USER BEHAVIOR MODELS FOR EXPLORATORY INFORMATION SEEKING

MICHAEL KOTZYBA

Dissertation

to achieve the academic degree

Doktoringenieur (Dr.-Ing.)

accepted by the Faculty of Computer Science
Otto von Guericke University Magdeburg

from Dipl.-Inform. Michael Kotzyba
born on April 20th, 1986 in Magdeburg

Supervisors:
Prof. Dr. Andreas Nürnberger
Prof. Dr. Ernesto William De Luca
Prof. Dr. Birger Larsen

Magdeburg, February 24th, 2022

Michael Kotzyba: *User Behavior Models for Exploratory Information Seeking,* Dissertation, February 24th, 2022

ABSTRACT

Nowadays it has become a crucial skill to gather information from the Internet to satisfy a current information need. To achieve this goal, the Web provides a wide range of information systems and sources, such as search engines, encyclopedias, forums, social networks, etc. Several theoretical and analytical models have been developed over decades to investigate and understand the activities associated to the corresponding information seeking behavior of the users. To support users' information seeking, several interest-based techniques, such as query suggestions and personalization, are utilized by the majority of search engine providers. However, individual user support regarding the type of search activity, a user is currently engaged in, is still neglected. Especially if the user's information need is complex and requires an exploration of a domain, current means can not provide adequate assistance. Hence, search engines should go beyond the simple fact-finding search paradigm and should support users during the search process itself. The goal of this thesis is the contribution to a deeper understanding of user's information seeking behavior to establish a stronger connection between the theoretical and analytical behavior models. For this purpose, the emphasis of the investigation is placed on the search paradigm of exploratory search. Taking into consideration the multiplicity of information behavior models and the related perspectives, exploratory search can serve as a connecting concept and therefore represents a promising object of investigation. This thesis reviews selected models of information (-seeking) behavior and integrates the paradigm of exploratory search. Furthermore, several user studies, analyzing exploratory search from different perspectives, have been implemented. The obtained study data is used to investigate the influence of users' personal characteristics but also to investigate the (exploratory) search process itself. This includes the analysis, modeling and classification of seeking behavior. Consequently, means to support several aspects of exploratory behavior for search systems will be proposed. Last but not least, a discussion regarding collaborative exploratory information seeking is given.

ZUSAMMENFASSUNG

Die Fertigkeit Informationen aus dem Internet zu akquirieren ist heutzutage immer wichtiger geworden. Für eine erfolgreiche Akquise bietet das Web ein breites Spektrum an Informations-Systemen und -Quellen, wie zum Beispiel Suchmaschinen, Enzyklopädien, Foren, soziale Netzwerke, usw. Über Dekaden hinweg wurden verschiedene theoretische und analytische Modelle entwickelt, um die Aktivitäten, welche mit dem jeweiligen Informationssuchverhalten von Nutzer assoziiert sind, zu Untersuchen und besser zu verstehen. Zur Unterstützung der Informationssuche von Nutzern verwendet die Mehrheit der heutigen Suchmaschinenbetreiber interessenbasierte Techniken, wie zum Beispiel Suchanfragenvorschläge oder Personalisierung. Die individuelle Unterstützung von Nutzern bezüglich einer aktuell angewendeten Suchaktivität findet jedoch noch immer wenig Beachtung. Insbesondere wenn das Informationsbedürfnis eines Nutzers komplexer ist und daher eine Exploration der Domäne erfordert, kann nahezu keine (systemseitige) Assistenz angeboten werden. Aus diesem Grund ist es erstrebenswert, dass Suchmaschinen über das gängige Paradigma der einfachen Faktensuche hinaus gehen und dem Nutzer eine Unterstützung während des Suchprozesses selbst anbieten. Das Ziel dieser Doktorarbeit ist es, zum tieferen Verständnis des Informationssuchverhaltens von Nutzern beizutragen, um eine stärkere Verknüpfung zwischen den theoretischen und analytischen Modellen zu etablieren. Zu diesem Zweck wird der Schwerpunkt der Untersuchungen auf das Paradigma der Explorativen Suche gelegt. Unter Berücksichtigung der vielfältigen Modelle des Informationsverhaltens und deren jeweiligen Perspektiven, dient die Explorative Suche als verbindendes Konzept und stellt damit ein vielversprechendes Untersuchungsobjekt dar. Diese Arbeit betrachtet verschiedene, ausgewählte Modelle des Informations-(Such)-verhaltens und integriert die Explorativen Suche in diese. Weiterhin wird die Umsetzung mehrerer Nutzerstudien zur Analyse der Explorativen Suche aus unterschiedlichen Perspektiven beschrieben. Die erhaltenen Studiendaten werden verwendet, um den Einfluss persönlicher Nutzercharakteristika, aber auch den (explorativen) Suchprozess selbst zu untersuchen. Dies beinhaltet die Analyse, Modellierung und Klassifizierung des Suchverhaltens. Infolgedessen werden Ansätze zur Unterstützung des explorativen Nutzerverhaltens für Suchsysteme vorgeschlagen. Zum Schluss wird die explorative Informationssuche auch im Kontext der Zusammenarbeit mehrerer Nutzer diskutiert.

PUBLICATIONS

The content of this thesis has already partially appeared in the following own publications (in alphabetical order):

[1] Peter Butka, Thomas Low, Michael Kotzyba, Stefan Haun, and Andreas Nürnberger. "A Framework for FCA-based Exploratory Web Search." In: *Proc. of the 1st Int. Symposium on Companion-Technology, ISCT '15*. 1. COST. Ulm, Germany, 2015, pp. 131–136.

[2] Peter Butka, Thomas Low, Michael Kotzyba, Stefan Haun, and Andreas Nürnberger. "Exploration of Web search results based on the formal concept analysis." In: *Semanitic Keyword-based Search on Structured Data Sources*. Vol. 10546. Third Int. KEYSTONE Conference, IKC '17. Gdańsk, Poland: Springer, 2017, pp. 123–135.

[3] Tatiana Gossen, Michael Kotzyba, and Andreas Nürnberger. "Anforderungen und Verhalten junger Nutzer bei der Informationssuche - Ein Überblick." In: *Information-Wissenschaft & Praxis* 66.1 (2015), pp. 57–64.

[4] Tatiana Gossen, Michael Kotzyba, and Andreas Nürnberger. "Knowledge Journey Exhibit: Towards Age-Adaptive Search User Interfaces." In: *Advances in Information Retrieval*. Vol. 9022. 37th European Conference on IR Research, ECIR '15. Vienna, Austria: Springer, 2015, pp. 781–784.

[5] Tatiana Gossen, Michael Kotzyba, and Andreas Nürnberger. "Search Engine for Children: User-Centered Design." In: *Datenbank-Spektrum* 17.1 (2017), pp. 61–67.

[6] Tatiana Gossen, Michael Kotzyba, Sebastian Stober, and Andreas Nürnberger. "Voice-Controlled Search User Interfaces for Young Users." In: *7th Annual Symposium on Human-Computer Interaction and Information Retrieval*. HCIR '13. 2013, pp. 2–5.

[7] Michael Kotzyba, Ingo Siegert, Tatiana Gossen, Andreas Wendemuth, and Andreas Nürnberger. "Exploratory Voice-Controlled Search for Young Users: Challenges & Potential Benefits." In: *Kognitive Systeme, DuEPublico, Duisburg-Essen Publication Online* 2015-1 (2015), pp. 1–10.

[8] Michael Kotzyba, Denis Ponomaryov, Thomas Low, Marcus Thiel, Birte Glimm, and Andreas Nürnberger. "Ontology - supported Exploratory Search for Physical Training Exercises." In: *Demonstrations Track at the 14th Int. Semantic Web Conference*. Vol. 1486. ISWC '15. 2015.

[9] Michael Kotzyba, Tatiana Gossen, Johannes Schwerdt, and Andreas Nürnberger. "Exploration or Fact-Finding: Inferring User's Search Activity Just in Time." In: *Proc. of the Conference on Conference Human Information Interaction and Retrieval*. SIGIR CHIIR '17. Oslo, Norway: Association for Computing Machinery, 2017, pp. 87–96.

[10] Michael Kotzyba, Tatiana Gossen, Sebastian Stober, and Andreas Nürnberger. "Model-Based Frameworks for User Adapted Information Exploration: An Overview." In: *Companion Technology*. Cognitive Technologies. Springer, 2017, pp. 37–56.

[11] Michael Kotzyba, Johannes Schwerdt, Tatiana Gossen, Martin Krippl, and Andreas Nürnberger. "The Effect of Motivational Goals on Information Search for Tasks of Varying Complexity Levels." In: *Int. Conference on Systems, Man, and Cybernetics*. SMC '18. Miyazaki, Japan: IEEE, 2018, pp. 2602–2607.

[12] Thomas Low, Nikola Bubalo, Tatiana Gossen, Michael Kotzyba, André Brechmann, Anke Huckauf, and Andreas Nürnberger. "Towards Identifying User Intentions in Exploratory Search using Gaze and Pupil Tracking." In: *Proc. of the Conference on Human Information Interaction and Retrieval*. SIGIR CHIIR '17. Oslo, Norway: Association for Computing Machinery, 2017, pp. 273–276.

[13] Johannes Schwerdt, Michael Kotzyba, and Andreas Nürnberger. "Inferring User's Search Activity using Interaction Logs and Gaze Data." In: *Int. Conference on Companion Technology*. ICCT '17. Ulm, Germany: IEEE, 2017, pp. 1–6.

[14] Johannes Schwerdt, Michael Kotzyba, and Andreas Nürnberger. "Fact-Finding or Exploration: Identifying Latent Behavior Clusters in User's Search Activities." In: *Int. Conference on Systems, Man and Cybernetics*. SMC '19. Bari, Italy: IEEE, 2019, pp. 1465–1471.

[15] Dominic Stange, Michael Kotzyba, and Andreas Nürnberger. "Professional Collaborative Information Seeking: Towards Traceable Search and Creative Sensemaking." In: *Transactions on Computational Collective Intelligence XXVI, TCCI*. Vol. 10190. Springer, 2017, pp. 1–25.

[16] Dominic Stange, Michael Kotzyba, Stefan Langer, and Andreas Nürnberger. "Interpreting and Leveraging Browser Interaction for Exploratory Search Tasks." In: *Proc. of the 19th Int. Conference on Enterprise Information Systems: ICEIS*. Vol. 3. Porto, Portugal: ScitePress, 2017, pp. 191–197.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

Part I

INTRODUCTION

> *"In thinking about this holistic view of people's worlds, the internet itself could serve as a metaphor for information behavior and the way our view of it has changed."*
>
> — Qu & Furnas [150], p. 1.

1 THESIS TOPIC AND GOALS

"Humans are explorers by nature, ...", as simple and true this proposition from White & Roth [189] sounds, as hard and vague it gets, at the same time, if an elaborated explanation or even a proof is desired. This thesis is dedicated to the topic of humans' exploration, and in particular users' exploratory search behavior in the vast domains of the Internet's digital information sources. To approach the topic, the thesis reviews the literature of the diversified area on human information behavior, the underlying information need and related theoretic information seeking behavior models. This facilitates a deductive consideration of various aspects to investigate (exploratory) search behavior. Furthermore, the thesis provides an overview of application oriented, analytical user models, which aim to represent and predict user's interactions, e.g., with search engines. This facilitates an inductive consideration of users' (exploratory) search behavior because concrete interactions with a given search system can be investigated. In between this two approaches and the related research areas with their well elaborated methodologies, the literature exhibits a gap. This gap avoids to have a seamless deductive, respectively inductive, view on human seeking behavior performed on specific search systems. Although the gap exists since the beginning of information science, it became more tangible with the rise of the Internet and the ability to log user's interactions with search engines. However, the gap still exists and hinders the development of adequate user support which is based on a holistic perspective of human's seeking behavior and necessary in complex search scenarios. To contribute bridging the gap, the investigation of exploratory search is a promising approach because the underlying information need is usually complex. Furthermore, the characteristics of exploratory search allow to investigate and establish a conjunction between the theoretical information behavior models and the analytical models at the same time. However, exploratory search itself is still not well understood and numerous questions regarding this multifaceted but downright natural behavior are currently open. Hence, in this thesis, some of the open questions will be answered by the mentioned literature reviews, by analyzing the data of several specially implemented user studies and by describing a suitable methodology for modeling exploratory search.

In the following, this first chapter provides a motivation for the thesis' topic from a user's perspective (Sect. 1.1). Afterwards, the research questions and hypothesis of this work are derived (Sect. 1.2). To provide an adequate entry point and overview for the reader, the outline of the thesis' structure and content is given (Sect. 1.3). Finally, some limitations of the corresponding investigations and findings of this work are discussed (Sect. 1.4).

1.1 MOTIVATION

With the proliferation of the Internet, the quantity of available information has increased enormously. Internet-based platforms, such as forums, encyclopedias, home pages, social networks and multimedia portals, have acquired more and more digital content. In addition to the growing availability and consumption of this content, the target group of Internet users has become very wide and heterogeneous. Users differ in age, knowledge, experience, cultural background, etc. For many users, the Internet itself but also the related interactions and the search for information have become a relevant aspect in daily life [190]. This leads to high requirements for search systems and tools which provide access to the information. The requirements also increase for related disciplines, such as Information Retrieval (IR) and Information Science (IS), in general.

In order to find relevant information, search systems not only have to retrieve web sites related to the provided search queries but also have to consider and adapt to the user's context. Especially for heterogeneous user groups it is challenging to precisely pick the relevant web documents since the individual user expectations may differ: While a tourist wants to know the "weather" in terms of the expected rain probability and temperatures to prepare the equipment and supplies for the next travel destination, a becoming meteorologist may rather be interested in the correlation of air-streams and altitudes to investigate how the "weather" develops. Current web search engines only partially provide supporting features to adapt the query suggestions or the search result set, e.g., by considering the user's location, previously used search queries or already visited result pages [13]. If the user's information need consists in a single fact-based information (e.g., the expected rainfall for the next day at a specific location), present search tools are usually helpful to solve the task relatively efficiently. However, if the user's information need can not be satisfied by a single fact or is even unclear, and thus the user has to explore the (search) domain, current means can not provide adequate assistance. Therefore, search engines should go beyond the simple fact-finding paradigm and should support the user during the search process itself. Unfortunately, a holistic model for user-adapted information exploration is currently missing. Already in 1991 Carol Kuhlthau [120] identified that:

> *"There appears to be a gap between the system's traditional patterns of information provision and the user's natural process of information use." (p. 1)*

To develop adequate search assistance further, search systems can benefit from literature's theoretical models (rooted in the library science) and from application-oriented, analytical user models (emerged in response to the amount of search related interactions performed in the Internet). An appropriate user model in the background enables the system to adapt to the user's interests, preferences and needs and therefore, might lead to a significant increase in retrieval performance and user experience. Unfortunately, the gap between both model types, the holistic theoretic models on the one hand and the application-oriented, analytical models on the other, is still huge. Though a deeper understanding of users' information seeking behavior on current search systems can help to narrow the gap and in particular if aspects of exploration are considered.

Both model types are extensively described in Chap. 2 and Chap. 3 resp.

Information seeking behavior models are described in Chap. 2.

1.2 RESEARCH QUESTIONS & HYPOTHESES

If humans want to know something, they ask questions and usually it is beneficial if they are able to specify the questions. On the one hand, it is beneficial for the person or information system that is asked to provide a preferably specific answer. On the other hand, the specification is also beneficial for the questioner to evaluate how well a given answer matches to the question. This enables the questioner to decide whether the underlying information need, that shall be satisfied, demands for more questions and clarification or not. If a user (e.g., in interaction with a search engine) is able to satisfy the perceived information need by answering clear fact related questions, the related search behavior is called lookup search. This kind of information acquisition is usually efficient and most of the current information systems are able to support the user adequately. However, a precise formulation of questions to satisfy an information need demands for knowledge about the domain, respectively the context of the need. If the information need can not be formulated precisely (e.g., due to little knowledge of the user or a high task complexity), the user needs to explore and investigate the domain. This is called exploratory search. From the research perspective it is still not well known how the process of exploration works; what kind of behavioral characteristics are relevant; and how exploratory search can be identified, respectively distinguished, from other search activities. Hence, the thesis' first central question emerges:

Def. Lookup (search): cf. Sect. 2.6.1

Def. Exploratory Search: cf. Sect. 2.6

Q1: *How do users behave during exploratory information seeking?*

Thesis Question Q1

As pointed out above, information(-seeking) behavior models are often theoretical in their nature what leads to the general demand for application-oriented, analytical approaches to capture user's seeking behavior in real world search environments. To extend and further contribute to the research in that area, the second central question of the thesis is:

Thesis Question Q2

*Q*2: *How can user's exploratory information seeking be modeled?*

If the user's seeking behavior is analyzed and therefore, can be modeled (and identified), the next step is to use the gained knowledge in practical context. Here the third question arises:

Thesis Question Q3

*Q*3: *How can user's exploratory information seeking be supported?*

The three questions *Q*1, *Q*2 and *Q*3 still cover a broad area and include a multitude of research questions, respectively hypotheses. Furthermore, the paradigm of exploratory search was developed rather in parallel to the theoretical and analytical model approaches, and thus the relation to both has to be examined in more detail[1]. To split the three prior questions into manageable parts and integrate the paradigm of exploratory search at the same time, five hypotheses are proposed: First, to investigate how users behave during exploratory information seeking (*Q*1), a theoretical foundation is necessary that links the paradigm of exploratory search to the models and methods of information (seeking) behavior:

Hypothesis H1

*H*1: *Exploratory search can be integrated from a theoretical point of view in different information -behavior and -seeking models.*

In particular, the theoretical framework and models from the information and library science provide an adequate background for this integration from different perspectives.

If the target group is heterogeneous, the user's individual characteristics become more relevant and represent potential influencing variables. This variables may contribute to a better understanding of user's exploratory information seeking (*Q*1). Hence, user's individual characteristics are covered by the second hypothesis:

Hypothesis H2

*H*2: *Personal characteristics (e.g., psychological, demographical, or experience) are influencing information seeking behavior and hence, exploratory search behavior.*

To address the thesis' question *Q*2, the third hypothesis implements the analysis and differentiation of exploratory search. This is accomplished by modeling and classifying two types of search activities, namely factual (as specification of lookup) and exploratory search:

1 In the Sections 2.6 and 3.3.1 the relation between exploratory search and the theoretical as well as the analytical models is explicated respectively.

*H*3: *Exploratory search can be analyzed, modeled and classified incorporating different model parameters and interaction levels.*

Hypothesis H3

In particular, the model parameters are utilized to find an adequate balance between the model's prediction and complexity. The interaction levels on the other hand are related to user's explicit and implicit interactions with the search systems.

With more knowledge about user's exploratory information seeking behavior (*Q*1) and with the ability to model and identify exploratory search (*Q*2), the next consequential step is to utilize the findings to the benefit of users. However, it is not well elaborated what kind of approaches at all are possible to provide beneficial supporting mechanisms for exploratory behavior. Therefore, the fourth hypothesis deals with the user's support during exploration (*Q*3):

*H*4: *Characteristics of exploratory search can be supported by different front- and back-end components of interactive information retrieval systems.*

Hypothesis H4

While hypothesis *H*4 considers a traditional scenario, where one user is performing an exploration on the individual level, the next and last hypothesis will consider a different user setting. Literature suggests that exploratory search is related to complex search tasks. In some cases it is assumed that the related search task(s) can reach a level of complexity and/or requires an amount of domain knowledge what can barley be handled by one person in time. To discuss this situation as well, the fifth hypothesis considers aspects of exploratory search on a collaborative level and thereby addresses *Q*3 as well:

*H*5: *The complexity of exploratory search allows and demands a discussion on collaborative information seeking to find appropriate approaches, especially in competitive (business) environments.*

Hypothesis H5

The three questions *Q*1 to *Q*3 and the related five hypotheses *H*1 to *H*5 are building the framework of the thesis' research and investigations to gain more insights into the area of exploratory information seeking. While *H*1 and *H*5 can and will be considered from a more theoretical perspective, the other three hypotheses, *H*2, *H*3 and *H*4 require a more empirical procedure. An outline of the thesis's structure and content as well as an overview, how the proposed hypotheses are processed, is given in the following section.

1.3 OUTLINE OF THE THESIS

This thesis consists of four parts, whereat each of the following paragraphs describe the parts and their chapters respectively: This first

Outline of Part i

Part i introduces the thesis and includes the first chapter. Chapter 1 is dedicated to describe the topic and goals of the thesis. It provides the motivation for the implemented investigations (Sect. 1.1), specifies the research questions and derives the corresponding hypotheses (Sect. 1.2). Furthermore, the thesis' structure and contents are outlined (Sect. 1.3). The first chapter concludes with a description of limitations regarding the thesis' objects of investigation (Sect. 1.4).

Outline of Part ii

The second Part ii provides the fundamentals that are necessary for the argumentation and investigations in this work. Part ii includes Chapter 2 and Chapter 3, which approach the thesis' topic from a deductive and an inductive view respectively. Chapter 2 provides an overview of humans' information behavior (Sect. 2.3), the relation to the underlying information need (Sect. 2.1); outlines the foundations of potential related user characteristics ($H2$); and provides an overview of related research areas (Sect. 2.2). The chapter continues with a review of selected theoretical models of information seeking (Sect. 2.4) and search behavior (Sect. 2.5); and introduces the search paradigm of exploratory search (Sect. 2.6). Furthermore, the integration of exploratory search from a theoretical point of view in different information -behavior and -seeking models ($H1$) is exemplified. Chapter 3 defines and discusses several application-oriented, analytical user behavior models and begins with a discussion of the model's application areas and challenges (Sect. 3.1). In the related research area(s), in particular user models to investigate the search behavior on the Search Engine Result Page (SERP) are popular and hence, are exemplified (Sect. 3.2). However, the investigation of exploratory search, as projected in this thesis, requires a more general approach than just analyzing the SERP behavior. Hence, Chapter 3 concludes with the discussion and definition of models that are appropriated for analyzing search activities and modeling information seeking (Sect. 3.3) in the context of exploration ($H3$). The approaches described in Chapter 3 usually lack for direct relations to the theoretical models described in Chapter 2, what causes the gap, addressed in this thesis. Nevertheless, the approaches enable the analysis, categorization or even prediction of users behavior and related search interactions. Some works bridging the theoretical and analytical perspective have already been published and are discussed in Chapter 3 as well. The here selected and defined models are used and further developed for the investigations in Chapter 5.

Outline of Part iii

The third Part iii of this thesis provides the procedures and methods applied to accomplish the proposed investigations. Part iii includes Chapter 4 and Chapter 5, which describe the generation and utilization of the user interaction data respectively. Chapter 4 provides a full description of all user studies which have been conducted to generate the necessary data for this thesis. The chapter begins with a motivation and a brief overview of all user studies before the details are exemplified (Sect. 4.1). The first user study *US-I* is dedicated to the user variable *age* and aims to investigate the (exploratory) search behavior of young users (Sect. 4.2). Since studies with users of lower ages bring their own challenges, *US-I* consists of two (sub-) user studies, namely *US-Ia* and *US-Ib*. In the second user study *US-II*, exploratory and fact-finding search activities of adults have been recorded (Sect. 4.3). The here obtained interactions serve as prior source to describe and analyze exploratory search on the one hand but also are used to exemplify the methodology for adequate user modeling (in Chapter 5) on the other hand (*H*3). To extend and confirm the investigations on exploratory and fact-finding search activities, the interaction of over one hundred uses have been recorded in the third user study *US-III* (Sect. 4.4). In addition to the interaction data, several user characteristics have been obtained to investigate the influence on exploratory but also fact-finding search (*H*2). In Chapter 5, the knowledge of all provided foundations, discussed approaches and defined models is utilized in combination with the generated data to investigate exploratory information seeking. At first, the here used (Markovian) models are defined and trained regarding the corresponding study data (Sect. 5.1). This allows to analyze and compare the models and in turn to reveal first (empirical) insights regarding the paradigm of exploratory search (*H*3). Afterwards, the models are used in a classification setting to investigate the ability to identify exploratory search (Sect. 5.2). Several approaches of parameter optimization are applied to tune the models and classifier but also to reveal relevant aspects of exploratory search behavior (*H*3). In a next step, the models are used in a clustering setting (Sect. 5.3). This approach is twofold. On the one hand, the study design can be validated and on the other hand, latent behavior clusters in user's search activities can be identified. Furthermore, Chapter 5 investigates influencing variables (*H*2) on exploratory search in terms of personal user characteristics (Sect. 5.4). Finally, the revealed findings are used to derive and propose promising approaches to support (*H*4) exploratory search (Sect. 5.5).

Outline of Part iv

The last Part iv provides a discussion and includes the sixth chapter. Chapter 6 discusses the revealed insights from different perspectives. At first, a discussion of exploratory search in a collaborative information seeking scenario (*H*5) is given (Sect. 6.1). Afterwards, the thesis' results in consideration of the hypotheses (*H*1 to *H*5) and research questions (*Q*1 to *Q*3) are reviewed and discussed (Sect. 6.2). Building on all revealed findings of the thesis, possible future work will be outlined (Sect. 6.3) and finally, a conclusion is given (Sect. 6.4). Last but not least, the Appendix v completes the thesis by providing several details which have been outsourced for the sake of comprehensibility.

1.4 LIMITATIONS

Most theories of information behavior and related models from the research area of IS describe users, their behavior and the used information sources on a quite abstract level. The information sources here can comprise (analog) books, (digital) messages, web pages or conversations with other people independent of the medium or environment. Certainly, the consideration of any kind of information source by the information behavior models is desirable and necessary for a general perspective and the model's validity. Though, for the sake of limitation, the scope of this thesis will be restricted to investigations of current IR systems (as information source) what still remains a challenging task. That is, the thesis focuses on formal, in particular, digital, Internet-based information sources retrieved by search engines, such as web pages, forums and encyclopedias. Accordingly, the investigations regarding the information behavior will be restricted to user interactions of (primary) individuals with search engines. Furthermore, it has to be mentioned that the revealed findings are limited to the specific user groups acquired for the corresponding user studies. Although care was taken on all participant acquisitions, and especially for user study *US-III* with it's over one hundred users, an argumentation for and the derivation of generalized statements have always be done with caution. An additional limitation is the following: The integration but also the modeling, analysis and identification of exploratory search are the main contribution of the thesis. The development of potential supporting means for exploratory search as consequential step is discussed extensively as well. Though, additional user studies to measure in how far this approaches benefit users' exploration would be an own thesis and hence, this thesis rather serves as an comprehensive entry point for such research directions.

Last but not least, a further important aspect of limitation has to be noted: User models and user behavior models are per definition restricted to the aspects they are made for. For example, a model to differentiate between user's search activities only estimates which activity is the most likely one under the given current user interactions, the previous training data and the given modeled (and used) variables. There can always be further (sub-)behavior or unexplored and not yet considered but relevant variables influencing the recorded behavior. Assuming a user is seeking for information and is in company. Than, a friend asks for a (small) different search task and/or even temporary is taking the control of the search system. Hence, the search activity (presumed to be from only one user with one task in mind) is not unbiased anymore. This can lead to erroneous estimations of the search activity by the user model. That is, it can always happen the (restricted) user models do not accurately represent individuals or groups. Hence, it remains under the (scientific) responsibility of those who apply and utilize the models to assess and act appropriately if essential decisions of any kind have to be made based on the calculated results. Nevertheless, of course the (here proposed and investigated) models shall be used to analyze users' (information) behavior; to reveal new insights; and to develop new means and technologies to support users in their daily (seeking) activities.

Part II

FUNDAMENTALS

"Information seeking is a fundamental human activity ... "
— White & Roth [189], p. 3.

2 THEORETICAL FOUNDATION AND RELATED WORK

Research regarding human's information behavior has a long tradition and a huge theoretical background rooted in the early empirical studies of library users and readership studies [17, 192]. Users' information (seeking-) behavior has been investigated in different ways by researchers, e.g., by observing the keywords applied during the search; recording the selected information sources that have been utilized; or considering how people used specific documents. Furthermore, self-reports have been made and eventually all collected observations and results were analyzed from phenomenological and intentional perspectives. The derived theories are empirical supported and cover a broad spectrum of human information acquisition. This chapter addresses the related work regarding theoretical aspects of humans' information behavior (Sect. 2.3) as well as related and relevant models of information seeking (Sect. 2.4), respectively search (Sect. 2.5) behavior. Furthermore, an introduction to and the integration of the paradigm of exploratory search (H1) is given (Sect. 2.6). Though, to provide a smooth entry to the theoretical background, first the origin and stimulus of almost all information (seeking) related behavior, namely the underlying information need (Sect. 2.1), will be discussed, followed by an overview of related research areas (Sect. 2.2). The content of this chapter is partially based on own relevant work concerning model based frameworks for user adapted information exploration that was published in [117].

