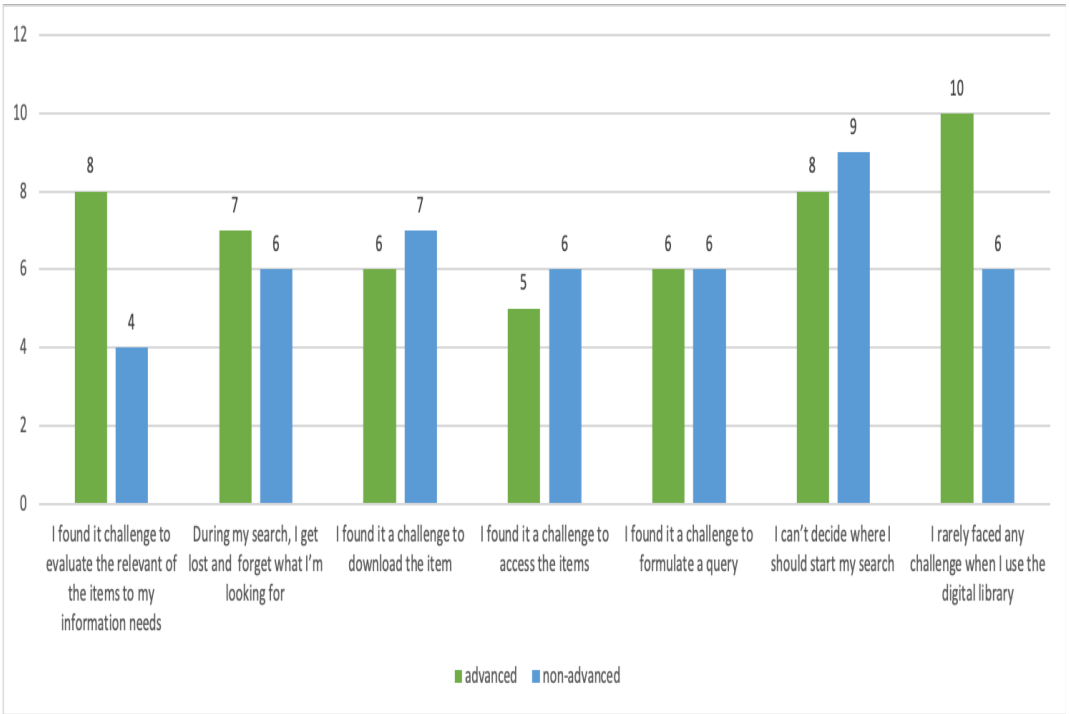


Figure 5.8. The Most Difficult Situations when Dealing with DL

5.4 Discussion

This study investigated a neglected aspect of the DL evaluation; assessing the user experience. As far as we know, there is no a holistic framework that can help the librarians and the DL system designers to evaluate the UX in the DL sector. Thus, this study contributes to the field by adapting Mahlke's interaction user experience framework. The framework helps in designing the research design by: (i) identifying the factors that affect the UX: system properties, user characteristics, and context metrics. And (ii) defining the main UX components: users' perceptions of the instrumental and non-instrumental quality, and the affective variables.

The former one, identifying the factors, was investigated by addressing RQ1: *to what extent the DL functions are usable by users with different expertise levels?*

To answer this question, we firstly categorized the participants into two groups: advanced and non-advanced searchers based on self-rating questionnaire. Accordingly, the searching behavior of both groups were analyzed with the aim of investigation any significant difference between both groups.

Overall, our results indicate that the DL functions were not fully used as they

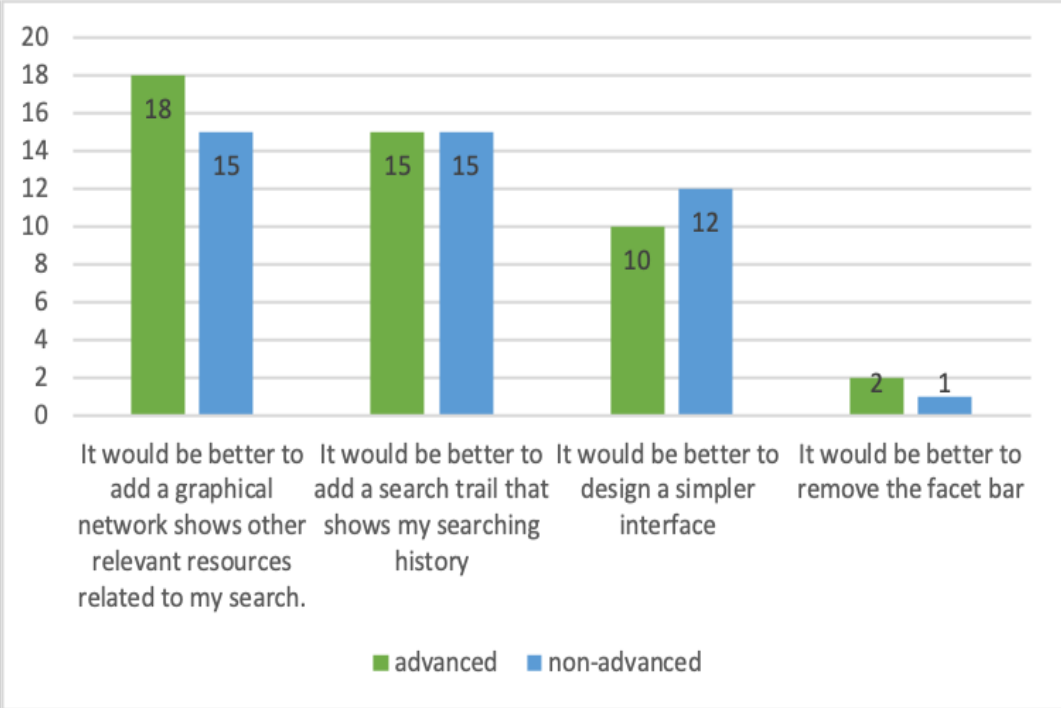


Figure 5.9. User Requirements to Improve the Interface

were supposed to be regardless of the expertise level. This was so obvious with the homepage functions where most of the participants started their searches from the default simple search function (search all) without modification. This finding is in line with Georgas [2014] who found that changed the default setting is very infrequently between the students. And Dahlen and Hanson [2017] who confirmed that the default settings in discovery systems have a substantial impact on student source choices. Similarly, Asher et al. [2013] found that many students are unlikely to change the default settings or use facets. This might indicate that the participants were not aware of the differences between the available information sources. Thus, special attention needs to be paid to the configuration of the default search parameters and the information architecture of the homepage interface.

Although both groups had spent most of their search time within the search result page, their usage of the functions was light. The low use of the search tools reinforces previous observations Asher et al. [2013]; Dempsey and Valenti [2016]. The findings of scrolling down behavior under the ten results contradicted the findings of the previous studies e.g. Georgas [2014]; Hamlett and Georgas [2019] where they confirmed that students are unlikely to go beyond

the first page, listed 10 results per page, of search results. The scrolling down behavior revealed the misuse of the sort function where the results could be rearranged chronologically based on the (newest, oldest, or relevance). This problem could also be resolved by using the facets and reduce a large result set to a manageable size as Sadeh [2007] suggested.

In summary, regardless the expertise levels, the participants did not take full advantage of the search tools that might aid them in their quest for information. The inadequate use of these tools suggests a lack of awareness of their importance, or even of conceptual understanding of the system's functions Trapido [2016]. Some studies e.g. Dempsey and Valenti [2016] and Dahlen et al. [2020] did not find significant impact of the DL instructional tutorials and the use of facets and search tools. Thus, a special attention needs to be paid to the interface design.

The subjective perceptions of the participants were also considered to understand if the users with different expertise levels shared the same perception of the DL; RQ2: *will the advanced participants share similar perception of the DL comparing to the non-advanced participants?* The results showed that both groups disclosed similar attitude towards five dimensions out of seven dimensions of the honeycomb framework. Three out of the five dimensions generated positive attitude "agree", those are: the usefulness of the DL, the desirability, and the credibility. The two other dimensions, findability and value, recorded an intermediary answer between disagreement and agreement. Both groups were neither agree-nor disagree that they can locate what they are looking for easily (findability).

This was also reflected in the searching behavior where the advanced participants kept scrolling under the first ten results, with rarely using search tools. This is also related to the participants perception of the value dimension of the DL when it was compared to the other academic search engines e.g Google Scholar, in terms of the difficulty and time requirements. The intermediary answer between disagreement and agreement that the DL is valuable could reveal the preference for the simplicity of search engines over the complexity as reported by Deodato et al. [2016].

The other two dimensions where the two groups disclosed different perceptions are: the usability and the accessibility dimensions. The advanced participants showed intermediary attitude toward the usability of the DL comparing to the negative attitude of the non-advanced participants. As none of the groups were agreed that the DL is usable, this might explain the light usage of the search tools and the usability issue of the DL. This also is applicable to the accessibility dimension where the non-advanced participants faced problems to access the

content or find the download icons. Despite the rationale of implementing the search tools within the DL interfaces, usability and accessibility reported in previous works are still existed. Students still struggle with the DL interfaces, the overwhelming of the search tools, or misuse of the search tools Blumer et al. [2014]; Debowski [2001].

Regards the RQ3: *what are the associated moods before and after the search experience in the DL?, and RQ4: what are the influential factors that might affect the moods before and after the search experience?* The results revealed two important findings, (1) the number of the pleasant moods increased after the experiment, compared to the unpleasant moods. And (2) the influential factors that affected the moods changed from being personal, contextual, and educational issues to be more search-oriented issues.

The significance of investigating the mood lays in its effect positively or negatively on the information seeking behavior. The findings of the previous works indicate that students' mood either positive or negative influence whether they seek out more information i.e. promote information seeking Trapido [2016], or inhibit information seeking Carver [2003]. Kuhlthau [2004] identified two types of moods that affect the information behavior: an invitational mood leads to expansive actions, and an indicative mood results in conclusive actions. In this research, we are more interested in understanding the factors that affect the participants' moods and might lead to positive or negative UX. Kuhlthau [2004] found that mood may shift during various stages of the information search process (i.e. beginning, the middle, and the end of the searching).

Overall, in our study the pleasant moods after the experiment were higher. Kuhlthau et al. [2008] found that negative emotions, particularly uncertainty and frustration, were experienced in the beginning and in the middle of the search episode, but at the end users felt satisfied and relieved. This is also the case with our participants where 46% of the participants experienced deactivated-pleasant (relaxed, calm) mood and 20% showed activated-pleasant (excited, cheerful) moods. Simultaneously, the unpleasant moods were dropped down to 33.9% after the experiment where 23% of the participants felt activated-unpleasant (tense, irritated), and 10.8% were deactivated- unpleasant (bored, sad) mood. Our findings contrast the findings of Gwizdka and Lopatovska [2009] and Flavián-Blanco et al. [2011] who found that better mood prior to the search resulted in increased regret and frustration after the search. However, by analyzing the reasons behind the unpleasant moods, we found that personal reasons and DL factors including content and usability play a significant role behind the unpleasant moods. Previous studies confirmed the relationship between the negative user experience and the poor website design, and they called for simpler

website design Bossaller and Sandy [2017].

In conclusion, there persists a need to improve the DL system and simplify its interface, this is in line with Markey [2007]. Designing simple, fast, and easy to use search systems and interfaces could be the way forward Trapido [2016]. In particular, focusing on an interface that increases users' awareness of the search tools and their accurate application remains a worthwhile goal. As a suggestion, and based on the gathered user requirements, visualized DL interfaces could better support the search experiences and speed up the time taken by struggling searchers. Interactive visualization techniques as in Nedumov et al. [2019]; Ruotsalo et al. [2013] might improve the UX and facilitate the search process.

5.5 Conclusion

This chapter attempts to explore the UX in a DL context and concludes that investigating the aspects of the UX provides us with better understanding of the user experience comparing to the log file analysis or usability studies. We found that Mahlke's interaction user experience framework can be used as an assessment tool to better design UX experiments. Mahlke's framework can be enhanced by including more practical tools e.g. Pick-a-mood scale and honeycomb model.

The use of Mahlke's framework as a validated UX framework was supposed to be validated by another user study conducted with other DL, but due to COVID-19 the study was cancelled. Instead a large-scale survey was run and we present it in the next chapter.

Chapter 6

Large-scale Attitudinal Survey

6.1 Introduction

So far, in this research the interactions with the DLs were examined implicitly by analysing the LFs, and explicitly by conducting online and laboratory user studies. In order to confirm or deny the commonality between the findings of the studies, a large-scale attitudinal survey was distributed to collect more data of the users' perspectives by considering a new DL with different population. Thus, this chapter attempts to answer the following research questions:

- RQ1: *How would user perceive Saudi Digital Library (SDL)?*, and
- RQ2: *What are the commonality between the studies?*

The following section (2) describes the DL, the study set up is explained in section (3), follows by presenting the results in section (4), and the chapter concludes by discussing the commonality between the studies of this thesis.

6.2 The Platform

The aim of this study is to collect more data of the users' perspectives with different population. Saudi Digital Library (SDL) [1] was chosen to conduct the study. SDL is a consortium of Saudi higher education institutions that was established by the Ministry of Education in 2010 to provide access to more than 169 global and Arabic databases, a variety of electronic information resources such as books, journals, theses, reports and multimedia via online subscriptions.

[1] https://sdl.edu.sa/sdlportal/en/publishers.aspx

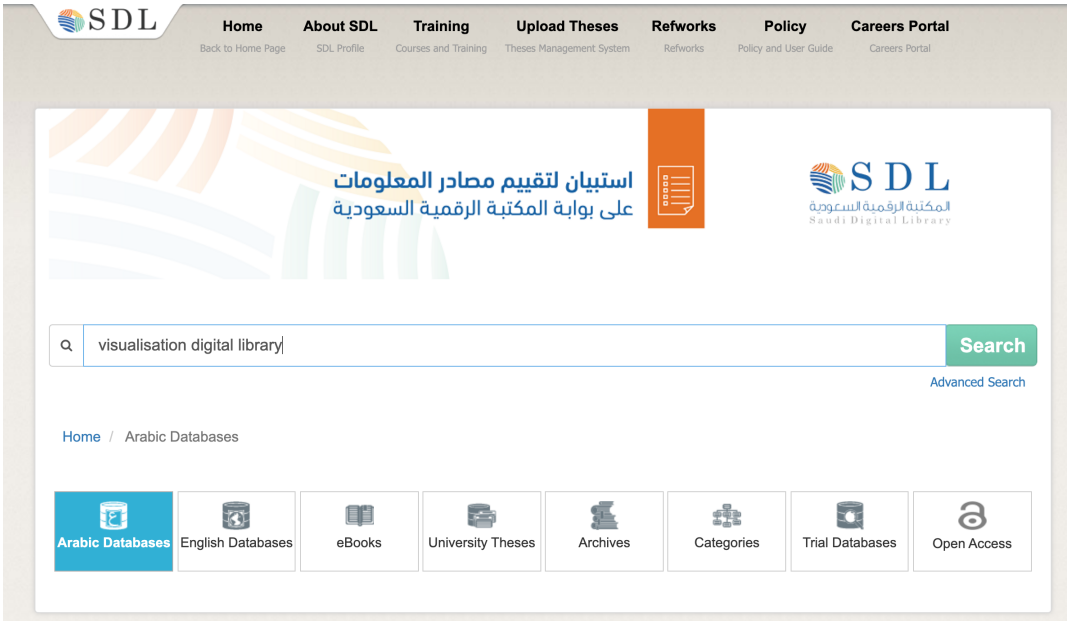


Figure 6.1. SDL Homepage

The information architecture of SDL is similar to other DL platforms, the home page, as in figure 6.1, consists of a simple search and advanced search functions and a list of the available databases. The search result page, as depicts in figure 6.2, has the searching functions (1), and it is divided into the facet bar (2), and the results section (3).

6.3 The Study

The questionnaire developed previously in the UX's study as in chapter 5 was used in this study too. The study was conducted online by sending the questionnaire to the faculty members of three Saudi universities, King Saud University [2], King Abdulaziz University [3], and Princess Nourah University [4]. Then, the faculty members sent the survey to the student lists with the purpose to collect user perspectives of the SDL. Those three universities were chosen because they are providing undergraduate and postgraduate programs, so we could have more representative sample.

[2]https://www.ksu.edu.sa/en/
[3]https://www.kau.edu.sa/home$_e nglish.aspx$
[4]https://www.pnu.edu.sa/en/pages/home.aspx

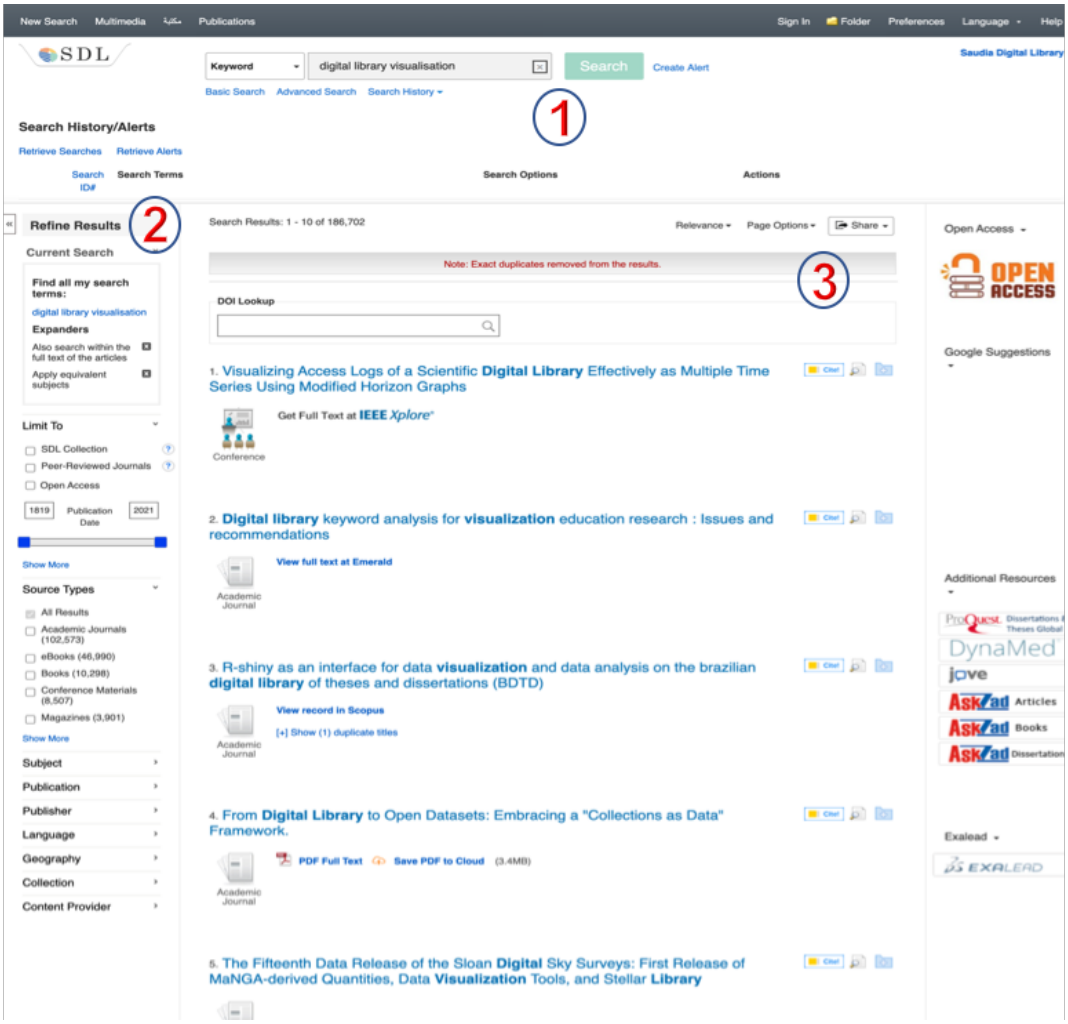


Figure 6.2. The Results Page of SDL

Departments			
Business 9	Medicine 13	Social Science 18	Information science 61
Educational levels			
Undergraduates 19	Postgraduate (Master and Doctorate) 72	Professors and researchers 10	
Gender			
Female 70		Male 31	

Table 6.1. Demographics Data of the Participants

6.3.1 Participants

Out of 150 participants, only 101 participation was completed. Table 6.1 shows the demographics data of the participants. The participants came from different disciplines; Business (9), Medicine (13), Social Science (18), and Information science (61). The majority were female 70, and postgraduate i.e master and doctorate, 72 students. 19 were undergraduate, and 10 were professors and researchers.

Regarding the expertise level of using the DL, the participants self-assessed their experiences as it depicts in table 6.2. Forty-four considered themselves as advanced searchers who are good at using advanced search functions e.g. Boolean operators and filter results by facets. Thirty were intermediate who need some help when searching DL, fourteen were experts, they were quite familiar with the searching tools in the DL, and had a adequate knowledge of the information retrieval techniques. Only thirteen participants were beginners. In terms of the frequency of using the DL, the majority, forty-three participants use the library at least once to three times a week, and twenty use it once to three times a month. Accordingly, the participants were divided into two groups: group 1 is the non-experts (43), and group 2 is the experts (58).

6.3.2 Instrument

In chapter 5, a questionnaire was developed based on the honeycomb UX facets as in figure 2.4. The same questionnaire is used in this attitudinal survey. The questionnaire consists of seventeen statements divided on the seven aspects of the honeycomb as in table 5.5 previously explained in section 5.3.3

Questions	Frequency of searching DL				
1	Never	1-3 times per term	1-3 times per month	1-3 times per week	Daily use
	6	22	20	43	10
2	**Familiarity with the search tools**				
	Not familiar at all	Slightly familiar	Somewhat familiar	Familiar	Extremely familiar
	15	30	22	24	10
3	**DL searching skills**				
	Beginner	Intermediate	Advanced	Expert	
	13	30	44	14	

Table 6.2. System's Expertise Level

6.4 Results and Discussion

This section presents the analysis of the results and the discussion of the two research questions: RQ1: *How would user perceive SDL?*, and RQ2: *What are the commonalities between the studies?*

6.4.1 Participants' Perspectives of the SDL

Both groups showed similar perspectives towards four aspects of the honeycomb aspects; desirability, findability, accessibility, and credibility. As it depicts in table 6.3, both groups expressed a positive attitude towards the *credibility* of SDL as an authentic information source.

Both groups revealed an intermediary attitude "neither agree-nor disagree" toward the *desirability* that measures the non-instrumental quality of the system concerning the aesthetic design elements. Probably more attention need to be given to the aesthetic design elements e.g. color and fonts.

The *findability* aspect investigates whether the users can locate what they are looking for. It depends on: the content coverage, the interface design, and the users experts to use the search tools. Thus, three elements were considered to measure the findability as follows: "the search results list shows too many results","The search results list shows too few results", and "the search results list shows non-relevant results". Both groups revealed intermediary answer between dis-agreement and agreement "neither agree-nor disagree" that the DL content is findable in terms of the quality and quantity, where the weighted averages were 2.52 with SD 1.03 for the experts participants, and 2.60 for the non-experts with SD 1.10. This might indicate that participants were not satisfied with the content available, also it might reveal the less utilization of the search tools.

Both groups showed similar attitude, intermediary attitude, towards the *accessibility* aspect. The accessibility dimension meant to measure the accessibility of the system or product to people with disabilities. Since our sample does not contain participants with disabilities, we evaluated the content accessibility by evaluating:"*I could easily access the item*" measures the login restrictions and links validate, and "*I could easily download the item*" investigates the easiness of finding the download icon. The intermediary attitude is in line with the previous aspects; desirability and findability, this might enforce the need to redesign the interface with paying special attentions to those aspects.

The participants showed different attitudes towards the other three aspects; usefulness, usability, and valuable of the DL. In terms of the *usefulness* aspect which investigates to what extent a system is useful to particular users. Two aspects of the DL were considered: "*the digital library is useful to find search-related materials comparing to other information sources*", and "*the search tools are useful tools to enhance their searching in digital library*". The weighted averages of group 1, non-experts, was 2.71 with SD 1.10 which is equal to "neither agree-nor disagree" comparing to the experts who were agreed on the usefulness and search tools. This might indicate the unfamiliarity and difficulty of the search tools to the non-experts participants.

This also was obvious on the *usability* aspect, where the non-experts were not satisfied with the usability dimension that examines to what extent the system is easy to use; "*interacting with the digital library interface is not easy*", and "*I feel frustration when I use the search tools*", and "*the search tools complicated the search process*". The non-experts were agreed with the statements comparing to the experts who showed an intermediary attitude.

Finally, the *valuable* dimension examines the overall satisfaction levels with a system or product. We measured this facet by comparing the DL with other information providers e.g. Google Scholar by "*searching in digital library requires time comparing to other academic search engines e.g Google Scholar*", and "*the digital library is too difficult, I usually search elsewhere*" the latter one is adopted from Rosenbaum et al. [2008].

The non-experts adopted a negative attitude with weighted average 2.30 and .89 as SD, comparing to the experts participants who showed intermediary attitude towards the value of the DL with weighted average 2.77 with SD: 1.11. This might indicate that the participants could use alternative system if they do not have to use the DL.

Aspect	Group	Mean	SD	Degree
Useful	G1	2.71	1.10	3 Neither agree nor disagree
	G2	2.51	1.14	2 Agree
Usable	G1	2.30	.81	2 Agree
	G2	3.15	1.11	3 Neither agree nor disagree
Desirable	G1	2.78	.94	3 Neither agree nor disagree
	G2	3.05	1.10	
Findable	G1	2.52	1.03	3 Neither agree nor disagree
	G2	2.60	1.10	
Accessible	G1	3.10	.97	3 Neither agree nor disagree
	G2	3.23	1.08	
Credible	G1	2.02	.81	2 Agree
	G2	2.20	.89	
Valuable	G1	2.30	.89	2 Agree
	G2	2.77	1.11	3 Neither agree nor disagree

Table 6.3. Perspectives of Both Groups to SDL

6.4.2 User Requirements and Challenges

Participants also answered two more multiple choice questions; to gather user requirements with the aim to improve the DL interface in the future, and to identify general challenges users face when they interact with the DL.

The first question was "*If you have the chance to modify the current digital library interface, what would you like to do?* This question aims to collect opinions for improving the interface. As depicted in 6.4, the majority of the participants from both groups went for: "*It would be better to add a citation function*", followed by "*It would be better to add a search trail that shows my searching history*", then "*It would be better to add a graphical network shows other relevant resources related to my search*", followed by "*It would be better to add more interactive search tools*", and finally "*It would be better to design a simpler interface*".

The second part of the survey was to measure the most difficult challenges that the participants usually face when they interact with the DL. The partici-

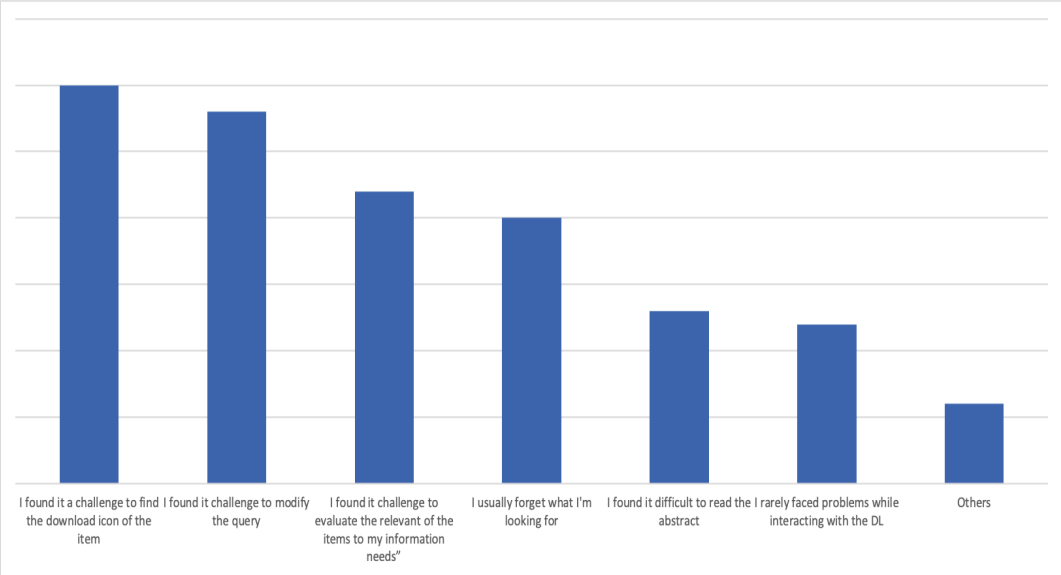


Figure 6.3. Challenges Faced by SDL Users

pants were encouraged to choose more than one option. Among six situations which identified from the previous works Blandford et al. [2001]; Rosman et al. [2016], and based on the researchers' evaluation of the interface, the participants reported that the most difficult situations, as it is shown in 6.3, were: "*I found it a challenge to find the download icon of the item*", "*I found it challenge to modify the query*", *I found it challenge to evaluate the relevant of the items to my information needs*", *I usually forget what I'm looking for*, and *I found it difficult to read the abstract*. In summary, most of the challenges reported in the second question complement the participants' attitudes collected from the questionnaire. Participants reported the accessibility to the DL material as the main challenge.

6.5 Commonality Between the Studies

By comparing the results of the two studies i.e. the laboratory study in chapter 5, and the attitudinal survey in chapter 6, we could confirm that regardless of the populations, there are some aspects that received similar attitudes of both DLs as it shows in table 6.4.

Participants in both studies confirmed that the DLs are considered as a credible source of information and scholarly materials. To maintain and boost this value, DL administrators need to pay attention to other UX fundamental aspects

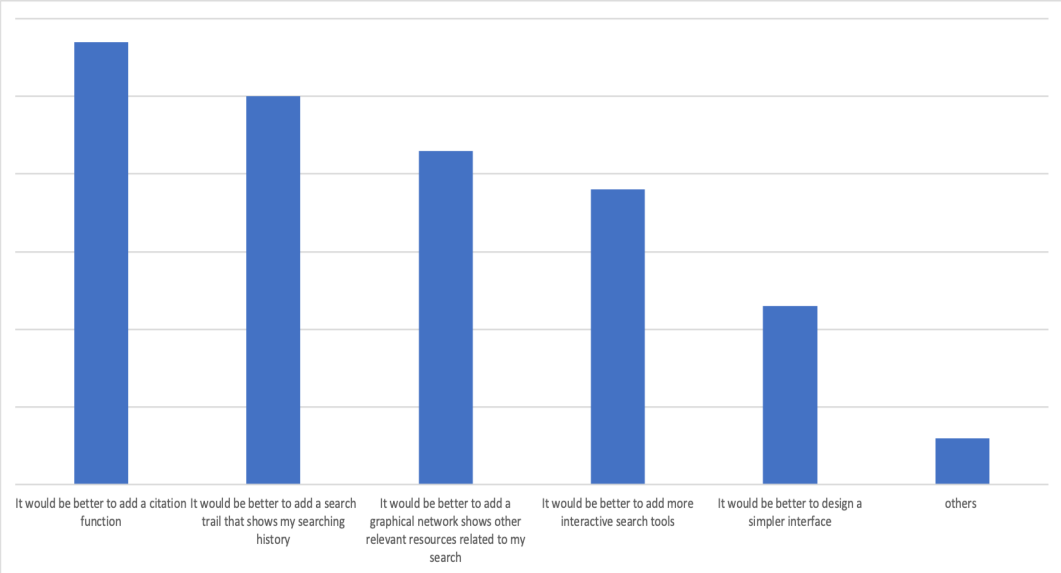


Figure 6.4. User Requirements and Challenges

that their absences might lead to users loose. For instance, both populations agreed that finding materials in the DL is not straightforward process. This is in line with Gaona-García et al. [2017] who confirmed that the use of the DL is not always straightforward, because their interfaces and search mechanisms do not always offer and support adequate searching strategies. Consequently, both populations showed either intermediary or negative attitude towards the value of the DLs, and to the accessibility aspects. Difficulties in finding resources, and navigation problems were reported in previous studies Khoo et al. [2012]. Similarly, Tammaro [2008]; Cignoli and Liu [2011] consider the findability features include i.e information architecture, interface design, accessibility, collection quality, collection's scope, availability of resources to public, speed and accessibility to many resources are important evaluation metrics.

Not surprisingly that the usability, as a cornerstone component of interface design, received different perceptions from both populations, non of them is a positive attitude. This is an indication that more effort is needed to improve the DL interfaces. This is in line with Tsakonas et al. [2013] who confirmed that searching in DL is time-consuming that requires great effort to understand the user interface. Thus, the authors called for the improvement of the DL interfaces. Similarly, the studies of Buchanan and Salako [2009]; Petrelli [2008] confirmed that search processes based on a limited list of filtered criteria are highly time-consuming for users.

Similar Perceptions (Positive)	**Similar Perceptions (intermediary)**	**Different Perceptions**
Credibility	Desirability	Usefulness
	Findability	Usability
	Accessibility	Valuable

(a) SDL

Similar perceptions (positive)	**Similar perceptions (intermediary)**	**Different perceptions**
Usefulness	Findability	Usability
Desirability	Valuable	Accessibility
Credibility		

(b) Effat DL

Table 6.4. Participants' Perceptions Towards Both DLs

The commonality between the thesis's studies is explained in table 6.5, and they can be summarised as follows:

- Facets intended to be a way of encouraging users to think about similar and interrelated words, provide a domain summary, broaden relevant vocabulary terms and pivot searches in new directions.
- The analysis of the LFs revealed that users did not take full advantage of the search tools.
- Thus, we assumed that this is might because of the level of familiarity and expertise of using the system.
- Based on the user studies, we concluded that the expertise level does not have an effect on using the DL's functionalities that were not used fully by both the experts and non-expert searchers.
- The main findings of those studies confirmed that there was a mismatch between the searchers' awareness, real interactions, and the functionality of the DL.
- The inadequate use of these tools suggests a lack of awareness of their importance, or even of conceptual understanding of the system's functions.

- Searchers might not understand how to craft an appropriate search strategy to meet their information needs.
- DL users are encountering the same search failure problems as reported in literature almost 20 years ago.
- Together with raising people awareness of how to use the system, we need systems that are better tailored to the users' needs.
- There is a need to improve the DL systems and simplify their interfaces.

Such conclusions are in line with the previous works where the systematic literature review conducted by Gaona-García et al. [2017] who identified the main challenges in the field of the DLs and repositories. Four main factors were identified as challenges in the DLs systems, they are:

Knowledge representation schemes play a significant role in the classification, categorization, linking and management of digital resources which usually are carried out by domain experts. Examples of the representation schemes are: hierarchical tree structure, and a faceted classification. These classification tools are not always used and exploited by the end-users. This is due to the lack of awareness of the context of the classification scheme and how terms are interrelated, thus users might not be aware of possibilities of such functions. Consequently, the end-users ignore the benefits of using these types of search tools.

The poor quality of the description of the resources through the use of metadata, this can be argued to be the most important factor that influences the location and access of digital resources. Examples of associated problems with describing the metadata are: ambiguity, inconsistency, redundancy, and lack of accuracy.

DLs offer different kinds of search and navigation mechanisms to facilitate access to relevant resources. *The misuse of the search and navigation methods* is another challenge that needs to be considered when designing and evaluating DLs. It encompasses the lack of a common and adequate terminology, problems in the accessibility of resources, and errors in the association of terms. Accordingly, users might face difficulties in finding resources, or returning to a previously accessed record.

All these listed challenges clearly affect the interfaces, therefore the final challenge relates to *the usability problems associate with the search interfaces*. Examples of the limitations associated with poor DL interfaces are: problems to locate and display resources, problems using the interface and interface's functionalities, combining search and navigation methods, higher learning effort and reliability of the search tools. In conclusion, the use of the DL is not always

straightforward, because their interfaces and search mechanisms do not always offer and support adequate searching strategies Gaona-García et al. [2017].

6.6 Conclusion

This chapter complement the previous user studies by comparing the results of the large attitudinal survey with previous conclusions. In the next chapter, as a response to the problems reported in the user studies, we proposed the interactive visualised interface and presented the evaluation study.

Study	Natural	No. of Participants	The Platform	Main Focus	Main Findings
1	LF	None	RERO Doc	Identify usage patterns	1. The results of the usage patterns show considerable variation in using the system functions. This might indicate the effect of the level of familiarity and expertise with the system functionality. 2. The insights gained from the UPs encouraged us to further investigate the relation of the system functionality to generate better interactions e.g. redesigning the interface based on the task types.
2	Online	32	RERO Doc	Identify failure causes	1. Users are encountering the same search failure problems as reported in literature almost 20 years ago. Thus, there is a need to improve the DL systems and simplify their interfaces. 2. There was a mismatch between the searchers' awareness, real interactions, and the functionality of the DL
3	Laboratory	65	Effat DL	UX	1. Investigating the aspects of the UX provides us with better understanding of the UX comparing to the log file analysis or usability studies. 2. Students still struggle with the DL interfaces, the overwhelming of the search tools, or misuse of the search tools. 3. Usability and accessibility problems reported in previous works are still existed 4. DL factors including content and usability play a significant role behind unpleasant moods. 5. Special attention needs to be paid to the information architecture of the DL interface
4	Survey	101	SDL	collect more data of the users' perspectives	1. Participants showed a positive attitude towards the DL in terms of the credibility. 2. Intermediary attitude towards desirability, findability, and accessibility 3. Negative and intermediary attitude towards usefulness, usability, and value of the DL and search tools.

Table 6.5. Summary of the Studies Conducted in this Thesis

Part IV

Prototyping

Chapter 7

Can Visualisation be a Solution?

> "It's true that there has been a good deal of research on visualisation of search interfaces, but I can't think of a single instance where the research has led to successfully implemented systems in general use. Why do you think it would be useful in the context you explored?"
>
> *A question received at ISIC 2020 by Thomas Wilson*

7.1 Introduction

Several studies have examined the information-seeking behavior of the DLs' users and recommended on one side to boost the information literacy in response to advancements in online discovery methods and on the other, to redesign the DL interfaces Buck and Mellinger [2011]; Daniels and Roth [2012]; Fawley and Krysak [2012]. Thus, we decided to take a step further and redesign the current DL information space by adopting Information visualisation (*Infovis*) method. Infovis is a multidisciplinary research areas, including HCI, computer graphics, and cognitive psychology. It is an alternative approach to present textual data and lists to reinforce human cognition. Infovis is as an opportunity to support different ISB and enhance the overall UX.

Björneborn [2010] stated that "the design of an information space shapes the

ways users can interact with this space", thus he claimed that the traditional design of the DL supports only the convergent information behaviour which refers to the goal-directed behaviour, and ignores the divergent or the explorative behaviour. Pérez-Montoro and Nualart [2015] confirmed that the available search tools of the DL including classification support users with only goal-directed information needs, while there is a need to design the DL interfaces to support the divergent information needs. Because the search behaviors and cognitive processes of users can be significantly influenced by the type of search interface Butcher et al. [2011], the goal of this chapter is to examine the effectiveness of the visualised interface as an alternative to the traditional DL interfaces.

The following section 2 describes the methods used to design the new interface and the evaluation study, the results of the study are reported in section 3, and the chapter concludes with the discussion and future directions in sections 4 and 5.