2.1 THE NEED FOR INFORMATION

Theories and models describing humans' information behavior often hypothesize and/or imply the emergence of a perceived need for information as one of the initial steps for any further informational actions. However, an extensively motivation to quote the information need as a first step in the models is rarely given by the authors. One exception is Wilson [193], who describes the environmental (e.g., political, economical, technological), social (role-related) and personal (e.g., personal traits) context of the information need and explains corresponding so-called barriers. He also argues that, depending on the given context, the information need may differ and he derives the need for information as a *"secondary need"* that is a consequence of more basic needs. Furthermore, Wilson [193] provides references to early studies from the 1960's, e.g., by Warner et al. [185], which confirm that the satisfaction of everyday needs also depends on information and hence, depends on the fulfillment of the corresponding information need. Warner et al. [185] even stated that:

> *"The individual without information is indeed powerless in a modern society."* (p. 10)

what highlights the value of information for any individual and the importance of information providing systems in general. That is, the need for information and it's satisfaction is an essential part of the information behavior. In the following, and in addition to the remarks above, a motivation for the information need in context of this thesis and in relation to the concept of exploration is given:

Any motile animal, and so also human, is inherently driven by the elementary and essential need to explore the surrounding environment by movement and perception.[1] This behavior, what can also be interpreted as curiosity, allows the individual to reveal and discover new places and opportunities and therefore, to gain new experience. During the discovery, the individual can accumulate information of the surrounding environment by observation and interaction. The gathered information is stored, processed, abstracted and used to build patterns that continuously adapt the individual's behavior for following exploration steps and situations.[2] This iterative process enables the individual to integrate and maybe understand it's environment; to derive associations, respectively knowledge; and to execute or even to plan for future actions. With the increasing amount of experience, the individual can perform more and more sophisticated actions, what allows the individual to survive and to satisfy further developed and perceived needs in general. A brief outline of these (general) needs is given in the following.

Maslow's hierarchy of needs

The topic of human needs has been investigated from different disciplines such as psychology, social science, philosophy, theology, etc. In 1943, Abraham Maslow developed one of the most popular approaches, called the classical hierarchy of needs [132]. According to Maslow's model, the physiological and safety needs and therefore, all actions to satisfy these needs, build the fundamental layers for an individual, cf. Fig. 2.1. Physiological needs represent the first basic survival oriented needs to keep the body alive and allow reproduction.

---

1 Marcia Bates derives a similar conclusion in her final discussion about the concept of *browsing* and it's inherent, initial motivation; cf. [16], Sect. *Summary and Conclusion*.

2 This process can be basically interpreted as the concept of *learning*. According to a definition from Richard E. Mayer: *"Learning is the relatively permanent change in a person's knowledge or behavior due to experience. This definition has three components: 1) the duration of the change is long-term rather than short-term; 2) the locus of the change is the content and structure of knowledge in memory or the behavior of the learner; 3) the cause of the change is the learner's experience in the environment rather than fatigue, motivation, drugs, physical condition or physiologic intervention."* -from Learning in Encyclopedia of Educational Research-. Furthermore, Carol Kuhlthau uses a similar argumentation as foundation of her *Information Search Process* (cf. [120], Sect. *Theoretical Foundation of the ISP*), which in turn partially builds on results of Howard Gardner [66].

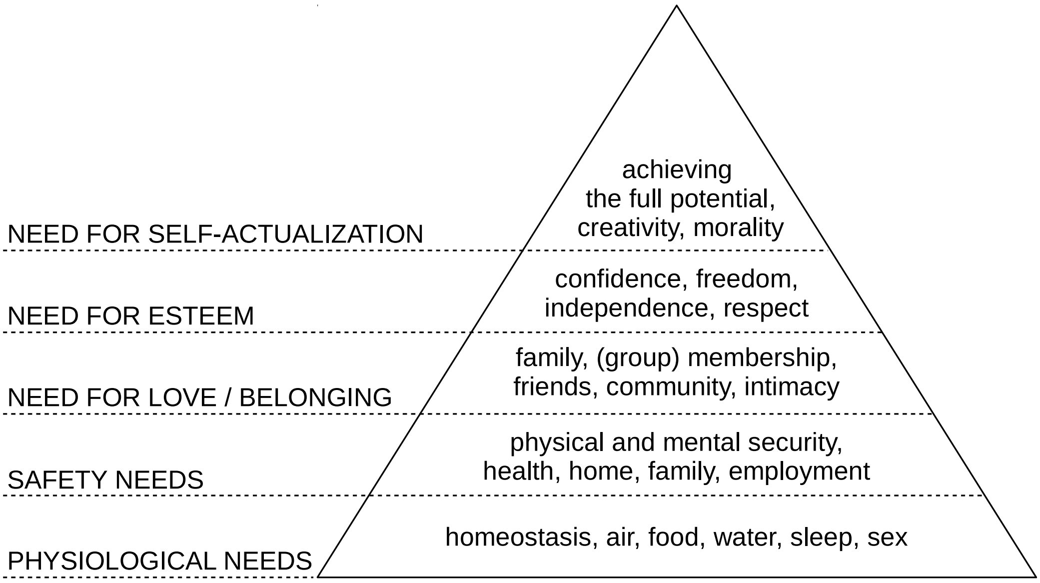


Figure 2.1: Illustration of Maslow's hierarchy of needs according to [132].

Safety needs, as the second layer of the hierarchy, are similar to the physiological needs but in less extend and with an additional aspect to prepare for the individual's future. Once the two basic layers are satisfied, individuals turn towards the next higher ones, the social and psychological needs. In a developed society, these higher level needs play an important role. They make humans to communicate with each other; to build families and other groups or communities; and they make us to be part of the society. With the digitization, the proceeding integration of data- and knowledge bases and the inclusion of social components (e.g., by social networks), also the need for and the value of information further increases, e.g., to satisfy the need for love/belonging, for esteem and/or for self-actualization[3]. The necessary information can be acquired in the own, analog environment (e.g., via books, by friends, etc.) or via digital platforms in the best case appropriately supported by the means of Information Retrieval (IR). In 2012, the study of Jean et al. [100] showed that the most stated information behavior, of people who frequently use the Internet, is reading and searching with 69.1%. Furthermore, the most stated intentions to use the Internet were to keep up-to-date (40.4%) and to gather data (35%).

Summing up, the need for information is a fundamental aspect and contributes to the fulfillment of the basic needs to guarantee surviving on the one hand but is also involved in the fulfillment of the higher layered needs on the other hand. Therefore, the information need runs (in some degree) in parallel to Maslow's hierarchy of needs and

3 In fact, also Bates [17] underlines the relevance of social aspects as context of information behavior research.

it becomes more clear why the need for information is often placed as one of the initial steps in the corresponding theories on humans' information behavior. However, to break away from the idea of a static hierarchical order of needs in general, Maslow himself noted human need's as being relatively fluid and that a hierarchical lower need not necessarily has to be fulfilled completely to desire a hierarchical higher one. This of course also depends on the individual's experience. Last but not least, even after decades of research, the concept of information need, as (first) part of information behavior, still remains as a challenging issue because it is identified as a subjective experience in a person's mind that can not (yet) be observed directly, cf. [195]. Though, the research of Moshfeghi et al. [136] showed first findings where fMRI data from human brain regions was successfully used to recognize differences in the brain activity depending on users who experienced a need for information or not.

2.2 MODELS & RELATED RESEARCH AREAS

Before discussing several models of information behavior in detail, some remarks regarding their perspectives, goals and relation to each other shall be made. Not all models of information behavior are eligible for all kinds of phenomena. That is, the underlying theories address behavior from different perspectives but usually the models are rather complementary than contradictory [196]. Some models describe information behavior from the problem-solving perspective, associated with different stages of a goal-directed process to integrate the research in the field; other models propose a more global perspective of the field; some include psychological and sociological aspects (e.g., personality) into the information seeking process; some are focused on decision-making; and some are driven by application-oriented areas, such as the analysis of consumer behavior, social network and multi media interactions or health related communication.

Although the models them self address different aspects of information behavior, the scope of the models' corresponding research areas can be arranged in a kind of hierarchy, as proposed by Wilson [196] and as illustrated in Fig. 2.2. Information Behavior (IB) models and their related objects of investigation represent the most general approach to describe humans' relation to information. They build the

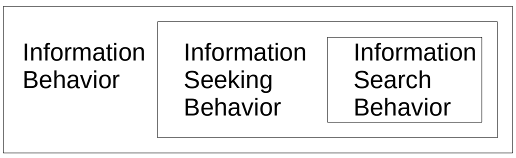


Figure 2.2: Wilson's nested model of information behavior research areas and their corresponding models [196].

framework to categorize a user attempting to satisfy an information need (cf. Sect. 2.1); include context information of the user; consider the general access to possible information sources (e.g., other people and/or the information system what in turn is addressed in more detail in the framework of information seeking models); as well as consider mechanisms to use the gained information for further processing. Information Seeking Behavior (ISB) models cover all methods which users conduct during a search to discover and gain information resources, i.e., all used strategies and tactics[4]. The models of ISB are included in the framework of IB and hence, can be considered as a subset. If a model is rather an instance or a class of ISB (as it is the case for exploratory search, cf. Sect. 2.6.3), then it is often categorized or described as *information activity*. The last and most concrete subset of the hierarchy are the models of Information Search Behavior (ISEB), sometimes also called *information searching behavior*. Here, all directly conducted interactions with the information system (e.g., mouse, keyboard or touchscreen interactions, eye movements or speech controls) and their related interface components, such as query input and result visualization, resp. comparison, are addressed. Regarding the areas of IB and ISB, Knight and Spink [111] come to a comparable conclusion:

Comparable to the nested model of Wilson [196]

> *"ISB represents one component of IB which can also include components such as the nature of the information, its specific context, format, or target audience, and other variables associated with its perceived usefulness or relevancy to the searcher, and searcher characteristics such as his or her cognitive level or efficacy."* (p. 209)

Furthermore, Knight and Spink [111] note that, depending on the authors, the term of information seeking is sometimes mistakenly used for the term information search. It is no wonder that this confusion happened because the field of Information Science (IS) is quit broad and is investigated by several, different disciplines that are furthermore influenced by the occurrence of new developments[5]. In addition to that, the terminology (e.g., IB, ISB and ISEB) has been developed and accepted only slowly. However, the hierarchy (and terminology) given above provides a means to assign the following IB models and their different research areas. A further possible step of integration in

4 Bates [14] discusses the concepts of and differences between (search) strategies and tactics. While a *strategy* is more related to an overall planning process for the whole search, *tactics* are rather moves to advance the search and to support short-term goals of finding desired information. Bates further distinguishes between four types of search tactics, namely (1) Monitoring, (2) File Structure, (3) Search Formulation and (4) Term tactics.

5 Major developments are the advent of early online IR systems and Web search engines, what also implicated a huge changing in the number and the type of users. Furthermore, the duration and frequency search systems are used has changed: from few professionals using the systems only at work to millions of common, daily users nowadays.

general is to embed the hierarchy above into the domain of human communication behavior and theory or to embed the research area of human computer interaction [88]. Though, a discussion regarding more (potential) influencing research areas would exceed the scope of this thesis, and thus will not continued here. Nevertheless, some entry points have been given.

2.3 INFORMATION BEHAVIOR

As described in the previous Section 2.2, IB models are the most general approach to describe a user during information acquisition and exploration. According to Wilson [192]:

> *"Information Behavior is the totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use. Thus, it includes face-to-face communication with others, as well as the passive reception of information as in, for example, watching TV advertisements, without any intention to act on the information given."* (p. 49)

This definition includes all actions to gather information and to satisfy the information need (cf. Sect. 2.1) of a user. One of the first models of IB on that general level was published by Wilson in 1981 [193]. An illustration is given in Fig. 2.3. The model describes a user who recognizes an information need and starts with ISB on different formal or

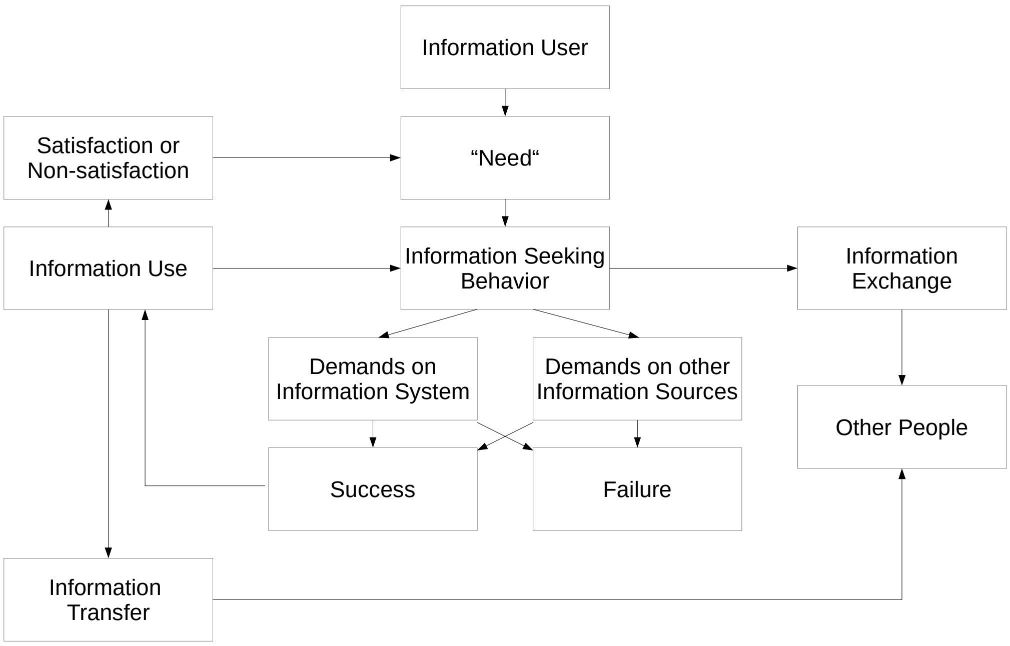


Figure 2.3: Wilson's first model of information behavior from 1981 [193].

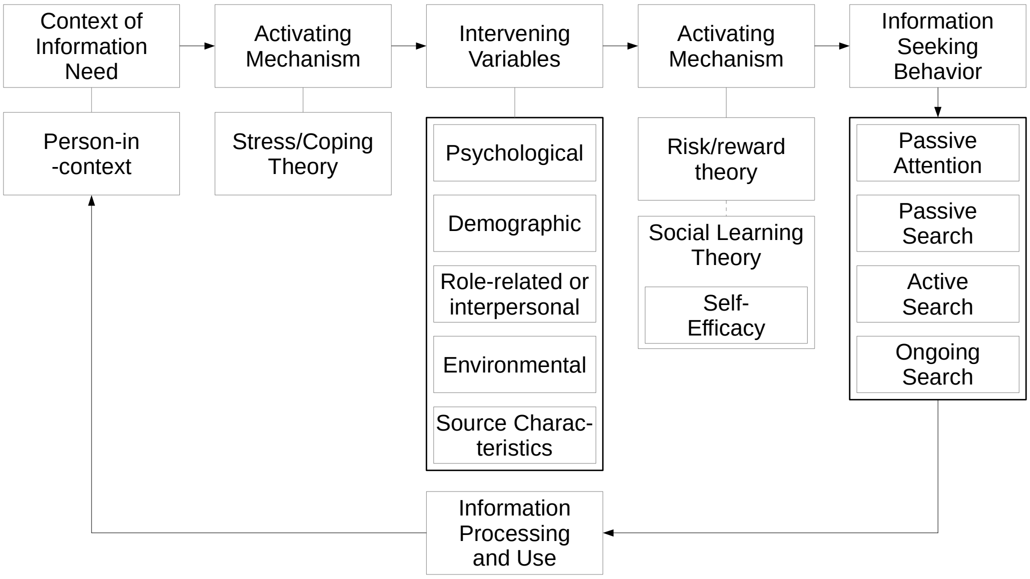


Figure 2.4: Wilson's extended version of information behavior from 1997 [195].

informal information sources. Alternatively, the user can seek information exchange with other people[6]. If successful, the user may use the gained information to further refine the ISB or to transfer the acquired information to other people. Simultaneously, successfully gained information may be used to evaluate the current state of satisfaction. If necessary, the user can (re-)formulate a (new) information need. The diagram in Fig. 2.3 illustrates the process of a user's IB but it is rather a representation of involved fields and activities than a model to derive the user's states or to illustrate the relationship in between. In 1997 an improved version of the model was published by Wilson [195]. An illustration is given in Fig. 2.4. The revision gives a more detailed view on stages around ISB and therefore, IB in general.[7] In particular between the stages of perceiving an information need and applying appropriate seeking behavior, the intervening variables are introduced. They represent the barriers of Wilson's previous model (cf. Sect. 2.1, p. 15) in a more elaborated manner:

Cf. Berry Picking in Sect. 2.5.2

- *Psychological and demographic* variables (i.e., personal barriers) represent the user's personal, cognitive and emotional characteristics, preferences, education, knowledge, age, sex and skills.

6 In fact, in her related work, Byström [35] summarized that the preference of users to ask, resp. consult other people increases with the difficulty level of the task they are confronted with. This could also be confirmed by Byström's final investigation [35].
7 Although the model is placed in Wilson's [196] summary under the headline of ISB, it clearly illustrates the context of ISB, and according to the nested model (cf. Sect. 2.2, Fig. 2.2), it should be (also) considered as model for IB. This example shows that the borders between models and their corresponding degree of abstraction is fuzzy and the models are (often) interwoven.

- *Role-related* variables (i.e., social or interpersonal barriers) become crucial if the information system has to switch from a single- to a multi-user scenario or if the information source is a person including several involved persons' attitude and motivation.

- *Environmental* variables (environmental barriers) help to derive the physical situation of the user and the context, such as time or location.

- *Source characteristics* (a further relevant and source related kind of barriers) are addressed if, e.g., different information sources (or channels of communication), and their properties such as accessibility and credibility are considered.

In his model description, Wilson also mentions further economical barriers that are related to the dimensions (perceived) *costs* and *time* in the information seeking process. However, this economic aspects are not directly listed in the model (cf. Fig. 2.4). In addition to the intervening variables, theoretical concepts, such as stress/coping, risk/reward and social learning theories, are included, which allow interdisciplinary researchers a more detailed hypothesis generation. Considering the information acquisition (i.e., the ISB), the model illustrated in Fig. 2.4 furthermore distinguishes between four different seeking modes:

- *Passive Attention*: Information that is passively consumed or acquired without intended perception or cognition and therefore, is not a direct (active) part of the information seeking process.

- *Passive Search*: Describes the incidental but for the user (potential) relevant information that was acquired in addition to the current search process.

- *Active Search*: Here the actively and conscious search and seeking-behavior to satisfy the current information need is considered. This search is addressed by most of the models of IS and ISB, e.g., Wilson, Ellis (cf. Sect. 2.4.3), etc.

- *Ongoing Search*: Describes all actions and motives to acquire information and derive knowledge that can be of value for following seeking processes.

While the active search mode is manifested in observable interactions between the user and the information system (e.g., mouse clicks or eye movements), the other modes require more advanced methods to estimate and investigate the related implicit attention. However, this thesis will not further focus on aspects of passive search or attention and hence, the investigations regarding the (exploratory) ISB models will be restricted here to users who actively seek for information.

2.4 INFORMATION SEEKING BEHAVIOR

As described in the previous Sections 2.2 and 2.3, ISB is integrated within the general framework of IB as a (major) component and hence, builds the first step from a macro- to a mesoscopic view on the user's search process[8]. According to Wilson [192]:

> *"Information Seeking Behavior is the purposive seeking for information as a consequence of a need to satisfy some goal. In the course of seeking, the individual may interact with manual information systems (such as a newspaper or a library), or with computer-based systems (such as the World Wide Web)."* (p. 49)

In the following, several empirically supported information seeking models are presented which are also partially used in the application-oriented, analytical literature as theoretical background.

E.g., cf. approach in, Sect. 3.3.1

2.4.1 *Dervin's & Belkin's Models*

Similar to Wilson's extended version from 1996 (cf. Sect. 2.3), Dervin's Sense-Making Theory [52] can also be placed at the border between IB and ISB models. It originates from the cognitive constructivist theory, was developed over several years and subsumes a bunch of assumptions, theoretic perspectives and methodological approaches to cope with perceived information need and performed IB. The model is illustrated in Fig. 2.5. Basically, the components are: (1) the user's current *situation* (in time and space) in which the need for information to solve a given problem or task arises; (2) an *outcome* describing the desired situation as result of the sensemaking process and problem solving; (3) the *gap* between the current and the desired situation; and finally (4) the concept of a *bridge*, describing the means to close the gap between the two situations.

Cf. Wilson's extended model: the entities between the perceived information need and applied seeking behavior, Sect. 2.3

A further quite fundamental example for a borderline case of a model between IB and ISB is Belkin's et al. [20, 21] proposition. They describe information seeking as a process of an emerging problem

8 From a historical point of view, most IB related models have been published under the headline of ISB but recently (from the 1990s), researchers more often used the term IB as general framework and associate ISB with *"explicit efforts to locate information"* [17].

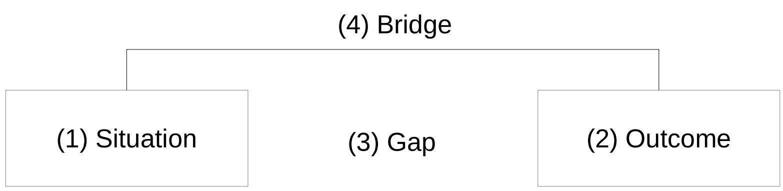


Figure 2.5: Dervin's Sense-Making Model [52].

and the user's inability to solve it, what is hypothesized as a so-called Anomalous State of Knowledge (ASK). Here, the information need (cf. Sect. 2.1) is understood as the gap (cf. Dervin) between the current knowledge of the user and necessary state of knowledge the user needs to achieve to solve the problem. Belkin et al. already postulated that the user's state of knowledge changes step by step and the user becomes more able to understand the situation to define the problem in more detail to solve it. Actually, the ASK can even include many information needs. Both, Dervin's and Belkin's et al. models serve as foundation of Kuhlthau's model (cf. Sect. 2.4.2) and therefore, serve as important cornerstones for the research areas of IB and ISB.

Cf. Berry Picking in Sect. 2.5.2

2.4.2 *Kuhlthau's Model*

Kuhlthau [120, 121] proposed an information seeking model (or actually a list) with six *stages* and corresponding activities in terms of the so-called Information Search Process (ISP). For the definition of the ISP, Kulthau uses the personal construction theory of Kelly [109] and considers the information seeking as a process of sensemaking (cf. Dervin [52]). Furthermore, the user's ability (here called level) to specify a given problem (cf. Belkin et al. [20, 21]) is identified as an crucial aspect of ISB. Kuhlthau's model describes the information acquisition from the user's feelings, thoughts and actions point of view and therefore, additionally has a phenomenological perspective. The underlying assumption of the model is that the perception of an information need (cf. Sect. 2.1) corresponds to a feeling of uncertainty what rises the motivation to satisfy the need. During the (partially successfully performed) ISP the user's feelings may change. The six stages of Kuhlthau's model imply an internal order[9]. The states and corresponding activities of the ISP are:

- *Initiation*: First awareness and recognition of a lack of knowledge or understanding what leads to the perceptions of an information need; possible feeling of uncertainty and apprehension.
- *Selection*: Identification of relevant search domains or approaches that apparently lead to success; ponder regarding available resources, information, requirements, and interests; optimistic feeling in case of quick, positive results or feeling of anxiety in case of (unexpected) delay of any kind.
- *Exploration*: Investigation of topic(s) in general to extend personal understanding; lack of knowledge to specify the information need; reading and listing facts as possible strategy, user may feel confused or uncertain.

Cf. relation of Exploratory Search to Kuhlthau's model in Sect. 2.6.3, p. 39

9 Wilson later picked up the six stages, ordered and combined them with of Ellis' feature model (cf. Sect. 2.4.3) and could derive further conclusions (cf. Sect. 2.4.4).

- *Formulation*: Turning point of the ISP, where the user feels confident and is able to focus the search and identifies respectively selects ideas.
- *Collection*: Most effective and efficient stage, where the user accumulates relevant information and identifies details to satisfy the information need and continues to perceive confidence.
- *Presentation*: Synthesis of the topic/problem and retrospective evaluation of the search process to estimate the satisfaction; if satisfied, the feeling of relief occurs.

In contrast to the other four states, the implied order between the *selection* and *exploration* stages in Kuhlthau's model is not binding. If the user has a promising seed of knowledge, the user can directly perform actions in terms of the selection state and (if necessary) explore afterwards to fill remaining gaps of knowledge. Otherwise, if the user has a lack of knowledge, first investigations are necessary to explore the topic(s) in general and afterwards more precise information resources during the selection stage can be utilized. A review by Kuhlthau of her own work and it's influences can be found in [122]. Here, the aspects of feelings in the ISP in addition to the cognitive (thoughts) and the physical (actions) aspects especially regarding the development of information systems in the last decades is underlined.

2.4.3 *Ellis' Model*

The model proposed by Ellis et al. [59–61] addresses behavioral patterns of search activities and is empirically supported by qualitative studies with, e.g., physicists, chemists or social scientists. The eight categories of Ellis' model are termed *features* instead of *stages* as in Kuhlthau's model. The features of Ellis are:

- *Starting*: All activities and means to initiate an information acquisition and seeking, such as asking other knowledgeable people or typing a query into a search engine and identifying a first key document(s).
- *Chaining*: Building forward and backward chains of (relevant) documents (and other information sources) by following references, such as hyperlinks on Search Engine Result Pages (SERPs), citations, footnotes or other index structures.
- *Browsing*: Performing a semi-directed or semi-structured search or exploration in a promising domain[10].

10 Cf. discussion regarding Browsing in Sect. 2.5.1 but also the relation of Exploratory Search to Browsing in Sect. 2.6.3, p. 41.

- *Differentiating*: Differentiation between several information sources to exploit their particular/specific characteristics as a filter to select appropriate content.
- *Monitoring*: Keeping awareness of developments in a (search) domain to stay up-to-date.
- *Extracting*: Identifying specific and relevant pieces of information that apparently lead to success.
- *Verifying*: Checking and evaluating the retrieved information.
- *Ending*: All activities that complete the information acquisition and seeking, such as validating or using the new gained information.

Ellis developed and refined his model over several years. Originally, only the first six categories have been proposed but after further investigations on the ISB of physicists and chemists, the two categories *verifying* and *ending* have been added. For a tabular comparison of the features in the several steps of development, see Knight and Spink [111]. Ellis' model does not attempt to define the connections and interactions between the features, thus the features not necessarily occur in a specific linear order in general. However, Wilson [196] suggested a relation between the features and in addition enriched the model with the stages and activities of Kuhlthau to complement both models (cf. Sect. 2.4.4). Wilson could also extend his own model incorporating the information need and corresponding barriers [193] by using Ellis' features, cf. [194].

2.4.4 *Complementary Models*

As mentioned in Section 2.2, models of information (seeking) behavior describe users information acquisition and the related phenomena from different perspectives and are rather complementary than contradictory [196]. In the following, two examples are given, where the combination of models unfold their relation to each other and therefore, allow to extends the particular models.

2.4.4.1 *Combination of Kuhlthau's & Ellis' Models*

Wilson [196] showed that the stages of Kuhlthau's model (cf. Sect. 2.4.2) can be aligned with Ellis' features (cf. Sect. 2.4.3). Consequently, he proposed a relation and an order between both model's elements in the combination, what implies the development of the ISP and the ISB over time. Nevertheless, the combination still includes Ellis' elements as possible outcomes in the ISP that may alternate within Kuhlthau's surrounding stages. Comparing both models, Kuhlthau's stages can be considered as slightly more general and phenomenological than

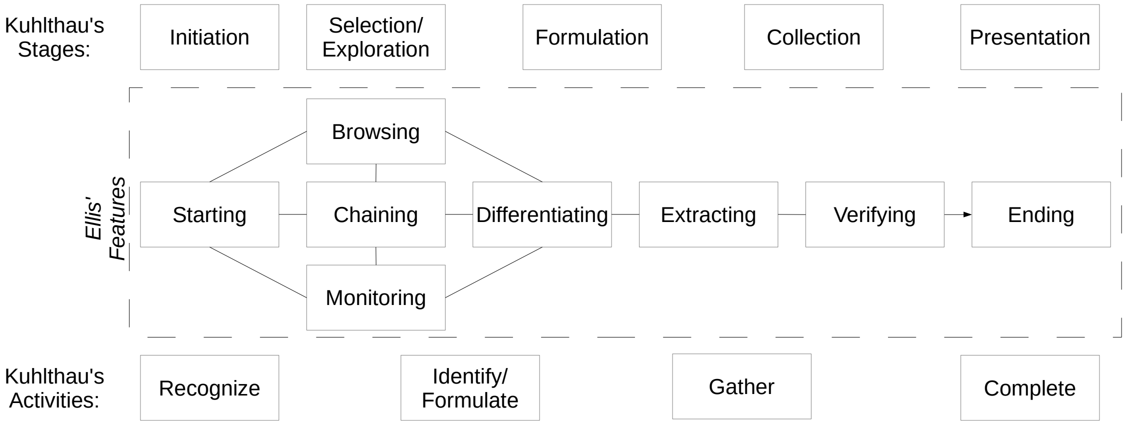


Figure 2.6: Wilson's aggregation of Kuhlthau's and Ellis' Models [196].

Ellis' features. Furthermore, Kuhlthau focuses more on user's affective aspects including learning and information processing, whereas Ellis highlights the interaction with the information system in terms of behavioral characteristics and the context in which the (missing) information is sought and explored. Hence, the combination of both models allows to look at information seeking from a user centered, behavioral and from an interaction centered perspective. In Fig. 2.6 the combined model is illustrated. The information seeking process is initiated by the recognition of an information need. Depending on the users knowledge, experience and confidence, the user can continue with Kuhlthau's stages selection and/or exploration. Here the user can perform a mixture of the alternating features, such as browsing through and differentiating between different information sources as well as following and monitoring promising results. If successful, the user is able to further focus the search by formulating the information need (e.g., by search query input) and gather the several pieces of information. By collecting the found information, the search process transforms into a semi-structured process within a confident phase where the user also verifies the found information. As described in the outline of IB models above (cf. Sect. 2.3), if the user is successful here as well and the information need (cf. Sect. 2.1) is satisfied, the search ends. Otherwise, the search proceeds.

2.4.4.2 *Combination of Ellis' & an Organization Oriented Model*

In their work, Choo et al. [39] present an integrated model of information seeking in terms of web interaction comprising two dimensions. The first dimension consists of four different, so-called, *scanning* modes which originates from the area of organization science and have been initially developed by Aguilar [4] and further expanded by Weick and Daft [50, 186]. These four modes are: *undirected viewing*, *conditioned viewing*, *informal search*, and *formal search* and they describe how peo-

Cf. discussion reg. relation between Scanning *and* Browsing *in Sect. 2.5.1*

ple in organizations seek, resp. "scan", . for information that appear to be relevant. While the undirected viewing is not based on a certain information need and purposes merely to "browse" broadly and (maybe) identify possible signals of change, conditioned viewing trails a selected, certain set of topics to identify and "learn" more about (for the organization) relevant information (sources). The goal of informal search is to gather and "select" more information regarding a certain topic set and to extend the knowledge but in a more unstructured manner, while formal search embodies the deliberated and projected "retrieval" of specific information to provide a foundation for (later) decision making. These four modes, as the first dimension, address a more motivational and purposeful perspective of organizational information seeking. The second dimension of the combined model in contrast is more activity-oriented, builds on the original six features of Ellis' model and associates each feature to common interactions that are nowadays applied using the web, e.g., via a web browser:

- *Starting*: Identify promising web documents for further information acquisition.
- *Chaining*: Following hyperlinks (forward or backward) on initial web pages and/or SERPs.
- *Browsing*: Scanning top-level structures (e.g., headlines, main tabs or lists) on web pages of promising domains.
- *Differentiating*: Filter relevant web pages and make them recallable by bookmarking or copying and pasting.
- *Monitoring*: Receiving update information of web pages or domains by (push) messages, mails, etc.
- *Extracting*: Search and select certain information found on web pages (or other sources) to satisfy the current information need.

The association of Ellis' features in the context of web browser moves is helpful in that it connects the more theoretical and abstract features of (Ellis') information seeking process (empirically derived from traditional studies with rather older information systems) to certain moves on more modern information systems. Therefore, the model of Choo et al. [39] serves as a kind of update or transfer although it is defined in a rather specific, namely organization science related, context. Table 2.1 illustrates the two dimensions of the combined model and lists possible web interactions[11]. Furthermore, the models and modes here have several worth mentioning parallels to the paradigm of exploratory search, which are outlined and discussed in more detail in Section 2.6.

11 In contrast to the source [39], the dimensions in Table 2.1 are transposed to reduce redundancy.

Table 2.1: Combined model of Choo et al. [39] using Ellis's original six features and four scanning modes of an organization oriented model in the context of web information seeking.

<table>
<tr><th></th><th>Undirected Viewing</th><th>Conditioned Viewing</th><th>Informal Search</th><th>Formal Search</th></tr>
<tr><td>Starting</td><td>Identifying, selecting, starting pages and sites</td><td></td><td></td><td></td></tr>
<tr><td>Chaining</td><td>Following links on initial pages</td><td></td><td></td><td></td></tr>
<tr><td>Browsing</td><td></td><td>Browsing entry pages, headings, site maps</td><td></td><td></td></tr>
<tr><td>Differentiating</td><td></td><td colspan="2">Bookmarking, printing, copying; Going directly to known site</td><td></td></tr>
<tr><td>Monitoring</td><td></td><td colspan="3">Revisiting 'favorite' or bookmarked sites for new information</td></tr>
<tr><td>Extracting</td><td></td><td></td><td colspan="2">Using search engines to extract information</td></tr>
</table>

2.5 Information Search Behavior

The step from a macro- to a mesoscopic perspective by describing the ISB within the IB, as given in the previous Section 2.4, shall now be continued by considering (rather theoretical) models of ISEB. However, the next step from ISB as mesoscopic to ISEB as a microscopic perspective is quite fuzzy an hence, a bumpy one. According to Wilson [192]:

> *"Information Searching Behavior is the 'micro-level' of behavior employed by the searcher in interacting with information systems of all kinds. It consists of all the interactions with the system, whether at the level of human computer interaction (for example, use of the mouse and clicks on links) or at the intellectual level (for example, adopting a Boolean search strategy or determining the criteria for deciding which of two books selected from adjacent places on a library shelf is most useful), which will also involve mental acts, such as judging the relevance of data or information retrieved."* (p. 49)

This primary user interaction related description of ISEB allows to cover many search activities actual performed with information systems. However, it makes the classification into the ISB models hard, because the continuum of (possible) interactions with current information systems and sources (e.g., dynamic web pages, applications, multi-medial search engine result pages, etc.) can not necessarily be clearly allocated to potential search activities (such as browsing) and can even consist of several information seeking stages resp. features (cf. Kuhlthau's and Ellis' models in Sect. 2.4.2 and 2.4.3). A further example for the entanglement between the mesoscopic and microscopic view can be found in Wilson's improved model (cf. Sect. 2.3 and Fig. 2.4). The four different seeking modes are introduced by Wilson under the headline of *information seeking* (and acquisition) but for the names of three of the modes the term *search* was used what, without question, fits to the description of the individual modes. In the following, several interaction related search paradigms are described to cover possible search activities within ISB.

As a reminder, the modes are: Passive Attention, Passive Search, Active search and Ongoing search.

2.5.1 *Browsing & Scanning Information*

In the area of IS, the concept of *browsing* has been discussed and developed long and exhaustively [91, 123, 159] and is often considered as a type of ISB [16, 130, 159]. As a result of the discussion, the term has a wide range of meanings [37], from unplanned, aimless viewing over unstructured information searching [17] to goal-directed, structured searching. Furthermore, browsing is used in differently scientific contexts on different technologies, from investigating the usage of books in library science to analyzing the surfing behavior in the web for IR. This wide interpretation actually hinders a comprehensive definition. However, a general consensus seems to be the description as semi-directed or semi-structured, informal [130] search activity that is performed as part of seeking behavior to select and/or explore (new) information. This characterization then is also in accordance with Wilson's model combining Kuhlthau's with Ellis' work (cf. Sect. 2.4.4).

Closely related to browsing is the concept of *scanning* [37]. Often, scanning is subsumed as part or dimension of browsing [159], where the user looks, examines and/or samples for information in a not necessarily linear manner and utilizes the user's knowledge regarding a given domain to identify relevant terms and develop attention for further seeking activities. Marchionini [130] further distinguishes between two modes of scanning: (1) *linear scanning*, in the sense of scanning a sequential list or arrangement of similar information objects (such as title lists) to identify potentially relevant sources presupposed (a) the objects can be recognized in a single glance, (b) the collection is reasonably small and (c) the user has some confidence regarding the search domain; and (2) *selective scanning* in the sense of scanning for lists of different information objects encompassing pertinent references with the goal to gain an overview of the search domain. In contrast to the interpretation of scanning as being a part of browsing, Bates [16] argues scanning requires elaboration that might not necessarily be given by the act of browsing and hence, scanning can also be interpreted as "not browsing" (p. 10). Clarifying, she argues further that scanning is a double-edged term (just as browsing) that can be used in terms of examine (something) closely but also in terms of look over (something) hastily. That is, both, browsing and scanning, are historical terms used in a broad spectrum which often refers to the degree of how structured or how purposeful the actions are to perform the current activity. Therefore, depending on the perspectives, both terms can be argued to be associated or contrariwise. Nevertheless, in all cases, scanning is considered more in terms of a perceptual recognition process. Closing her investigation, Bates [16] accumulates different definitions and perspectives and proposes browsing as an episode of possible four steps that are iterated until the browsing ends:

- *Glimpsing*: Browsing consists at least of glimpsing or acquiring a field of vision, abandoning it and glimpsing again.
- *Selecting or Sampling*: Selecting resp. latching on a physical, informational or representational object within the field of vision.
- *Examining*: Examining the object.
- *Acquiring the Object*: Physically or conceptually acquiring the object or abandoning it.

With the exception of glimpsing, the above listed steps do not necessarily need to be contained in each browsing episode. Furthermore, Bates argues that browsing can be seen as an instance or a behavioral expression of exploratory behavior [16] and thereby, she states the relation to the concept of exploratory search.

The indisputable presence of browsing and it's broad spectrum of activities related to the seeking process has caused it to be part of several information seeking models. Ellis, for example, emphasized browsing as an important part of information seeking [59] and integrated it as one of the model's features (cf. Sect. 2.4.3). Also Choo et al. [39] used browsing as part of their combined model (cf. Sect. 2.4.4) and that even twofold: On the one hand, it is used to describe the activities in the undirected viewing mode. On the other hand, browsing entry pages is the key (web-) interaction in the conditioned viewing mode. This usage of browsing in the model of Choo et al. has similarities to the discussion of the allied term scanning above, because undirected viewing is less structured or purposeful than conditioned viewing. Notably, the four modes in the model of Choo et al. (based on Aguilar [4], Weick and Daft [50, 186]) are called scanning modes.

As a reminder, undirected viewing is the first and conditioned viewing the second mode in the scanning dimension of Choo et al. [39].

Last but not least, a discussion of Marchionini [130] regarding different browsing strategies and tactics resulted in three different types addressing the underlying degree or focus of the performed browsing activity: *undirected browsing*, *semidirected browsing* and *directed browsing*.

2.5.2 *Picking Information*

Cf. Wilson's IB models, Sect. 2.3; but also Dervin's & Belkin's models, Sect. 2.4.1

According to the models of IB but also ISB, during information acquisition and exploration, users retrieve new pieces of information with each iteration. The new information will influence the degree of satisfaction of the current information need but the information need itself can also change because a more specified piece of information turned out to be relevant or the information need got more general because related topics have been revealed and therefore, a domain overview has become more important. That is, the search and the related information need are not static and can evolve over time. This process is described by Bates [15] under the concept of *Berry Picking* and is illustrated in Fig. 2.7. The information space is usually high dimensional but for simplicity, in Fig. 2.7, only two dimensions are illustrated. The search starts with an initial query formulation denoted by $Q(t_0)$. During the search, the user's knowledge and information need evolve by finding (new) documents (D), resp. information and the user reflect (T) on them before (re-)formulate the former query or specify a new one. Its is also possible that a former query variation

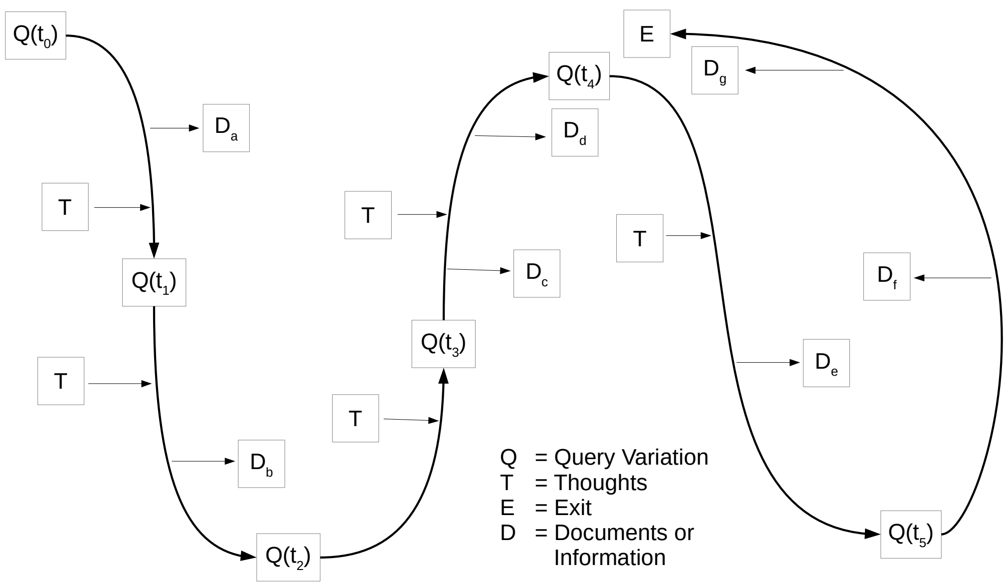


Figure 2.7: Bates' Berry Picking Model [15] to illustrate a sequence of search behavior. The arrows, connecting the query variations, describe the user's search path (in space) and are time-correlated.

and/or corresponding, selected documents have been closer to the desired document or piece of information than later ones. This can cause loop ways. For example, in Fig. 2.7 the retrieved document(s) D_d via $Q(t_3)$ and D_g via $Q(t_5)$ are closer to the end of the search (in space) as the document(s) D_e.

Bates [15, 16] also emphasized the difference between browsing and Berry Picking. Berry Picking characterizes information search or seeking as a whole with evolving queries, knowledge and information need(s) and hence, provides a wide variety of possible search techniques that can be applied. These techniques can be more related to a standard (lookup) search and/or can involve further search strategies, such as browsing. That is, Berry Picking rather serves as a framework highlighting the variability of information seeking and can include (steps of) the browsing process or other search activities.

2.6 Exploratory Search

The way users interact with information systems during seeking depends on the users' experience with the information systems but also on the users' knowledge about the present search domain. For instance, an expert for a specific (search) domain is usually able to define his or her information need precisely [131]. Furthermore, if the expert already has experience with the available information systems, i.e., with the user interface (the front end) and maybe even with the mechanisms in the background (the back end), the more precisely the expert's information need can be expressed in terms of promising search queries and available search parameters. Therefore, the desired relevant document(s) or piece(s) of information can be located relatively easy and without circumstances. In contrast to that, if the user has little knowledge about or is new to the (search) domain and hence, the context of the information need is vague, the user can only formulate imprecise search queries and has to discover resp. explore the domain. If the user in addition has little knowledge about the utilized information system, more experience with the usage of the system, it's search parameters and result representation has to be gained.

A further aspect that influences users' interaction with (information) systems is the type of the underlying information problem or (search) task that has to be solved[12]. If the task (and the corresp. information need) can be answered by a single fact-based information, it is called lookup search where no extensive seeking is required. As already pointed out in the thesis' introduction in Chapter 1, current search systems are relatively successful in providing adequate answers to

12 At this point, a discussion about the origin of information problems, respectively (search) tasks, goals and their relation to motivation and satisfaction of (information) needs could be conducted but this would be out of the scope of this chapter. However, aspects of motivation and goals as intervening variables to the ISB (cf. Wilson's improved model in Sect 2.3 and Fig. 2.4) are investigated in Section 5.4.

such kind of tasks. However, if the search task (and the corresp. information need) gets more complex and/or uncertain; if the search task can not be answered by a single fact; and if (thus) also browsing is required [131, 190], the user has to explore. In case of complex search tasks, even experienced users may have to perform several search iterations (maybe even in collaboration, cf. Sect. 6.1) to approximate a sufficient solution. Consequently, exploration is driven by the intention to reduce uncertainty [188, 190] to achieve a better understanding of the given domain and to advance the current (search) task.

That is, the necessity and the degree to explore and discover a domain depends on the users' corresponding domain knowledge, the experience with the available means but also on the type of search task behind. This characteristics are central and agree with Marchionini's as well as White's (et al.) definition of Exploratory Search (ES) as illustrated next (cf. Sect. 2.6.1). Along with the research field of IS, the concept of ES has been also addressed by several related (major) research fields, such as Human Computer Interaction and IR, as well as several sub-fields, such as Information Foraging [144–146], Sensmaking (cf. Dervin's model Sect. 2.4.1) and Information Visualization (IV)[13]. In particular, one work from the area of IV shall be highlighted here. It describes ES in terms of three dimensions, fits nicely to the perspective of ISB and is described in Section 2.6.2. Afterwards, ES is discussed from the perspectives of the proposed information (seeking and search) behavior models (cf. Sect. 2.1 to 2.5) to illustrate their relation to each other and to integrate ES (*H*1) into the area of IS (cf. Sect. 2.6.3).

2.6.1 *Characteristics of Exploratory Search*

Marchionini explicates the paradigm of *exploratory search* [131] by integrating it into a framework of three relevant categories, namely: *lookup*, *learn* and *investigate*, which in turn are supported by several search related activities, cf. Fig. 2.8. ES is understood as an extension of a standard lookup (search) and includes learning and investigation. In addition to the interpretation as extension, ES can also be performed alternating (or in parallel) to lookup and may include aspects of blending querying and browsing behavior (as an exploratory expression; cf. Sect. 2.5.1).

Lookup

Lookup can be considered as a basic interaction that primary encompasses standard fact-finding search processes with a specified query, e.g., to retrieve or verify a certain factual information, answer a factual question or to navigate to an already known item. That is, lookup search is an elementary, conscious and purposeful action to satisfy a need for a specific piece of information. However, to

13 Of course, these research fields are interconnected and overlap to a certain degree. An recommendable illustration of several research fields related to ES can be found in the work of White and Roth [189], (p. 39).

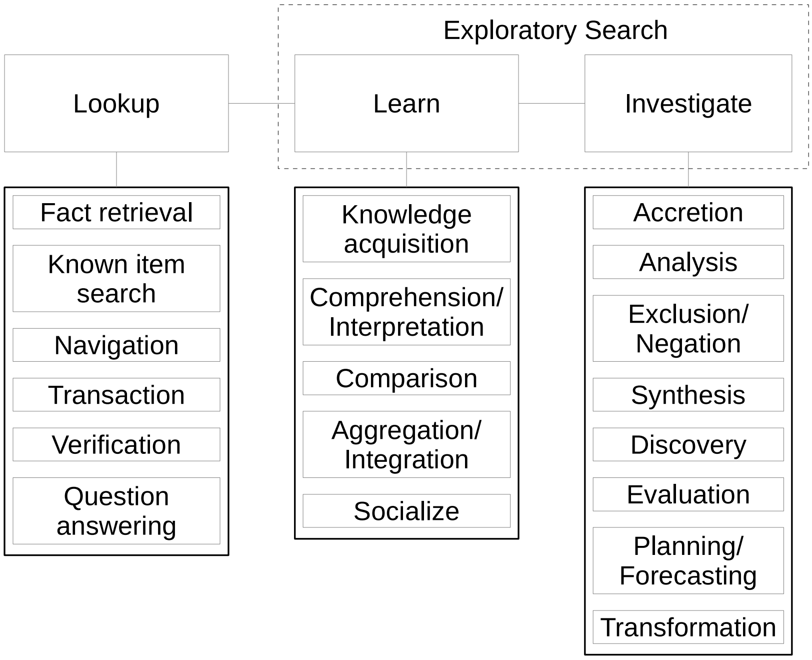


Figure 2.8: Illustration of exploratory search embedded in several (search) activities according to Marchionini [131].

execute that kind of search, it is necessary (1) to have knowledge about the search domain the individual is acting in; (2) to have knowledge about the available means the individual can utilize; and (3) the type of (search-) task should be answerable by a single fact. If this is not the case for one or more of the characteristics, exploration becomes necessary. Exploration includes aspects of learning and investigation as iterative processes what includes actions, such as acquisition, comparison, aggregation and integration of (new) information, but also the analysis, synthesis, forecasting, evaluating and interpretation of the (new) knowledge. Under review of the literature, White and Roth [189] furthermore identify and confirm characteristics of ES as:

- Open-ended, persistent and multifaceted because of the kind of the underlying information need, respectively search task, and thus long-lasting, e.g., for days, weeks or even month;
- Performed in multiple search sessions using multiple query iterations;
- Associated with learning and understanding (cf. Marchionini's explication above);
- A combination of focused searching (similar to "lookup search") and browsing (cf. Marchionini's definition) and;

- A search paradigm that may involve or require collaboration[14] in a synchronous or asynchronous manner.

On the one hand, Marchionini's and White and Roth's descriptions are abstract enough to align the paradigm of ES with the models of ISB (cf. discussion in Sect. 2.6.3). On the other hand, the descriptions are specific enough to investigate concrete informatory interactions performed with a certain information system, i.e., to address the perspective of ISEB (cf. Sect. 3.3.1, Sect. 3.3.2.3, and Chap. 5). Therefore, ES is selected and serves as a promising and central objective of investigation for this thesis.

2.6.2 *Three Dimensions of Exploratory Search*

While exploring new (search) domains, it becomes important to provide an appropriate support to assist users by their investigation process[15]. If the user discovers domains that differ from their common interests and knowledge, the system can use this as indication, can derive a possible demand on exploration and can provide adequate means. For example, to get familiar with a new domain, some types of information sources, like encyclopedias, can be more helpful than other sources, such as specific topics discussed in forums. Therefore, one possibility is to capacitate the information system to ask the user whether the result ranking should be adapted or not. Furthermore, additional information snippets about current information objects, e.g., extracted from knowledge networks such as DBpedia[16], can (if requested) enrich the exploration. To support but also illustrate users during their exploration, Noël et al. proposed a model (originally in the application area of tourism information systems) that describes the characteristics of ES by the following three dimensions [139, 140]:

- *Vertical Axis*: The user changes the level of focus for the relevant information (sub-) space, e.g., by tools that allow zooming in and out the current search domain.
- *Horizontal Axis*: The user differentiates (cf. Ellis' features in Sect. 2.4.3) between the retrieved results to identify those that best match to the current information need. On the horizontal axis, the user furthermore derives information about the domain what facilitates action on the focus level via the vertical axis.
- *Transversal Axis*: The user changes the perspective on the retrieved information pieces what allows to derive further knowledge and to identify relation(s) between the perspectives.

14 In Sect. 6.1 the topic of collaborative information exploration is discussed.
15 In Sect. 5.5, search systems to support ES are discussed in more detail.
16 https://wiki.dbpedia.org/

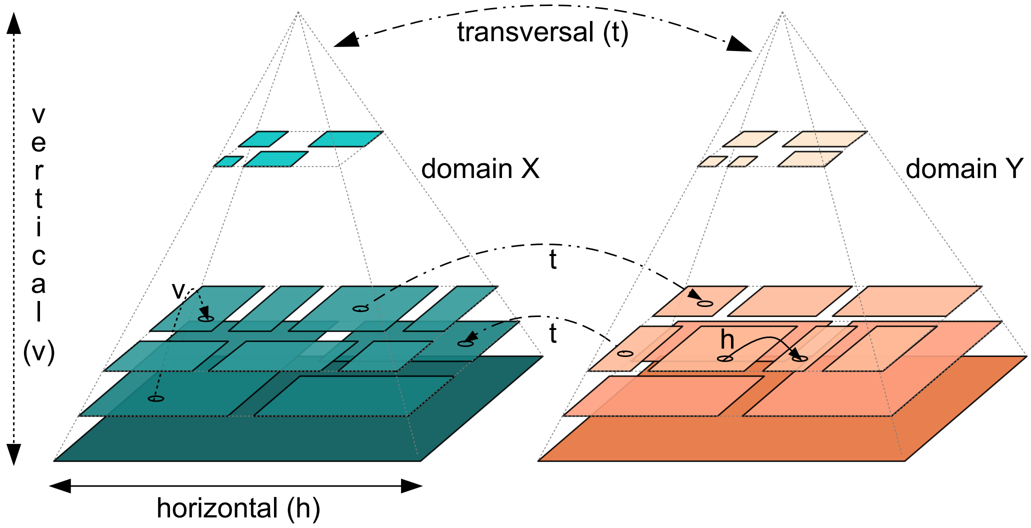


Figure 2.9: Illustration of the three-dimensional character of exploratory search according to Noël et al. [139, 140].

An illustration of the three-dimensional character of ES is given in Fig. 2.9. The user can explore the information space related to the search domain X; can change the focus levels on the vertical axis (i.e., the level of detail); or can differentiate between the sub-domains of X on the horizontal axis. On the transversal axis, the user can switch between different domains (e.g., domain X and Y) to change the perspective to the same or similar piece of information. In the following, an example shall further illustrate the three-dimensional character: Methods of clustering in computer science are often used in the domain of scientific data analysis (X). The complexity of the methods or the details of discussions about single methods are related to the vertical axis. Several clustering methods in the domain of data analysis (X) can also be divided into sub-domains (cf. horizontal axis), which are represented by different web results. In contrast to the domain of data analysis (X), methods of clustering are also used to structure retrieved (web) results in the domain of interactive information systems (Y), e.g., to provide an overview of all retrieved results for the user. By switching between the two domains, the transversal axis is addressed. That is, the same clustering method can be used in both domains for different application areas with different purposes.

The three axes imply that the exploration of a domain goes along with three tasks, namely (1) to adjust the focus; (2) to differentiate between the sub-domains (and focus levels); and (3) to change the perspective. Indeed, these implied tasks can be interpreted as a possible outcome of the exploratory search's characteristics pointed out in the previous Section 2.6.1: (1) To change the focus level (on the vertical axis), at least some knowledge regarding the given search domain is necessary, what corresponds to lookup and it's actions. (2) To differentiate between retrieved results and their related sub-domains (on the horizontal axis), actions of learning and investigation appear to be relevant to acquire the necessary knowledge, what is covered by the concept of ES in general. In addition, an interplay between the vertical and horizontal axis, and thus between lookup and exploration, seems to play an important role what was also highlighted by Marchionini and White et. al. Last but not least, (3) to changes the perspective (on the transversal axis) is a high-level skill that first has to be enabled by a successful conducted (initial) knowledge acquisition that leads to understanding of the domain(s) as result of a comprehensive exploration.

2.6.3 *Integration of Exploratory Search into Information Seeking Behavior*

With the description of ES, it's characteristics and possible outcomes in terms of three dimensions, now the integration of ES into the models of IB and ISB can be discussed, as it is postulated by the thesis' first hypothesis $H1$ (cf. Sect. 1.2):

Relation of ES to the information need

Section 2.1 described the process of discovery and exploration of an individual's environment as a very fundamental but also general process to learn, to understand and therefore, to satisfy (basic) needs by purposeful executed actions. If the individual has only little knowledge about the current surrounding environment, it has first iteratively to accumulate information by observation and interaction. In essence, this exactly represents the point of origin but also the approach applied during the process of ES, as described in this section. Users who already have the knowledge to solve a given (fact-based) task and are experienced with the available information system(s), i.e., the environment on a technical level, are able to perform a lookup search in terms of a conscious and purposeful action to acquire a certain piece of information. Otherwise, the users have to learn and investigate to understand the domain (figurative the environment on an intellectual level) what is covert by ES.