Hence, we conducted a user-centred evaluation study to provide answers to the following research questions:

- RQ1: *How do users perceive a visualised interface for a DL?*
- RQ2: *To what extent the new visualised interface can be adopted and understood by users?*

7.2 Methods

This section explains the design of the new interface and the evaluation study.

7.2.1 System Implementation

RERO Doc DL [1] was adopted in this study to be the experimental system.

Instead of building a system, a clone website was developed. As it shows in figure 7.1, basically, users interact with the clone website through its interface (1), as the users request a specific page from a server (2), our DLS sends the same request of the users to the original website doc.rero.ch (3). In response to the request, the original website sends the required data to the DLS to be stored in a specific controller (4). Finally, the controller processes the data (5), and passes it to the clone interface (6).

[1] http://doc.rero.ch/

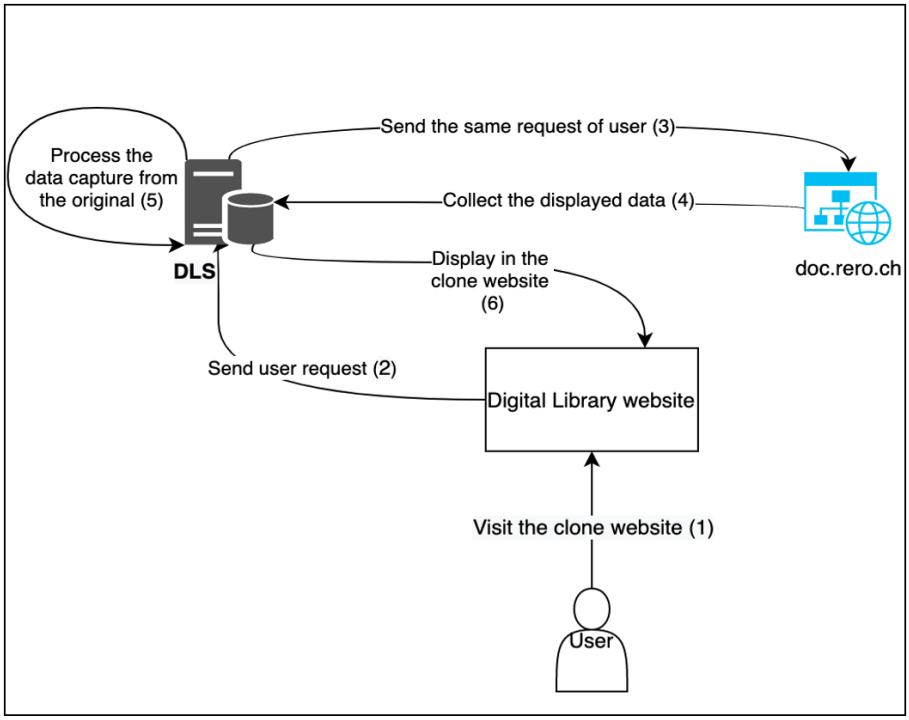


Figure 7.1. The System Architectural

7.2.2 Interface Design

The design of the new interface was informed by the studies of Wilson [2011]; Huurdeman et al. [2019] who identified four groups of interface features that have effects on the information search stages, those are: (1) *input* and (2) *informational features* that aid users in expressing their needs and provide results; the search box and results list. A set of assistive (3) *control* features that allow users to restrict or modify their input, e.g. search filters, tags and query suggestions, and (4) *personalisable* features e.g. query history and a feature to save results that are tailored to the search experience.

Accordingly, the new interface consists of:

- The *input* and *informational features*: the search functions and search results list.
- The *control* feature: the interactive *radar function* located on the left side of the search results, and
- The *personalised* features: (1) *searched query* function that showed the query history, and the (2) *saved articles* function that enables users to save

the relevant documents and to build a personal work space for storing links for future reference.

Figure 7.2 shows both interfaces, before (7.2a) and after the visualisation as in figure (7.2b).

The implementation of the new interface was based on the General Visualisation Reference Model (GVRM) proposed by Card [1999].

Several models of infovis were proposed to establish a common ground that help researchers and designers to develop and describe the visualisations process, and the transformation from data to a user-visible view. One of the most famous works in infovis that formalising the information visualisation process and explained the overall information visualization process is the "reference model" proposed by Card [1999]. The model is a refinement of the Data State Model described by Chi and Riedl [1998]. The model has been widely used and advocated by many researchers as it assumes that most of the techniques implemented in the infovis systems and toolkits share similar operating steps that can easily be reused.

The reference model, as in figure 7.3, describes three subprocesses from data space to graphic space namely: *data transformations*, *visual mappings* and *view transformations*. Basically, the implementation went through three subprocesses as follows:

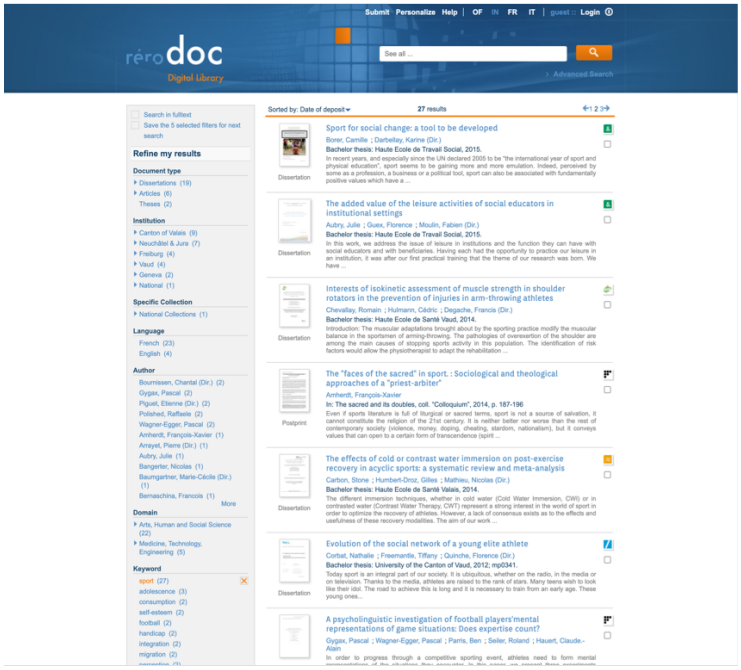
1. *Data Transformations:* As we aim to boost the explorative information behavior, and design interface that might enhance the berrypicking ISB, we decided to visualise the keywords only.

   The data was already processed and indexed by RERO Doc. RERO Doc implemented four indexing systems: RERO indexing , which is used by most of the libraries in the network and covers practically all fields except engineering sciences; MeSH indexing , which is used by medical libraries; and LCSH indexing (in English), which appears in some records imported from other networks [2]. The indexing process enable to search for documents or information by words or phrases.

2. *Visual Mappings:*

   In the second phase, the radial layout technique was chosen and used as a control feature. The radial layout is described by Gaona-García et al. [2017] as one of the visualisation techniques with higher ratings for aesthetics and methods of classification in knowledge representation schemes

[2] https://www.biblio-geneve.ch/matieres/index.php



(a) The Original Interface

(b) The Experimental Interface

Figure 7.2. The Interfaces Before and After Visualisation

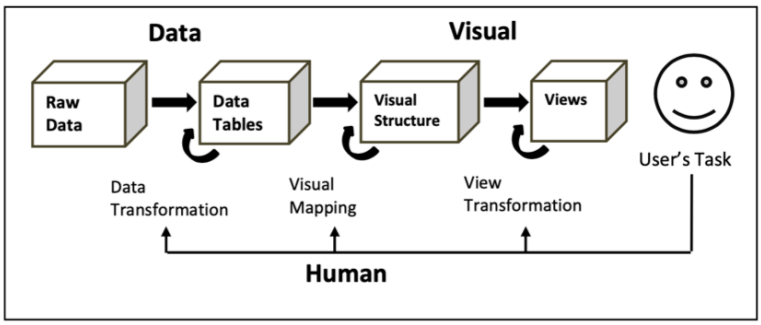


Figure 7.3. General Visualisation Reference Model Card [1999]

along with relation and tree techniques. As an example, the study of Ruotsalo et al. [2013] applied the radar technique.

To present the keywords on the radar layout based on their ranking relevance and number of time the query appeared in the collections, the following formula was applied:

$$\text{x\_coor} = \text{OA}\ \ (1\text{-Y/X})\ \ \cos(\text{Z})$$
$$\text{y\_coor} = \text{OA}\ \ (1\text{-Y/X})\ \ \sin(\text{Z})$$

For example, as figure 7.4 depicts a radius OA = OB where the most relevant keyword is (X), other associated keywords are (Y).

The radial layout allows to organize the keywords in the information space according to their relevance and time appearance on the collections. The keywords in the inner circle represent the most relevant search query, and the surrounding keywords represented the associated queries with the inner keywords which can be placed randomly at any points M, P or N for examples.

3. *View Transformations:* finally, in the third sub-processes *view transformations*, visual structures are transformed into human views, which involves graphic adjustment e.g. adjust the size of the shape, decide about the aesthetics aspects i.e. color, and the positions of the other functions i.e. the query history, and the save article functions.

After reviewing the literature about visualised interfaces, the decision of adopting the *Radar* technique was based on the following reasons: (i) Radial visualization will allow users to deploy techniques of berrypicking in the search process Bates [1989], where the results can be refined with new suggestions. This is

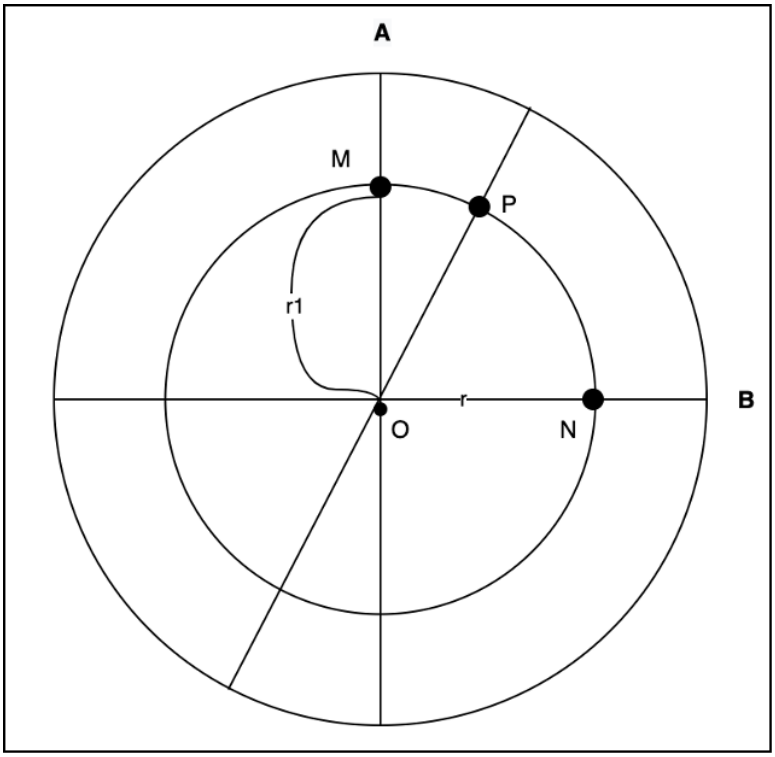


Figure 7.4. The Radial Layout

to overcome the limitation of the classic DL design, facet-based systems, where the navigation options allow user to access only very narrowly defined content based on user's initial query which might force repeating typed ad-hoc queries to explore beyond the initial query scope. (ii) it supports a serendipity-fostering environment. Empirically grounded evidence is to found in McCay-Peet and Toms [2011] who proposes four dimensions of design to foster serendipity, all of them are supported in the new design as follow:

1. The interface design *enable connections* between content, and this is promoted by the unique layout of the radar along with the position of the labels.

2. *Support exploration and presenting a variety* of information; the radar function helps the users by letting them explore the most and the least related topics to their information needs.

3. *Trigger divergence* by capturing user's attention and initiating divergent thinking by coloring the labels, and distributing them around the center, and

4. *Induce curiosity* by incorporating in the new design the query history and saved article functionality. These functions could trigger curiosity by encouraging the users to browse, stop, and assess found resources during a search session.

In summary, as the information space boosts the information seeking experience, we believe that the interactive radar could support different types of ISB including the convergent and divergent information behaviour.

7.2.3 Design of the Evaluation Study

Due to the COVID-19 international pandemic circumstances, an online user study was conducted by sending email invitations to postgraduate and undergraduate students. The participants came from three different departments: 8 were from Computer Science, 9 were from Information Science, and 5 were from Business school.

7.2.4 The Task

The evaluation of the new interface is based on a task-based approach which is recognised by the HCI community as an important tool in user interface design

Topics	Examples
Machine learning	*'I'd like to find more about using machine learning techniques to improve result ranking'* *'I would like to know more about applications and results of applying topic models on patents, what kind of algorithms exist, what tools are available for the same, if there are any open source implementation of those models etc.'*
Interactive application	*'The Applications of Voice Assistants in Health Care sector, I need to find more about the privacy issues, credibility and their design'* *'I'm looking for more resources for designing educational platforms'*
Query expansion	*'I need to find more about the latest models for citation analysis and mining'* *'I'm interested in finding more about improving the performance of the health information retrieval systems techniques, how PubMed thesaurus are indexing ..etc'*
Information seeking	*'I need to look for the current web-based therapies applications, I'm more interested in elderly habits when using such applications'* *'Are there any established information seeking models that help in designing kids interactive applications'*

Table 7.1. Examples of the Participants' Information Needs

Meyyappan et al. [2004]. Basically, the task-based approach takes the existing work related tasks performed by the users as the starting point of the design process as these contribute to a deeper understanding of users' needs. In our previous user study in chapter 4, we asked the participants to use RERO Doc DL to search for their own personal information needs and questioned them to give a description of the information that they would like to search for as follows:

> "*Suppose you are writing a paper on a topic of your interests and you wish to gather more information to enhance your existing knowledge on a topic or discover new knowledge. Kindly use RERO Doc digital library as the information source.*"

We used those self-generated tasks, table 7.1 shows examples of the participants' information needs, to generate a user-tailored task for this study. Accordingly, the four most frequent topics were extracted and proposed to the participants as imposed queries, that is information seeking externally imposed rather than self-generated by the users Gross [1999]. The imposed queries related to: *machine learning*, *query expansion*, *interactive application*, and *information seeking*.

The participants were asked to choose two imposed queries according to their interests, and they were encouraged to explore each of those topics by identifying three subtopics related to their interests. As the participants started exploring the topics, they were asked to add relevant documents to their personal list i.e the *saved articles* function. Once the participants finished their tasks, they were asked to fill up a post-experiment questionnaire.

7.2.5 Instruments

Both subjective and objective data were examined. Two different data sources were utilized; log file (LF) analysis and post-experiment questionnaire. The LF was analysed with the aim of understand to what extent the participants depended on the radar function and/or on the query-search function. The post-experiment questionnaire was used to collect the perceptions of the participants after the experiment. The questionnaire was divided into two sections: *task-level* part that measured search task performance, and *test-level* part to quantify users' overall impression of the usability and experience with the visualised DL.

Because usability, that deals with the functionality of the system and whether the system is usable, is not the only aspect that impacts users' experiences, a

Aspects	Metrics	Questions	Source	Level
Interaction	Satisfaction	1.Overall, I'm satisfied with the ease of completing the tasks in this scenario 2.Overall, I'm satisfied with the amount of time to complete the tasks in this scenario	ASQ	Task
	Performance	3.How successful were you in accomplishing what you were asked to do?	NASA-TLX	
	Effort	4.How hard did you have to work to accomplish your level of performance?		
	Frustration	5.How insecure, discourage, irritated, stressed, and annoyed were you in completing this task?		
Appeal	Aesthetics	6. I found the website to be attractive. 7. The website has a clean and simple presentation.	SUPR-Q	Test
	Fun	8. I found interacting with the radar interface is enjoyable	Researchers	
Impact	Utility	9. Comparing to the traditional digital library interface, this interface is useful 10. The radar interface enhances my searching experience 11. I found the query history feature is a useful feature		
Usability	Value	I think that I would like to use this system frequently. I found the system unnecessarily complex. I thought the system was easy to use. I think that I would need the support of a technical person to be able to use this system. I found the various functions in this system were well integrated. I thought there was too much inconsistency in this system. I would imagine that most people would learn to use this system very quickly. I found the system very cumbersome to use. I felt very confident using the system. I needed to learn a lot of things before I could get going with this system.	SUS	

Table 7.2. UX Evaluation Metrics and Questions

combination of standard questionnaires were utilized to measure different aspects of the UX. Thus, the evaluation considered four aspects: interaction, appeal, impact, and usability. Each one of those aspects included different metrics. Table 7.2 summaries the metrics along with the questions and their sources.

7.3 Results

Twenty-two participants succeed to complete the experiment. Although the number of the participants is comparatively low considering the study nature i.e. online, the sample size is very close to the number of the participants (20) that was reported in similar study e.g. Ruotsalo et al. [2013] study. The outcome of the results are summarized as follows.

7.3.1 Task-level Evaluation

For the task-level evaluation, After-Scenario Questionnaire (ASQ) and NASA-TLX Task Load Index were used. ASQ designed by Lewis [1991] used to quantify perceived ease and satisfaction, where NASA-TLX created by NASA Hart and Staveland [1988], as an assessment tool for measuring subjective workload of interfaces.

Regarding the satisfaction level, as it is indicated in table 7.3, the items were combined into a single scale through averaging to allow for easier interpretation of results. The cumulative average of ASQ statements was 3.54 with SD 1.14 which lies in the interval [3.40-4.19], that is equal to "agree" on the five Likert scale. That means that most of the participants were satisfied with the visualise interface to accomplish the task, and found the interface usable in terms of time completion and ease of use.

Regarding the task performance, effort, and frustration, the means of the three metrics were 2.36 with SD 0.95, 3.46 with SD 1.04, and 2.18 with SD 0.9, respectively. This indicates a success level with the task performance, an easy level of task effort, and low frustration level while conducting the task. Overall, most of the participants were satisfied with the ease of interacting with the visualised interface, and showed a positive attitude towards it.

7.3.2 Test-level Evaluation

The test level evaluation was measured by considering three aspects; appeal, impact, and usability. The two questions about appearance were taken from

Statement	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	**Mean/SD**	CM/SD
1	3 13.6%	3 13.6%	1 4.5%	12 54.5%	3 13.6%	3.41/ 1.29	3.54/1.14
2	1 4.5%	3 13.6%	4 18.2%	8 36.4%	6 27.3%	3.68/1.17	Agree
3	Very Successful	Successful	Success to some extent	Failed	Completely fail		
	4 18.2%	9 40.9%	6 27.3%	3 13.6%	0	2.36/ 0.95	Successful
4	Extremely difficult	Difficult	Neither easy nor difficult	Easy	Extremely easy		
	0	0	7 31.8%	11 50%	2 9.10%	3.46/ 1.04	Easy
5	Very low	Low	About the same	High	Very high		
	5 22.7%	10 45.5%	5 22.7%	2 9.1%	0	2.18/ 0.9	Low

Table 7.3. Task-level Evaluation

Statement	Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	**Mean/SD**	CM/SD
6	1 4.5%	3 13.6%	4 18.2%	11 50%	3 13.6%	3.55/1.05	3.73/0.85
7	0	2 9.1%	4 18.2%	12 54.5%	4 18.2%	3.82/.85	Agree
8	1 4.5%	2 9.1%	4 18.2%	8 36.40%	7 31.8%	3.82/ 1.14	
9	0	1 4.5%	2 9.1%	9 40.9%	10 45.5%	4.27/.82	4.08/0.74
10	1 4.5%	3 13.6%	3 13.6%	10 45.5%	5 22.7%	3.68/1.12	Agree
11	0	0	3 13.6%	10 45.5%	9 40.9%	4.27/0.70	

Table 7.4. Test-level Evaluation

SUPR-Q Sauro [2015], Standardized User Experience Percentile Rank Questionnaire which consists of eight items to measure four aspects of the quality of the website user experience, these are: usability, credibility, loyalty and appearance. In our study, only the appearance questions were adopted.