Relation of ES to models of IB and ISB

As pointed out in Section 2.2, models of IB and ISB are driven by different perspectives of human information interaction and consequently depend on the research, resp. application area. Furthermore, the models are differently suited to describe a given problem context. Nevertheless, a (problem) description and/or investigation from diverse perspectives can be helpful and the fact that models of IB and ISB are rather complementary than contradictory turned out to be beneficial. Keeping in mind the different perspectives of information (seeking) models, indeed a similar situation can be identified for ES: The process of ES is a one of sensemaking (cf. Dervin's model in Sect. 2.4.1) in which a perceived information need, to solve a given problem, arises. The solution can not be reached immediately due to a knowledge gap; less experience with the available means; or the neces-

Relation of ES to Dervin's Sense-Making model sity to find several interlaced answers for the (complex) task. Though, iteratively acquired new knowledge provides means (in terms of a "bridge") to converge to a sufficient solution. ES can also be interpreted as an anomalous state of knowledge (cf. Belkin's et al. model regarding ASK in Sect. 2.4.1), caused by the user's inability to solve a given problem. If the knowledge changes (by exploration) and the user becomes more capable to understand the domain and the problem context, the more the user can define the corresponding information need and can apply lookup related activities.

Relation of ES to Belkin's model

Considering Kuhlthau's model (cf. Sect. 2.4.2) and it's related stages, *exploration* as placed in the ISP aims to solve uncertainty, minimize confusion and to (re-)achieve orientation in an (unfamiliar) domain by collecting facts and learn about the domain. Kuhlthau's concept of exploration (as a stage) definitely matches to ES as defined in this section but rather in terms of a sub-set than the identity. The reason is that in Marchionini's and White's et al. definition of ES, the perception of missing knowledge, the (purposeful) acquisition and collection of information, but also the confident lookup in sub-domains are covered but this is not the case in Kuhlthau's exploration stage. However, these characteristics are not missing in Kuhlthau's model but are partially covered under the remaining stages such as *initiation*, *formulation* and *collection*. Furthermore, lookup (as in Marchionini's model) can be understood as a standard fact-finding search with a specified query that leads to Kuhlthau's *selection* stage. Thus, ES can also easily be covered by the ISP as an instance. Additionally, examining Kuhlthau's model in it's original setting, namely (also) considering the user's feelings, each stage has it's justification since each stages allows to correspond to an own set of emotional outcomes. Continuing the integration of ES into the framework of ISB, Ellis' model (cf. Sect. 2.4.3) and it's features now excellently fit for two reasons:

Relation of ES to Kuhlthau's ISP model

Relation of ES to Ellis' model

1. The first reason is a quite pragmatic one using the relation of ES to an already described ISB model: Ellis' and Kuhlthau's models can be combined (cf. Sect. 2.4.4) and consequently, the relation between ES and Ellis' model is given by the model of Kuhlthau.

2. The second reason builds on Ellis' features and their composition as process by Wilson: The features of Ellis, interpreted as possible outcomes within the ISP, describe various patterns of search activities that may appear during information seeking. The first four features, connected by Wilson in his aggregation[17], essentially describe the interplay between lookup and ES (as described in Sect. 2.6.1) in an early stage or part of ES by: *Starting* (and repeating) to formulate a query; if necessary *Browsing*[18]

17 As a reminder, the first four connected features from Ellis according to Wilson are: *Starting*, *Chaining*, *Browsing* and *Differentiation*, cf. Sect. 2.4.4.

18 Ellis' feature *Browsing* itself is defined by using the term "exploration" but Ellis' understanding here is rather similar to Kuhlthau's stage of "exploration" and does

and following (i.e., *Chaining*) promising information objects to learn about and investigate the current sub-domain; and *Differentiating* the revealed information and/or documents to filter and therefore, to proceed the seeking process. The remaining feature *Monitoring* becomes more important if the ES is targeted on staying up to date regarding domain knowledge, e.g., in context of so-called technology scouting [197], in multiple search sessions maybe even in a collaborative setting over a longer period of time. Thus, the feature *Ending* here has to be understood as a temporal ending of seeking that is continued later, not least to fulfill the open-end characteristic of ES. The features *Extraction* and *Verification* are also relevant for the process of ES but are rather connected to an advanced, resp. later stage or part of ES that incorporates understanding a current sub-domain, where the "extracted" and "verified" information is first "formulated" and "collected" afterwards (cf. Kuhlthau) to proceed to the next sub-domain[19].

White and Roth [189] likewise argue that Ellis' model is able to cover *"Most situations involving information seeking"* (p. 33) but they also note that the model not fully captures all aspects of ES because of: (1) missing external causative factors; (2) a not guaranteed identical ISB process; (3) not supported tasks; and (4) the originally missing relationship between the features. However, the discussion given above attempted to address and resolve these issues: The positioning of ISB as a subset and a central part of IB, as shown in Section 2.2 and 2.3, allows to derive statements about external factors (1), such as intervening variables or the information need itself, and their effect on ISB and hence, Ellis' model. The ISB as process over time is a highly interdependent iterative sequence of several (search) activities that not necessarily consists of all possible actions (resp. features considering Ellis' model). This is exemplary shown above where the feature *Monitoring* only appears if necessary. Therefore, an identical ISB process (2) is debatable for several models of ISB. Task support (3) is only indirectly addressed by Ellis' but in context of Kuhlthau's stages and the integration of ES, the procedure to solve a (complex) task is generally given. Last but not least, the relationship between the features (4) was illustrated by Wilson's combination (cf. Sect. 2.4.4) and in the discussion about the integration of ES in Ellis' model above.

not cover completely the characteristics of ES as defined in Sect. 2.6.1. However, Marchionini as well as White (et al.) highlight the combination of focused search (i.e., lookup and it's activities) and browsing as important characteristic of ES what is inherently given by the first four features of Ellis.

19 Here, a connection to the three dimensions of ES (cf. Sect. 2.6.2) can be demonstrated where the first four features from Ellis correspond to the vertical and horizontal axis and may even count as kind of requirement for the later features which than allow for further interactions on the horizontal axis (to new sub-domains) or even allow a change of perspective on the transversal axis.

That is, with the help of Wilson's combination of Kuhlthau's and Ellis' model, the picture of ISB and the subsumption of ES becomes more clear and the derivation of ES as a kind of class, as suggested in the literature, gets more evident. The conclusion that ES is a class of ISB is also supported by White and Roth [189]. Furthermore, White and Roth [189] describe ES as a specialization of information seeking that (in accordance to Marchionini [131]) uses a combination of querying and browsing to acquire information.

Relation of ES to Browsing

The relation between ES and browsing is a close one. Marchionini [131] states that exploration blends querying and browsing strategies. White and Roth [189] describe ES as a process that combines focused searching browsing. Furthermore, Bates [16] concluded that:

> *"General exploratory behaviour in humans is manifested in a number of ways, with many of the activities being similar, though not necessarily identical to browsing information ..."* (p. 15) and further *"Browsing is a cognitive and behavioural expression of this exploratory behaviour."* (p. 19).

Relation of ES to Berry Picking

Reminding the concept of Berry Picking (cf. Sect. 2.5.2) and it's relation to browsing as a framework that provides a wide variety of possible search techniques which also may include episodes of browsing, Berry Picking describes information (search and seeking) more as process, where the user's information need but also knowledge evolves over time. The same approach is identified by ES, highlighting the procedural character of both and is supported by White and Roth's [189] conclusion that:

> *"Berrypicking is a commonly used strategy in exploratory searches, ..."* (p. 29).

To close this section, a proper integration of ES into the given IB and ISB models was necessary because ES is the primary focus of investigation in this thesis and hence, the connection to the theory serves as crucial foundation. The argumentation for the integration given in this section is supported by the literature but unfortunately most of the supporting statements from the literature are given just in fragments and/or only in relation to a limited set of ISB models to motivate a certain application related and/or domain specific investigations. Therefore, this sections is proposed as a contribution to integrate ES into IB and ISB in a more holistic level. Since ES could be identified as an instance of ISB, the entirety of actions within the concept of ISB and in context of ES is considered as so-called Exploratory Information Seeking (EIS). As a further proposed contribution to the current literature, one goal of this thesis is to analyze the nature of ES, reveal intervening variables and to investigate possibilities for modeling to identify ES on the interaction level. By reviewing, analyzing, modeling and classifying ES based on interaction data with real world search

Def.: Exploratory Information Seeking (EIS) engines, the relation of ES to the area of ISEB is given (cf. Sect. 3.3.1, Sect. 3.3.2.3 and Chap. 5). In doing so, the investigations in this thesis attempt to minimize the number of constrains in terms of application scenarios or domain related result sets by using common search engines as information systems and utilizing search tasks of rather general than domain specific topics. However, to execute the investigations on interaction data, first, the necessary analytical foundations are described in Chapter 3 and second, the required data generation is outlined in Part iii, Chapter 4.

2.7 FURTHER INFORMATION (SEEKING) BEHAVIOR MODELS

This chapter gave a (rough) overview about selected information seeking and search behavior models which allows to construct an adequate and solid foundation to address IB from a theoretical point of view but also to integrate the search paradigm of ES. However, the list of presented models is not complete and can be extended in the direction of various research areas: Models addressing the ISB from the user- and search process perspective, can be found in Ingwersen's cognitive model [96, 97] and Belkin's model of information seeking episodes [23][20]. To gain more insight into the ISB and in particular to investigate less experience biased seeking, studies with young users have been conducted. Young users have less expert knowledge [94] and the emotional, respectively affective, state often plays a large role for their ISB [43, 44, 119]. Nesset [137] proposed two representations of children's ISB in terms of the so-called *preparing*, *searching* and *using* (PUS) model and a simpler representation, called the *beginning*, *acting* and *telling* (BAT) model. Besides the illustration of ES [131], Marchionini also investigated a model-based perspective on information seeking in electronic environments [130]. He proposed a model of the information seeking process that consists of eight sub-processes[21] and accordingly several probability transitions between the sub-processes. Related work on information search activities and the information retrieval process can be found in Saracevis's stratified [165, 166] and Spink's [176] models. Besides Wilsons' summary of IB models [196], Knight and Spink [111] but also Al-Suqri et al. [7] provide a recommendable overview.

The given overview of the information seeking and search behavior models in this chapter was accompanied by the motivation to select representative concepts which introduce and cover the broad area of IS and the related sub-areas on the one hand but also facilitate a

20 Since Belkin's episodic model is strongly related to the user's activities resp. interactions performed with information systems, it can be argued to consider the model also as ISEB model.

21 The eight sub-processes of Marchionini's information seeking process [130] are: (1) Recognize Accept, (2) Define Problem, (3) Select Source, (4) Formulate Query, (5) Execute Query, (6) Examine Results, (7) Extract Information and (8) Reflect Stop.

clear and demonstrative integration of the paradigm of ES on the other hand. The (further) alternative models listed above, exhibit (as ISB models) connections to ES as well but an all-embracing exemplification would exceed the focus of the thesis and would not complement the remaining investigations in this work. Nevertheless, the listed further models can serve as point of entry for further investigations and attempts of integration.

2.8 CHAPTER SUMMARY

The goal of this chapter was to provide an overview of theoretical aspects of human information behavior. At first, the origin and role of the information need, as initial component in most of the behavior models, was discussed. Afterwards, several information (seeking and search) models have been described and set into relation to each other. While most of the models can be arranged in a kind of hierarchy, some of them also have to be considered from a complementary perspective. By discussing the models and their individual components, it has been revealed that aspects of exploration are often already involved namely to advance the search process in case of uncertainty. This further fortifies the procedure for this thesis to utilize an exploration related search paradigm as central concept to investigate users' information seeking in more detail.

Besides the discussion of the selected approaches to represent human information behavior in general, the models in this chapter have also been set into context to the thesis' topic as fundamental framework for the investigations in the next chapters. In particular Wilson's aggregation of Kuhlthau's and Ellis' Models (cf. Sect. 2.4.4) but also Wilson' intervening variables which can influence the information (seeking) behavior (cf. Sect. 2.3) serve as basis for the (main) investigations in Chap. 5. The aggregation provides the context for Exploratory Search (ES) but also highlights the difference to Kuhlthau's "selection" stage (also cf. the integration of ES and the relation to the ISP model in Sect. 2.6.3), what leads to the classification scenario in Sect. 5.2. Wilson' intervening variables (namely psychological and demographic variables) lead to the investigations in Sect. 5.4. The investigations finally result into promising approaches to support users' information behavior, resp. ES search (cf. Sect. 5.4).

The rather theoretical models, described in the beginning of this chapter, facilitate a deductive consideration of users' (seeking and search) behavior, but the relation to concrete interactions of users with given search systems is only partially considered and hence, can not be modeled directly from a technical point of view. However, to overcome this issue, the search paradigm of ES was identified as a promising object of investigation because of it's multifaceted and natural behavior; the ability to model and analyze concrete exploration related search activities; but also because of the implicit relations of ES to the existing IB and ISB models. Therefore, the concept of ES was laid-out (Sect. 2.6) and could be integrated into several information (seeking) behavior models. In case of concrete interaction with information systems, ES was identified as an instance of information seeking what established the introduction to EIS.

> *"Search is an important way to access the ever growing amounts of information available on the Web, but it is rarely a stand-alone activity. Instead it is typically part of a complex process to accomplish some larger task."*
>
> — Case & Given [37], p. 4.

3 ANALYTICAL FOUNDATION AND RELATED WORK

With the emergence of the Internet, the amount of available digital information sources for users, provided by search engines and web sites, has been multiplied. Consequently, the number of user's interactions with the sources and the corresponding (search) systems increased as well. Logging the interaction data stimulated a prosperous branch of research providing a large number of analysis and models to investigate the user's Information Search Behavior (ISEB), for instance, to identify user groups in different contexts; generate adaptive query suggestions; analyze purchase behavior; provide individual Search Engine Result Pages (SERPs); etc. Usually, the related studies are application- or task-oriented to reveal and exploit the limits of the available user data and the models themselves, e.g., for economic reasons. Though, the relation to theoretical aspects of Information Behavior (IB) or Information Seeking Behavior (ISB), as described in the former Chapter 2, is not always given. Fortunately, the number of studies, which focus on the investigations and understanding of user's information (seeking-) behavior on that analytical level, increase. This chapter provides an overview and a discussion of analytical models and related research, the underlying assumptions, purposes and advantages. Taking into account the thesis' topic of Exploratory Information Seeking (EIS), also the limitations of the approaches in context of Exploratory Search (ES) are depicted. The chapter begins with an introduction to analytical models, a discussion of challenges and a brief outline of application areas (Sect. 3.1). Afterwards, several models to investigate the user's search behavior on SERPs are exemplified (Sect. 3.2). Since research on ES requires the consideration of the search process in it's entirety, the chapter continues with a discussion of approaches that are appropriated for analyzing search activities and ISB in general and concludes with the definition of models used in this thesis to investigate EIS (Sect. 3.3).

3.1 ANALYTICAL MODELS

While the purpose of qualitative investigations regarding ISB is rather to illustrate the whole picture of the user's search process in general, quantitative investigations (e.g., log file based), usually aim to analyze certain episodes or snippets of the search process but this with a high degree of detail and significance. Although log file investigations relatively often restrict themselves to certain behavioral aspects of user interaction, the revealed insights can contribute to the research area of ISB in that they facilitate to focus on individual relevant user and system variables which in turn enable to develop new interfaces, methods and user models. A crucial point of the restriction to certain behavioral aspects is the requirement to make several assumptions regarding the remaining user behavior. On the one hand, these assumptions enable the calculation of the underlying models. On the other hand, sometimes, the assumptions appear to be questionable from a more theoretical, holistic perspective. For example, the so-called *Cascade Model* (cf. Sect. 3.2.2.2) is grounded on the premise that users iteratively process the SERP from the top to the bottom and for each result decide to click or not before they proceed to the next item on the list. This assumption is truly a strong and restrictive one because a casual user may also (re-)view (upper) results or skip several ones. However, even under restricting assumptions the utilization of analytical models to investigate certain episodes of the search behavior enables to reveal precise findings regarding the user's search process. Furthermore, the analytical approach allows to reveal weaknesses in the models themselves and enables to investigate the impact of the assumptions to the corresponding models. Thereby, the models themselves can be analyzed and improved again afterwards.

3.1.1 *Challenges*

To apply quantitative investigations, analytical models rely on corresponding data bases, e.g., extracted from log files. However, if log files of users are collected by (huge) search engine providers over a long period of time including thousands of daily life search interactions from an unknown user base, additional challenges rise in contrast to a controlled investigatory environment. This includes:

- *Challenge 1*: to investigate and understand the user's seeking behavior without further knowledge about the (intrinsic motivating) information need, because the underlying need for information is not necessarily derivable from the log files;
- *Challenge 2*: to draw conclusions regarding possible relevant and intervening variables, such as task type (cf. Sect. 3.3.1) or demography (cf. Sect. 2.3), because these variables may not be known or be ambiguous;
- *Challenge 3*: to allocate performed and recorded interactions to the "correct" corresponding seeking behavior, because (a) the seeking may actually last over several sessions (cf. characteristics of ES), (b) the seeking may be executed on different machines by the same user and/or (c) one machine may be used by different users in the same session.

An example for the presence of these challenges can be reviewed in White and Drucker [187], who identified different classes of user search behavior, in particular "navigatory" or "exploratory" behavior, based on recorded interaction logs. The user's underlying information need, that caused the interactions with the search engine, was not known (*Challenge 1*). The allocation of several search sessions to a given (maybe evolving) information seeking behavior was created under the assumption that each search session encompasses a (full) episode of seeking behavior (*Challenge 3*). Furthermore, the authors decided to not collect data about intervening variables (*Challenge 2*), in their case for the sake of privacy and to minimize biases that may be caused by the user's worries regarding possible profiling. That is, the study of White and Drucker [187] had to deal with all of these challenges and the authors had to consider the resulting effects in their data analysis. Nevertheless, the results of the study demonstrated the ability to differentiate between seeking behavior and especially revealed (at least a variant of) *exploratory* behavior to be identifiable in interaction logs. This findings from the area of analytical models serve as further motivation for the investigations in this thesis.

3.1.2 *Research & Application Areas*

As pointed out above, analytical models involve several restrictions and assumptions to establish a calculus that enables advanced modeling at all and facilitates the investigation of certain behavioral aspects. If interaction data is extracted from (huge) collections of log files, this property, namely to calculate with the recorded data, gets quite important. Especially the analysis of behavioral aspects on SERPs has become an increasing research area of interest. One reason is that users often utilize SERPs as an entry point for their seeking. A further reason is the potential semantic within the SERP interaction data, e.g., the manner how users search for information or what topics users are interested in. This in turn can be used to tune search engine's parameters [101] from the providers perspective or can be utilized to improve personalized ranking [3] what is a relevant factor as well. Nowadays, SERPs have taken the role of information resp. knowledge hubs where users not only begin their seeking but also return, generate new SERPs (by reformulating the query) and evolve their information need and seeking over time. This allows search engine providers to cover but also to support an essential part of user's ISB. With their popularity and relative high coverage of the user's search process, investigations on SERPs (including mere empirical observations but also modeling aspects), serve as an adequate entry point to the analytical foundations of the thesis and is outlined in the next section.

3.2 ANALYZING & MODELING BEHAVIOR ON SEARCH ENGINE RESULT PAGES

The previous section motivated and justified the huge number of investigations which are focused on analyzing and modeling user's behavior on SERPs. The purposes of the investigations are manifold and comprise for example: analyzing the influence of result's ranking; examining the user's gaze behavior on SERPs (measured by eye-tracking or estimated by the usage of the cursor and item hovering); or predicting the click probabilities for the several SERP items. For each of this purposes, several models have been proposed under their own specific assumptions. For instance, assumptions made for models of SERP related investigations are: *'the examination of a result depends on the rank'*, *'individual items in the SERP are independent to each other'* or *'after clicking on a result the search stops'*. However, even with this restricting assumptions, the studies, analysis, resulting models and findings have a huge value for the Information Retrieval (IR) and IB research field. Furthermore, they are relevant for this thesis as well because they provide insights to ISEB and ISB and allow to derive possible interaction features for the design and development of new interfaces. In the remainder of this section, investigations and models regarding user search and click behavior in SERPs are presented and discussed. Afterwards, in Section 3.3, a step back is made and the analysis of search activities as well as the modeling of ISB in general, inherently considering SERP related behavior, will be described.

3.2.1 *User's Click Behavior*

The analysis of user's click behavior basically aims to answer the question, how users interact with lists of web search results provided by a search engine as response to a query. This facilitates the understanding of user's behavior on SERPs in more detail and partially user seeking behavior in general. According to Granka et al. [77], the results of such investigations are beneficial for advances in the interpretation of implicit user feedback; the development of improved user interfaces; but also can lead to suggestions for more metrics to evaluate search engines' retrieval performance. In the following, several investigations of user's click behavior are presented. The findings motivate the approximation and modeling of the related click behavior on SERPs, as presented in Section 3.2.2, and contribute to (more general) modeling approaches of user's ISBs, as presented in Section 3.3.

While searching for information in the Internet utilizing search engines, users are confronted with the task to identify the most promising items in the result list which have a high relevance, i.e., a good matching to the user's underlying information need, expressed by a query. Usually, search engines provide the results in a ranked list and to generate such a ranked list is one of the fundamental problems in IR [3]. However, the better the IR algorithms, the better the matching and rankings and the more likely it is to find relevant results in the higher ranks. Hence, users can minimize their effort by (first) visiting the higher ranked items in the SERP. This leads to the so called *position bias* where the probability of a user's click decays with the rank [56]. The position bias was investigated, modeled and confirmed by several studies, such as [47, 49, 56, 102, 103]. Nevertheless, Agichtein et al. [3] showed that, also with position bias, users are able to identify relevant results even though the relevant results are not in the first positions.

From the perspective of modeling (cf. Sect. 3.2.2), the first and simplest approach is just to calculate the probability of a click on a (relevant) result given a query and therefore, simulating an unbiased[1] user. Granka et al. [77] investigated the user's eye-tracking behavior on SERPs before the very first click on a result and could confirm as well that users tend to scan the result list from the top to the bottom. The authors also showed: Users spend almost the same time on viewing the first and second result (which together takes up the majority of the search time) but eventually users mostly click on the first result (not the second). Furthermore, if users start to scroll down in the SERP (typically after the fifth to sixth link), the influence of the rank to the user behavior appears to decrease and users, who clicked on a document with a low rank, scanned proportionately more snipes. This could be a small indicator for more exploratory related search behavior. Cutrell and Guan [48, 49] conducted a SERP eye-tracking study with users who had to answer several informational and navigational tasks. They found that the amount of additional information about a search result, in particular the length of the snippets, has an influence to the user's search performance but also depends on the task type users have to solve. Furthermore, the authors refer to a so called *hub and spoke* search pattern where users on SERPs click on a web document, return to the same SERP (using the web browser's back button) and click the next target document. This rather empirical observations gave a first insight regarding users' click behavior on SERPs. In the following, modeling approaches regarding click behavior are reviewed.

Hub and spoke search pattern on SERPs

3.2.2 *Models to Predict the Click Behavior*

Similar to the empirical observation of users' click behavior, the goal of click models on SERPs is to calculate resp. predict the probability that a user clicks on a certain web document in the result list. To calculate a click (frequently denoted by a binary random variable c), on a web document u that is listed on a SERP, often the corresponding query q

1 "Unbiased" here is used in the sense that the models do not include biases explicitly as parameter. However, the (real world) data used to train the models may still contain several biases.

issued by a user is used. Relative simple[2] versions of click behavior models just strait use the conditional probability $P(c|u,q)$ with:

$$P(c|u,q) = \frac{P(u,q|c) \times P(c)}{P(u,q)} \tag{3.1}$$

Since it was revealed that the position (i.e., the rank) of a web document u in the SERP also plays an intervening role and biases users, the probability that a click decays with the rank [56] was added to several models, usually by extending the parameter set by the rank variable r what leads to $P(c|u,q,r)$ with:

$$P(c|u,q,r) = \frac{P(u,q,r|c) \times P(c)}{P(u,q,r)} \tag{3.2}$$

Obviously, between the presentation of a SERP to the user and the user's decision to click on a certain web document in the SERP list, there is happening more seeking related behavior what is not yet represented by Eq. (3.2). That is why there has been a huge interest for investigation on further (physically) observable user variables, such as the gaze behavior (cf. Sect. 3.2.1). Joachims et al. [102, 103] analyzed the gaze behavior and decision making of users on SERPs in comparison to the click behavior. They could show that clicks are a good indicator for relative relevance judgments but are not sufficient for absolute judgments and therefore, methods of eye-tracking can deliver additional beneficial data.

3.2.2.1 *User Browsing Model*

A popular and still relative simple model for users' click behavior in SERPs that also considers and confirms findings of eye-tracking research is the User Browsing Model (UBM) from Dupret and Piwowarski [56]. It's goal is to estimate the probability that a user *examines* (i.e., looks at) a web document on a SERP estimated by the document's rank r but also by the distance d (in terms of ranks) to the last clicked (i.e., probably relevant) web document. Furthermore, the UBM builds on the observation that user's decision to click a link to a web document depends on it's so called *attractiveness*, what is estimated by the available information about the document, e.g., by the URL or the spinet provided by the search engine on the SERP. Both, the importance of the distance to the last clicked web document but also the attractiveness of a web document have been investigated and highlighted by the eye-traking studies of Joachims et al. [102, 103].

2 Of course, a simpler version can be achieved by ignoring the query and just attempt to derive the click probability given a certain web document u by $P(c|u)$ (interpreted as the document's relevance), cf. [47] or even simpler by ignoring certain documents and only use a universal click probability $P(c)$, cf. [78].

As a result, the probability $P(c|u,q,r)$ is extended by a distance d to the last clicked web document in the same SERP. Finally, in the UBM, $P(c|u,q,r,d)$ is calculated by the product of the probability for a web document being examined (indicated by the binary variable e) and the probability representing a web document's attractiveness (indicated by the binary variable a):

$$P(c|u,q,r,d) = P(e|r,d) \times P(a|u,q) \tag{3.3}$$

The probability of examining a web document $P(e|r,d)$ by, e.g., viewing it's URL or snippet, depends on the rank r and the rank distance d to the last clicked web document on the same SERP[3]. If there is no such last web document, d is approximated by the distance to a virtual position zero. The probability of a link's attractiveness $P(a|u,q)$ in turn depends on the web document u and the query q as in Eq. (3.1) and can be interpreted as the document's relevance regarding q that leads to a click. For simple (position related) models, such as in Eq. (3.1) and Eq. (3.2), there have been critic that the web documents u_i in the SERP are considered to be independent and hence, the models do not capture the interaction between. By considering the distance d between the last clicked and the current (examined) document, the UBM addresses this issue even though two different examined but not clicked documents u_i and u_j (with $i \neq j$) are still not correlated.

3.2.2.2 *Cascade Model*

In the following, a further traditional click model, namely the *Cascade Model* (already mentioned in Section 3.1 because of it's strong assumptions), is explained. The model, proposed by Craswell et al. [47] and inspired by the work of Joachims (et al.) [101, 102], is grounded on the assumption that users iterative view the result list from the top to the bottom and for each result first decide whether they click on that result or not before they move to the next result on the list. In the model's most basic version, a web document u can only be clicked with the probability of $P(a|u,q)$ or skipped with $1 - P(a|u,q)$. Finally, the calculation ends if a document is clicked, i.e., users coming back to the SERP and proceed their search can not be modeled. The calculus for the Cascade Model is:

$$P(c|u_r,q) = P(a|u_r,q) \times \prod_{i=1}^{r-i}(1 - P(a|u_i,q)) \tag{3.4}$$

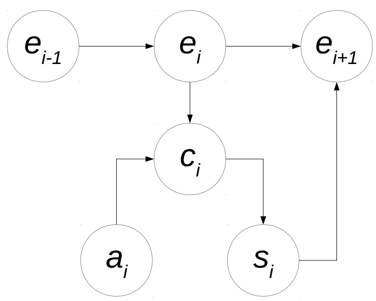
3 As a matter of fact, it could be shown that the probability of a click on a web document, considering it's rank, decays faster than the probability of examining the document [102, 103].

where $u_{<rank>}$ indicates the rank of document u in the SERP. That is, to calculate the probability $P(c|u_r, q)$ that document u_r at rank r is clicked, all documents u_i in the SERP before have to be skipped by the user. Although the Cascade Model performs better than the models from Eq. (3.1), (3.2) and (3.3) on higher ranks [47], it's assumptions are to strong for a holistic perspective. That is, the model can not be used for searches with more clicks [38] or to explain EIS because it can be assumed that back tracking (to the SERPs) or even more complex interactions are relevant aspects of exploratory search (activities)[4]. Furthermore, on lower ranks, the Cascade Model is outperformed by models using Eq. (3.1) to (3.3). However, to investigate and approximate the position bias (it's actual goal), the Cascade Model had a huge impact to it's field of research. In addition to that, the model is (still) utilized as foundation for many extensions and/or is used as baseline for comparative evaluations.

3.2.2.3 *Dynamic Bayesian Network*

A click model that extends the Cascade Model using the Bayesian network approach [67] was proposed by Chapelle and Zhang [38] and is called Dynamic Bayesian Network (DBN). As in the Cascade Model, the DBN assumes that users iterative view the SERP list beginning from the top but the DBN does not stop if a document was clicked and therefore, allows to proceed the search and perform several clicks. This enables the DBN to distinguish between a perceived relevant document, that was clicked but not satisfied the user's information need, and an actual relevant document, that ends the calculation because it satisfied the user's information need. A simplified version of the DBN as graphical model and set of equations is illustrated in Fig. 3.1. Similar to the previous click models in this chapter, the probability of a *click* c_i on the i'th, *examined* web document u_i depends on the

4 In fact, back tracking respectively the *hub and spoke* pattern, mentioned by Cutrell and Guan [48, 49] (cf. Sect. 3.2.1), could be confirmed in the investigations of this thesis (cf. Sect. 5.1.3).



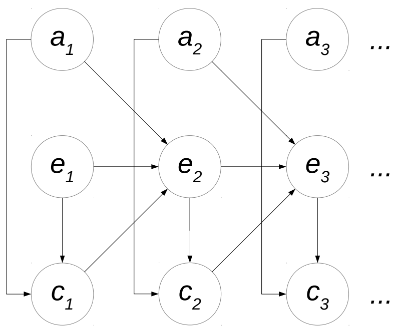
$$
\begin{aligned}
&a_i = 1, e_i = 1 \Leftrightarrow c_i = 1\\
&c_i = 0 \Rightarrow s_i = 0\\
&s_i = 1 \Rightarrow e_{i+1} = 0\\
&P(e_{i+1} = 1 | e_i = 1, s_i = 0) = \gamma\\
&e_i = 0 \Rightarrow e_{i+1} = 0
\end{aligned} \tag{3.5}
$$

Figure 3.1: Illustration of the DBN as graphical model (left) and it's corresponding (simplified) set of equations that defines the DBN (right) according to Chapelle and Zhang [38].

attractiveness indicated by the binary variables e_i and a_i respectively. Specific to the DBN is that in addition, the i'th web document can satisfy the information need ($s_i = 1$) or not ($s_i = 0$) and if not, the search proceeds. Equation set (3.5) has some more implications. The probability of attractiveness $P(a_i = 1)$ depends only on the URL of a web document u_i; and the probability that a web document satisfied the user's information need depends on a prior click on that documents $P(s_i = 1|c_i = 1)$. Furthermore, the probability that the user examines the next document e_{i+1} (in case of a non-satisfying current document) is denoted by γ. That is, the user can also just abandon the search with the probability of $1 - \gamma$. Thus, the Cascade Model can be seen as a special case of the DBN where $P(s_i = 1|c_i = 1) = \gamma = 1$. Basically, the DBN is a Hidden Markov Model (cf. Sect. 3.3.2), expressed as a Bayesian Network, with an additional conditional dependency between the observation c_i and the hidden state e_{i+1}. Such as the UBM, also the DBN does not consider the documents in the SERP list as independent, simply because the examination of one document depends on the probability of satisfaction of the previous document in the list. Finally, the approaches to estimate the users satisfaction by certain web documents are still manifold. In addition to the DBNs original proposal, the query-specific feature set investigated by Agichtein et al. [3] resp. the more general query-independent model of Fox et al. [64] could be used to extend the DBN model.

3.2.2.4 *Click Chain Model*

The last model of this sub section, the Click Chain Model (CCM) [79], also was influenced by the way how the Cascade Model addresses the position bias. The available information about web documents in the SERP are examined by the user step-by-step. The predecessor of the CCM is the Dependent Click Model (DCM) from Guo et al. [78]. As the DBN, the CCM uses a solid Bayesian background and (as extension to the Cascade Model) allows the user to continue the search after



$$
\begin{aligned}
&P(c_i = 1|e_i = 0) = 0 \\
&P(c_i = 1|e_i = 1, a_i) = a_i \\
&P(e_{i+1}|e_i = 0) = 0 \\
&P(e_{i+1} = 1|e_i = 1, c_i = 0) = \alpha_1 \\
&P(e_{i+1} = 1|e_i = 1, c_i = 1, a_i) \\
&= \alpha_2(1 - a_i) + \alpha_3(a_i)
\end{aligned}
\tag{3.6}
$$

Figure 3.2: Illustration of the CCM as graphical model (left) and it's corresponding set of equations that defines the CCM (right) according to Guo et al. [79].

a click on a web document. The CCM can not distinguish whether a clicked document was satisfying or not (as the DBN) but differentiates the probability of examining the following web document depending on a click or skip regarding the current document by using three specific variables (α_1 to α_3). An illustration of the CCM as graphical model and the set of corresponding equations is given in Fig. 3.2. After examining e_i a document at position i, the user chooses to click c_i on the document if it appears to be relevant a_i or the user chooses to skip it. If skipped, the examination e_{i+1} of the next document at position $i+1$ only depends on the parameter α_1 that in turn depends on user behavior parameters. If the user chooses to click c_i, the examination e_{i+1} of the next document at position $i+1$ depends on the perceived relevance a_i of the clicked document and ranges between the two parameters α_2 and α_3 using: $\alpha_2(1-a_i)+\alpha_3(a_i)$. As in the DBN, only the click variable(s) c_i are observed by the log files. Comparing the CCM to the UBM and it's predecessor, the DCM, the CCM performs better on several evaluation metrics [79].

3.2.3 *Limitations of Search Engine Result Page related Approaches*

Although the investigations and findings of SERP related approaches extensively contributed to the research fields of IB and ISB, there are several limitations that have to be taken into account.

First, there are limitations regarding the modeling of click behavior itself considering the multifaceted findings of the analysis on SERPs (cf. Sect. 3.2.1). For example, in addition to the position bias, Granka et al. [77] revealed that users primary examine the first and second results in the SERP (before clicking most likely on the first result). An interaction behavior where users do not progressively examine (and maybe click) a result, i.e., also examine or even (re-)click on previous results is not allowed by the here describes click models and is rarely addressed in the research. Furthermore, the findings that the influence of the position bias decreases on lower ranks [77] is only implicitly addressed by the models, because the difference in the click probability between two lower rank items is small if their click probability is small anyway (because of the modeled position bias). Similar to the inability to examine and click SERP items in arbitrary order, the *hub and spoke* pattern (cf. Sect. 3.2.1) would require the SERP models to (re-)click on results in arbitrary order, what is not the case. That is, only a part of the findings revealed by the SERP analyses are implemented by the models. The impact and reciprocal interplay to (not implemented) behavioral aspects and the models performance however remains uncharted.

Second, there are limitations of the click behavior models regarding the investigation focus, because a holistic perspective on the information (seeking) behavior, as depicted in Chapter 2, which considers aspects of pre- post- but also several peri- search engine usage, is neglected. SERP related models do and can not consider the rise and development of the user's information need or the consolidation of single queries to answer a (maybe more complex) search task. Furthermore, they (usually) ignore the executed behavior (and duration) on clicked web pages, neglect the interaction between web pages in general, etc. To give some more specific examples, Lorigo et al. [127] showed that users spend only about half of the time or less (depending on the current search task) on SERPs and the remaining time on web pages. This highlights the importance to include web pages but also information about (the possible current) search task into the modeling as well. The study of Claypool et al. [41] could show that the time users spend on web pages indicates the level of the user's interest. Konstan et al. [112] analyzed implicit measures for user behavior in a collaborative filtering setting. There findings revealed a strong relationship between reading time of articles and user's interest. Hence, dwell (and also reading) time are important factors for seeking behavior. To take this knowledge and further IB and ISB aspects into account, the modeling approaches from this section (cf. Sect. 3.2) have to been extended. In the following, analysis and user models in a more general context are described. That as well includes approaches, which already have partially connections to the theoretical models from Chapter 2.

3.3 analyzing search activities & modeling information seeking behavior

The models described in Section 3.2.2 largely contributed to the interpretation and understanding of user behavior on SERPs as a probabilistic process. Nevertheless, as pointed out, investigation on ISB, utilizing analytical models, should not only consider SERP interaction but seeking behavior on a more general level. In the field of IR, this usually leads to the representation of the users' entire search activities as a process over time (e.g., by utilizing Kuhlthau's model as underlying framework, cf. Sect. 2.4.2). That is, seeking behavior is considered as a sequence of search (engine) related interactions where the users traverse different stages. Thereby, interactions on SERPs become a part of the process. This approach not only allows but also requires the investigation and analysis on a multitude of interactions performed during the search to reveal crucial behavioral aspects and to identify relevant variables. The variables here can be user-, tasks- or search system related. User related variables can have a multifaceted impact to the seeking behavior and, e.g., may include age, gender, experience, etc. The (search) tasks related variables may include characteristics, such as task difficulty or complexity[5]. The task's type, (e.g., catego-

5 Task difficulty and task complexity can influence the search behavior. According to Li and Belkin [125], the difficulty of a task is considered as subjective and depending

rized as a fact-finding or exploratory search task) is a further tasks related variable that represents crucial key aspects for current investigations and can influence the development for future search interfaces and systems. Furthermore, the search system itself and the provided means may have an impact to the search and it's outcome.

In the following Section 3.3.1, related work on the analysis of users' search activities and relevant variables for ISB are discussed. Afterwards, the (Markovian) models, as used in this thesis to investigate user's ISB in context of ES, are motivated and defined in Section 3.3.2. In addition, related work using the (Markovian) models is reviewed and similarities as well as differences to the here applied approaches are discussed. Last but not least, alternative approaches to model ISB are outlined in Section 3.3.3 to conclude this chapter. It shall be noted that depending on the focus of research, (user) models can predict respectively classify several (user) aspects and hence, the models can largely differ in size, complexity, number of considered variables but also in their theoretical foundations and assumptions.

3.3.1 *User's Search Activities*

Attempts to identify the stages of Kuhlthau's ISP

Kuhlthau's model [120] (cf. Sect. 2.4.2) and it's corresp. Information Search Process (ISP) frequently serves as inspiration and framework to investigate users' search behavior and search activities. Shah and González–Ibáñez [172] proposed an approach to analyze the six stages of the ISP over two search sessions in a collaborative setting (where two people are seeking together). In their study, exploratory search tasks have been given and the participants utilized the author's experimental *Coagmento* [170] web browser plugin that was designed for collaborative information seeking. To operationalize the several ISP stages, the authors used a mapping of certain logged iterations from the search system to the individual phases. It could be shown that after reading the task description (ISP initiation) discussing and devising the strategy (ISP selection) could be detected using the mapping and hence, this ISP stages could be differentiated relatively clearly. Afterwards the users often switched between the hereafter ISP stages. That is, for this collaborative information seeking setting, the distinction between the ISP stages formulation, exploration and collection stages was vague. Using the authors' mapping, the last ISP stages (presentation) could be identified as well. In contrast to their work, this thesis investigates (among other things) the classification of search activities within a single session, not two or several sessions. Furthermore, the tasks used

on the user's assessment, i.e., how easy or hard the user perceives resp. estimates the task completion. The objective task complexity addresses whether just a single or (significantly) more paths are included in the task solving. In addition to potential multiple ways, Campbell [36] also associates uncertainty, conflicting interdependence between the paths and the number of possible task outcomes as relevant attributes for task complexity.

in [172] initially fulfill the requirements for exploratory search tasks in context of this thesis but the investigators simplified the task's challenge by splitting the exploratory tasks into single (almost fact based) sub-tasks what may influence an unbiased exploratory task processing and actually changes the task's type. Last but not least, the utilized search system was designed for collaborative seeking. This addressed the influencing variables of system provided means but the users had the additional challenge to get familiar with the functionality of the system. However, the investigation of Shah and González–Ibáñez [172] showed how (collaborative) seeking can be supported, they could illustrate the dense interplay between several ISP stages and thereby, again confirm the necessity and challenges of their differentiation. Being able to differentiate between search activities, a user is currently engaged in, is especially crucial for adaptive information retrieval systems in order to provide users with appropriate support at the time of the search. This challenge was already mentioned by Belkin [22] in 2008 but is still not sufficiently solved.

Differences between "open" and "closed" search tasks

Further research to investigate differences in ISB was done by Marchionini [129] in that he analyzed influencing user variables but also the task type. In particular, Marchionini investigated the differences between young[6] users' search behavior caused by "open" (i.e., imprecise) tasks where the information need can not be specified precisely and "closed" (i.e., precise) tasks that lead to a clear derivable information need. In his studies, Marchionini was able to show that especially novices need more time and have more difficulties to specify the queries for "open" tasks. In contrast, older users had more success and needed less time. In general, all users needed more time and moves for "open" tasks than for "closed" tasks. That is, the results could confirm that the user's seeking behavior differs regarding the type of the current (search) task and further depends on the user's experience in terms of domains knowledge as well as the familiarity with the available means to access the information sources[7] and consequently the ability to formulate an adequate information need by search queries. The "open" tasks used by Marchionini basically fulfill the requirements for exploratory search tasks and vice versa, the "closed" tasks basically fulfill the requirements for fact-finding search tasks. Similar to Shah and González–Ibáñez [172] (see above), Marchionini used a mapping of certain system iterations for the operationalization of the two categories *lookup* and *examine*. Lookup here is associated to actions in terms of query (re-)formulation. Examine encompass information gaining actions like showing titles or text. That is, the two categories lookup and examine, used here by Marchionini,

Cf. discussion on exploratory search tasks, as used for user study US-II *and US-III, in* Sect. 4.3.2.2

6 The study was conducted with elementary school children consisting of two groups: First, a novice group (28 third and fourth graders) and second, an experienced group (24 sixth graders).

7 This is in accordance with the conclusions of exploratory search's characteristics described in Sect. 2.6, p. 34.

are early concepts of the search activities lookup and exploratory search as used for this thesis' investigations (cf. Chap. 5). Anyway, the results of Marchionini showed that to solve "open" tasks, more moves respectively actions (in contrast to "closed" tasks) were necessary. Younger users needed more time to refine their queries and older searchers more frequently used examination related actions than lookup related actions.

The differentiation between search behavior (or patterns), and in particular, the identification of exploratory search behavior, is a central research question in this thesis (cf. research question *Q*2, resp. hypothesis *H*3). In contrast to the work of Marchionini [129] and Shah and González–Ibáñez [172], in this thesis no mapping for predefined specific interactions to operationalize the several search activities are necessary. The models in this work can handle sequences of interactions (of arbitrary length) and identify search activities by state transition probabilities and selected features even independent of specific semantic actions made by the user. Furthermore, Marchionini simplified the analysis by only consider staying or switching in the corresponding two search categories (resp. stages). For the investigations in this thesis, each search activity may consist of several (user) states, incorporating several feature information.

Differences between easy and difficult closed search tasks

Focusing on easy and difficult closed informational tasks, Aula et al. [11] presented the results of a lab- and a large-scale study with the goal to detect differences in users' seeking behavior if they have difficulties to find certain information. In turn, this differences can allow to derive the task difficulty based on the performed ISB. In their work, Aula et al. could show that users increase the number of queries and spend more time and a larger proportion on SERPs if they had difficulties to solve the tasks. Furthermore, the authors reported that if keywords from the task description are used as query and did not showed satisfying results (for unsuccessful tasks), users started to utilize full phrases, e.g., full questions. For easy tasks, successful users often started with a general (few keywords) query and afterward specified the request (using more keywords) to find the exact answer. In their study, Aula et al. [11] primary focused on performed and adapted search strategies regarding the utilized queries. The interaction and differences on SERPs were discussed as well but the interaction between SERP and web pages were mostly not discussed. Furthermore, possible influencing features, such as scrolling, and a discussion in context of information seeking as process was not given. Especially the behavior on web pages and possible differences in the sequences (query, SERP, web page) would be helpful to understand the differences from a more general perspective. However, that users examine more results (i.e., web pages) in general if they are confronted with difficult tasks could be shown by Gwizdka and Spence [82] as well as by Kim [110]. Finally, in the study of Aula et al. [11], no further evaluation whether the difficulty level of the tasks correspond to the perceived success resp. to the user rated success was given. Although this was not the main goal of the investigation, to analyze user's seeking behavior (induced by certain search tasks), such an evaluation could give insights about what kind of tasks lead to what behavior and whether there are border line tasks resp. behavior.

Cf. Sect. 5.3 providing an investigation to identify borderline search behavior

Identifying the tasks difficulty

Detecting the difficulty of the task –a user is currently engaged in and given several behavioral indicators– was the goal of the investigation of Liu et al. [126]. They investigated the influence of tasks with different difficulty levels and different types on the user behavior. The authors criticize that most of the previous work regarding behavioral aspects have been focused on the overall task level. Under circumstances, this can prevent a consideration and investigation of certain important user resp. system variables until the end of the tasks or generates measures which are (too) dependent on the task perspective. Therefore, for real time systems, more dynamically approaches are necessary, for example, to detect the users' current search activity or to detect whether users have troubles as early as possible and not if the task is already solved or abandoned. The analyzes of Liu et al. [126] confirmed previous studies by showing that for difficult but also for open-ended (exploratory) tasks users need more completion time, formulate more queries and visit more web pages[8]. Furthermore, their results showed that dwell time alone is not a reliable measure for prediction whether users are performing difficult tasks but additional knowledge, e.g., about the number of viewed (content) pages, can increase the detection rate[9]. The results of Liu et al. are promising and also serve as an inspiration and motivation for several investigations of this thesis. However, in addition to the study of Liu et al. [126], in this thesis several relevant variables (features) are used to identify the users' current search activity. Furthermore, the used framework of Markovian models allow to classify interaction chains of arbitrary length (cf. Sect. 5.2) but also to identify new previously unknown search behavior using a clustering approach (cf. Sect. 5.3).

Similarities between exploration and struggling to find certain information

Hassan et al. [85] noted that there are similar aspects in between users who are exploring and users who are struggling to find certain information they are seeking for. The argumentation is that in both, exploration but also having difficulties during the search, several

8 In fact, Kim [110] showed that for exploratory search tasks even the previously perceived difficulty, i.e., a user's pre-task estimation of how difficult the task will be, correlates with the number of pages viewed and saved. For factual tasks this was not the case.

9 This is consistent with the findings in this thesis. The dwell time turned out to be a reliable feature for search activity classification. The number of web page views, recommended by Liu et al. [126], are included in this thesis models' as well but not as total number but as probability based transitions between the state "web page" and the remaining states. This also allows a broader investigation on the user's information behavior and not focusing on the (pure) tasks level, as criticized by Liu et al. [126].

characteristics such as long(er) dwell times, or the number of queries per session, are observed. Hence, the goal of their investigation is to analyze similarities and to build a classifier that is able to distinguish between both scenarios. The analysis of the click behavior showed that exploratory search sessions have more result clicks than sessions where struggling occurs. At the beginning of a session, this difference is small but increases if the session continues. Furthermore the authors could show that the dwell time on pages in case of struggling increases during the session but the dwell times in exploratory sessions are always higher. Considering the session topics, Hassan et al. revealed in their data set that exploration on certain topics, such as "traveling" and "entertainment", is more likely whereas topics as "software" and "download" are more likely for struggling sessions. However, an analysis of different search features could show that the search topic is only helpful if no other information is given (e.g., at beginning of the session). The importance of query related features turned out to be moderate in the beginning and decreases slightly over time. On the contrary, the importance of click related features increases quickly and stays moderate over time. This confirms again the advantages of analyzing and modeling SERP related behavior as an relevant part of information seeking behavior, as described in Section 3.2. The investigation of Hassan et al. also showed that the need for advanced methods to distinguish between the users situation is high. However, the major limitation of the study is the generation of the ground truth. The labels for exploring or struggling but also the estimation of the search success labels was done on the base of log data without further knowledge about the performed tasks or motivation of the users.

Cf. discussion on challenges of online log file based investigations, Sect. 3.1.1

Differentiating exploratory and lookup tasks

Athukorala et al. [10] conducted a user study with 32 computer science researchers and investigated different search behavior indicators, such as maximum scroll depth, proportion of browsing and query related parameters, that can help to distinguish exploratory from lookup tasks. They found that the length of the first query is shorter, users spend more time on (reading the) documents and scroll significantly deeper in exploratory tasks than in lookup tasks. To operationalize exploratory and lookup tasks, the two facets *preciseness of the search goal* and *objective task complexity* are used. On the basis of these facets, exploratory and lookup tasks could be further divided into six sub-categories: *knowledge acquisition*, *planning* and *comparison* for exploratory search; and *fact-fining*, *navigation* and *question answering* for lookup search tasks. For example, knowledge acquisitions has usually an open-ended search goal with a high complexity whereas pure fact-finding is usually a closed informal task with low complexity. The sub-categories' comparison and question answering turned out to be borderline cases since their parameters preciseness of the search goal and objective task complexity are closer to the other main category respectively. Recalling Marchionini's framework from Section 2.6.1, the six sub-categories are also part of his framework. The relation between the six sub-categories from Athukorala et al. [10] to further search activities (and hence, potential sub-categories) listed by Marchionini, such as verification, interpretation or synthesis remain open. Considering the possibility that lookup is partially contained in the process of exploratory search makes an analytical differentiation or even automated classification of sub-categories decidedly challenging if not even impossible for (short) single search sessions and without exhausting long term personalization. That is why the investigations for this thesis are focused on the user modeling and user seeking behavior classification on the broader categories exploratory and lookup search (as described by Marchionini) but in contrast allow identification after only a few interactions (cf. Sect. 5.2.5). Furthermore, the broader categorization enables the connection to the theoretical foundations of information (seeking) behavior from Chapter 2 as a further thesis' goal. Nevertheless, the findings of Athukorala et al. [10] revealed insights of and the relation between a set of selected, possible sub-categories. However, the findings are limited to information search tasks related to the machine learning domain in scientific documents utilizing a scientific search engine with, e.g., up to 40 search results visualized per SERP what exceeds the usual number of results for search engines these days in general.

Analyzing sub-sequences in exploratory search

Hendahewa and Shah [90] conducted a sequential analysis of user actions performed during an exploratory search task. To find similarities between sub-sequences, a sliding window approach to segment the sequences was used. Afterwards, a hierarchical agglomerative clustering (using a Hamming distance as similarity measure) to identify four sub-sequence clusters representing exploratory search episodes was applied. While in one cluster users strongly tend to view resp. read the content on web pages, in another cluster user almost only spend their time on collecting snippets. However, in the study, SERPs have been neglected for the analysis what seems to be disadvantageous considering the impact and dominance of this pages highlighted and confirmed a multiple time by the literature, cf. Section 3.2. Furthermore, the relation of the identified four clusters to empirically identified, individual phases, stages or tactics of the ISP, i.e., the relation to the ISB models is not given. Nevertheless, the investigation of Hendahewa and Shah [90] showed again the diversity of user actions exploratory search can involve.

The results of the different investigations above point out the complexity of user's seeking behavior and the need for appropriate user support in relation to the current search activity. A search system that is able to distinguish between fact-finding and exploratory search would be a first and solid step to provide enhanced methods for an adaptive user support in the future. Therefore, this aspects are investigated in Chapter 5. Finally, a comprehensive and recommendable literature review about exploratory search and related studies is given by Shah et al. [173].

3.3.2 *Markovian Models for User's Information Seeking*

The analysis of users' information seeking behavior allows to reveal relevant variables that influence the search process and to decide what aspects should be considered in the corresponding user models. Conversely, user models can be used to further analyze relevant parameters of user's information seeking but also provide a powerful means to detect certain user behavior in terms of classification (e.g., with the goal to customize and adapt a search system to the user's needs) or to reveal new search patterns of search behavior. By nature, (user) models represent always an abstraction resp. simplification of the real world (users). That is, the findings and possible conclusions always have to be considered carefully and may not necessarily reflect the current user state or the predicted future behavior correctly. However, for certain aspects and cases, user models facilitate sufficiently performance, even outperform alternative approaches and allow to include behavioral facets which can not be represented otherwise. Since user behavior can and actually should be considered as a sequence of interactions with the system over time (i.e., as time series), methods that are able to handle sequential data have several advantage over static methods which may be neglect possible interference between sequence units (also called states, situations or interactions). At this point, (Hidden) Markov Models [58] have been established as a popular and rich framework that allows to represent several aspects of sequential data. Originally used to detect vowels in words, Markov models (actually their extension to Hidden Markov Models) have been used in all kinds of text analysis, such as part of speech or named entity tagging. But also in automatic speech recognition; speech enhancement and synthesis; and other natural language processing tasks, such as language understanding or machine translation, this models have gained huge popularity. Reasons for this status are the model's adaptable structure, performance and their ability to assign probabilities to unambiguous sequences. With the availability of data about users' interaction in the Internet, (Hidden) Markov Models also found their way into the area of user modeling and ISB modeling respectively (interactive) IR. In the following sub-sections, (Hidden) Markov Models are introduced, defined and related work in the area of user ISB modeling is discussed.