Appeal and Impact:

Table 7.4 explains the participants' perceptions towards the appeal and impact. The cumulative mean was 3.73 with SD 0.85 of the appeal includes the aesthetics and fun which indicates that most of the participants had an enjoyable experience, and agreed that the new interface is interactive and has clean and simple presentation.

In terms of the impact, the participants agreed that the new features; the radar, query history, and the saved functions enhanced their searching experiences comparing to the traditional digital library interface with overall mean 4.08 and 0.74 SD.

Usability:

To measure usability, as a cornerstone procedure of interface design, System Usability Scale (SUS) Sauro [2020] was used. The analysis of SUS was done by converting the five-likert scale points into 100 score to get a general score that can be compared to the standards degrees (excellent, good, okay, poor, awful). Figure 7.5 shows SUS scale, and table 7.5 presents the scores along with the adjective ratings where the highest SUS Score is 100, and the average is 68.

The score of SUS in our experiment is 80.1 which represents B grade and Good as the adjective ratings. Figure 7.6 indicates participants' rating for the 10 dimensions of SUS. Overall, the participants showed a more positive attitude towards the system being user friendly, easy, well integrated, learn quickly, and confident to use, while they did not agree on any of the negative attributes such as complex, in need of support, inconsistency, cumbersome, and need to learn a lot to use.

7.3.3 Log File Analysis

Along with the subjective data drawn from the post-experiment questionnaire, we analysed the LF with the aim of quantifying the interactions, investigating the utility of the radar function, and examining the number of nodes or paths accessed by the participants. Such analysis helped us to answer RQ2: *To what*

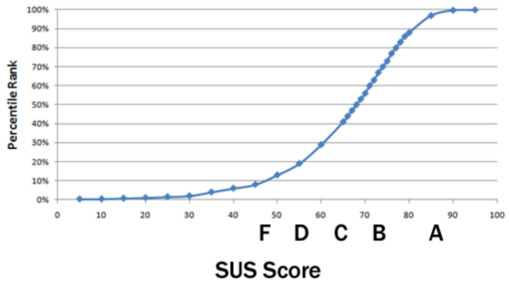


Figure 7.5. SUS Scale Sauro [2020]

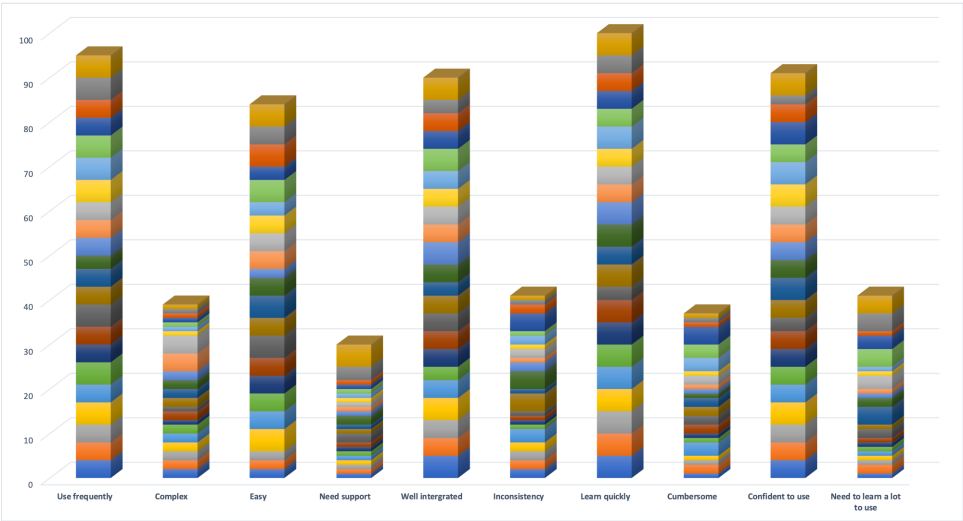


Figure 7.6. Participants' Perceptions of SUS Dimensions

SUS Score	Grade	Adjective Rating
>80.3	A	Excellent
68-80.3	B	Good
68	C	Okay
51-68	D	Poor
<51	F	Awful

Table 7.5. SUS Scores and the Interpretation Sauro [2020]

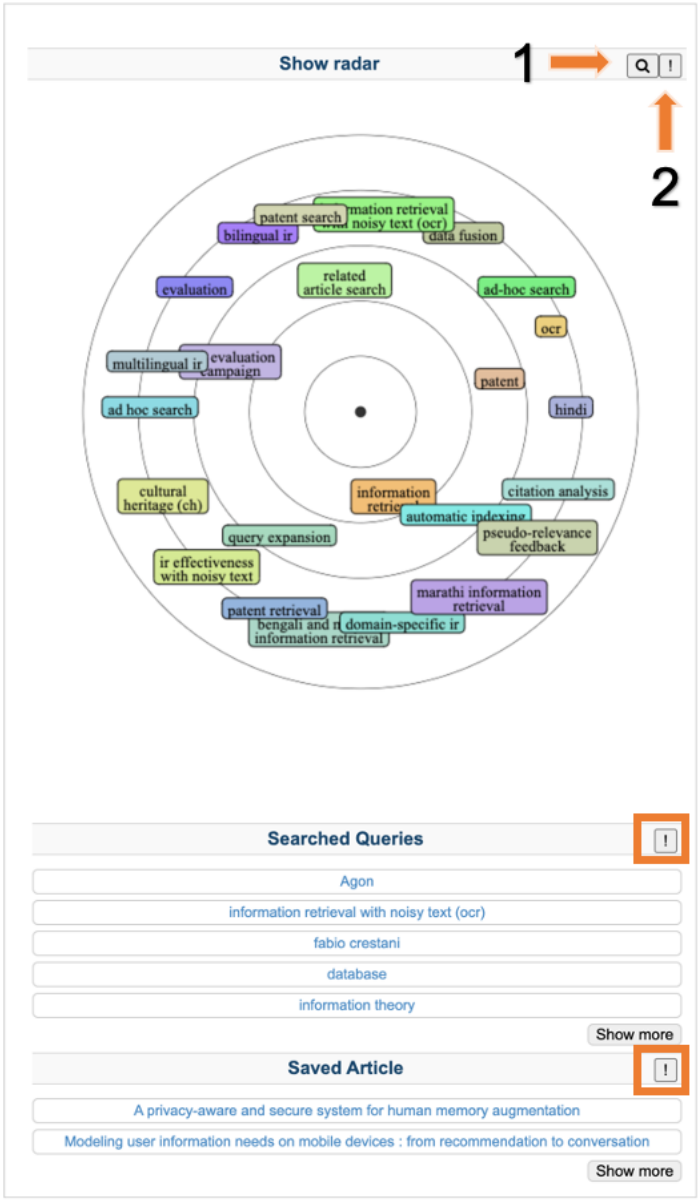


Figure 7.7. The Interface After the Modification

Metrics	Score
Session duration (in average)	10 minutes
Number of nods (in average)	4 nods
Saved articles (in average)	3 articles
# of sessions using the search function	10 sessions
# of sessions utilizing the radar functions	34 sessions

Table 7.6. Metrics Used to Quantify the Interactions

extent the new visualised interface can be adopted and understood by users?. Table 7.6 shows the metrics we used to quantify the interaction.

We found that the participants spent almost 10 minutes in average to conduct the task that consists of two exploratory sub-tasks, as previously mentioned in section 3.3. Interestingly, although we encouraged the participants to explore each topic by identifying only three subtopics, we found that the average nodes of the subtopics were made of four topics. This might indicate that participants were engaged and kept exploring the topics.

In terms of how used the radar function was, we found that the majority of the participants, 80% utilized the radar function more frequently in their search sessions compared to the search function.

The main reason behind implementing the imposed queries method, as it was explained in section 3.3, was to have a baseline that enabled us to detect the participants' interests and their search paths. Eventually we could examine the effectiveness of the visualised interface in supporting the divergent or the explorative behaviour by examining the number of nodes accessed by a participant, and considering the variety of the paths of the same query. Three examples of each imposed query are exported from the LF and presented below to show the diversity of the paths.

Machine learning →*assistive technology* →*facial expressions*→ *deep learning*

Machine learning→*pattern recognition*→*patent retrieval*→ *corpus linguistics*→*text categorization*→*classification*

Machine learning→*opinion detection*→*sentiment analysis*→ *interactive topic detection and tracking (iTDT)*→*natural language processing*

Interactive application→*authoring tool*→*visual programming* → *computer vision*→ *visual feedback*

Interactive application→*augmented reality*→ *computer assisted surgery* → *computer designed implants*

Interactive application→ *audience response systems*→ *question practice*→ *an-*

swer depth format

Query expansion→*open-domain IR*→*information interaction*→ *3D graphics*

Query expansion→ *pseudo-relevance feedback*→*data mining*

Query expansion→*query formulation*→*information retrieval*→ *Fabio Crestani*

Information seeking→*information services*→ *help-seeking behaviour*→ *information-seeking conversations*

Information seeking→*information theory*→*collaborative work* → *knowledge production*

Information seeking→*context-aware recommendation*→
location-based services →*mobile information retrieval*

7.3.4 Challenges and Solutions

The participants were also asked to reflect on their experiences by answering open questions: *How was your overall experience with the digital library?, if we could improve it, what would you like to add or modify?* We classified the participants' suggestions and challenges into three categories; usability issues, implementations, and content coverage. Table 7.7 presents examples along with the changes made to address them.

7.4 Discussion

Due to the nature of the human mind, an overview of high-resolution display environments is recommended to support users in their information discovery Shen et al. [2019]. Therefore, in this study we redesigned the traditional DL interface with the aim of evaluating the impact of the interactive visualised interface on the UX with the DL users. Bargas-Avila and Hornbæk [2011] identified two aspects of the user experience: the usability and the hedonic qualities that assess affective aspects, to include emotion, engagement, aesthetics, and enjoyment.

To answer the RQ1: *How do users perceive a visualised interface for a DL?* a post-experiment questionnaire was used to measure the subjective perceptions of the participants. As Toms et al. [2004] stated that DL users are not only seeking information while interacting with IR and DL systems, but they are enjoying the process. Thus, instead of assessing the system performance alone, different aspects of UX were considered too, these included interaction, appeal, impact, and usability. By taking into consideration all these UX elements we built a fuller picture of the search experience, in comparison with what we achieved in our previous studies Barifah et al. [2020] and *(another paper accepted at ISIC 2020*

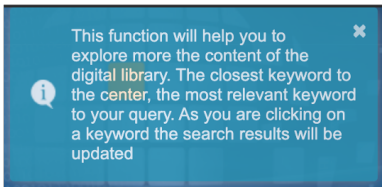
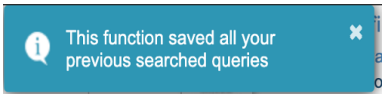
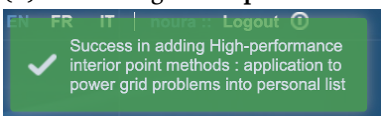
Challenges	Examples	How we responded
Usability issues	*"I think the interface should have some tips to make the experience more pleasant"* *"The system should have provided me with some information about whether the topic's color and position were arbitrary or not. I just assumed as a user that they must mean something, but the system neither confirmed nor denied."* *"Some functions were not obvious. For example, it was not obvious that the list of terms under the radar was my search history"* *"I just had problem adding to my personal list for the first time, there was no feedback if the action was completed or not"*	Pop up windows were used to provide explanation for unclear functions, these would appear by clicking on "!" labelled buttons, as in figure 7.7. (1) For the **radar** function  (2) **Searched queries**;  (3) For the **saved articles**;  (4) For adding to the **personal** list; 
Implementation	*"Sometimes there is an overlap between words on the left of the screen which make it hard to read. Otherwise, I liked it!"*	1-Decrease number of displayed labels from 30 to 25 2- Add a magnifier icon to support different screen resolutions
Content Coverage	*"The search result obtained with my two queries were not so long"* *"The results I found are very few, and that the related subtopics are somewhat unrelated to the topic"*	Feedback to RERO Doc about content coverage

Table 7.7. Participants' Suggestions and Solutions

but not published yet), and we could see how the interactive visualised interface influenced positively the UX. This was visible in the positive gauging of the participants' satisfaction, together with experiential values that invoke low level of frustrations and high enjoyment. This is in line with other studies e.g. Gaona-García et al. [2016] who confirmed the benefits of the visualised interfaces to enhance the exploratory search that might improve the chance to discover new areas as well as supporting the information serendipity.

This was also noticed in the LF analysis, where different participants explored the topics in different directions based on their interests. The radar technique helped the participants to explore different semantic relationships between the documents retrieved and other related topics. This is in line with Bauer [2014] who confirmed that the visualization techniques facilitate the searching process by providing hidden information in more rapid search times.

In terms of generating or reformulating queries, we found that most of the participants used the radar function as a mean to generate related queries. This is in line with the findings of Choo et al. [2013] who confirmed the usefulness of the visualised interface in generating useful keyword sequences regardless of the level of user expertise.

Regarding RQ2: *To what extent the new visualised interface can be adopted and understood by users?* Our aim was to investigate how intuitively usable the new interface was, according to the definition provided by Naumann et al. [2007] where "the particular user is able to interact effectively, not-consciously using previous knowledge". The intuitive interface is characterised by being easy to use and learn, requires little to no previous knowledge, and just feels "natural" in its use. The intuitive use largely depended on the users as it emerges through interaction, and takes place with a minimum of mental effort. In summary, interfaces that tap into lower knowledge levels are more intuitive to use Stößel et al. [2009].

By analysing the LFs, we found that the majority of the participants 80% utilized the radar function more frequently in their search sessions comparing to the search functions. This might indicate that the visualised interface is a promising solution that drew the attention of the users, increased their curiosity, and promoted resource exploration processes.

Although SUS score was above 80 which indicates a good usability, some of the participants reported challenges faced when they interacted with the new interface. Basically, three types of challenges were reported by the participants as in table 7.7. Those concerned: usability issues, implementation problems, and content shortage coverage.

The main reported challenge, not surprisingly, was usability this is in line

with Chen [2005] where usability was identified as the first problem among the top 10 unsolved infovis problems. Recently, Shen et al. [2019] confirmed the shortage of the usability studies in the infovis field comparing to the infovis methods and techniques. In our study, the main problems with the usability were reported in the labelling and explanation of the new functions. For example, a comment by one of the participants was "*the system should have provided me with some information about whether the topic's color and position were arbitrary or not*". This issue was confirmed by Chen [2005] who stated that a major obstacle in the usability of the infovis is the lack of users' understanding of the underlying analytic process of raw data and its representation in the interface i.e. the relation and logic.

Thus, we responded to those issues by adding pop up window to provide explanation as explained in table 7.7. For the implementation problem, some of the participants found that the resolution of the interface is not fitting all possible screens, and this issue was reported by other studies as mentioned in the systematic review of Shen et al. [2019]. Therefore, to respond to this challenge, a magnifier icon was added to allow users to expand the radar window. The shortage coverage of the DL collections is a well known challenge face most of the DLs Trapido [2016].

7.5 Conclusion and Limitation

In conclusion, this study applied the radar technique in the DL platform, and evaluated the experience of users who interacted with the visualised interface while performing search tasks. We compared these findings with those emerging from a related study where a comparable sample of users was engaged via similar type of tasks with a traditional DL interface. Thus, we observed how users demonstrated overall a more positive attitude towards the visualised interface. Still, there is more research to be done in this area, and we will start from addressing some of the limitations of the user study here described.

The main limitation lies in the use of the simulated work task that might hide the real effect of the visualised interface, thus in the future, it will be interested to consider a more naturalistic tasks where users search for their own information needs. Moreover, the evaluation was conducted for short session searches, thus, there is a need to investigate the effectiveness of the visualised interface over a longer period of time by conducting a longitudinal study.

Besides, the application of the radar technique to the design of an interface for a DL cannot resolve all of the issues inherent in the search process and its

complexity. As a suggestion for future work, other parts of the DL can be visualised, e.g. co-author recommendation and some hidden filtering functions that are supposed to help users in sorting the results.

Finally, it is important to distinguish between two different types of infovis evaluations: one looking at the usefulness and usability of visualization components, and another focusing on the level of complexity of the perceptual-cognitive task undertaken by users when making sense of the process and activities behind their interaction with the visualised interface. A further investigation is necessary, in order to measure the cognitive load and effort required for the users to understand the underlying analytic process involved in most infovis systems. Literature agrees this is a major obstacle for users as they fail to see how raw data are magically turned into colorful images. A more in depth user study, possibly supported by user data collected with eye tracking technology could help us shade some light on this still unexplored research area.

Part V

Conclusion

Chapter 8

Conclusion and Future Work

This part consists of two sections: summary of the thesis, and some suggestions for possible directions for future work.

8.1 Summary of the Thesis

DL systems are extensively used to facilitate access to collections of digital resources. The content of the DL is organized and structured by using the classification schema that is based on the description of contents' metadata. Such systems provide a vertical list of results along with the clickable labels and facets that are supposed to facilitate information locations by filtering different criteria such as author, year of publication, journal or publisher and subject.

The architectural elements of the traditional representation of contents in DL systems are criticized by different researchers for two main reasons: firstly, it does not allow the user to deploy techniques of berrypicking in the search process and does not support other ISB than the goal-directed behaviour. Secondly, the knowledge organisation schemes adopted to build DL systems were created by experts in resource classification, this results in inadequate use of the search tools i.e. facets and other DL functions by the conventional users who are not familiar with the purpose of such techniques.

To overcome these challenges many researchers recommended on one side to boost the information literacy of the users, and on the other, to redesign the DL interfaces. Thus, we decided to take a step further and redesign the current DL information space.

The decision of redesigning the interface was informed by empirical evidence based on subjective and objective data collected over a series of studies. This

study is an exploratory research which aimed to discover more about ISB in DL, and proposes a visualised interface as an alternative to the facet-based one.

In **chapter 3**, this research started by exploring the usage patterns of a large scale DL with heterogeneous users. This was done by conducting a log file analysis. Discovering information from LFs is not a trivial process; it requires a strategic plan to gain a better understanding of the hidden information.

In this dissertation, as a contribution, we defined a framework for analysing DL log files to help and encourage DL administrators, information science researchers and students to conduct LF analysis. Initially, the author explored two questions: **RQ(1)** *What are the potential UPs of a large-scale DL?*, and **RQ(2)** *What are the main features that characterise the above UPs?*.

Three main UPs were identified: item seekers, navigators and searchers. Within those main patterns, sub-patterns were recognised. UPs can be defined by four main features: session starting points, content discovering actions, types of functions used (if any), termination actions, and session duration. Identifying the features as in the second research question helped us to conclude that regardless of the UP type, the DL functions i.e. search tools were not fully utilized across the all users.

Results from **chapter 3**, helped us to identify aspects to be further investigated via user studies. One of those aspects is the considerable variation in using the system functions (e.g. facets). This leads us to investigate deeply the real usage of the DL functions; i.e. to what extent the DL's search tools are used?

Researchers believed that facets-based DL interfaces should assist users to achieve higher task accuracy and satisfaction, as the facets and other DL search functions are supposed to facilitate information location.

Accordingly, in **chapter 4**, we started our investigation by running an online study with real users in their natural settings, taking into consideration their level of expertise in using the DL. Without declaring the real aim of the study, participants were asked to use the digital library and run one specific and one exploratory search task, using their own topics. Accordingly, different outcomes for both tasks were recorded i.e. success and failure searches. Our ultimate aim was to explore to what extent the DL's search tools help searches to locate the information sources. Thus, only the failed searches were considered in the analysis. As we were concerned with identifying the failed searches from the users' perspectives, participants' answered pre- and post-questionnaires.