Origin of Markov Models

Markov Models for user interactions

3.3.2.1 *Markov Models*

Markov models, also called Markov Chains, are probabilistic sequence models or sequence classifiers. For a given sequence of units they calculate it's probability. The units of the sequence can be any kind of elements over a given finite and discrete alphabet. For example, a word is a sequence of letters over a natural language's alphabet; or a search session of a user can be a sequence of (a finite set of possible) interactions with a search engine. The latter example will be used in this thesis to model and classify the users' ISB. In context of Markov models, the units of a sequence correspond to the model's states consisting of a finite and discrete state space (to represent the alphabet for all possible sequences). To traverse from one state to another, transitions with corresponding probabilities are used. That is, Markov models are generative models which embody a sequence generation processes by traversing their states with certain probabilities. For a solid mathematical background and to derive several properties for the Markov models, the states can be represented as random variables. Furthermore, (1st-order) Markov models are often represented as a directed graph model where it's nodes represent the states and the links represent the transition probabilities between the states. As usual, the graph's link structure can be represented as a matrix. According to the notation of Huang et al. [93], Jurafsky and Martin [106] as well as Rabiner and Juang [153], for this thesis, Markov models are defined by the following components Q, A and π:

Def.: Markov Model

$Q = \{q_1, q_2, ..., q_N\}$	A set of N states, the finite and discrete state space
$A = a_{11}, a_{12}, ..., a_{1N}, ..., a_{NN}$	A transition probability matrix where each a_{ij} represents the probability to traverse from state q_i to state q_j; furthermore: $a_{ij} \geq 0$ and $\sum_{j=1}^{N} a_{ij} = 1$ for $1 \leq i \leq N$
$\pi = (\pi_1, \pi_2, ..., \pi_N)$	A vector of N start probabilities denoting the model will start in state q_i; furthermore: $\pi_i \geq 0$ and $\sum_{i=1}^{N} \pi_i = 1$ for $1 \leq i \leq N$

That is, Markov models are a specification of weighted finite state automatons with stochastic state transitions [152] where for each state the sum of all outgoing probability links equals to one. Using the *Bayes Theorem*, the probability of a sequence $S = S_1, S_2, ..., S_L$ of random variables and with length L is calculated by:

$$P(S) = P(S_1, S_2, ..., S_L) = P(S_1) \times \prod_{l=2}^{L} P(S_l | S_1, ..., S_{l-1}) \tag{3.7}$$

where each S_l is associated to a certain value resp. state q_i from the finite state space Q. From Eq. (3.7) it follows that reaching each S_l in sequence S depends on the full history of predecessor states. A common way to simplify that strong requirement but still keep the ability of memorization is the so called *Markov Assumption*:

$$P(S_l|S_1, ..., S_{l-1}) = P(S_l|S_{l-1}) \tag{3.8}$$

Markov Assumption

That is, reaching each S_l in sequence S depends only on its predecessor S_{l-1}. Applying Eq. (3.8) to Eq. (3.7) leads to:

$$P(S) = P(S_1, S_2, ..., S_L) = P(S_1) \times \prod_{l=2}^{L} P(S_l|S_{l-1}) \tag{3.9}$$

If the time index l is discarded, Markov models can be used to model time invariant events:

$$P(S_l = q_j|S_{l-1} = q_i) = P(q_j|q_i) \tag{3.10}$$

Here it has to be noted that the $P(q_j|q_i)$ in Eq. (3.10) corresponds to the definition of the transition probabilities a_{ij} in the Markov model. That is, the a_{ij} are used to model a sequence (a time series) with L states from Q, where each step at l is denoted by S_l, with:

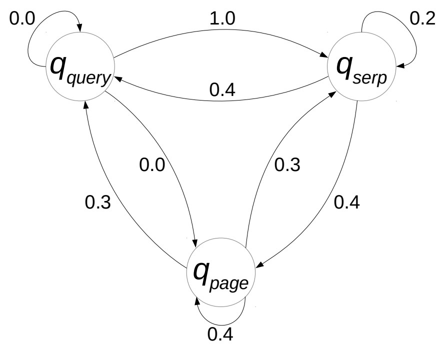
$$a_{ij} = P(S_l = q_j|S_{l-1} = q_i) \quad 1 \leq i, j \leq N \tag{3.11}$$

and

$$\pi_i = P(S_1 = q_i) \quad 1 \leq i \leq N \tag{3.12}$$

for the start probabilities. Using the Markov Assumption as in Eq. (3.8), the probability that the Markov model transits to a certain state q_j at a given step l depends only on its predecessor state q_i. This models are called 1st-order Markov models. In the following, an example of a 1st-order Markov model is given: Assuming a search session with a given search engine is a sequence that can consists of the three interactions resp. states: “user is typing a query” (q_{query}), “user is viewing a SERP” (q_{serp}) and “user is reading a web page” (q_{page}). A graphical and formal description of the corresponding 1st-order Markov model including example transitions and start probabilities is given in Fig. 3.3. Further assume a recorded user search session $S = S_1, S_2, S_3, S_4$ of length $L = 4$ where a user typed a query (S_1), viewed the SERP (S_2), clicked and read a web page (S_3) and finally

1st-order Markov Model



$$Q = \{q_{query}, q_{serp}, q_{page}\}$$

$$A = \{a_{ij}\} = \begin{bmatrix} 0.0 & 1.0 & 0.0 \\ 0.4 & 0.2 & 0.4 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \end{pmatrix}^T = \begin{pmatrix} 1.0 \\ 0.0 \\ 0.0 \end{pmatrix}^T$$

Figure 3.3: Illustration of an example 1st-order Markov model as graph (left) and it's corresponding components Q, A and π (right).

returned to the SERP (S_4). The probability of the session $P(S)$ using the Markov model can than be calculated by:

$$\begin{aligned} P(S) &= P(S_1, S_2, S_3, S_4) \\ &= P(S_1 = q_{query}, S_2 = q_{serp}, S_3 = q_{page}, S_4 = q_{serp}) \\ &= P(S_1 = q_{query}) \times P(S_2 = q_{serp}|S_1 = q_{query}) \times \\ &\quad P(S_3 = q_{page}|S_2 = q_{serp}) \times P(S_4 = q_{serp}|S_3 = q_{page}) \\ &= \pi_{query} \times a_{query\ serp} \times a_{serp\ page} \times a_{page\ serp} \\ &= 1.0 \times 1.0 \times 0.4 \times 0.3 = 0.12 \end{aligned}$$

Furthermore, the probability of being in a particular state $S_l = q_i$ (independent of a certain sequence) can be calculated by the multiplication of $\pi \times A^{l-2}$. This corresponds to the calculation and summation of all possible sequences, generated by the Markov model, where the probability of being in $S_l = q_i$ only depends on the initial situation denoted by π. Reviewing the example above, starting at the state $S_1 = q_{query}$, there are exact three possible sequences ending in state $S_4 = q_{serp}$, namely: $q_{query} - q_{serp} - q_{query} - q_{serp}$, $q_{query} - q_{serp} - q_{serp} - q_{serp}$ and $q_{query} - q_{serp} - q_{page} - q_{serp}$ with the probabilities 0.4, 0.04 and 0.12 respectively, what sums up to 0.56 for the probability of being in $S_4 = q_{serp}$. Using $\pi \times A^{4-2}$, results into $(0.2, 0.56, 0.24)$, showing the same probability 0.56 for $S_4 = q_{serp}$. This example for a 1st-order Markov model illustrated the probability calculation of sequences with arbitrary lengths L considering only the last state. Given a data set $\underline{S}$ of $|\underline{S}| = K$ sequences (e.g., search sessions) indexed by S^k, the calculation of the transition probabilities a_{ij}, i.e., the training of an 1st-order Markov model can be implemented by:

$$a_{ij} = \frac{\sum_{k=1}^{K} \#(\text{transitions in } S^k \text{ from states } i \text{ to } j)}{\sum_{k=1}^{K} \#(\text{transitions in } S^k \text{ from state } i)} = \frac{\sum_{k=1}^{K} \sum_{l=1}^{L-1} \varphi(S_l^k, S_{l+1}^k, i, j)}{\sum_{k=1}^{K} \sum_{l=1}^{L-1} \psi(S_l^k, i)}$$

(3.13)

with

$$\varphi(S_x, S_y, i, j) = \begin{cases} 1 & \text{if } S_x = q_i \quad \text{and} \quad S_y = q_j \\ 0 & \text{otherwise} \end{cases} \tag{3.14}$$

and

$$\psi(S_x, i) = \begin{cases} 1 & \text{if } S_x = q_i \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

Higher-order Markov Model

The # here denotes the counting function to calculate the number of certain transitions from unit i to unit j resp. from unit i (independent of it's successor unit) in a given sequence S^k. Of course, the extend of considered history (i.e., the size of the model's memory) can be increased, what leads to higher-order Markov models. For example, the probability of a sequence S of length L using a 2nd-order Markov model would be:

$$P(S) = P(S_1, S_2, ..., S_L) = P(S_1) \times P(S_2|S_1) \times \prod_{l=3}^{L} P(S_l|S_{l-2}, S_{l-1}) \tag{3.16}$$

Increasing the order of an Markov model, and thus it's complexity, not necessarily results into better classification rates but this will be discussed in Chapter 5. Ignoring the previous state history leads to a 0-order Markov model with the sequence probability:

$$P(S) = P(S_1, S_2, ..., S_L) = P(S_1) \times \prod_{l=2}^{L} P(S_l) \tag{3.17}$$

Here the probability to reach a state only depends on the fraction of time the state was reached resp. traversed in the data set. As mentioned above, Markov models can be used to analyze and classify any kind of sequential data but in this thesis the analysis of ISB as a sequence of interactions with a given search system is focused.

Using Markov Models to simulate users' book search

Tran and Fuhr [182] used a Markov model to investigate users who sought for books of certain topics in a book data base of *Amazon*[10]. The provided user interface consisted of four elements, which correspond to four states in their model: query input (1), SERP (2), a detail area for a chosen result (3) and a shopping basked to save relevant books (4).

10 https://www.amazon.com/

The model parameters in this study showed a strong tendency for the users to: stay in the SERP state; mostly place books in the basked after viewing their details; and most probably return to the SERP afterwards. During this application-oriented study with the given book data base and an individual interface, the strong similarity between the retrieved data base entries (the books) turned out to be disadvantageous because users were forced to check differences between items in the SERP, their details and the items already placed in the basket what may have influences to the users seeking behavior in contrast to a common search session with a known search engine interface. However, the authors also investigated the duration, users spent in the states. Most of the time, users viewed the book's details (ca. 15 sec.), about five sec. users needed to formulate a query and only about two sec. to view a result in the SERP or to view the basket.

As investigated by Tran and Fuhr [182] but also as shown in Section 3.3.1, several search variables, such as duration or scrolling depth, are relevant as well and may reveal further insights into users' seeking behavior. In addition to merely modeling the user behavior via search engine states (e.g., being on the SERP or typing a query), this search variables may represent each of the states in a more natural and comprehensive manner. This implicit (in the sense of rather unconscious) user interactions can be integrated into Markov models as additional emissions or features for particular model states. Formally, this naturally extend the Markov models to Hidden Markov Models, which are defined in the next section.

3.3.2.2 *Hidden Markov Models*

Markov models, as described in the previous section, sometimes are also called observable Markov models because at each time step each state resp. the value of each random variable is known. Thus, Markov models are not able to represent problems that are inherently ambiguous. There are several real life processes, however that do not have full transparency, which means that those processes only have observable (output) variables but the underlying sequence generation process and the (possible) internal states of the process behind are not or only partially known and hence, "hidden". In the beginning of Section 3.3.1, several examples for this not (directly) observable processes are given which are often related to sequence labeling (classification) problems. In contrast to Markov models, Hidden Markov Models (HMMs) can represent non-deterministic, ambiguous processes which can generate observable (output) variables from any internal (hidden) state. Therefore, each generated observation is a random variable generated by the probabilistic function of a state. As with Markov models, according to the notation of [93, 106, 153], for this thesis, HMMs are defined by the following components Q, A, O, B and π:

Def.: Hidden Markov Model

$Q = \{q_1, q_2, ..., q_N\}$	A set of N states, the finite and discrete state space
$A = a_{11}, a_{12}, ..., a_{1N}, ..., a_{NN}$	A transition probability matrix where each a_{ij} represents the probability to traverse from state q_i to state q_j; furthermore: $a_{ij} \geq 0$ and $\sum_{j=1}^{N} a_{ij} = 1$ for $1 \leq i \leq N$
$O = \{o_1, o_2, ..., o_M\}$	A set of M possible observation symbols
$B = b_i(o_j)$	An output observation probability distribution, also called emission probabilities; each represents the probability of an observation o_j being generated from state q_i; further: $b_i(o_j) \geq 0$ and $\sum_{j=1}^{M} b_i(o_j) = 1$ for $1 \leq j \leq M$
$\pi = (\pi_1, \pi_2, ..., \pi_N)$	A vector of N start probabilities denoting the model will start in state q_i; furthermore: $\pi_i \geq 0$ and $\sum_{i=1}^{N} \pi_i = 1$ for $1 \leq i \leq N$

That is, HMMs are basically Markov models where the output observations sequence of random variables not necessarily corresponds to one internal state sequence. A simplifying assumption often applied to HMMs is that the random output variable X_l (later associated to observations from O) only depends on the S_l that caused the observation. That is:

$$P(X_l|S1, S2, ..., S_L, X_1, X_2, ..., X_L) = P(X_l|S_l) \tag{3.18}$$

Now using Eq. (3.18), the *Bayes Theorem*, and recalling Eq. (3.7), the probability of a sequence $Z = Z_1, Z_2, ..., Z_L$ of random variables with $Z_i = (X_i, S_i)$ under a sequence $S = S_1, S_2, ..., S_L$ of state variables and under a sequence $X = X_1, X_2, ..., X_L$ of observation variables is calculated by:

$$\begin{aligned} P(Z) &= P(X, S) = P(X_1, S_1, X_2, S_2, ..., X_L, S_L) \\ &= P(X_1|S_1) \times P(S_1) \times \prod_{l=2}^{L} P(X_l|S_l) \times P(S_l|S_1, ..., S_{l-1}) \end{aligned} \tag{3.19}$$

where each S_l is associated to a certain value resp. state $q_i \in Q$ and each X_l is associated to a certain value resp. observation $o_i \in O$. The Markov Assumption (Eq. 3.8) is a further simplifying assumption also applied to HMMs leading to the adaptation of Eq. (3.19) to:

$$P(Z) = P(X, S) = P(X_1|S_1) \times P(S_1) \times \prod_{l=2}^{L} P(X_l|S_l) \times P(S_l|S_{l-1}) \tag{3.20}$$

In addition to Eq. (3.10), Eq. (3.11) and Eq. (3.12) which apply to HMMs as well, the time index l can also be discarded from the observations for time invariant events:

$$P(X_l = o_j | S_l = q_i) = P(o_j | q_i) \tag{3.21}$$

Here the $P(o_j|q_i)$ in Eq. (3.21) corresponds to the definition of the observation probabilities $b_i(o_j)$ in the HMM. That is, $b_i(o_j)$ is used to model a sequence (a time series) of output variables with L observables from O, where each output at l is denoted by X_l, with:

$$b_i(o_j) = P(X_l = o_j | S_l = q_i) \quad 1 \le i \le N, 1 \le j \le M \tag{3.22}$$

Analog to Markov models, the extend of considered history (i.e., the size of the model's memory) can be increased what leads to higher-order HMMs.

3.3.2.3 *Using Hidden Markov Models for ISB Modeling*

HMMs to predict the search success

In order to model users interactions, HMMs in versatile variations are applied to address and investigate various challenges in the research area of IR. For example, Hassan et al. [84] have gone beyond the evaluation of web pages' relevance (what is commonly applied in IR) and aimed to predict the success of a web search session from recorded log files. To segment single search sessions and to derive their success, the authors use a manual labeling process. Each search session consists of pre-defined states (corresp. to interactions with the search engine) which than correspond to the states of a Markov model. Additionally, the authors integrate several features such as transition times (what technically makes the Markov model a HMMs) and showed that this approach performs significantly more accurate than traditional relevance-based (static) models for predicting user search goal success. Their approach is similar to the ones in this thesis. However, there are several methodical and technical differences. For instance, the data set used by Hassan et al. [84] is from a large commercial web search engine that does not contain a ground truth. Therefore, a subsequent manual labeling was necessary and the users' original information need; the search session segmentation; and the search session success had to be estimated. The log files for this thesis have been generated in controlled lab studies with carefully selected search tasks to induce the desired information need. A further technical deference is the integration of dwell times. In Hassan et al. [84], each transition is described by a corresp. distribution. This may allow a higher degree of detail for the representation of certain user behavior but drastically increases the demand on data. In this thesis, feature distributions are assigned to the HMMs' states what reduces the model's (parameter) complexity and hence, size of the required data. Finally, the goal to predict the search session success differs from this thesis' goal to detect the search activity (e.g. fact vs. exploratory search behavior).

HMMs and CRFs to model the success for factual search

Similar to Hassan et al. [84], Ageev et al. [2] aim to predict user's success but explicitly for fact-finding search tasks applying different graph based approaches such as (Hidden) Markov Models and Conditional Random Fields. They define different types of search success and propose their Query-Result-Answer-Verification (QRAV) model that is primarily designed for factual tasks with a specific information need. Some specifications of the QRAV model even subsume definitions of other related work such as Aula et al. [11] and Hassan et al. [84]. For their investigation, Ageev et al. [2] uses an own game oriented search interface (named *UFindIt*) that allows to choose a (common) search engine and to submit the question's answer and URL. As baseline model the approach of Hassan et al. [84] is used. The results could confirm that including the dwell time into the user models increases the performance for search success prediction. The main model applied by Ageev et al. [2] is a Conditional Random Field (CRF) [124, 181] using the Mallet implementation[11]. Basically, CRFs are an extension (resp. an abstraction) of HMMs where the (strong) independence assumption is relaxed. Furthermore, as conditional models, CRFs have the ability to specify the probabilities of (hidden) states given an observation[12]. However, because of it's origin, the Mallet implementation allows only nominal features what demands a feature discretization and the usage of value bins for continuous features such as dwell time. For their analysis, Ageev et al. [2] use a rich feature set and the results show that, e.g., high search session success correlates with shorter queries, more queries per question, more page views, deeper SERP browsing and faster SERP clicks. Furthermore, for most of the QRAV model's specifications, CRFs performed better if all features (in [2] up to 17) are used. In addition to the differences between the approaches used in this thesis and the approaches from Hassan et al. [84] (cf. paragraph before), Ageev et al. [2] solely applied their model on factual questions to detect the success rate. In this thesis factual question of varying difficulty are considered and the more challenging differentiation between exploratory search and successively conducted fact-finding searches on multiple topics is investigated.

Using Markov Models to reveal search task's complexity, specificity and obtained type of information

Cole et al. [42] used as well a sequential approach, i.e., the framework of Markov models, to study interaction patters for search tasks

11 The Mallet implementation from the University of Massachusetts, Amherst was originally designed for natural language processing and further diverse document and text related IR tasks such as classification and information extraction. The toolkit is available at: `http://mallet.cs.umass.edu`

12 Formally, CRFs model a conditional distribution of $p(S|X)$ what leads to undirected graphs as representation instead of modeling a joint probability distribution $p(S,X)$ what leads to directed graphs as representation as in HMMs.

with different characteristics, such as complexity, specificity or the type of information obtained. They utilized interaction data and gaze data for a cognitive representation of textual information acquisition. To record the user interaction and gaze data, the authors implemented their own logging framework [25] and showed that those two data types can be used independently to distinguish between the search tasks. The four used search tasks are typical work tasks for journalism and the participants were students from that domain as well. One of the tasks was considered as factual, the other three tasks as mixed, i.e., they include intellectual but also factual components. Notable in context of the thesis is the result of Cole et al. [42] that: the one factual related search task differed in both data types at most from the three other more intellectual challenging tasks. The one factual related search tasks is also be considered as the one with least complexity regarding the user's interactions induced by the tasks. Since the three mixed tasks partially fulfill the requirements of exploratory search tasks, the mentioned difference is a further indicator for the possibility to successful distinguish between factual and exploratory search behavior.

HMMs to analyze the number of search tactics

Han et al. [83] directly applied a HMM to log files and analyze the optimal number of search tactics (as hidden states) in relation to the model's parameter complexity. According to the conclusion of Han et al. [83], their results have a notable agreement with the sub-processes of the information seeking process proposed by Marchionini [130] (cf. Chap. 2.7). A further notable result is that the model implies a state wise tactic change of the users. That is, at each time step, the user changes the current tactic. This is in accordance with Marchionini's default path of transitions in his information seeking process and also agrees with the Bates [14] description of search tactics as moves performed by the user (cf. Chap. 2.2). However, the data base used for the investigation was generated from log files of only seven students solving two search tasks. Furthermore, several sub-processes are subsumed to one hidden state and other comprehensive sub-processes are shrunk to simple actions. For example, Marchionini's sub-processes "Extract Information" originally includes the execution of reading, scanning, classification of information ect. but in the model of Han et al. [83] it is primarily substituted by the bookmarking action. Nevertheless, the investigation contributes to the overall goal of closing the gap between the theoretical ISB models and concrete actions performed on information systems. Furthermore, the significance and potential of HMMs in this areas is confirmed. In a follow-up study, Yue et al. [201] modeled collaborative exploratory search using HMMs and compared it to individual exploratory search behavior. The collaborative search turned out to be more complex because the model's parameter optimization (i.e., the model selection) indicated more (precisely six) hidden states as necessary. Furthermore and in contrast to Han et

HMMs in collaborative exploratory search settings al. [83], instead of Marchionini's model of the information seeking process [130] the sensemaking framework of Qu and Furnas [150] appeared to be more appropriated as theoretical background.

Extending HMMs to POMMs to simulate users gaze behavior

Last but not least, Wang et al. [184] and He and Wang [87] (in a follow-up study), treat the user's information seeking as a partially observable process and extend the HMM to a so called Partially Observable Markov Model (POMM). The extension is motivated by the fact that not always all relevant events resp. states are observable or even emit an observation. In case of Wang et al. [184] as well as He and Wang [87], user's gaze behavior on SERPs is taken as example application for data that is not given in each user study and therefore, unobservable. That is, the authors aim to derive and hence, to simulate users gaze behavior if no eye-tracking data is available and finally compare the estimated gaze behavior (generated by the POMM) to real eye-tracker data. The results show that POMMs are sufficient for certain aspects if gaze behavior is originally not avaliable in the recorded data. However, room for optimization on the model's parameter as well as it's assumptions is still given. In He and Wang [87] the POMM has been extended (in contrast to Wang et al. [184]) by including state duration into the models what leads to Partially Observable Markov Model with Duration (POMMD). This addition lead to an enhancement on the part of the POMMDs and again confirmed that dwell time in ISB analysis play an important role.

3.3.3 *Further Approaches to model Information Seeking*

In addition to (Hidden) Markov Models, as described in the previous Section 3.3.2, Bayesian Networks are further representatives of graph models which can be used to model time series and in particular user seeking behavior. Piwowarski et al. [147], for instance, showed how to use layered Bayesian Networks to estimate the relevance of documents without considering the documents' content but regarding the users' search activities. Downey et al. [54] developed an own language for so called *Search Activity Models* (SAM) building on possible states and events during a search session with the goal to predict the user's next move. The authors use their SAM language to encode action sequence in a framework of Bayesian Networks and thereby include both actions and time. Boldi et al. [28] use (query) search logs to generate a so-called *query-flow graph*. This directed, weighted graph representation of query (re-)formulations of all users and search sessions from a given search engine log illustrates "sequence(s) of queries with a similar information need" [154] (called *chains*) as well as "sequence(s) of all the queries of a user in the query log, ordered by timestamp" (called *supersession*). Hence, the interest as well as query related aspects of the user's search behavior are represented. The model is used for the segmentation of sessions in interrelated query chains and for the generation of query recommendations.

Besides Markov and other graph related models, diverse machine learning approaches have been used as well in order to make predictions of user's behavior and search success in search. Shah et al. [173] applied Support Vector Machines to forecast how well users will perform in later stages of the ES process based on the actions currently applied. That is, a n-step-ahead prediction is implemented. For this, the authors use a vector of seven different features for each user and each time step (here one time steps corresp. to one minute). Athukorala et al. [10] used Random Forests, a machine learning method for classification, to distinguish exploratory from lookup searches applied on a full search session. Most of the related work investigates search activities in specific domains and/or with specific search systems. In contrast to that, for this thesis, investigations considering multiple topics with a common web search engine will be conducted.

3.4 CHAPTER SUMMARY

The aim of this chapter was to provide an overview of analytical approaches to investigate human's ISEB. Several analysis of search activities but also models for ISB have been reviewed, described and set into context of the thesis' topic of EIS. Thereby, the underlying assumptions, purposes, advantages but also limitations of the approaches have been discussed. At first, several models to investigate the user's search behavior on SERPs have been exemplified. Research witnessed an increasing interest of such models because almost each user in the Internet utilizes SERPs. These result pages carry potential semantic within the interaction data so that SERPs nowadays have taken the role of information resp. knowledge hubs where users evolve their information need and seeking. However, user behavior models which only consider the interactions on SERPs are limited to the pure search engine perspective (i.e. do not consider interactions on web pages) and further are often biased by strong assumptions, such as neglecting users who (re-)click on previous search results. Therefore, models and findings of search activity analyses on a more general level have been discussed (Sect. 3.3.1). While the investigations individually showed pioneering insights for the field, several open questions (and thus differences to the investigations in this thesis) remain. These open points are often related to (a) missing: exploratory (not factual) search activities in the related studies; clear differentiation/identification of the (exploratory) search activities of arbitrary length within a single search session; ability to identify new, previously unknown search behavior; search tasks on different domains (not only one specific domain); ground truth regarding search tasks, search success and/or the user base (because of the kind of the recording). In contrast to the investigation of the literature, all of these aspects will be considered in this thesis. Finally, this chapter identified the mathematical framework of Markovian Models as the most promising approach to implement the thesis's objects of investigation. In addition to the necessary definitions of the (Hidden) Markov models, related work regarding these type of models and the differences to the present work was outlined.

Part III

INVESTIGATING EXPLORATORY INFORMATION SEEKING

"Knowledge must come through action ..."
— Sophocles

4 CAPTURING USER'S SEEKING BEHAVIOR

The former Part ii and it's two Chapters 2 and 3 outlined the topic of Information Seeking Behavior (ISB) as well as Information Search Behavior (ISEB) from a theoretical and analytical perspective and provided necessary foundations to address the thesis' hypotheses. Furthermore, the relation to the search paradigm of Exploratory Search (ES) has been depicted. While hypothesis *H*1 could already be elaborated in Section 2.6.3, the hypotheses *H*2, *H*3 and *H*4 require a more empirical and analytical procedure. To execute the corresponding investigations, appropriate data is required and therefore, the target of this chapter is to describe how data for investigations on Exploratory Information Seeking (EIS) can and have been acquired. In particular, to investigate exploratory but also factual search activities, several own user studies were designed and implemented. The participants with their different individual characteristics have been considered as well. The procedures to obtain the data in the studies and a description of the data characteristics are also exemplified. After providing a motivation and a brief overview of all user studies (Sect. 4.1), the details of each user study are described (Sect. 4.2 to 4.4). The content of this chapter has already partially appeared in the following own publications: [76, 114, 116, 118].

4.1 USER STUDIES TO ACQUIRE DATA FOR EXPLORATORY INFORMATION SEEKING

As motivated in the introduction of this chapter, to investigate the hypotheses *H*2, *H*3 and *H*4, appropriate data sets are required. Available sources, e.g., from TREC[1], provide a multifaceted repertoire of user interactions with search engines but usually: do not have a detailed interaction record (i.e., no step by step record of each interaction); or only have a limited number of user parameters (e.g., no eye-tracking); or have utilized search tasks for the user studies which do not (fully) satisfy the conditions for ES. Therefore, an important part of this thesis comprises the planning and implementation of several own User Studies (USs) to generate data sets and subsequently answer the research questions of the thesis. The planning and implementation of the user studies comprises the study procedure, search tasks, Search User Interface (SUI) as well as spatial and technical setting. In particular, the choice and design of the search tasks is a crucial aspect because the

Cf. H2 to H4 in Sect. 1.2

1 Text REtrieval Conference, co-sponsored by the National Institute of Standards and Technology (NIST): `https://trec.nist.gov/`

tasks are the implemental key to induce the (exploratory or factual) search behavior that is desired. However, the provided SUI, the implemented study procedure but also the spatial and technical setting are no less important since all of these aspects have influence to the study and therefore, may also influence the seeking behavior. All of these mentioned aspects will be discussed in detail for the individual studies in their corresponding sections. In the following, the purpose of the individual user studies are summarized:

- *US-I*: According to Wilson's model (cf. Sect. 2.3) and following the results of several investigations regarding user's search activities (cf. Sect. 3.3.1), the user variable *age*, as demographic variable, certainly represents a relevant aspect for user's seeking behavior (*H*2). That is why two (smaller) user studies to investigate the search behavior of young users of the age eight to ten have been conducted. Since young users have difficulties with typing [31], a voice-controlled SUI was provided for the participants. The first user study (*US-Ia*) aims to analyze the interaction of young users in general. The second study with young users (*US-Ib*) involves an ES task to investigate the children's interaction on that kind of tasks. Furthermore, differences and similarities to *US-Ia* will be identified.

- *US-II*: To provide adequate user support, search systems have to be able to distinguish between search activities, a user is currently engaged in (cf. Sect. 3.3.1). Therefore, a user study that consists of fact-finding and exploratory search behavior was conducted. Interactions with the search engine and the gaze behavior of the users have been recorded. The collected data enables the analysis and modeling of ES on the one hand and allows investigations regarding the classification and hence, differentiation of search activities on the other hand (*H*3). User study *US-II* also serves as an important precursor for user study *US-III*. The projected number of participants is 20. With this straightforward number of participants, the data set was also extended by an annotation of users' reading states. This enables the analysis of reading as a further seeking related variable.

- *US-III*: This large scale user study of over one hundred users covers additional personal user characteristics that may influence the (exploratory) seeking behavior (*H*2). Several psychological variables have been obtained, such as personality and aspects of intelligence. The social context will also be considered by a competition between users as one sub-part of the study. To validate and extend the findings of study *US-II*, fact-finding and exploratory tasks have been utilized again (*H*3). The recorded data but also the search tasks are conform with *US-II* what makes *US-III* an extended and valuable super-set of *US-II*.

In the following four Sections 4.2 to 4.4, a detailed description of each of the user studies (*US-I* to *US-III*), their implementation and the recorded data is given which in turn is analyzed and used in the next Chapter 5 to investigate and model user's EIS.

4.2 USER STUDY I: EXPLORATORY BEHAVIOR OF YOUNG USERS

User Study US-I: *Cf. summary: p. 78; Cf. overview: p. 112*

Young users usually have less search experience than adults, have little domain-knowledge [94] and have difficulties to evaluate the relevance of retrieved documents to satisfy their information needs [104]. Furthermore, children can get frustrated easily if they are not able to find relevant information, do not understand the displayed search engine output or if a failure during the seeking emerges [26]. Therefore, if young users want to find an answer to their questions and execute a search, they need to be motivated and supported.

Besides the potential challenges, children may be confronted with, analyzing young user's seeking can be a promising source of elementary search patterns and can reveal additional insights into human's (exploratory) information behavior. One reason for this is that little domain-knowledge leads to an enhanced need to explore (new) topics, e.g., to answer (even simple appearing) questions. A second reason is that users with less search experience, as children, are also less biased in contrast to adults. For example, experienced users can be biased by long term usage of common search engine (paradigms); own developed technical know-how or search procedures learned since the first contact with search systems; or be biased by the domain-knowledge gained in professional education, hobbies, etc.

However, a crucial aspect considering young users is that most children have difficulties with typing [31]. In contrast to that, interaction with a voice-controlled system does not require typing interactions and can be more intuitive and motivating for young users since they do not have to learn the cumbersome interaction with mouse and keyboard. Voice-control is also considered to be a natural way to interact with computer systems. That is why two user studies with young users, utilizing a voice-controlled SUI, have been conducted. While the first study (*US-Ia*) aims to investigate the interaction of young users performing a free search in general, the second user study (*US-Ib*) applies and investigates an exploratory search task setting. In the following, the user study design as well as the data generation of both studies is described in detail. The content of this section has already partially appeared in the own publications [76] and [114].

4.2.1 *Study Procedure: A Wizard-of-Oz-Setting*

To implement a voice-controlled search, both user studies (*US-Ia* and *US-Ib*) have been conducted in form of a *Wizard-of-Oz-Experiment*[2]. This method allows to study user's seeking behavior without any technical limitations and allows the investigators to react appropriately to unexpected situation and user actions. In both studies, a voice interaction in both directions was allowed. That is, on the one side, the children just had to articulate the desired controls. The system (the wizard) received the user's voice input and reacted accordingly, e.g., by formulating the queries or perform interactions on Search Engine Result Pages (SERPs) and web pages (such as scrolling or clicking on links). On the other side, the system (the wizard) could use a voice output to prompt the young users for interaction, e.g., if the wizard could not understand the control command. In both studies, two investigators were involved, one investigator to play the role of the wizard, i.e., to operate the whole system, and one investigator to conduct the user study with the participants, i.e., to interview, give instructions, stay on reach in the event of difficulties, etc. In the following, the study procedure (consisting of four steps for both studies) is described:

1. *Pre-Interview:* In the first step, a pre-interview to gather the user's demographic information and their experience with computer systems and the Internet was conducted.
2. *Introduction:* In the second step, the participants have been introduced to the SUI and the interaction with it's elements have been explained. No information about the usage of voice control was given to receive a most natural interaction which also may include learning aspects. However, after the introduction, the users could shortly test some commands and the resulting response of the system. In the *Introduction*, the only difference between the two studies *US-Ia* and *US-Ib* was that in *US-Ib* the investigator additionally explained the ES task that had to be performed after this second step.
3. *Search Experiment:* In this step, the actual search was performed.
   - In *US-Ia* the participants could execute a free exploration, i.e., they could look for everything they liked and use the SUI's elements how they want but only using voice commands. If a child had no idea for what or how it should use the system, the investigator gave some suggestions, e.g., *"Currently it's Christmas time and there are a lot of things one can do during this time. Maybe you can search for these things?"*.

2 In a Wizard-of-Oz-Experiment users interact with a program, here a SUI, that seems to act autonomously but actually is remotely controlled by a hidden investigator, the so-called *wizard*.

- In *US-Ib* the participants had to execute the given ES task (cf. Sect. 4.2.2). If a young user had no idea how to use the system, the investigator gave assistance in terms of *"Simply try to say the system what you want to do."* or *"You can try to say it in any other way."*. Of course, if the users had any other problem, the investigator helped as well but always with a minimal amount of intervention.

If a child used an ambiguous command or the system (the wizard) could not understand the command, a prepared audio message *"I cannot understand you."* was triggered. The duration of this third *Search Experiment* step in *US-Ia* was approximately 10 to 15 min. In *US-Ib* the ES took about 20 min.

4. *Post-Interview:* The last step was a post-interview to evaluate the users' attitude towards the system and if they would use voice-controlled SUIs in the future. Questions regarding likes and dislikes as well as recommendations to improve the search engine's interface have been asked as well.

4.2.2 *Search Task(s): Free & Exploratory Task*

As mentioned in the previous section, in *US-Ia* no specific search task was given. The participants could execute a free exploration and could search for everything they were interested in. Only if they had no idea what they could search for, the investigator made suggestions to search for Christmas related topics or items[3].

For the second user study with young users, the goal was to conduct an ES. That is, in *US-Ib*, a pre-defined ES task was given to all participants. According to the discussion of ES in Section 2.6, to design a task that intends to induce exploratory search behavior on children, at least the three following aspects need to be considered: The young users should not be too familiar with the search domain and/or the available search tools (1). With a sophisticated knowledge about the domain, the young users would be able to formulate precise factual queries, and thus several search activities of the categories learning and investigation, as described by Marchionini (cf. Sect. 2.6.1), may not occur. The next aspect is that the task should be open-ended (2). This typical ES characteristic enables, respectively requires, more thought-out search interactions since a single fact can not sufficiently answer the task. Furthermore, it increases the necessary time to solve the task. Although an ES can take hours or days over multiple search sessions in general, at least for a user study, as conducted here, open tasks may cause multiple search iterations on one topic. Finally, the search domain of the task should motivate the young users, awake their inter-

3 The topic of Christmas was considered to be appropriated and motivating because user study *US-Ia* was conducted in December.

est and preferably should not have any gender specific allocation (3). With these three requirements in mind, to induce an ES in *US-Ib*, the following task was chosen:

> *"Imagine you are a zoo director and you start your own small zoo. Your zoo shall provide different animal species a new home. To create an environment where all your animals are fine, you have to maintain an adequate animal housing. Use the Internet to read up on the needs of your animals. Use our Knowledge Journey to search for information and control it via speech."*

The needs of zoo animals are known to some extent by most children. However, only a few children should be able to name specific facts about adequate animal housing. Hence, learning and investigation activities are more likely to appear during the search and the domain aspect (1) for this task should be basically fulfilled. For user studies, especially with children, the constraint addressing the search time is difficult to fulfill. On the one hand, for a search that only takes a few minutes, it is almost impossible to comprise the complete spectrum of the user's seeking behavior. On the other hand, to conduct a search that would take several hours, is not realistic for users, in particular for children, because after some time they would get exhausted, bored or distracted. Therefore, for *US-Ib*, a 20 minutes time frame was chosen for each search session as a trade-off. Together with the open character of the given search task –the number of possible animals and details about adequate animal housing is virtually infinite– the open-end aspect respectively the aspect regarding the time to solve the task (2) is fulfilled as well. Finally, the remaining requirements for ES tasks for young users (3) are addressed by taking a topic (about animals) that in general is of interest for the majority of children, should motivate the young users to perform an adequate search and intrinsically has no clear preference regarding the user's gender.

At this point it should be noted that the three requirements, explained above, basically are also fulfilled for the search experiment in *US-Ia*. The young users could search for everything they want and/or for Christmas related topics or items, what may further increase their domain knowledge. The search experiment was in theory open-ended and at least the topic of Christmas should be motivating, interesting and is gender independent. However, the goal and setting of *US-Ia* was a bit different to *US-Ib* and the search experiment in *US-Ia* was not intended to induce exploratory search behavior.

4.2.3 *Search User Interface: Knowledge Journey*

For both studies, *US-Ia* and *US-Ib*, the *Knowledge Journey* [73], a SUI designed for children, was used. The SUI makes use of the metaphor of a treasure hunt. That is, the young users take a journey to gather

Figure 4.1: Screenshot of the *Knowledge Journey's* SUI consisting of a guidance figure (a penguin pirate) and a bookmark item (a treasure chest) on the right, the query input element on the top, a navigation menu on the left and the SERP visualized as a coverflow in the center.

new information and find relevant web documents. In the thesis of Gossen [68], the whole development of the *Knowledge Journey's* interface in context of search engines for children is discussed in depth. The interface can be adapted to the user's preferences in terms of search result visualization type, font, available means for guidance, etc. However, for *US-Ia* and *US-Ib* a SUI configuration was chosen that represents *Knowledge Journey's* original concept, namely a journey on the sea (*US-Ia*) or in space (*US-Ib*). Furthermore, the used configuration was also accepted resp. preferred by the majority of young users [75]. Figure 4.1 illustrates the interface of the child-centered search engine as it was used in *US-Ia*. Basically, the interface consists of five groups of elements: a guidance figure (here a penguin pirate), a treasure chest to store favorite web pages, query elements for keyword search, a pie-menu for a navigation in several pre-defined categories and a coverflow visualization of the search engine's results. The coverflow allows to clearly separate the current, central search result from the remaining results, and thus children can concentrate on one item at a time what leads to a reduced cognitive load. Furthermore, the SUI contains diverse multimedia elements, good readable fonts and buttons to make it more attractive for young users. The purpose of the guidance figure is to support the children's search process and to avoid frustration. Features of the figure are, e.g., providing a spelling correction if a misspelled query is submitted. In case of a *Wizard-of-Oz-Experiment*, the figure represents the search system's voice if the hidden investigator wants to interact with the child, e.g., to ask what to do next.

For study *US-Ib*, the *Knowledge Journey* [73] in it's voice-controlled version was used as well but with two differences: The first difference to *US-Ia* was the chosen theme. Here, the space theme including an

Figure 4.2: Screenshot of the *Knowledge Journey's* SUI consisting of a guidance figure (an alien) and a bookmark item (a logbook) on the right, the query input element on the top and the SERP visualized as a coverflow in the center.

alien as guidance figure was taken because this theme was more preferred by children than the original pirate theme [75][4]. The second difference to *US-Ia* was the deactivated navigation menu. The reason to omit the navigation menu was that this element provides general categories for the users to initiate a search on predefined domains if a user needs suggestions for exploration. The given ES task regarding animal housing however makes the navigation menu unnecessary. Furthermore, if a certain task is given, the navigation menu could even cause undesired distraction. Figure 4.2 illustrates a screenshot of the *Knowledge Journey's* SUI as used in *US-Ib*.

Overall, the *Knowledge Journey* and it's elements provide different options for interactions for the users, as illustrated in Fig. 4.3. Formulating a query in the corresponding field but also choosing a pre-defined category or sub-category (in *US-Ia*) results in a coverflow visualization of the retrieved search results. Search input and search result list (SERP) are visualized in the same page. Next and previous search result items can be scrolled by clicking them. A click on a current search result or on the treasure chest, resp. the log-book, opens the corresponding items. Finally, if a web page is opened, the user can save and find it in the treasure chest, resp. the log-book.

4 The reason for the two different themes in *US-Ia* and *US-Ib* is a pragmatic one. The *Knowledge Journey* itself was in a process of development and improvement. For *US-Ib*, just the even more preferred settings (in general) have been used. However, the theme is not expected to cause a significant difference in the seeking behavior.

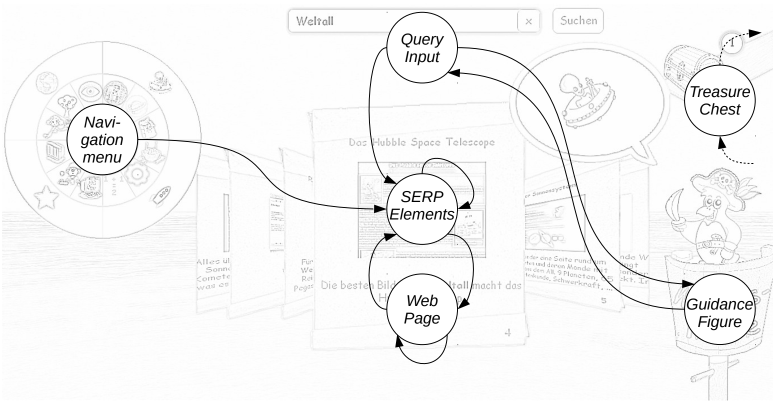


Figure 4.3: Interaction graph of the *Knowledge Journey*. A user can retrieve search results by using the query input or the navigation menu, can activate the currently centered web page in the SERP elements or scroll through the results, may receive feedback from the guidance figure and can store favorite web pages in the bookmark item.

4.2.4 *Spatial & Technical Setting*

In the following, the spatial conditions and utilized technical equipment including the recorded data for both studies (*US-Ia* and *US-Ib*) are described.

4.2.4.1 *User Study US-Ia*

The study was conducted at the trilingual international elementary school in Magdeburg, Germany. In the study room, a class room, the search system was arranged in a way that participants could not see the second investigator (the wizard). Only one participant and the two investigators were present. However, sometimes other pupils entered the room to get personal belongings but only for a very short time. Nevertheless, since *US-Ia* was conducted in an elementary school, during the experiment, typical school sounds such as voices of other pupils and teachers or the sound of the school bell occurred and hence, have been partially recorded as well. This issue was avoided in the second study with young users (cf. Sect. 4.2.4.2). To investigate the young users interactions but also to better comprehend possible (unexpected) incidents during the study, a technical setting comprising of audio, video and eye-tracking recording devices have been used for *US-Ia*. The specifications are given in the following:

- Eye-Tracker: Tobii T60
- Microphone: One built in Eye-Tracker
- Camera: One built in Eye-Tracker with 640x480px

Furthermore, the Eye-Tracker software *Tobii-Studio* allowed to record the user screen and supports an annotation of the data afterwards. For the voice-controlled interactions, as in this kind of study, the annotation function is important since the utterance of a child to execute an interaction and the actual performed execution by the wizard can take several seconds and may be ambiguous. Therefore, a subsequent annotation of the individual interactions of the young users was done. Here, each interaction was set exactly if the child's utterance (to perform an interaction) ended. That is, the recorded data for *US-Ia* consist of the participants acoustic utterances to control the SUI, a user video, a screen record, eye-tracking data and an annotation of the users interactions with the SUI.

4.2.4.2 *User Study US-Ib*

In contrast to *US-Ia*, the second study with young users was conducted at the Otto von Guericke University, Magdeburg, Germany. To enhance the quality of the study and it's recorded data in favor of the thesis' investigation but also to increase the reusability for further research, for *US-Ib*, an appropriated physical environment (i.e., a dedicated user study room) was utilized. In the study room, (young) users can perform a search, increase their knowledge about a certain domain and investigate new sub domains over a longer time without any external influence. This includes an adequate setting with respect to ambiance and installation. Figure 4.4 depicts the spatial and technical setting of study *US-Ib*. Child and wizard are spatially separated. The young user can notice only the SUI on the screen, the investigator and the different recording devices. The study room was designed to provide a comfortable environment. It had different furniture, such as cupboards and a couch. The colors of the painted walls and furniture are bright and the fanlight has a low neutral white luminous color (about 4000 K). To avoid (side) noises on audio records, a sound-proof

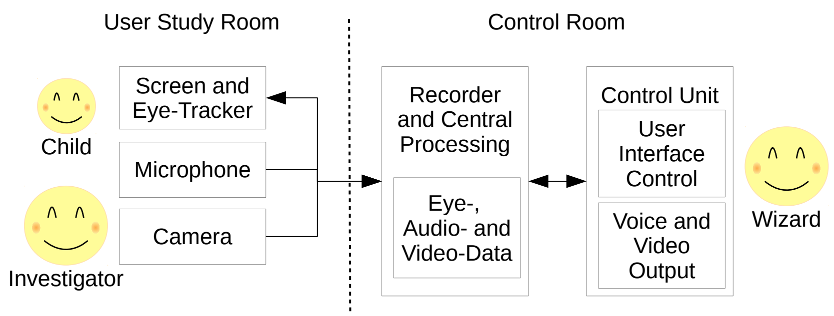


Figure 4.4: User study setting of *US-Ib* with it's technical implementation. User study room and control room are spatial separated by a wall. The video and audio streams allow the wizard to understand the current situation in the next room to act appropriately.

door was installed. For the study, the following recording devices have been used:

- Eye-Tracker: Tobii T60
- Microphone: Two Sennheiser headset microphones: HSP 2-EW-3
- Camera: Two AVT Pike F145C cameras with 1388x1038px and Tevidon objective 1.8/16.

That is, the recorded data for *US-Ib* consist of the participants acoustic utterances to control the SUI, the investigators acoustic utterances to support the participant, two user videos, a screen record, eye-tracking data and an annotation of the users interactions with the SUI. The wizard in the neighboring room was controlling the entire system, i.e., controlled the SUI, could see the user and investigator in the front of the user screen (via the Pike cameras), heard the user and investigator and had an overview over the whole scene in the user study room. To synchronize the different video and audio streams in this advanced setting, the audio interface Yamaha Steinberg MR 816X and its internal world clock trigger with 44 kHz was used. One small drawback was that the Eye-Tracker could not be triggered externally but this could be solved by synchronizing the data afterwards.

4.2.5 *Participants & Data Characteristics*

In the *US-Ia*, 10 young users (3 male and 7 female) participated[5]. The pupils were of age eight to ten (average 8.8 years) and were in the third (8 children) or fourth (2 children) grade. All children had experience with computer systems and the Internet. In particular, the usage of the Internet was: 1x everyday; 3x two-four times a week; 1x once a week; 4x once a month; and 1x less than once a month. No significant correlation regarding the frequency of Internet usage and age respectively school grade was found. Most of the children use the Internet to play online games, watch videos on the video-sharing platform *YouTube* or search information for school, e.g., with *Google*. In total, user interactions that correspond to 475 transitions in the free search experiment have been recorded.

In the *US-Ib*, 5 children (2 male and 3 female) participated[5]. The pupils were of age eight to nine (average 8.8 years) and were in the third (4 children) or fourth (1 child) grade. All children had experience with computer systems and the Internet. Three of the pupils use the Internet multiple times per week and two of them approximately once a week. All children stated to use the Internet to search information for school or for themselves. In particular, four pupils use *Google*

5 All the relevant agreements from the caretakers had been obtained. The Parents agreed in advance, that their children can participate in the study.

Table 4.1: Average time and standard deviation (SD) in sec. for the *search experiment* in *US-Ia* and *US-Ib*.

Study	Time in sec.	SD in sec.
US-Ia	951	315
US-Ib	1402	238

and one pupil uses a German search engine dedicated for children. Furthermore, two of the children also write e-mails and listen to music in the Internet. In total, user interactions that correspond to 317 transitions regarding the ES tasks on the "zoo animal housing" topic have been recorded. The time distribution for the *search experiment* in *US-Ia* and *US-Ib* is shown in Tab. 4.1. Finally, Tab. 4.2 lists the major characteristics of user study *US-I* and differences between *US-Ia* and *US-Ib* to provide an overview.

4.3 USER STUDY II: EXPLORATORY & FACT-FINDING SEARCH (PART A)

User Study US-II*: Cf. summary: p. 78; Cf. overview: p. 112*

The ability to distinguish search activities, users currently are engaged in, is crucial for search systems to provide adequate user support in an adaptive information retrieval setting (cf. Sect. 3.3.1). The goal of this study is to obtain data of users who perform factual as well as exploratory search tasks (1) to investigate models that are able to differentiate between this two search activities but also (2) to distinguish and better understand ES as class of information seeking (cf. Sect. 2.6.3). In contrast to the work of Shah and González–Ibáñez [172] (cf. Sect. 3.3.1), the focus here is on the identification of search activities within a single session (not several sessions). A further goal is to analyze the minimal number of necessary interactions for the identification of search activities, i.e., using only parts of a single search session. This goal is motivated by the practical reason that an early identification of the user's current search activity enables the search system to provide appropriate support in time and not if the search is in an advanced state or (almost) over.

To differentiate between a user who is performing an ES and a user who is performing a single fact-finding search is relatively easy because factual searches are shorter in nature than exploratory ones [10]. That is, just considering the time, a user needs to solve a given task, would be sufficient for the identification but requires knowledge about the beginning and the ending of the task processing. However, in reality the chance is high that a user performs multiple searches. In that case, it becomes not clear when one search ends and a new begins. Furthermore, the search activities (factual or exploratory search) can even alternate or be interwoven (cf. Sect. 2.6.1).

Table 4.2: Overview of user study *US-I*'s major characteristics and differences between *US-Ia* and *US-Ib*. Used abbreviations are *Knowledge Journey* (*KJ*) and Otto von Guericke University (OVGU).

	US-Ia	US-Ib
Procedure	Pre-Interview → Introduction → Search Experiment → Post-Interview	
Search Task	Free search; if necessary "Christmas"	ES: "zoo animal housing"
Search Interface	*KJ*: "Pirates" with navigation menu	*KJ*: "Space" without navigation menu
Spatial Setting	Elementary school, class room	OVGU, sound proof study room
Technical Setting	Eye-Tracker, 1x voice rec., screen and user video rec., interactions	Eye-Tracker, 2x voice rec., screen and 2x user video rec., interactions

This study considers users who are performing different ESs. In addition to that, successively conduced fact-finding searches on multiple topics are considered as well. This search behavior is known as multitasking search [177] and is considered as a common human seeking behavior. Multitasking search sessions are longer than single topic sessions which makes this type of seeking more challenging to distinguish it from exploratory search sessions. In the following, user study *US-II*, it's procedure, the utilized search tasks, logging techniques, spatial setting as well as the recorded data are described. The content of this section has already partially appeared in the own publication [116].

4.3.1 *Study Procedure*

In order to acquire the necessary data to analyze and model user's search behavior on different search activities, in *US-II* participants had to solve two ES tasks and one task consisting of several factual search tasks, i.e., a multitasking assignment. The study implements a three step procedure consisting of a *questionnaire*, an *introduction* and a subsequent *search experiment*:

1. *Questionnaire:* The questionnaire was used to gather user's demographic data and their experience with the Internet and search engines. The participants have been asked for what reasons, how often and how long they use the Internet, respectively diverse search engines.
2. *Introduction:* In this step, the participants have been introduced to the *search experiment* afterwards, i.e., where they can see the

search tasks assignments; the given time to solve the tasks; how they can submit their answers; which search engine they should use, etc. Furthermore, the calibration of recording devices such as eye-tracking and user camera was done in this step.

3. *Search Experiment:* In this step, the actual search was done. The system provided multiple search tasks for the users and recorded their interactions during the task processing. Only if a participant had problems which potentially compromised a frictionless data record, e.g., if the system did not respond or the Internet connection was interrupted, the investigator was consulted.

As mentioned above, there are two different search task types in this study. One type are exploratory search tasks(Expl). The other type is a sequence of fact-finding search tasks which create a multitask assignment (Fact). The maximum number of single fact-finding search tasks, a Fact could consist of, was twelve because a sequence of twelve fact-finding tasks was estimated to have approximately the same extent like a single exploratory task for the given study setting. However, the task processing of Expl and Fact was limited by a pre-defined time frame. Therefore, the time to solve all twelve factual tasks in Fact just needed not to be significantly shorter than the pre-defined time frame, what guarantees that users do not (often) finish Fact earlier. This increases the probability that the time to process Expl and Fact was approximately equal and eventually worked out well. Each participant had at most 20 minutes to perform an Expl and at most 20 minutes to solve a Fact (with up to twelve factual tasks). Two exploratory search tasks with different topics (indicated by Expl_1 and Expl_2) and one Fact have been used. Therefore, each search session in the *search experiment* of the study was limited to one hour for each participant. To vary the order of task types between participants, a Latin Square study design on the corresponding three task blocks (each max. 20 min.) was used in the *search experiment*:

- $Design_1$: Fact $\rightarrow$ Expl_1 $\rightarrow$ Expl_2
- $Design_2$: Expl_1 $\rightarrow$ Expl_2 $\rightarrow$ Fact
- $Design_3$: Expl_2 $\rightarrow$ Fact $\rightarrow$ Expl_1

This three designs have been equally distributed among the participants. To illustrates the whole procedure of *US-II*, Fig. 4.5 depicts all steps exemplified on $Design_1$. The factual tasks within the multitasking Fact block were randomized for each user at the beginning of the *search experiment*. To avoid biases regarding the ES task topic in the model generation later, two different exploratory tasks with different topics have been used (cf. Sect. 4.3.2.2).

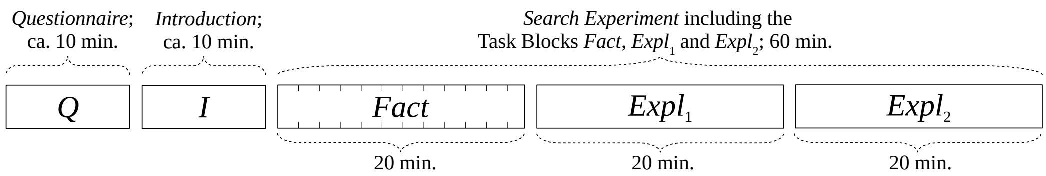


Figure 4.5: Illustration of the user study procedure of *US-II* exemplary incorporating *Design*$_1$.

4.3.2 *Search Tasks: Factual & Exploratory Tasks*

In the following, the creation of the two search task types fact-finding multitask assignment Fact and exploratory search task Expl as used in *US-II* are described in detail.

4.3.2.1 *Factual Search Tasks*

For the Fact block, twelve fact-finding tasks from different domains and with two difficulty levels (easy and hard) have been designed. Six of twelve tasks are easy and six are hard according to the following constrains:

- *Easy Factual Task*: A factual search task is considered to be *easy*, if the answer to the task can be found in (or derived from) the search results or result snippets on the first SERP using search queries only consisting of terms emerging in the search task formulation.
- *Hard Factual Task*: A factual search task is considered to be *hard*, if all the results or result snippets on the first SERP for search queries only consisting of terms emerging in the search task formulation are non-informative, so that a user has to evaluate several documents or at least read one (longer) document to find the answer.

In Tab. 4.3 all factual search tasks for the two difficulty levels are listed. One reason to use this two level categorization was that in reality, fact-finding search tasks usually have different difficulties and therefore, this attribute should also be considered for the data acquisition to further promote natural search behavior. A second reason was that the fact-finding search block Fact will therefore be a random[6] composition of easy and hard factual search tasks for each participant, what makes Fact more challenging to differentiate from an Expl. All twelve factual search tasks have been pre-tested several times by two researchers.

6 Random in the sense of an arbitrary order of twelve *easy* and *hard* factual search tasks per Fact block.

Table 4.3: Difficulty levels and corresponding fact-finding search tasks as used in *US-II*, translated from German.

Nr.	Level	Task description
1	Easy	*In what year did the Google search engine went online for the first time?*
2	Easy	*How old is Mickey Mouse today?*
3	Easy	*According to current information, how many rooms are in the Buckingham Palace in London?*
4	Easy	*What are sciaphobs afraid of?*
5	Easy	*How many men were on the moon until June 2004?*
6	Easy	*What is the name of the largest passenger aircraft?*
7	Hard	*In what period of the Paleozoic era the first reptiles appeared?*
8	Hard	*What percentage of German men aged between 70 and 79 suffer from diabetes (data for the year 2011)?*
9	Hard	*Which word is coded in Morse code as: "—.—. — — — — — — .—..”?*
10	Hard	*Which actor won the same year the Golden Raspberry Award for Worst Actor, Worst Supporting Actor and Worst Supporting Actress?*
11	Hard	*What is the largest known planet in the binary star system Kepler-47?*
12	Hard	*How many years passed between the first flight of the Kitty Hawk Flyers and Neil Armstrong's moon landing?*

Regarding the search task attribute *difficulty*, as briefly discussed in Section 3.3 (cf. footnote, p. 55), the following has to be remarked: Although task difficulty is ascribed as subjective, the constrains above, for easy and hard tasks, allow an argumentation for the difficulty assessment of the users. *Easy* tasks here are considered to have a low difficulty because the corresponding conditions for this categorization can be fulfilled with single copy past interactions (i.e., copy terms from the task formulation and use them as query) and a minimum amount of search engine interactions (the question's answer may be found on the results and/or snippets of the first SERP). *Hard* tasks here are considered to have a high(er) difficulty (than *easy* tasks) because the corresponding conditions for this categorization can only be fulfilled with at least some thought-out query (re-)formulations (queries only consisting of terms emerging in the task formulation are non-informativ) and at least some search engine resp. web page interactions (the question's answer could not be found on the results and/or snippets of the first SERP). Although hard tasks are not designed to be very challenging for the users (otherwise the tasks could provoke users to get stuck for the remaining factual search), they are more demanding than easy tasks. Finally, this two level difficulty categorization provides an approach of operationalization to investigate the user's factual task processing in more detail because easy and hard tasks (as defined here) are expected to induce a different course of action from the user's side[7].