Very little research has looked at failed searches, specially from the user perspective. Thus, this thesis contributes to research by explicitly investigating what is the user perception of failed searches.

A combination of subjective data extracted from the post-questionnaire, and

objective data from the behavioural analysis of the interactions recorded in LF was used to answer **RQ (3)** *what are the reasons behind search failures from the searchers' perspective?*

Users identified the lack of coverage and poor usability as main causes behind failed searches. We examined search behaviour along with declared perception of the causes behind failures and realised that digital library's functionalities were not fully used. Poor awareness of DL functionalities could be an unreported cause for search failures.

A large body of research suggests that experiencing failure has marked emotional and psychological consequences across a range of individuals and settings. Thus, we also were interested to explore the associated emotions with the failed searches. For this reason, and in order to answer **RQ (4)** *what are the primary emotions individual experienced in the failed searches?*, a textual analysis was conducted by using the Linguistic Inquiry and Word Count (LIWC) software.

Sixty-four excerpts reporting search failures were extracted from the answers provided by the participants in the open-questions. We asked the participants to reflect on their experiences by answering the following opened questions: (1) *If you are not satisfied with your searching, can you please explain why?*, and (2) *what do you think of your overall experience with the digital library?*. In general, users had a positive attitude toward the digital library, expressing trust, joy, and anticipation. Anger and sadness were linked specifically to failed searches.

We concluded that regardless of the expertise levels, DL users are still encountering the same search failure problems as reported in literature almost 20 years ago. Thus, there is a need to improve DL systems and simplify their interfaces.

The result of the study in **chapter 4** motivated us to redesign the current DL interface by following the user-centered design approach. The user-centered design approach requires a deep understanding of the UX before the processing of the redesign.

Therefore, in **chapter 5** a controlled laboratory study was conducted to validate the results of the online study reported in **chapter 4** and to explore the use of a different method to collect more affective data.

UX is mainly based on ethnographic methodologies that aim to explore the interactions of the real users along with other user attributes for example, experience level, attitude, and affective variables including emotion and mood. This study contributes to the DL sector by investigating a neglected aspect of the DL evaluation; assessing the UX. As far as we know, there is no a holistic framework that can assist the librarians and the DL system designers to evaluate the UX in the DL sector. Thus, this study contributes to the field by adapting Mahlke's interaction user experience framework. The framework helps in designing the

research by: (a) identifying the factors that affect the UX: system properties, user characteristics, and context metrics. And (b) defining the main UX components: users' perceptions of the instrumental and non-instrumental quality, and the affective variables.

A primer objective of the UX study was to confirm the results presented in **chapter 4** which indicated that the DL functions were rarely used regardless of the expertise levels. Therefore, we included **RQ (5)** *to what extent the DL functions are usable by users with different expertise?* Overall, our results confirmed that the DL functions were not fully used as they were supposed to be regardless of the expertise level.

The subjective perceptions of the participants were also considered in order to understand if the users with different expertise levels shared the same perception of the DL; **RQ (6)** *will the advanced participants share similar perceptions of the DL compared to the non-advanced participants?* With the help of the honeycomb framework, the results showed that both groups disclosed similar attitude towards five dimensions out of seven dimensions of the honeycomb framework. Three out of five dimensions received a positive response, those are: the usefulness of the DL, the desirability, and the credibility.

The participants also showed similar attitude to the findability and value aspects. The findability inspects whether the users can find what they are looking for, and the value examines the overall satisfaction levels of the users with a system. Both groups were neither agree-nor disagree that they can locate what they are looking for easily (findability). This was also reflected in the searching behavior where the advanced participants kept scrolling under the first 10 results, with rarely using the search tools. This is also related to the participants perception of the (value) dimension of the DL when it was compared to the other academic search engines for example, Google Scholar, in terms of the difficulty and time requirements. The intermediary answer between disagreement and agreement that the DL is valuable could reveal the preference of the simplicity of search engines over the complexity.

The other two dimensions where the two groups disclosed different perceptions are: the usability and the accessibility dimensions. The advanced participants showed intermediary attitude toward the usability of the DL comparing to the negative attitude of the non-advanced participants. As none of the groups were agreed that the DL is usable, this might explain the light usage of the search tools and the usability issue of the DL. This also is applicable to the accessibility dimension where the non-advanced participants faced problems to access the content or find the download icons. Despite the rationale of implementing the search tools within the DL interfaces, usability and accessibility reported in pre-

vious works are still existed.

The third aim of the UX study was to investigate if the moods were affected by the nature of the interactions and the search outcome, thus we examined **RQ (7)** *What are the associated moods before and after the search experience in the DL?*, if so what are the main factors behind such changes: **RQ (8)** *What are the influential factors that might affect the moods before and after the search experience?*. The "pick-a-mood" scale was selected as a mood tracker before and after the experiment.

The results revealed two important findings, (a) the number of the pleasant moods increased after the experiment, compared to the unpleasant moods. And (b) the influential factors that affected the moods changed from being personal, contextual, and educational issues to be more search-oriented issues.

In conclusion, the interaction user experience framework can be used as an assessment tool to better design UX experiments. However, the framework can be enhanced by including more practical tools e.g. Pick-a-mood scale and honeycomb model.

So far, in this research the interactions with the DLs were examined implicitly by analysing the LFs, and explicitly by conducting online and laboratory user studies. In order to confirm or deny the commonality between the findings of the studies presented in the previous chapters, a large-scale attitudinal survey was distributed, as in in **chapter 6**, to collect more data of the users' perspectives by considering a new DL with different population. Thus, the same questionnaire developed in **chapter 5** was used in this attitudinal survey presented to answer **RQ (9)** *how would user perceive SDL?* Regardless of the expertise levels, both groups revealed similar perspectives towards four aspects of the honeycomb aspects; desirability, findability, accessibility, and credibility. The participants showed different attitudes towards the other three aspects; usefulness, usability, and value of the DL.

By comparing the results of the two studies i.e. the laboratory study in **chapter 5**, and the attitudinal survey in **chapter 6**, we could confirm that regardless of the populations, there are some aspects that received similar responses for both DLs. Participants in both studies confirmed that the DLs are considered as a credible source of information and scholarly materials. To maintain and boost this value, DL administrators need to pay attention to other UX fundamental aspects, particularly the usability and findability aspects.

RQ (10): *what are the commonality between the studies?* aims to summarize the findings of the studies reported in the previous chapters.

In a nutshell, users did not take full advantage of the search tools, and the expertise level does not have an effect on using the DL's functionalities. The need

to improve the DL system and simplify its interface still persists.

Users' search behavior and cognitive processes can be significantly influenced by the type of search tools and interfaces they interact with. Search tools traditionally available in the digital libraries that are based on the classification, and they are specifically support users with goal-directed information needs, while the divergent information needs are not catered for. Several studies examined the ISB of the DL users pointed out the need for the redesign of available DL interfaces.

From our user studies, we argue that there is a need to provide alternative modes of access to information than the traditional facet-based DL interfaces. Therefore, one contribution of this thesis is to propose the interactive visualisation interface and evaluate its effect as in **chapter 7**. Mainly, we examined to what extent a radar search tool and its visualised DL interface can be adopted by users. The decision of adopting the visualised interface approach was based on the fact that visual search interface has the possibility of increasing user retention via the use of design aspects, once the visualisation is implemented accurately. Thus, to answer **RQ (11)** *How do users perceive a visualised interface for a DL?*, an online evaluation study was run.

A post-experiment questionnaire was used to measure the subjective perceptions of the participants. The questionnaire consists of different aspects of UX, these included interaction, appeal, impact, and usability

The prototypes we have built showed that the interactive visualised interface influenced positively the UX. This was visible in the positive gauging of the participants' satisfaction, together with experiential values that invoke low level of frustrations and high enjoyment.

The last research question of this thesis was **RQ (12)** *To what extent the new visualised interface can be adopted and understood by users?* Our aim was to investigate how intuitively usable the new interface was.

By analysing the LFs, we found that the majority of the participants utilized the radar function more frequently in their search session compared to the search function solely. This might indicate that the visualised interface is a promising solution that drew the attention of the users, increased their curiosity, and promoted resource exploration processes.

Nevertheless, this thesis has some limitations that can be summarized as follows: firstly, in the analysis of the LF as in chapter 3, we aimed to gain a general overview of the user interactions. Thus, the analysis excluded the fine-grained analysis covering: the semantic or query analysis, and the sequential behaviour of the information patterns which can be detected by applying Markov chain techniques. Secondly, a cornerstone of this study is to evaluate the interactions

of the real users in their own natural settings. Without experimental incentives or rewards as in chapter 4, the number of the participants were comparatively low. This low sample size was one of the hinders of this study to generate statistical result that can be generalized to a larger population. Another limitation concerns the information needs used in the evaluation study in chapter 7 was not linked to the participants' personal information needs, but instead it was a user-tailored task that was generated from our previous study presented in chapter 4.

Thirdly, the evaluation of the visualised interface in chapter 7 excluded the perceptual-cognitive load of the users when making sense of the process and activities behind the visualised process i.e. how the user behaves and reacts to the representative visualisation and understands the relationship behind the data visualisation. Such investigation required more precise data tools e.g. eye tracking.

The outcomes of this research suggested various possible future directions that need further investigation as in the following section.

8.2 Future Works

Digital library systems succeeded in the last decade to provide unprecedented access to library content that were confined to their physical spaces. However, with the rapid development in the technology and web sectors, access cannot remain the only function of the digital library systems. Instead DL designers need to design innovative information spaces that enable scholarly functions beyond the traditional focus. Recently, DLs have begun to give some attention to creating a more visual space, ACM [1] as an example. Visualisations might not be the only solution for handling the inherent complexity of DL systems. Thus, some of the possible future directions are suggested below.

8.2.1 Technical Development:

- The application of the radar technique to the design of an interface for a DL cannot resolve all of the issues inherent in the search process and its complexity. As a suggestion for future work, other parts of the DL can be visualised, e.g. co-author recommendation and some hidden filtering functions that are supposed to help users in sorting the results.

[1] https://dl.acm.org/

- A well known challenge faced the DLs is the issue of the information resources building which is strongly related to the openness and access restrictions by the digital content providers. This challenge affects the creation and innovation of the information space of the DL systems as Harkema [2015] stated that:

> "*The construction of a space like this requires openness – an open source, open data, open everything environment that simply may not be possible in the current digital economy.*" Harkema [2015] (p.10)

The author believes that with the growing development in the emerging technologies domain, solutions for the openness issues might be provided.

An example of an emerging technology whose applications can be beneficial for the DL sector is the blockchain technology. As the blockchain technology can be used to store information in a distributed and tamper-resistant setting, Hoy [2017] suggested that blochchain technology might be used to create timestamped, verifiable versions of journal articles. Blochchain technology also can be adopted to develop better digital rights management (DRM) tools. For instance, to avoid copying of their content, publishers have placed draconian, often unworkable DRM tools on libraries and consumers.Thus, with the help of the blockchain, a unique verifiable record can be created to allow digital materials to be uniquely identified, controlled, and transferred. Accordingly, the publishers could be ensured that no copies were being made. This solution might contribute to the openness issue by decreasing the subscription prices Hoy [2017].

As blockchain technology is opening new opportunities, it also brings limitations and problems, one of them is the lack of unified standards for managing metadata, process, and others Zhang [2019]. Such challenges might open new research directions that need to be addressed. Currently, the author is not aware of any libraries' project that adopt the blockchain technology, probably in the future, this technology will be more visible in libraries sector. Thus, to exploit this technology, further studies will be required to develop new information architecture of the DL systems and interface design beyond the traditional DL interfaces.

8.2.2 Methodological Development:

- The ISB is affected by different individual factors. In this research, due to the complexity of human experience, only one aspect of the user characteristics was considered which is the expertise level of using the DL systems. The decision of examining this variable was based on the related works which confirmed that the current design of the DL requires adequate knowledge of information retrieval mechanism.

  However, in the future works, it could also be interesting to consider locus of control (LoC) as an influential individual factors in ISB. LoC is a personality variable in psychology that refers to one's perception of control over life events. According to the theory of Rotter [1966], individuals fall into two major categories: *internals* who believe that outcomes depend primarily on their own efforts, or *externals* who strongly believed that the outcomes determined by forces beyond their personal control e.g. the design of the interface, and the available search tools.

  To the best knowledge of the author, LoC has not yet received much attention from the researchers as a significant factor either in the UX studies or ISB studies in the DL domain. Thus, it might be interesting to examine whether LoC influences the search for information process, or it might have an affect on the UX results.

- The evaluation of the visualised interface was conducted online due to the COVID-19 international pandemic circumstances. As a future work, an A/B testing between the facet-based interface and the interactive interface can be run as a comparative study.

- It is important to distinguish between two different types of infovis evaluations: one looking at the usefulness and usability of visualization components, and another focusing on the level of complexity of the perceptual-cognitive task undertaken by users when making sense of the process and activities behind their interaction with the visualised interface.

  A further investigation is necessary, in order to measure the cognitive load and effort required for the users to understand the underlying analytic process involved in most infovis systems. Literature agrees this is a major obstacle for users as they fail to see how raw data are magically turned into colorful images. A more in depth user study, possibly supported by user data collected with eye tracking technology could help us shade some light on this still unexplored research area.

Bibliography

Agosti, M., Crivellari, F. and Di Nunzio, G. M. [2012]. Web log analysis: A review of a decade of studies about information acquisition, inspection and interpretation of user interaction., *Data Mining and Knowledge Discovery* **24**(3): 663–696.

Akhigbe, B. I., Afolabi, B. S., Udo, J. I. and Adagunodo, E. R. [2011]. An evaluative model for Information Retrieval system evaluation: A user- centered approach, *International Journal on Computer Science and Engineering (IJCSE)* **3**(7).

Alben, L. [1996]. Quality of experience: defining the criteria for effective interaction design, *interactions* **3**(3): 11–15.

Andrews, K., Gutl, C., Moser, J., Sabol, V. and Lackner, W. [2001]. Search result visualisation with xfind, *Proceedings Second International Workshop on User Interfaces in Data Intensive Systems. UIDIS 2001*, pp. 50–58.

Andrews, K., Kienreich, W., Sabol, V., Becker, J., Droschl, G., Kappe, F., Granitzer, M., Auer, P. and Tochtermann, K. [2002]. The infosky visual explorer: exploiting hierarchical structure and document similarities., *Information Visualization* **1**(3-4): 166–181.

Antell, K. and Huang, J. [2008]. Subject searching success: Transaction logs, patron perceptions, and implications for library instruction, *Reference & user services quarterly* pp. 68–76.

Antelman, K., Lynema, E. and Pace, A. K. [2006]. Toward a 21st century library catalog, *Information technology and libraries* **25**(3): 128–139.

Appleton, L. [2016]. User experience (ux) in libraries: let's get physical (and digital), *Insights the UKSG journal* **29**(3): 224–227.

Arshad, A. and Ameen, K. [2015]. Usage patterns of punjab university library website: a transactional log analysis study, *The Electronic Library* **33**(1): 65–74.

Asher, A. D., Duke, L. M. and Wilson, S. [2013]. Paths of discovery: Comparing the search effectiveness of ebsco discovery service, summon, google scholar, and conventional library resources, *College & research libraries* **74**(5): 464–488.

Azzopardi, L. [2011]. The economics in interactive information retrieval, *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 15–24.

Bargas-Avila, J. A. and Hornbæk, K. [2011]. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience, *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 2689–2698.

Barifah, M. and Landoni, M. [2019a]. Exploring usage patterns of a large-scale digital library, *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pp. 67–76.

Barifah, M. and Landoni, M. [2019b]. Interactive search profiles as a design tool, *International Conference on Human-Computer Interaction*, Springer, pp. 18–30.

Barifah, M. and Landoni, M. [2019c]. Obstacles to conducting log file studies in saudi arabia, *With an Eye to the Future: HCI Research and Practice in the Arab World* pp. 40–45.

Barifah, M., Landoni, M. and Eddakrouri, A. [2020]. Evaluating the user experience in a digital library, *Proceedings of the Association for Information Science and Technology* **57**(1): e280.

Bates, M. J. [1989]. The design of browsing and berrypicking techniques for the online search interface, *Online review* **13**(5): 407–424.

Bauer, S. [2014]. Interactive visualizations for search processes, *IEEE GSC*.

Behnert, C., Behnert, C., Lewandowski, D. and Lewandowski, D. [2017]. A framework for designing retrieval effectiveness studies of library information systems using human relevance assessments, *Journal of Documentation* **73**(3): 509–527.

Behnert, C. and Lewandowski, D. [2017]. Known-item searches resulting in zero hits: Considerations for discovery systems, *The Journal of Academic Librarianship* **43**(2): 128–134.

Belkin, N. [2010]. On the evaluation of interactive information retrieval systems, *The Janus Faced Scholar. A Festschrift in Honor of Peter Ingwersen. Det Informationsvidenskabelige Akademi (Royal School of Library and Information Science, Copenhagen)* pp. 13–22.

Belkin, N. [2015]. People, interacting with information, *SIGIR Forum* **49**(2): 13–27.

Belkin, N. J., Cole, M. and Bierig, R. [2008]. Is relevance the right criterion for evaluating interactive information retrieval?, *Proceedings of the ACM SIGIR 2008 Workshop on Beyond Binary Relevance: Preferences, Diversity, and Set-Level Judgments. http://research. microsoft. com/˜ pauben/bbr-workshop*.

Bhargav, A. and Bhargav, M. [2014]. Pattern discovery and users classification through web usage mining, *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pp. 632–636.

Bilal, D. [2000]. Children's use of the yahooligans! web search engine: I. cognitive, physical, and affective behaviors on fact-based search tasks, *Journal of the American Society for information Science* **51**(7): 646–665.

Bilal, D. and Bachir, I. [2007]. Children's interaction with cross-cultural and multilingual digital libraries. ii. information seeking, success, and affective experience, *Information processing & management* **43**(1): 65–80.

Björneborn, L. [2010]. Design dimensions enabling divergent behaviour across physical, digital, and social library interfaces, *International Conference on Persuasive Technology*, Springer, pp. 143–149.

Blandford, A., Stelmaszewska, H. and Bryan-Kinns, N. [2001]. Use of multiple digital libraries: a case study, *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*, pp. 179–188.

Blumer, E., Hügi, J. and Schneider, R. [2014]. The usability issues of faceted navigation in digital libraries, *JLIS. it* **5**(2): 85–100.

Bollen, J. and Luce, R. [2002]. Evaluation of digital library impact and user communities by analysis of usage patterns, *D-Lib Magazine* **8**(6): 1–13.

Borgman, C. L. [1996]. Why are online catalogs still hard to use?, *Journal of the American society for information science* **47**(7): 493–503.

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

Borlund, P. [2003]. The iir evaluation model: a framework for evaluation of interactive information retrieval systems, *Information Research* **8**(3).

Borlund, P. [2016]. A study of the use of simulated work task situations in interactive information retrieval evaluations: A meta-evaluation, *Journal of Documentation* **72**(3): 394–413.

Borlund, P. and Dreier, S. [2014]. An investigation of the search behaviour associated with ingwersen's three types of information needs, *Information Processing & Management* **50**(4): 493–507.

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

Börner, K. and Chen, C. [2002]. Visual interfaces to digital libraries, *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pp. 425–425.

Bossaller, J. S. and Sandy, H. M. [2017]. Documenting the conversation: A systematic review of library discovery layers, *College & Research Libraries* **78**(5): 602.

Brave, S., Nass, C. and Hutchinson, K. [2005]. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent, *International journal of human-computer studies* **62**(2): 161–178.

Buchanan, S. and Salako, A. [2009]. Evaluating the usability and usefulness of a digital library, *Library Review* **58**(9): 638–651.

Buck, S. and Mellinger, M. [2011]. The impact of serial solutions' summon™ on information literacy instruction: Librarian perceptions, *Internet Reference Services Quarterly* **16**(4): 159–181.

Butcher, K. R., Davies, S., Crockett, A., Dewald, A. and Zheng, R. [2011]. Do graphical search interfaces support effective search for and evaluation of digital library resources?, *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, pp. 315–324.

Byström, K. and Hansen, P. [2005]. Conceptual framework for tasks in information studies: Book reviews, *J. Am. Soc. Inf. Sci. Technol.* **56**(10): 1050–1061.

Byström, K. and Järvelin, K. [1995]. Task complexity affects information seeking and use., *Information processing & management* **31**(2): 191–213.

Cabrerizo, F. J., López-Gijón, J., Ruiz, A. A. and Herrera-Viedma, E. [2010]. A model based on fuzzy linguistic information to evaluate the quality of digital libraries, *International Journal of Information Technology & Decision Making* **9**(03): 455–472.

Capra, R., Marchionini, G., Oh, J. S., Stutzman, F. and Zhang, Y. [2007]. Effects of structure and interaction style on distinct search tasks, *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, ACM, pp. 442–451.

Card, M. [1999]. *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann.

Carevic, Z., Schüller, S., Mayr, P. and Fuhr, N. [2018]. Contextualised browsing in a digital library's living lab, *arXiv preprint* pp. 89–98.

Carver, C. [2003]. Pleasure as a sign you can attend to something else: Placing positive feelings within a general model of affect, *Cognition and Emotion* **17**(2): 241–261.

Casarosa, V. [2010]. A conceptual model for digital libraries. NMPLIS Summer School,Tbilisi, Georgia.

Catarci, T. and Kimani, S. [2012]. Human-computer interaction view on information retrieval evaluation, *Information Retrieval Meets Information Visualization*, Springer, pp. 48–75.

Chan, L. M. and O'Neill, E. T. [2010]. Fast: Faceted application of subject terminology, *Prnciples and Applications Libraries Unlimited, Santa Barbara* p. 261.

Chen, B. [2018]. Usage pattern comparison of the same scholarly articles between web of science (wos) and springer, *Scientometrics* **115**(1): 519–537.

Chen, C. [2005]. Top 10 unsolved information visualization problems, *IEEE computer graphics and applications* **25**(4): 12–16.

Chen, H.-M. and Cooper, M. D. [2001]. Using clustering techniques to detect usage patterns in a web-based information system, *Journal of the American Society for Information Science and Technology* **52**(11): 888–904.

Chen, H.-M. and Cooper, M. D. [2002]. Stochastic modeling of usage patterns in a web-based information system, *Journal of the Association for Information Science and Technology* **53**(7): 536–548.

Chi, E. H.-h. and Riedl, J. T. [1998]. An operator interaction framework for visualization systems, *Proceedings IEEE Symposium on Information Visualization (Cat. No. 98TB100258)*, IEEE, pp. 63–70.

Choo, J., Lee, C., Clarkson, E., Liu, Z., Lee, H., Chau, D. H. P., Li, F., Kannan, R., Stolper, C. D., Inouye, D. et al. [2013]. Visirr: interactive visual information retrieval and recommendation for large-scale document data, *Technical report*, Georgia Institute of Technology.

Cignoli, N. and Liu, Y. Q. [2011]. Evaluation of digital libraries systems' usability and performance through user perception, *International Digital Libraries Perspectives ILS, Special Topics: Study Abroad in China* pp. 599–670.

Clough, P. and Sanderson, M. [2013]. Evaluating the performance of information retrieval systems using test collections., *Information Research* **18**(2).

Cole, C. [2015]. Information need and the beginning of information search, *Encyclopedia of Information Science and Technology, Third Edition*, IGI Global, pp. 4117–4128.

Cronin, G. [2014]. What is a ux librarian anyway?
URL: *https://www.infotoday.eu/Articles/Editorial/Featured-Articles/What-is-a-UX-Librarian-anyway-98012.aspx*

Cugini, J. V., Laskowski, S. and Sebrechts, M. M. [2000]. Design of 3d visualization of search results: evolution and evaluation., *Visual Data Exploration and Analysis VII*, Vol. 3960, International Society for Optics and Photonics, pp. 198–210.

Cushman, W. H. and Rosenberg, D. J. [1991]. Human factors in product design, *Advances in human factors/ergonomics* **14**.

Dahlen, S. P., Haeger, H., Hanson, K. and Montellano, M. [2020]. Almost in the wild: Student search behaviors when librarians aren't looking, *The Journal of Academic Librarianship* **46**(1): 102096.

Dahlen, S. P. and Hanson, K. [2017]. Preference vs. authority: A comparison of student searching in a subject-specific indexing and abstracting database and a customized discovery layer, *College & Research Libraries* **78**(7): 878–897.

Daniels, J. and Roth, P. [2012]. Incorporating millennium catalog records into serials solutions' summon, *Technical Services Quarterly* **29**(3): 193–199.

Debowski, S. [2001]. Wrong way: go back! an exploration of novice search behaviours while conducting an information search, *The Electronic Library* **19**(6): 371–382.

Dempsey, M. and Valenti, A. M. [2016]. Student use of keywords and limiters in web-scale discovery searching, *The journal of academic librarianship* **42**(3): 200–206.

Denton, W. and Coysh, S. J. [2011]. Usability testing of vufind at an academic library, *Library Hi Tech* **29**: 301–319.

Deodato, J., Gambrell, K. and Frierson, E. [2016]. One size doesn't fit all: Tailoring discovery through user testing, *Varnum, Kenneth J. (Hg.): Exploring discovery. The front door to your library's licensed and digitized content, London* pp. 183–199.

Desmet, P. M., Vastenburg, M. H. and Romero, N. [2016]. Mood measurement with pick-a-mood: review of current methods and design of a pictorial self-report scale, *Journal of Design Research* **14**(3): 241–279.

Drabenstott, K. M. and Weller, M. S. [1996]. Failure analysis of subject searches in a test of a new design for subject access to online catalogs, *Journal of the American Society for Information Science* **47**(7): 519–537.

Eason, K., Richardson, S. and Yu, L. [2000]. Patterns of use of electronic journals, *Journal of documentation* **56**(5): 477–504.

Fagan, J. C., Mandernach, M. A., Nelson, C. S., Paulo, J. R. and Saunders, G. [2012]. Usability test results for a discovery tool in an academic library, *Information technology and libraries* **31**(1): 83–112.

Fast, K. V. and Campbell, D. G. [2004]. "i still like google": University student perceptions of searching opacs and the web, *Proceedings of the American Society for Information Science and Technology* **41**(1): 138–146.

Fawley, N. and Krysak, N. [2012]. Information literacy opportunities within the discovery tool environment, *College & undergraduate libraries* **19**(2-4): 207–214.

Flavián-Blanco, C., Gurrea-Sarasa, R. and Orús-Sanclemente, C. [2011]. Analyzing the emotional outcomes of the online search behavior with search engines, *Computers in Human Behavior* **27**(1): 540–551.

Fox, E. A. [2002]. Overview of digital library components and developments, *Technical report tr-02-23*, Computer Science, Virginia Tech.

Frias-Martinez, E., Chen, S. Y., Macredie, R. D. and Liu, X. [2007a]. The role of human factors in stereotyping behavior and perception of digital library users: a robust clustering approach, *User Modeling and User-Adapted Interaction* **17**(3): 305–337.

Frias-Martinez, E., Chen, S. Y., Macredie, R. D. and Liu, X. [2007b]. The role of human factors in stereotyping behavior and perception of digital library users: a robust clustering approach, *User Modeling and User-Adapted Interaction* **17**(3): 305–337.

Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., Klas, C.-P., Kovács, L., Landoni, M., Micsik, A. et al. [2007]. Evaluation of digital libraries, *International Journal on Digital Libraries* **8**(1): 21–38.

Fulton, C. [2009]. The pleasure principle: the power of positive affect in information seeking, *Aslib Proceedings: New Information Perspectives*, Vol. 61, Emerald Group Publishing Limited, pp. 245–261.

Gaona-García, P. A., Martin-Moncunill, D. and Montenegro-Marin, C. E. [2017]. Trends and challenges of visual search interfaces in digital libraries and repositories, *The Electronic Library* **35**(1): 69–98.

Gaona-García, P. A., Stoitsis, G., Sánchez-Alonso, S. and Biniari, K. [2016]. An exploratory study of user perception in visual search interfaces based on skos, *KO Knowledge Organization* **43**(4): 217–238.

Georgas, H. [2014]. Google vs. the library (part ii): Student search patterns and behaviors when using google and a federated search tool, *portal: Libraries and the Academy* **14**(4): 503–532.

Gooding, P. [2016]. Exploring the information behaviour of users of welsh newspapers online through web log analysis, *Journal of Documentation* **72**(2): 232–246.

Gross, J. and Sheridan, L. [2011]. Web scale discovery: the user experience, *New library world* pp. 236–247.

Gross, M. [1999]. Imposed queries in the school library media center: A descriptive study, *Library & Information Science Research* **21**(4): 501–521.

Guo, F., Liu, C. and Wang, Y. M. [2009]. Efficient multiple-click models in web search, *Proceedings of the second acm international conference on web search and data mining*, pp. 124–131.

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

Gwizdka, J. and Spence, I. [2007]. Implicit measures of lostness and success in web navigation, *Interacting with Computers* **19**(3): 357–369.

HAJEK, P. and Stejskal, J. [2012]. Analysis of user behavior in a public library using bibliomining, *Advances in Environment, Computational Chemistry and Bioscience* pp. 339–344.

Hajek, P. and Stejskal, J. [2017]. Library usage mining in the context of alternative costs: The case of the municipal library of prague, *Library Hi Tech* **35**(4): 558–576.

Hall, C. E. [2016]. *Facets in Library Catalogs: The Beliefs, Behaviors, Policies and Practices That Guide Implementation*, Drexel University.

Hamlett, A. and Georgas, H. [2019]. In the wake of discovery: Student perceptions, integration, and instructional design, *Journal of Web Librarianship* **13**(3): 230–245.

Han, H. and Wolfram, D. [2016]. An exploration of search session patterns in an image-based digital library, *Journal of Information Science* **42**(4): 477–491.

Harkema, C. [2015]. A site for scholarly primitives: Exploring the digital library interface, *International Journal of Technology, Knowledge & Society: Annual Review* **10**(1): 1–13.

Hart, S. G. and Staveland, L. E. [1988]. Development of nasa-tlx (task load index): Results of empirical and theoretical research, *Advances in psychology*, Vol. 52, Elsevier, pp. 139–183.

Hienert, D., Sawitzki, F., Schaer, P. and Mayr, P. [2012]. Integrating interactive visualizations in the search process of digital libraries and integrating interactive visualizations in the search process of digital libraries and ir systems, *European Conference on Information Retrieval*, Springer, pp. 447–450.

Hoeber, O., Patel, D. and Storie, D. [2019]. A study of academic search scenarios and information seeking behaviour, *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pp. 231–235.

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

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

Hoy, M. B. [2017]. An introduction to the blockchain and its implications for libraries and medicine, *Medical reference services quarterly* **36**(3): 273–279.

Huang, C. [2014]. *Understanding novice users' help-seeking behavior in getting started with digital libraries: Influence of learning styles*, PhD thesis, University of Wisconsin-Milwaukee.

Hügi, J. and Schneider, R. [2013]. A framework for evaluating the usefulness of digital libraries, *Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten: Information Science between Virtual Infrastructure and Material Lifeworlds: Proceedings des 13. Internationalen Symposiums für Informationswissenschaft (ISI 2013)*, pp. 1–12.

Huurdeman, H., Aamodt, M., Låberg, K. T. and Heggø, D. M. [2018]. The collection in a new light–towards visual exploration and navigation of academic library resources, *International Information & Library Review* **50**(3): 219–235.

Huurdeman, H. C., Kamps, J., Wilson, M. L. et al. [2019]. The multi-stage experience: the simulated work task approach to studying information seeking stages, *BIIRRR@ CHIIR*, pp. 7–13.

Huvila, I., Enwald, H., Eriksson-Backa, K., Liu, Y.-H. and Hirvonen, N. [2019]. Information behaviour and practises research informing technology and service design, *Proceedings of the Association for Information Science and Technology* **56**(1): 541–545.

Ingwersen, P. [2000]. Users in context, *Lectures on information retrieval*, Springer, pp. 157–178.

Ingwersen, P. and Järvelin, K. [2005]. *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*, Springer-Verlag New York, Inc.

Islamaj Dogan, R., Murray, G. C., Névéol, A. and Lu, Z. [2019]. Understanding pubmed® user search behavior through log analysis, *Database* .

Ives, B., Olson, M. H. and Baroudi, J. J. [1983]. The measurement of user information satisfaction, *Communications of the ACM* **26**(10): 785–793.

Jansen, B. J. and Spink, A. [2006]. How are we searching the world wide web? a comparison of nine search engine transaction logs, *Information processing & management* **42**(1): 248–263.

Jeffrey, R. and Chisnell, D. [2008]. *Handbook of usability testing: How to plan, design, and conduct effective tests*, John Wiley & Sons.

Johnson, J. D., Donohue, W. A., Atkin, C. K. and Johnson, S. [1995]. A comprehensive model of information seeking tests focusing on a technical organization, *Science Communication* **16**(3): 274–303.

Johnson, J., Panagioti, M., Bass, J., Ramsey, L. and Harrison, R. [2017]. Resilience to emotional distress in response to failure, error or mistakes: A systematic review, *Clinical psychology review* **52**: 19–42.

Johnson, M. [2013]. Usability test results for encore in an academic library, *Information technology and libraries* **32**(3): 59–85.

Joho, H., Villa, R. and Jose, J. [2007]. Interaction pool: Towards a user-centered test collection, *SIGIR'07 Web Information-Seeking and Interaction Workshop*.

Joo, S. [2013]. *Investigating User Search Tactic Patterns and System Support in Using Digital Libraries*, PhD thesis, University of Wisconsin Milwaukee.

Julien, H., McKechnie, L. E. and Hart, S. [2004]. A content analysis of affective issues in library and information science systems work, *Information Research* **10**(1): 10–1.

Kadury, A. and Frank, A. J. [2007]. Harvesting and aggregation of digital libraries using the oai framework, *WEBIST (2)*, pp. 441–446.

Karat, J. [2002]. Beyond task completion: evaluation of affective components of use, *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, L. Erlbaum Associates Inc., pp. 1152–1164.

Keim, D. A., Barro, H., Panse, C., Schneidewind, J. and Sips, M. [2004]. Exploring and visualizing the history of infovis, *IEEE Symposium on Information Visualization*.

Kelly, D. [2009]. Methods for evaluating interactive information retrieval systems with users, *Foundations and Trends in Information Retrieval* **3**(1-2): 1–224.

Kelly, D. and Sugimoto, C. [2013]. A systematic review of interactive information retrieval evaluation studies, 1967–2006, *Journal Of The American Society For Information Science And Technology* **64**(4): 745–770.

Khoo, M., Kusunoki, D. and MacDonald, C. [2012]. Finding problems: When digital library users act as usability evaluators, *2012 45th Hawaii International Conference on System Sciences*, IEEE, pp. 1615–1624.

Kovacevic, A., Devedzic, V. and Pocajt, V. [2010]. Using data mining to improve digital library services, *The electronic library* **28**(6): 829–843.

Kuhlthau, C. C. [1991]. Inside the search process: Information seeking from the user's perspective, *Journal of the American society for information science* **42**(5): 361–371.

Kuhlthau, C. C. [1993]. A principle of uncertainty for information seeking, *Journal of documentation* **49**(4): 339–355.

Kuhlthau, C. C. [2004]. *Seeking meaning: A process approach to library and information services*, Vol. 2, Libraries Unlimited Westport, CT.

Kuhlthau, C. C., Heinström, J. and Todd, R. J. [2008]. The 'information search process' revisited: Is the model still useful, *Information research* **13**(4): 13–4.

Kumar, V., Furuta, R. and Allen, R. B. [1998]. Metadata visualization for digital libraries: interactive timeline editing and review, *Proceedings of the third ACM conference on Digital libraries*, pp. 126–133.

Lau, E. P. and Goh, D. H.-L. [2006]. In search of query patterns: A case study of a university opac, *Information processing & management* **42**(5): 1316–1329.

Lazar, J., Feng, J. H. and Hochheiser, H. [2017]. *Research methods in human-computer interaction*, Morgan Kaufmann.

Lazar, J., Jones, A., Hackley, M. and Shneiderman, B. [2006]. Severity and impact of computer user frustration: A comparison of student and workplace users, *Interacting with Computers* **18**(2): 187–207.

Leckie, G. J., Pettigrew, K. E. and Sylvain, C. [1996]. Modeling the information-seeking of professionals: A general model derived from research on engineers, health care professionals and lawyers, *Library Quarterly* **66**(2): 161–193.

Lee, B., Czerwinski, M., Robertson, G. and Bederson, B. B. [2005]. Understanding research trends in conferences using paperlens, *CHI'05 extended abstracts on Human factors in computing systems*, pp. 1969–1972.

Lewis, J. R. [1991]. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the asq, *ACM Sigchi Bulletin* **23**(1): 78–81.

Li, Y. [2009]. Exploring the relationships between work task and search task in information search, *Journal of the American Society for Information Science and Technology* **60**(2): 275–291.

Liu, J. and Zhang, X. [2019]. The role of domain knowledge in document selection from search results, *Journal of the Association for Information Science and Technology* **70**(11): 1236–1247.

Lopatovska, I. [2014]. Toward a model of emotions and mood in the online information search process, *Journal of the Association for Information Science and Technology* **65**(9): 1775–1793.

Lueg, C. and Twidale, M. [2018]. Designing for humans, not robots (or vulcans), *Library Trends* **66**(4): 409–421.

Mahlke, S. [2008]. *User experience of interaction with technical systems. Theories, methods, empirical results, and their application to the design of interactive systems.*, PhD thesis, Saarbrücken, Germany: VDM Verlag.

Manning, C., Raghavan, P. and Schütze, H. [2008]. *Introduction to Information Retrieval*, Cambridge University Press.

Mansourian, Y. [2008]. Contextualization of web searching: a grounded theory approach, *The Electronic Library* **26**(2): 202–214.

Marchionini, G. [2006]. Exploratory search: from finding to understanding, *Communications of the ACM* **49**(4): 41–46.

Markey, K. [2007]. Twenty-five years of end-user searching, part 1: Research findings, *Journal of the American Society for Information Science and Technology* **58**: 1071–1081.

Marks, L., Hussell, J. A., McMahon, T. M. and Luce, R. E. [2005]. Activegraph: A digital library visualization tool, *International Journal on Digital Libraries* **5**(1): 57–69.

Massis, B. [2018]. The user experience (ux) in libraries, *Information and Learning Science* **119**(3/4): 241–244.

Maxwell, D. and Azzopardi, L. [2016]. Agents, simulated users and humans: An analysis of performance and behaviour, *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management series, CIKM '16*, number 10 in *1*, ACM, pp. 731–740.

Mayer, J. D., Roberts, R. D. and Barsade, S. G. [2008]. Human abilities: Emotional intelligence, *Annu. Rev. Psychol.* **59**: 507–536.

McCay-Peet, L. and Toms, E. [2011]. Measuring the dimensions of serendipity in digital environments, *Information Research: An International Electronic Journal* **16**(3): n3.

Mckie, I. A. S. and Narayan, B. [2019]. Enhancing the academic library experience with chatbots: An exploration of research and implications for practice, *Journal of the Australian Library and Information Association* **68**(3): 268–277.

Meyyappan, N., Foo, S. and Chowdhury, G. G. [2004]. Design and evaluation of a task-based digital library for the academic community, *Journal of Documentation* **60**(4): 449–475.

Morris, W. N. [2003]. 9 the mood system, *Well-being: Foundations of hedonic psychology* p. 169.

Morville, P. [2004]. User experience design.
URL: *https://semanticstudios.com/user*$_e$*xperience*$_d$*esign/*

Nagy, A. [2011]. *Analyzing the next-generation catalog*, American Library Association.

Nahl, D. [2005]. Affective and cognitive information behavior: Interaction effects in internet use, *Proceedings of the American Society for Information Science and Technology* **42**(1).

Naumann, A., Hurtienne, J., Israel, J. H., Mohs, C., Kindsmüller, M. C., Meyer, H. A. and Hußlein, S. [2007]. Intuitive use of user interfaces: defining a vague concept, *International Conference on Engineering Psychology and Cognitive Ergonomics*, Springer, pp. 128–136.

Ndumbaro, F. [2018]. Understanding user-system interactions: An analysis of opac users' digital footprints, *Information Development* **34**(3): 297–308.

Nedumov, Y., Babichev, A., Mashonsky, I. and Semina, N. [2019]. Scinoon: Exploratory search system for scientific groups, *IUI Workshops*.

Nicholas, D., Huntington, P., Jamali, H. R. and Tenopir, C. [2006]. Finding information in (very large) digital libraries: a deep log approach to determining differences in use according to method of access, *The Journal of academic librarianship* **32**(2): 119–126.

Nielsen, J. [1993]. *Usability Engineering*, Morgan Kaufmann.

Nielsen, J. [2020]. Usability 101: Introduction to usability.
URL: *https://www.nngroup.com/articles/usability-101-introduction-to-usability/*

Niu, X. and Hemminger, B. [2015]. Analyzing the interaction patterns in a faceted search interface, *Journal of the Association for Information Science and Technology* **66**(5): 1030–1047.

Nualart, J., Pérez-Montoro Gutiérrez, M. and Whitelan, M. [2014]. How we draw texts: a review of approaches to text visualization and exploration, *El Profesional de la Información, 2014, vol. 23, num. 3, p. 221-235* **23**: 221–235.

Nualart, J. and Pérez-Montoro, M. [2013]. Texty, a visualization tool to aid selection of texts from search outputs, *Professor TD Wilson* .

O'Brien, H. L., Dickinson, R. and Askin, N. [2017]. A scoping review of individual differences in information seeking behavior and retrieval research between 2000 and 2015, *Library & Information Science Research* **39**(3): 244–254.

Pääkkönen, T., Kekäläinen, J., Keskustalo, H., Azzopardi, L., Maxwell, D. and Järvelin, K. [2017]. Validating simulated interaction for retrieval evaluation, *Information Retrieval Journal* pp. 1–25.

Padmaja, S. and Seshasayee, A. [2016]. Clustering of user behaviour based on web log data using improved k-means clustering algorithm, *International Journal of Engineering and Technology (IJET)* **8**(1): 305–310.

Papatheodorou, C., Kapidakis, S., Sfakakis, M. and Vassiliou, A. [2003]. Mining user communities in digital libraries, *Information Technology and Libraries* **22**(4): 152.

Pérez-Montoro, M. and Nualart, J. [2015]. Visual articulation of navigation and search systems for digital libraries, *International Journal of Information Management* **35**(5): 572–579.

Peters, T. A. [1989]. When smart people fail: an analysis of the transaction log of an online public access catalog, *Journal of academic librarianship* **15**(5): 267–73.

Peters, T. A. [1993]. The history and development of transaction log analysis, *Library hi tech* **11**(2): 41–66.

Petrelli, D. [2008]. On the role of user-centred evaluation in the advancement of interactive information retrieval, *Information processing & management* **44**(1): 22–38.

Plaisant, C., Grosjean, J. and Bederson, B. B. [2002]. Spacetree: Supporting exploration in large node link tree, design evolution and empirical evaluation, *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.*, IEEE, pp. 57–64.

Poddar, A. and Ruthven, I. [2010]. The emotional impact of search tasks, *Proceedings of the third symposium on Information interaction in context*, pp. 35–44.

Priestner, A. and Borg, M. [2016]. *User experience in libraries: applying ethnography and human-centred design*, Routledge.

Renaud, J., Britton, S., Wang, D. and Ogihara, M. [2015]. Mining library and university data to understand library, *The Electronic Library* **33**(3): 355–372.

Repke, T. and Krestel, R. [2020]. Visualising large document collections by jointly modeling text and network structure, *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, pp. 279–288.

Rijsbergen, C. J. V. [1979]. *Information Retrieval*, 2nd edn, Butterworth-Heinemann Newton, MA, USA.

Robertson, G. G., Mackinlay, J. D. and Card, S. K. [1991]. Cone trees: animated 3d visualizations of hierarchical information, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 189–194.

Romero, C. and Ventura, S. [2007]. Educational data mining: A survey from 1995 to 2005, *Expert systems with applications* **33**(1): 135–146.

Rosenbaum, S. E., Glenton, C. and Cracknell, J. [2008]. User experiences of evidence-based online resources for health professionals: user testing of the cochrane library, *BMC medical informatics and decision making* **8**(1): 34.

Rosenfeld, L. and Morville, P. [2002]. *Information architecture for the world wide web*, O'Reilly Media, Inc.

Rosman, T., Mayer, A.-K. and Krampen, G. [2016]. On the pitfalls of bibliographic database searching: Comparing successful and less successful users, *Behaviour & Information Technology* **35**: 106–117.

Rotter, J. B. [1966]. Generalized expectancies for internal versus external control of reinforcement., *Psychological monographs: General and applied* **80**(1): 1.

Rousseeuw, P. J. [1987]. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of computational and applied mathematics* **20**: 53–65.

Ruotsalo, T., Athukorala, K., Głowacka, D., Konyushkova, K., Oulasvirta, A., Kaipiainen, S., Kaski, S. and Jacucci, G. [2013]. Supporting exploratory search tasks with interactive user modeling, *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries*, Vol. 50, Wiley Online Library, pp. 1–10.

Russell-Rose, T., Clough, P. and Toms, E. G. [2014]. Categorising search sessions: some insights from human judgments, *Proceedings of the 5th Information Interaction in Context Symposium*, pp. 251–254.

Sadeh, T. [2007]. Time for a change: new approaches for a new generation of library users, *New library world* **108**(7/8): 307–316.

Sanderson, M. [2010]. *Test Collection Based Evaluation of Information Retrieval Systems*, Now Publishers Inc.

Santos, H. and Vieira, R. [2017]. Pln-pucrs at emoint-2017: psycholinguistic features for emotion intensity prediction in tweets, *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 189–192.

Saracevic, T. [2000]. Digital library evaluation: Toward an evolution of concepts, *Library Trends* **49**(2): 350–369.

Sauro, J. [2015]. Supr-q: A comprehensive measure of the quality of the website user experience, *Journal of usability studies* **10**(2): 68–86.

Sauro, J. [2020]. Measuring usability with the system usability scale (sus). **URL:** *https://measuringu.com/sus/*

Savolainen, R. [1995]. Everyday life information seeking: approaching information seeking in the context of "way of life", *Library & information science research* **17**(3): 259–294.

Savolainen, R. and Kari, J. [2006]. User-defined relevance criteria in web searching, *Journal of Documentation* **62**(6): 685–707.

Scherer, K. R. [2001]. Appraisal considered as a process of multilevel sequential checking, *Appraisal processes in emotion: Theory, methods, research* **92**(120): 57.

Schmidt, A. [2016]. Usability and desirability | the user experience. **URL:** *https://www.libraryjournal.com/?detailStory=usability-and-desirability-the-user-experience*

Secretary, I. C. [1998]. Ergonomic requirements for office work with visual display terminals (vdts)–part 11: Guidance on usability, *Standard 22*, International Organization for Standardization.

Sharp, E. C., Pelletier, L. G. and Lévesque, C. [2006]. The double-edged sword of rewards for participation in psychology experiments., *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* **38**(3): 269.

Shen, H., Bednarz, T., Nguyen, H., Feng, F., Wyeld, T., Hoek, P. J. and Lo, E. H. [2019]. Information visualisation methods and techniques: State-of-the-art and future directions, *Journal of Industrial Information Integration* **16**: 100102.

Sheth, N. and Cai, Q. [2003]. Visualizing mesh dataset using radial tree layout, *Citeseer* .

Shneiderman, B., Feldman, D., Rose, A. and Grau, X. F. [2000]. Visualizing digital library search results with categorical and hierarchical axes., *Proceedings of the fifth ACM conference on Digital libraries*, pp. 57–66.

Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J. and Cattrysse, D. [2015]. Literature review of data mining applications in academic libraries, *The Journal of Academic Librarianship* **41**(4): 499–510.

Slone, D. J. [2000]. Encounters with the opac: On-line searching in public libraries, *Journal of the American Society for Information Science* **51**(8): 757–773.

Srivastava, J., Cooley, R., Deshpande, M. and Tan, P.-N. [2000]. Web usage mining: Discovery and applications of usage patterns from web data, *Acm Sigkdd Explorations Newsletter* **1**(2): 12–23.

Stößel, C., Wandke, H. and Blessing, L. [2009]. Gestural interfaces for elderly users: help or hindrance?, *International Gesture Workshop*, Springer, pp. 269–280.

Tamine-Lechani, L., Boughanem, M. and Daoud, M. [2010]. Evaluation of contextual information retrieval effectiveness: overview of issues and research, *Knowledge and Information Systems* **24**(1): 1–34.

Tammaro, A. M. [2008]. User perceptions of digital libraries: a case study in italy, *Performance measurement and metrics* **9**(2): 130–137.

Tang, M.-C. [2007]. Browsing and searching in a faceted information space: A naturalistic study of pubmed users' interaction with a display tool, *Journal of the American Society for Information Science and Technology* **58**(13): 1998–2006.

Thayer, R. E. [1997]. *The origin of everyday moods: Managing energy, tension, and stress*, Oxford University Press, USA.

Toms, E. G., Dufour, C. and Hesemeier, S. [2004]. Measuring the user's experience with digital libraries, *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries*, IEEE, pp. 51–52.

Trapido, I. [2016]. Library discovery products: discovering user expectations through failure analysis, *Information Technology and Libraries* **35**(3): 9–26.

Tsakonas, G., Kapidakis, S. and Papatheodorou, C. [2004]. Evaluation of user interaction in digital libraries, *Notes of the DELOS WP7 workshop on the evaluation of Digital Libraries, Padua, Italy*.

Tsakonas, G., Mitrelis, A., Papachristopoulos, L. and Papatheodorou, C. [2013]. An exploration of the digital library evaluation literature based on an ontological representation, *Journal of the American Society for Information Science and Technology* **64**(9): 1914–1926.

Vaughan, J. [2011]. Chapter 1: Web scale discovery what and why?, *Library technology reports* **47**(1): 5–11.

Wang, Y., Liu, J., Mandal, S. and Shah, C. [2017]. Search successes and failures in query segments and search tasks: A field study, *Proceedings of the Association for Information Science and Technology* **54**(1): 436–445.

Watson, D., Clark, L. A. and Tellegen, A. [1988]. Development and validation of brief measures of positive and negative affect: the panas scales., *Journal of personality and social psychology* **54**(6): 1063.

Wei, F., Liu, S., Song, Y., Pan, S., Zhou, M. X., Qian, W., Shi, L., Tan, L. and Zhang, Q. [2010]. Tiara: A visual exploratory text analytic system, *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, Association for Computing Machinery, pp. 153–162.

Wildemuth, B., Freund, L. and Toms, E. [2013]. Designing known-item and fact-finding search tasks for studies of interactive information retrieval, *Proceedings of the second association for information science and technology ASIS&T (European Workshop)*, pp. 131–162.

Wilson, M. L. [2011]. Search user interface design, *Synthesis lectures on information concepts, retrieval, and services* pp. 1–143.

Wilson, T. [1999]. Models in information behaviour research, *Journal of Documentation* pp. 249–270.

Wilson, T. D. [1981]. On user studies and information needs, *Journal of documentation* **62**(6): 658–670.

Wirth, W., Sommer, K., Von Pape, T. and Karnowski, V. [2016]. Success in online searches: differences between evaluation and finding tasks, *Journal of the Association for Information Science and Technology* **67**(12): 2897–2908.

Wong, W., Chen, R., Kodagoda, N., Rooney, C. and Xu, K. [2011]. Invisque: intuitive information exploration through interactive visualization., *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, Association for Computing Machinery, pp. 311–316.

Xie, I. [2012]. Information searching and search models, *Understanding information retrieval systems: management, types, and standards* pp. 31–46.

Xie, I. and Cool, C. [2009]. Understanding help seeking within the context of searching digital libraries, *Journal of the American Society for Information Science and Technology* **60**(3): 477–494.

Xu, B. [2013]. Clustering educational digital library usage data: A comparison of latent class analysis and k-means algorithms, *Journal of Educational Data Mining* **5**(2): 38–68.

Xu, W., Esteva, M., Jain, S. D. and Jain, V. [2011]. Analysis of large digital collections with interactive visualization, *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, IEEE, pp. 241–250.

Yang, S. Q. and Wagner, K. [2010]. Evaluating and comparing discovery tools: How close are we towards next generation catalog?, *Library Hi Tech* **28**(4): 690–709.

Yeh, S.-T. and Liu, Y. [2011]. Integrated faceted browser and direct search to enhance information retrieval in text-based digital libraries, *Intl. Journal of Human–Computer Interaction* **27**(4): 364–382.

Zavalina, O. and Vassilieva, E. V. [2014]. Understanding the information needs of large-scale digital library users, *Library resources & technical services* **58**(2): 84–99.

Zha, X., Wang, W., Yan, Y., Zhang, J. and Zha, D. [2015]. Understanding information seeking in digital libraries: antecedents and consequences, *Aslib Journal of Information Management* **67**(6): 715–734.

Zhang, L. [2019]. Blockchain: The new technology and its applications for libraries, *Journal of Electronic Resources Librarianship* **31**(4): 278–280.

Zhang, X., Anghelescu, H. G. and Yuan, X. [2005]. Domain knowledge, search behaviour, and search effectiveness of engineering and science students: An exploratory study, *Information Research: An International Electronic Journal* **10**(2): n2.

Zuva, K. and Zuva, T. [2012]. Evaluation of information retrieval systems, *International Journal of Computer Science and Information Technology (IJCSIT)* **4**(3): 35.

Appendices

.1 The online study

The materials of chapter 4 consists of: informed consent of the online study shows in figure 1, the pre-questionnaire in figure 3, and the post questionnaire in figure 3.

.2 The laboratory study

The ethical approval of the user study presented in chapter 5 shows in 4.

.3 The visualisation study

The questionnaires used to evaluate the visualized interface in chapter 7 of the task completion presents in figure 5, and figures 6, 7,and 8 show the test questionnaire.

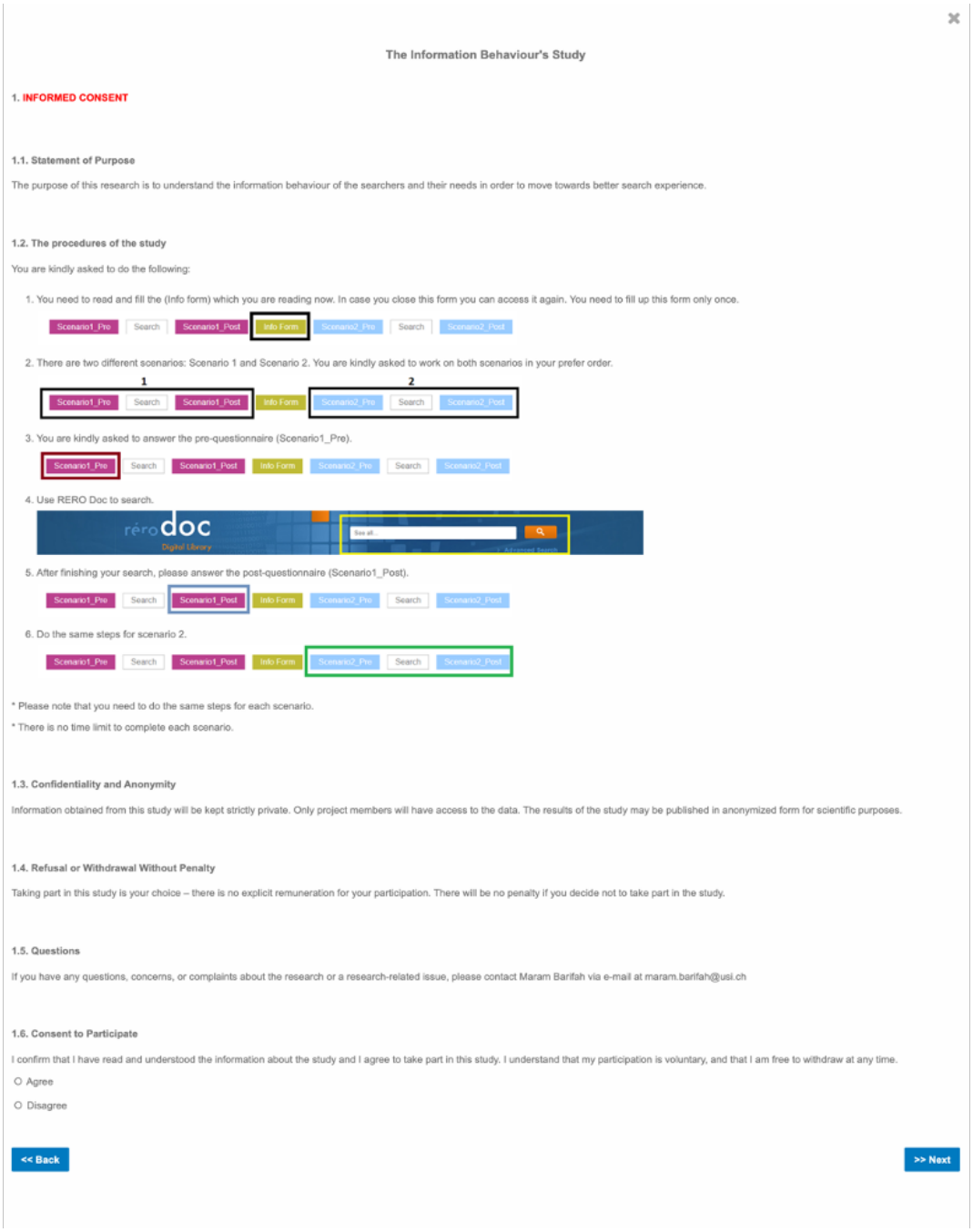


Figure 1. The Informed Consent

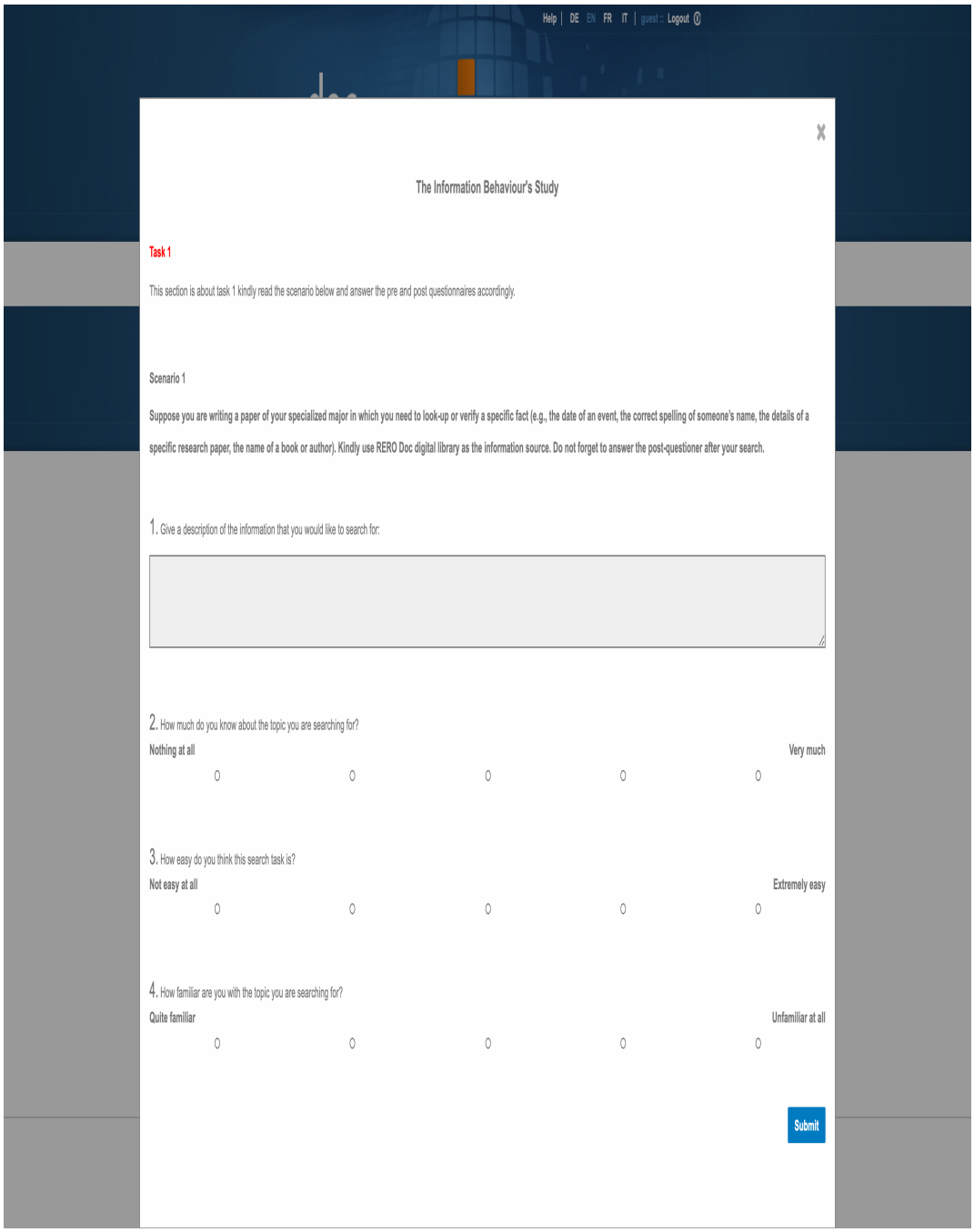


Figure 2. The Pre-Questionnaire

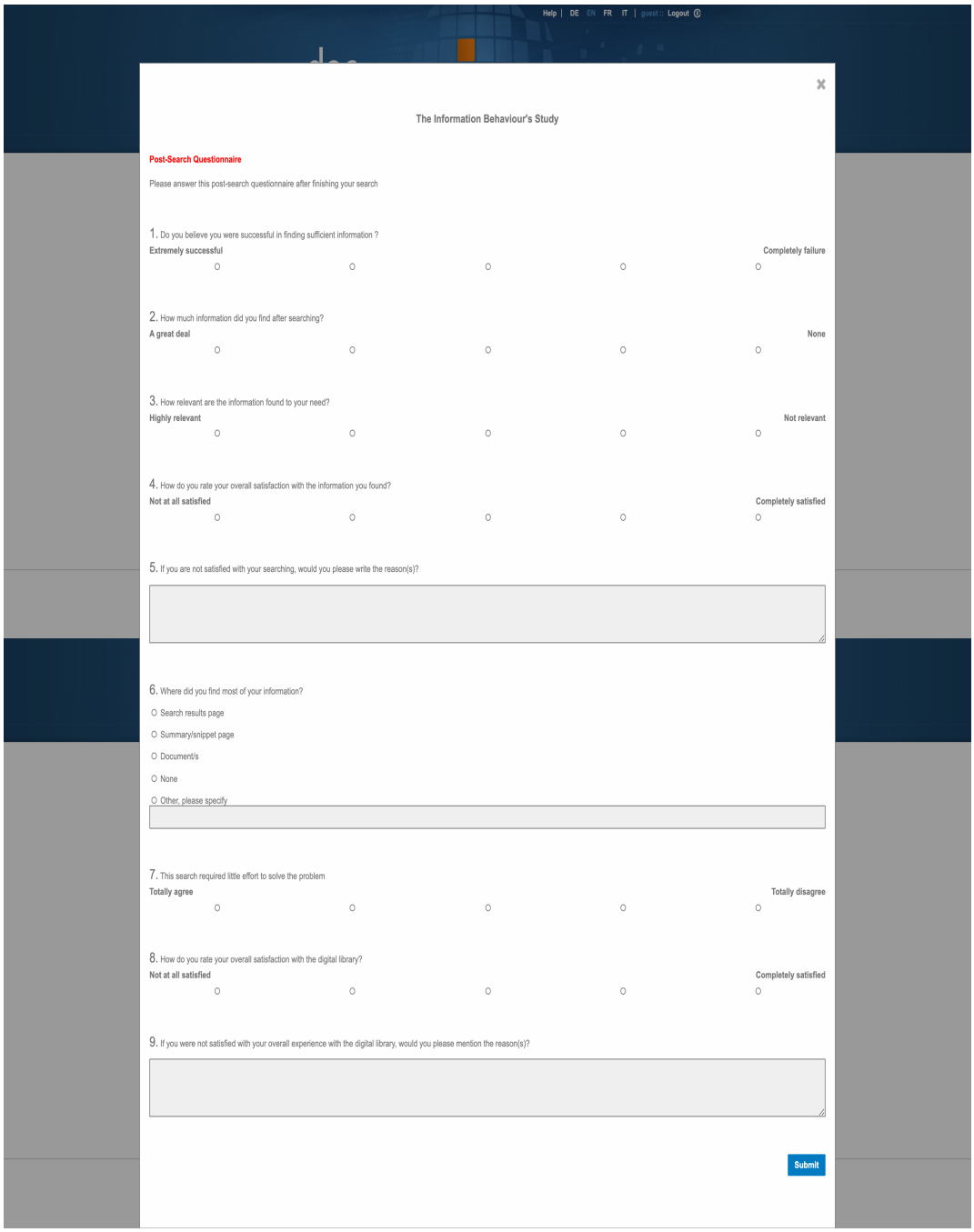
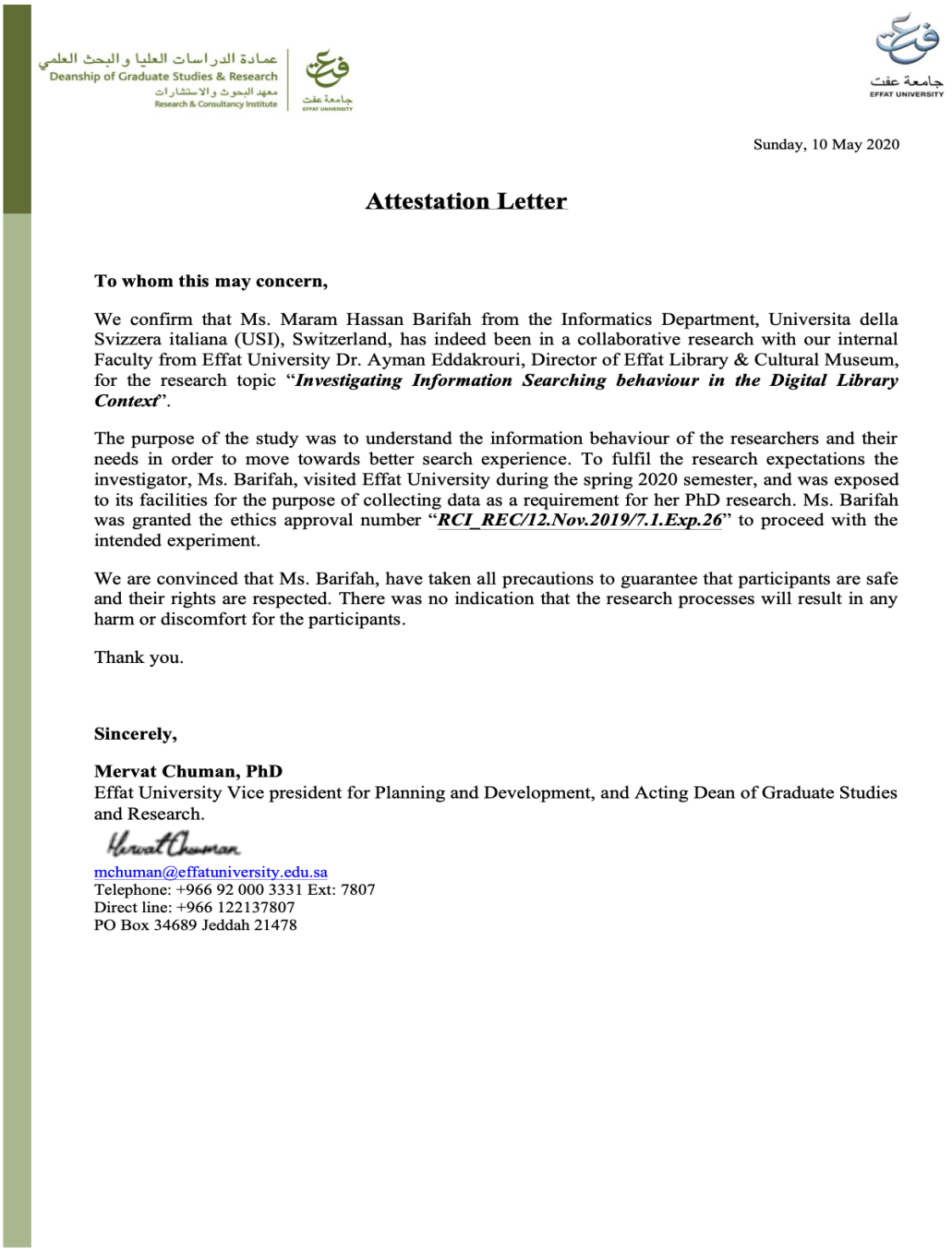


Figure 3. The Post Questionnaire



عمادة الدراسات العليا والبحث العلمي
Deanship of Graduate Studies & Research
معهد البحوث والاستشارات
Research & Consultancy Institute
جامعة عفت
EFFAT UNIVERSITY

جامعة عفت
EFFAT UNIVERSITY

Sunday, 10 May 2020

Attestation Letter

To whom this may concern,

We confirm that Ms. Maram Hassan Barifah from the Informatics Department, Universita della Svizzera italiana (USI), Switzerland, has indeed been in a collaborative research with our internal Faculty from Effat University Dr. Ayman Eddakrouri, Director of Effat Library & Cultural Museum, for the research topic "***Investigating Information Searching behaviour in the Digital Library Context***".

The purpose of the study was to understand the information behaviour of the researchers and their needs in order to move towards better search experience. To fulfil the research expectations the investigator, Ms. Barifah, visited Effat University during the spring 2020 semester, and was exposed to its facilities for the purpose of collecting data as a requirement for her PhD research. Ms. Barifah was granted the ethics approval number "***RCI_REC/12.Nov.2019/7.1.Exp.26***" to proceed with the intended experiment.

We are convinced that Ms. Barifah, have taken all precautions to guarantee that participants are safe and their rights are respected. There was no indication that the research processes will result in any harm or discomfort for the participants.

Thank you.

Sincerely,

Mervat Chuman, PhD
Effat University Vice president for Planning and Development, and Acting Dean of Graduate Studies and Research.

mchuman@effatuniversity.edu.sa
Telephone: +966 92 000 3331 Ext: 7807
Direct line: +966 122137807
PO Box 34689 Jeddah 21478

Figure 4. The Ethical Approval of the User Study

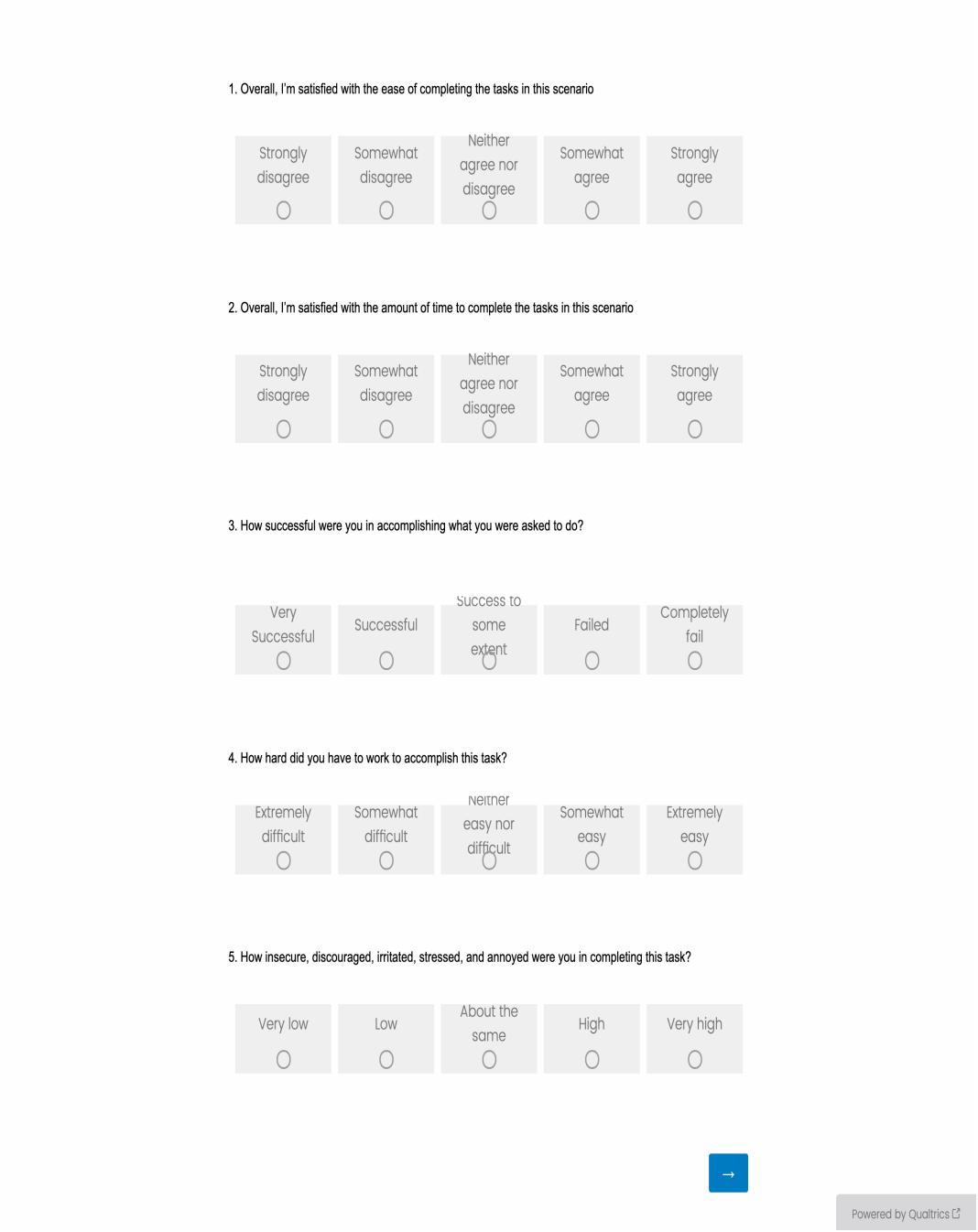


Figure 5. Task Questionnaire

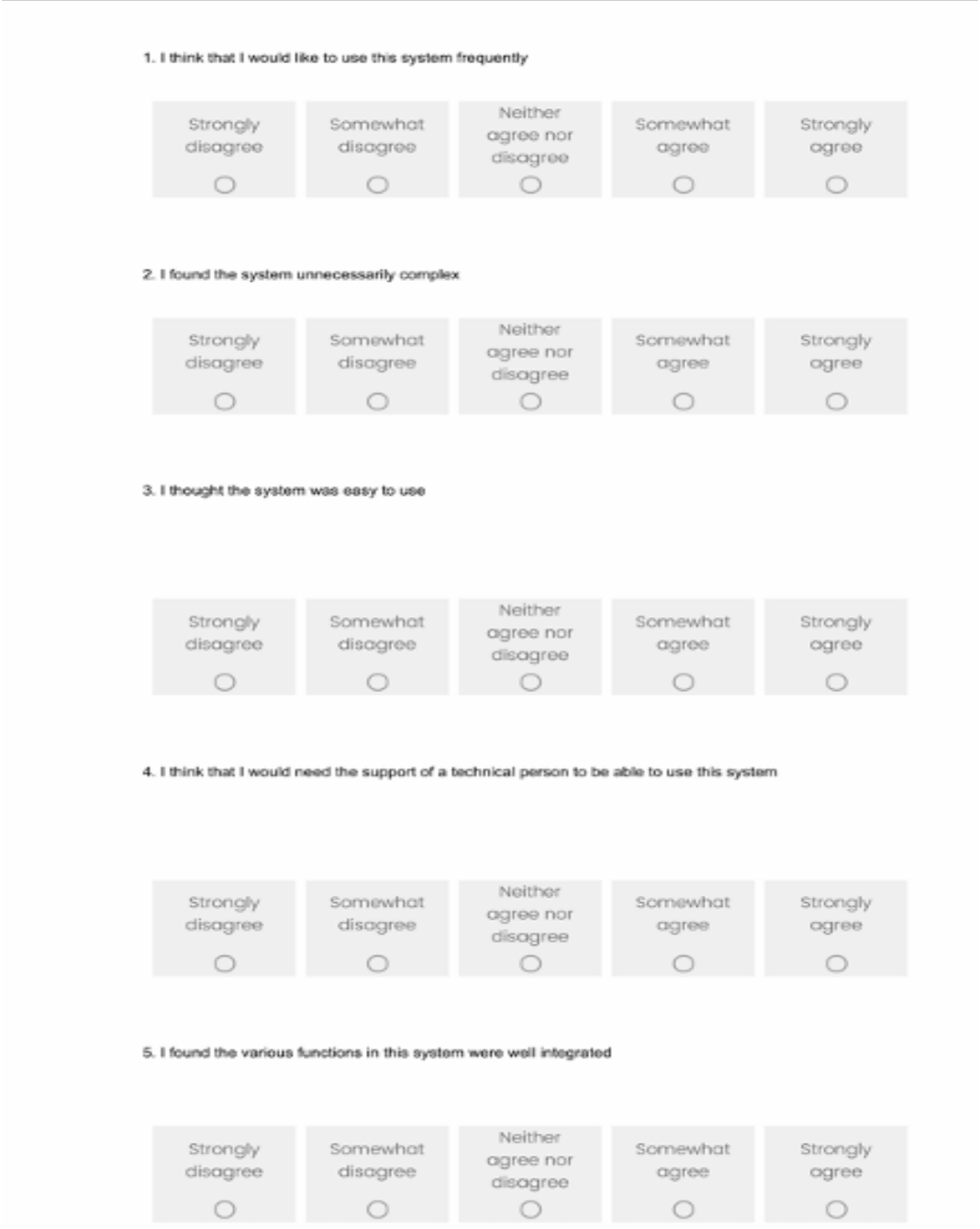


Figure 6. Test Questionnaire Part 1

Figure 7. Test Questionnaire Part 2

Figure 8. Test Questionnaire Part 3