The two task difficulty categorizations, used for Fact, can likewise be considered from the perspective of task *complexity*, also briefly discussed in Section 3.3 (cf. footnote, p. 55), if *easy* tasks are considered as "simple" (in the sense of low complexity) and if *hard* tasks are considered as "complex" (in the sense of higher complexity). The answers to "simple" factual tasks, used in *US-II*, can be found on the first SERP (or can be extracted relatively fast from a resulting web page of the first SERP), i.e., there is at least one direct path that can be identified to solve it. Furthermore, if during the design of the tasks before, different information sources to solve a "simple" task have been found (on the first SERP), then it was taken care to ensure that the information sources provide identical solutions, to avoid conflicting paths. Hence, the *easy* tasks here can be considered as "simple" also from the perspective of Li and Belkin [125] as well as Campbell [36]. Using an analog argumentation, for "complex" factual tasks, several paths could and sometimes had to be followed to solve the task. For example, the user had to use queries with different terms (not only terms emerging in the task formulation) and/or had to execute a click on a next SERP and/or had to extract more information (e.g., by reading) from web pages to find the solution. Furthermore, "complex" tasks require to identify at least two information and/or required the user to comprehend and process the found information (e.g., to calculate the difference between two dates for task Nr. 12 or know how the Morse Code works for task Nr. 9, cf. Tab. 4.3). This can lead to different and maybe even conflicting paths during the task answering and therefore, the *hard* tasks here can be considered as "complex". However, the *hard* factual search tasks used in Fact are certainly less complex than the ES tasks Expl_1 and Expl_2 used in *US-II* (and *US-III*).

4.3.2.2 *Exploratory Search Tasks*

To design the two ES tasks Expl_1 and Expl_2, the attributes discussed in Section 2.6 have been considered[8]: For the ES tasks of this study, the following three aspects have been identified as relevant: (1) The users should not be too familiar with (the current state of) the domain[9].

7 This was done on the extended factual search task set of user study *US-III*, cf. Sect.5.4.2
8 Similar to the design of the ES task for *US-Ib*, Sect. 4.2.2.
9 For both ES, participants have been asked for their expertise regarding the search task's topic.

Therefore, the tasks have to address topics that are under frequent development, e.g., regarding the associated technology. This results in learning and investigation as task goals. In case the user partially performs factual search related activities (e.g., because the participant already has some domain knowledge), a sufficiently solution of the search tasks should not be achieved immediately what leads to the next aspect. (2) The task should be general and address less structured and open-ended problems. As in *US-Ib*, this ES characteristic usually results in more thought-out search interactions and increases the time that is necessary to solve the task. (3) During the task assignment, uncertainty regarding the task's solution is involved. That is, the task should allow to be answered in different ways according to different, maybe even contradicting, arguments. Although uncertainty is not explicitly involved in Marchinonini's [131] definition of ES, White and Roth [189] stated that uncertainty is related to open-endedness. Furthermore, they state that over time, ES and browsing facilitate to resolve uncertainty (cf. Sect. 2.6). Also Kuhlthau [120, 121] mentioned uncertainty as part of information seeking and includes this aspect in terms of the user's feelings in the stages *initiation* and *exploration* of her model (cf. Sect. 2.4.2). Kuhlthau [120] stated uncertainty as *"..., a natural and necessary aspect of the early stages of the ISP, ..."* (p. 364). Recalling the task attribute *complexity* (as briefly discussed in Sect. 3.3, footnote on p. 55), uncertainty may lead to different paths, which are followed during the task solving, in turn may be conflicting and hence, according to Campbell [36], the ES tasks can (and probably should) be considered as complex as well. Finally, the following two ES tasks (assigned in German) have been used in the study *US-II*:

- $Expl_1$ (adapted from [81]): *"Your friends are planning to build a new house and have heard that using solar energy panels for heating can save a lot of money. Since they do not know anything about home heating and the issues involved, they have asked for your help. You are uncertain as well, and do some research to identify some issues that need to be considered in deciding between more conventional methods of home heating and solar panels. Afterwards you want to discuss this topic with your friends and, therefore, make some notes."*

- $Expl_2$ (adapted from [198]): *"You are flying to Moscow next month. During the travel arrangements you learn that body scanners are being used in many airports as part of routine security procedures. You start thinking about health issues related to their use. Your friends want to calm you down and say that people are exposed to different kind of radiation every day. You want to learn more and start a research to gather a range of information about radiation and health. Afterwards you want to discuss this topic with your friends and, therefore, make some notes."*

This two tasks fulfill the three identified aspects (see above) for ES tasks to induce exploratory behavior (see discussion below). Furthermore, the two tasks correspond to the aspects for exploratory tasks stated in the work of Wildemuth and Freund [191], according to which ES tasks should:

- be associated with learning and/or investigation as goals
- have a general, ill-structured and open-ended problems context
- involve multiple/multi-faceted items and uncertainty as well
- are not "too easy", and are dynamic

The topics "Home Heating" ($Expl_1$) and "Radiation" ($Expl_2$) are known to some extent by most adults. However, only a few people should be able to name specific, and in the best case scientifically supported, arguments pro and con regarding the two topics. Furthermore, both topics are subject to constant technical development which require up-to-date information resp. knowledge for a proper argumentation. Therefore, learning and investigation activities are more likely to appear during the research for the topics what addresses and fulfills the domain aspect (1) for the two tasks. Collecting several arguments regarding the topics surely advances the task processing but this alone does not sufficiently solve the tasks because of their general, open-ended character –the methods and aspects of (efficient) house heating but also the number of possible risks of radiation are multifaceted–, what addresses the second aspect (2). That is, the participants also have to compare and comprehend the found information, apply the knowledge, compare the facts and synthesize the resulting conclusion(s) for the hypothetical discussion with the friends afterwards (cf. last part of the ES task's description). Hence, even if users partially perform factual search activities, e.g., to clarify (factual) questions, emerged during the search, the tasks can not be solved quickly and furthermore, are not considered being "too easy". Last but not least, to research for the unspecified (and therefore, uncertain) topics of home heating and radiation leads to multiplicity of articles of diverse sub-topics in the web with different quality (cf. *source characteristics* as intervening variable in Wilson's improved Model [195] in Sect. 2.3). Therefore, the search (most likely) evolves dynamically. To order the collected information (reducing the uncertainty), the tasks require cognitive resources and the tasks require also to differentiate between the acquired information. Without checking the credibility of the sources, the search can result into contradicting tasks solutions. From that perspective, also the aspect (3) of (decreasing) uncertainty during the tasks processing is fulfilled by both, $Expl_1$ and $Expl_2$.

According to Gwizdka and Lopatovska [81], the source of $Expl_1$, the task was categorized as *"simple information gathering task"*. However, given the argumentation above but also the experience regarding the user study preparation and conduction, Expl_1 is better to be considered as ES tasks instead. The second ES task Expl_2 was categorized as *"exploratory information problem"* in it's source, i.e., by Wu et al. [198], what (also) fits to the requirements on ES tasks for this user study. In addition, both exploratory tasks have been extended with the final claim to the participant to support friends and prepare a discussion with them afterwards. This demand was used to further motivate the participants to undertake an ES in responsibility for other people, the participants feel connected to, with the goal to achieve more natural seeking behavior.

4.3.3 *Search User Interface*

The *search experiment* was conducted using the *Mozilla Firefox*[10] web browser. To provide an usual search environment, a vanilla installation of *Firefox* was used including standard interaction elements, such as "home", "back", "forward" and "refresh" buttons, URL- and search bar, tabs, etc. Logging of the user's interaction was activated by the investigator at the begin of the *search experiment* by a corresponding button of a specially developed browser add-on. To avoid distracting advertisement during the search, an ad-blocker was installed in addition. The task assignment was programmed as a quiz which was placed and fixed in the first tab of the web browser, also called *quiz tab*. The interface with activated quiz tab is illustrated in Fig. 4.6. The user

10 https://www.mozilla.org/firefox/

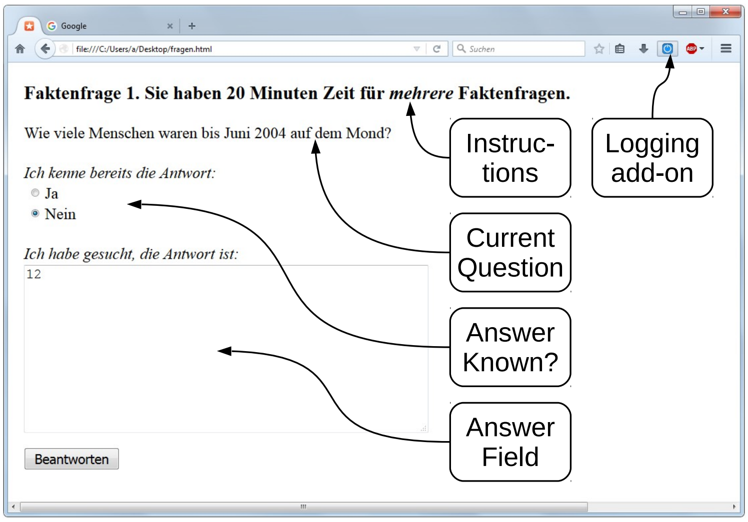


Figure 4.6: Opened *quiz tab* during a fact-finding search in the prepared *Firefox*. The major elements are described by the balloons. The *quiz tab* was available by clicking the first tab with the star symbol in the upper left corner.

could see whether the current task is exploratory of fact based. If fact based, also the number of the current task was shown. Furthermore, the given time limit was displayed and a hint whether the current search block consists of one (in case of Expl) or several (in case of Fact) task(s) was provided. In addition to the display of actual current search tasks, the quiz tab also provided a field to submit the task's answer. For each task, the participants were asked to state whether they already know the answer (Fact) or how much expertise they have regarding the task's topic (Expl). The users have been requested to perform a search for every task even if the answer was already known. In the *introduction* the motivation was given to confirm the already known answer by the search if that case applies. The participants have been instructed (insistently) that the correctness (Fact) resp. the quality (Expl) of the answers is a priority over the number of tasks solved. The reason for that was to further promote a natural search behavior of the users. The maximum number of factual tasks in the Fact block (twelve) was not mentioned to avoid biases and do not challenge the users to answer all of them in the given time (20 min.). To perform the search and answer the tasks, the *Google*[11] search engine in it's German version was provided. Once a participant had submitted an answer, the system provided the next question in the quiz tab. That is, the user could not correct respectively adapt answers later. Since all questions were from different domains and the participants should always search only for the current task, it was not necessary to allow an adaptation of the previous answers. If the time limit for a search block was reached, a corresponding message was shown and the next search block started. After a completed study session with a participant, the browser history was deleted to avoid highlighting of previously clicked search results and personalization.

4.3.4 *Spatial & Technical Setting*

The study was conducted at the Otto von Guericke University, Magdeburg, Germany. To facilitate a good quality of the study and it's recorded data, the lab of the *DKE* research group[12] was chosen as study room. In the lab, only one participant and one investigator were present. After the *introduction* step, the investigator stayed in the background and avoided any interference to the study. As mentioned in the study procedure above, only if a participant had technical problems regarding the study conduction, the investigator was consulted. The search system was equipped with the following recording devices:

- Eye-Tracker: Tobii X2-60
- Camera: One USB-CAM-152H from Phytec with 1280x960px

11 https://www.google.com/
12 http://www.dke-research.de/

The users' interactions have been recorded by a logger developed in the *DKE* research group. The logger was implemented as *Firefox* web browser add-on. It enables to record the user's interactions with the search engine (e.g., used queries, result clicks, dwell times on all web pages, ect.) but also with the web browser (e.g., scrolling, tab and window activation). That is, the recorded data for *US-II* consist of the participant's interactions with the SUI, eye-tracking data, a screen record and a (high resolution) user video that can be used for troubleshooting but also for further investigations regarding face-based user variables.

4.3.5 *Participants & Data Characteristics*

In the study *US-II*, 19 users (13 men and 6 women) participated. The majority of the users (14) had a computer science background (PhD students). They were recruited via mailing lists. All participants use the Internet on the daily basis (several times), to search for information and search with *Google*. The Internet is used for work by 17 participants. All but two of the participants did not know any answers to the factual tasks before the search. Each of those two participants only knew the answer to the (easy) question relating to the largest passenger aircraft (cf. search task Nr. 6 in Tab. 4.3). As described in Section 4.3.3, for the exploratory tasks, the participants had to indicate how familiar they were with the topic. Here a 5-point Likert scale from 1 (Expert) to 5 (Not familiar) was used. The participants had only little information about the topics for both exploratory tasks (Expl_1: $median = 4$, $Avg = 4$; Expl_2: $median = 5$, $Avg = 4$). They knew less about the topic of the second exploratory task Expl_2. Overall, the participants had comparable pre-conditions for the search induced by the fact-finding and exploratory tasks.

Participants spent on average 128 seconds ($min = 18$, $max = 693$, $median = 93$) to answer a factual question. It took them on average 2.6 times longer to answer difficult questions (in sec.: $Avg = 184$, $min = 43$, $max = 693$, $median = 147$) than easy ones (in sec.: $Avg = 69$, $min = 18$, $max = 229$, $median = 51$). On average, the participants answered to 8.8 factual questions ($min = 4$, $max = 12$, $SD = 2.6$). Considering all fact-finding search tasks, 80% of the answers were correct. The

Table 4.4: Average time and standard deviation (SD) in sec. for the search blocks Fact, Expl_1 and Expl_2.

Block	Time in sec.	SD in sec.
Fact	1177	59
Expl_1	1029	194
Expl_2	901	319

participants made more errors answering difficult questions (28.6% errors) than easy ones (10% errors). The time distribution for the three task blocks is shown in Tab. 4.4.

In total, user interactions that correspond to 4022 transitions (2130 in Fact and 1892 in Expl) in 19 fact-finding and $2 \times 19 = 38$ exploratory search sessions have been recorded. Considering the fact-finding search sessions, this results in the median of 111 interactions with a 25-% and 75-%-quantil of 93 and 122 interactions respectively. In contrast, the exploratory search sessions results in a median of 50 and quantils of 34 and 65 interactions respectively. That is, considering the average time spend on the task blocks Fact and Expl (cf. Tab. 4.4), in Fact, each 10.6 seconds and in Expl each 19.3 seconds an interaction was made.

4.4 USER STUDY III: EXPLORATORY & FACT-FINDING SEARCH (PART B)

User Study US-III*:*
Cf. summary: p. 78;
Cf. overview: p. 112

User Study *US-III* pursues the same goals as *US-II* in general but allows to extend the object of investigation because of a greater data base. That is, the goal of *US-III* is to generate data of users who perform factual and exploratory search tasks but to a higher extend in terms of (a) participants, (b) search tasks and (c) additional user variables that have been obtained during the study execution. In particular, *US-III* comprises data of (a) 115 participants who performed (b) four fact-finding multitask assignments (Fact) and two exploratory search tasks (Expl) and (c) provided personal characteristics by answering several psychological questionnaires. These personal characteristics are user variables which may have influence to factual and exploratory search behavior and therefore, are included as relevant research parameters for the investigations regarding this user study[13]. For example, variables regarding *personality*, *sensation seeking* and *motivation* are obtained. The social influence (as aspect of *motivation*) is also considered by implementing a competitive search setting between two users. Corresponding to this extended set of user variables, the study procedure was slightly adapted in contrast to *US-II*. Nevertheless, the two main objectives (as in *US-II*) remain, namely (1) to investigate (validate and extend) user behavior models that can differentiate between fact-finding and exploratory search activities and (2) to distinguish and better understand the search paradigm of exploratory search. While preparing *US-III*, a statement, given by Ageev et al. [2] in their investigation of user's search success (cf. Sect. 3.3.2, p. 70), was considered

13 As a reminder, Wilson's Model [195] describes several intervening variables, such as demographic but also psychological variables, that influence user's information seeking behavior (cf. Sect. 2.3). Hence, this variables are also relevant for factual and exploratory search, in particular if the relation between factual and exploratory search (cf. Sect. 2.6.1) and the interpretation of ES as class of information seeking behavior (cf. Sect. 2.6.3) is considered.

to achieve an adequate and maintainable trade-off regarding the user study size:

> *"However, research of search behavior has been struggling with the tension between the relatively small-scale, but controlled lab studies, and the large-scale log-based studies where the searcher intent and many other important factors have to be inferred."* (p. 345).

On the one hand, *US-III* fulfills the requirements for a controlled lab study incorporating detailed logged (not inferred) interactions with the search system and several user parameters acquired via questionnaires. On the other hand, with eventually 115 participants and the recorded amount of approx. 10 TB data with about 35340 user interactions, the study can be considered as a large-scale study allowing for some statements of generalization, at least for the given user group. Therefore, the user study is considered as an adequate trade-off regarding the statement of Ageev et al. [2]. As in the description of the previous user studies, in the following, the details of user study *US-III* are given. However, since the selection and inclusion of the additional, personal user characteristics require a separate justification and description, first their role for the thesis is addressed in the next sub-section. Afterwards, the section continues with the familiar structure of describing the study's procedure, utilized search tasks, logging techniques, spatial setting, recorded data, etc. The content of this section has already partially appeared in the own publication [118].

4.4.1 *The Role & Integration of User Characteristics*

In this sub-section, reasons to include the chosen user characteristics for this web search related user study shall be motivated more:

4.4.1.1 *User's Motives*

In his research regarding human's information behavior and the origin of the underlying information need, Wilson [195] (cf. Sect. 2.3) identified a work of Morgan and King [135]. In that work, three primary motives (physiological, unlearned and social motives) for human needs are emphasized. This three motives possess noticeable similarities to Wilson's [193] proposition that needs are cognitive, affective and physiological and are related to corresponding barriers (cf. Sect. 2.3)[14]. Consequently, Wilson argued that the concept of motives and the integration of motivational factors for the investigation of ISB and

14 The origin and number of primary human needs (cf. Maslow's Hierarchy of Needs, Sect. 2.1) and motives but also the relation between needs and motives is discussed extensively in the psychological research, e.g., see Olson and Chapin [142]. However, that discussion is not focus of this thesis and should therefore not be continued here.

information needs is a relevant aspect. Given that background, user's motives are also part of consideration and investigations in this work.

According to the Motivation Theory of Rheinberg (et al.) [53, 157, 158], motivation is related to goals or results. These goals can be of any type but for this thesis, goals will correspond to the number of (correctly) answered factual tasks (cf. study procedure in the next sub-sect.). Rheinberg (et al.) [53, 157, 158] proposes three different types of motives which are related to goals and which can be distinguished, each depending on a so called Reference Norm (RN), based on the work of Heckhausen (et al.) [30, 89]. These RNs are goal related stimuli or standards that emerge by a comparison of the individual's performance. The three RNs addressed in user study *US-III* are:

- Criterion-Oriented Reference Norm (CRT): Goals which are defined by guided content, i.e., a given, (easy) identifiable situation or measurably quantity as criterion that can be achieved. In other words, a given external stimulus in relation to a certain natural matter motivates the human being to act. For user study *US-III*, a guided, pre-defined number of fact-based questions, which have to be solved correctly, will be given as criterion to address this type of RN.

- Individual Reference Norm (IND): Goals defined in reference to the own (former) performance. This norm is based on the individual (or "ipsative") comparison between the previously and current resp. prospective achievements. For user study *US-III*, this intra-individual (longitudinal) comparison is implemented by allowing users to choose the required number of correct factual answers by themselves.

- Social Reference Norm (SOC): A comparison between the own performance and the performance of others, cf. Festinger [63]. For user study *US-III*, this inter-individual (cross section) comparison (originally applied on groups) is implemented by asking users to answer more questions correct than another participant vis-à-vis.

Each of this RNs serves a certain purpose especially in the context of motivation, performance and learning, as shown by Rheinberg [156], and the aspect of comparison between the achieved goals allows humans to assess the (own) performance[15]. However, since goals (or results) require an assessment of their achievement, i.e., goals have to be mensurable, utilizing the correctness of fact-finding task's answers, resp. the number of correct answers for a single factual search task block (Fact) is an appropriate mean to address the three

15 Applying this three *reference norms* (conscious or unconscious) to evaluate the performance, e.g., the learning performance if teachers rate pupils in class, is called *reference norm orientation*.

RNs. In contrast, an ES (task) is too complex to evaluate it's (degree of) "correctness" during the study execution for participants and investigators. Furthermore, to compare the given answers for the ES tasks between the participants (to address the SOC) was not feasible at least for the given study setting. Consequently, in *US-III*, each of the three RNs is addressed by a factual search task block separately. Furthermore, a fourth factual search task block (without a goal) is used that serves as reference for comparison.

4.4.1.2 *Further Personal Characteristics*

In contrast to the user's motives, which require an advanced study setting, other personal user characteristics can be obtained more easily, e.g., by psychological measuring instruments in form of questionnaires. Following the same argumentation regarding the relevance of intervening variables for the user's information seeking behavior (cf. Sect. 2.3), the inclusion of user's personality, aspects of intelligence and sensation seeking will be motivated respectively:

Five Factors of Personality

User's *personality* can be operationalized by a relative stable composition of traits which, e.g., are obtained by the *NEO Five Factor Inventory* [45]. The derived five factors of personality, also called *Big-Five*, are a result of several factor analyses and comprise a notable consistency [151]. As a substantial personal user characteristic, personality may represent an invervening variable to ISB, thereby to ES, and therefore, is also considered in this work. Applying the *NEO Five Factor Inventory*, here in it's German translation [29], obtains the five factors, namely[16]:

- *Neuroticism (N)*: The person's tendency to experience negative feelings. People with high values on neuroticism have low resources to cope with stress, are emotionally reactive, impulsive, anxious and are prone to depression.
- *Extraversion (E)*: The person's tendency to be sociable, talkative and adventurous. People with high values on extraversion enjoy interacting with others and are often perceived as full of energy.
- *Openness to experience (O)*: The person's tendency to seek for experience, be (intellectually) curious, tolerant, open to emotions and sensitive to beauty. People with high values on this factor are willing to try new things, have unusual ideas and high imagination.
- *Agreeableness (A)*: The person's tendency to get along with others. People with high values on agreeableness are friendly, cooperative, altruistic and considerate. They trust others and are perceived trustworthy and sympathetic.

16 The one letter abbreviations for the factors (N, E, O, A, C) will be used later in the corresp. analysis, cf. Sect. 5.4.2.2.

- *Conscientiousness (C)*: The person's tendency to show discipline, be duteous, like order and control own impulses. People with high values on conscientiousness act preserving, responsible but also are perceived as scrupulous, tidy and less spontaneous.

The Five Factor Inventory enables to investigate the relation between user's traits and their exploratory (but also factual) seeking behavior. For instance, the obtained factors allows to test assumptions like "Differences in exploratory search behavior between extroverted and introverted users can be identified." or "Users with high values on conscientiousness spend more time on web pages and SERPs during ES in contrast to users with low values on conscientiousness.".

User characteristics regarding *intelligence* also appear to be relevant for (exploratory) information seeking. The Wechsler Adult Intelligence Scale (WAIS), here in it's German translation [9], allows to test adults' intelligence according to the two categories *verbal* and *performance* intelligence. The test consists of ten (to optionally fifteen) sub-tests and is devised for adults from 16 to 89 years. However, to reduce the total number of tests for the participants, and thus to avoid exhausting effects in the *Search Experiment*, only the three following sub-tests of the WAIS have been used:

Wechsler Adult Intelligence Scale

- *Similarities*: Test regarding verbal comprehension. Participants have to describe in which way two words are related to each other. This sub-test purposes to measure the participant's abilities regarding abstract verbal reasoning, build concepts, resp. semantic knowledge. For all search activities in this thesis, this aspect is important, e.g., to formulate appropriate queries or to extract relevant information from SERPs and web pages. That is why this test was used.

- *Symbol Search*: Test regarding processing speed. Participants have to view and compare rows of abstract symbols and target symbols and mark whether the target symbols appear in each row or not. To solve the tasks, a time limit is given. Since the speed to process information is considered as relevant for answering the search tasks in the given study setting, this test was used.

- *Letter-Number Sequencing*: Test regarding working memory. Participants hear a combination of numbers and letters. Afterwards, they have to recall and order them: First the numbers in increasing order and second the letters in alphabetical order. This sub-test purposes to measure the participant's abilities regarding the working memory, attention and mental control. Since the number of pieces of information (also called *chunks*), a user can remind during search tasks processing, is considered as relevant aspect, this test was used.

The WAIS assess the participant's intelligence by measuring individual abilities even though the overall intelligence is not supposed to be equal to the sum of the individual abilities. However, according to the test-recension of Molz et al. [134], the WAIS enables a reliable measurement of relevant components of intelligence.

Sensation Seeking Scale

Last but not least, the personal characteristic of *sensation seeking* is addressed. Originally created by Zuckermann [202], the underlying conception of sensation seeking is that humans (inter-individual) differ in their needs for situations they feel comfortable with. Consequently, humans seek differently strong for the corresponding situation which cause reinforcing effects of stimuli. The effects in turn, are related to the unfamiliarity and complexity of the (new) situation. Since the paradigm of ES is inherently accompanied by aspects of complexity (of the search tasks) and unfamiliarity (of the search domain), user attributes of sensation seeking are relevant and promising variables. To measure the user's sensation seeking, the so called Sensation Seeking Scale (SSS), here in it's German translation [18, 19], can be used. Analysis of sensation seeking [202, 203] repetitively revealed four sub-factors, which are (also) obtained by the corresponding SSS:

- *Thrill and Adventure Seeking*: The person's tendency or desire for exciting and risky (physical) activities and adventures.
- *Experience Seeking*: The person's tendency to achieve new impressions, experiences and personal development.
- *Disinhibition*: The person's tendency to get stimulation by social activities often accompanied by executing disinhibited behavior.
- *Boredom Susceptibility*: The person's tendency to refrain from any kind of repeating experiences and routine, e.g., regarding activities, social interactions and entertainment.

This curiosity related traits have not been investigated yet in context of (exploratory) search activities but might reveal some interesting insights. In particular, the second sub-factor, *Experience Seeking*, appears to be relevant for investigation on ES[17].

4.4.2 *Study Procedure*

In user study *US-III*, participants had to solve the same two exploratory search tasks Expl_1 and Expl_2 as in *US-II* (cf. Sect. 4.3). Since fact-finding search activities can partially be interwoven in exploratory search activities (cf. Sect. 2.6.1), in addition to the differentiation also the

17 Since the two sub-factors *Thrill and Adventure Seeking* as well as *Disinhibition* are less related to the experimental setting of *US-III*, strong relations are not expected. However, the analysis of the SSS revealed positive correlations between the four sub-factors, mostly around 0.3 [8].

relation between factual and exploratory search activities is a crucial issue of investigation. Therefore, for *US-III* a higher fraction of factual tasks during the *search experiment* seemed reasonable. In accordance to the discussion regarding user *motives* above and the three associated Reference Norms (RNs): CRT, IND and SOC, three corresponding factual task blocks Fact_{CRT}, Fact_{IND} and Fact_{SOC} have been used, each with a different goal condition regarding the number of questions to be answered correctly. In that way, the user's motives can be addressed in the user study and the fraction of factual tasks can be increased at the same time. Furthermore, a fourth factual task block Fact_{NON} is used where no specific number of correctly answered fact-finding search tasks have been asked to have a reference for comparison to the three other Fact blocks. User study *US-III* implements a four step procedure consisting of *pre-search questionnaires*, an *introduction* the *search experiment* and *post-search questionnaires*:

1. *Pre-Search-Questionnaires:* The questionnaires have been used to gather user's demographic data and personal characteristics. Tests which include more demanding test items have been applied here, i.e., before the *search experiment*, to gain an unbiased baseline regarding each users' individual personal characteristics. In particular, the two following psychological measuring instruments have been utilized in this first step: WAIS and SSS (cf. Sect. 4.4.1.2)

2. *Introduction*: In this step, the participants have been introduced to the *search experiment* afterwards, i.e., where they can see the search tasks assignments, the given time to solve the tasks, how they can submit their answers, which search engine they should use etc. Furthermore, the calibration of recording devices such as eye-tracking and user camera was done in this step.

3. *Search Experiment*: In this step, the actual search was done. The system provided multiple search tasks for the users and recorded their interactions during the task processing. The users had to answer the two ES tasks and several factual tasks in the four Fact blocks. The *search experiment* was executed by two participants simultaneously to enable the awareness of the other participant's presence in the study room. That was important for the Fact_{SOC} block to address the *Social Reference Norm*. In the study room, only the two participants were present. Only if a participant had problems which potentially compromised a frictionless data record, the investigator was called.

4. *Post-Search-Questionnaire*: In this last step the final questionnaire was used to obtain the remaining user variables (cf. prev. sect.). The *NEO Five FactorInventory* to get user's *personality* traits (the

Big-Five) was considered as less demanding than the tests used in the *pre-search-questionnaires* step and hence, was applied here.

For each of the factual task blocks Fact_{NON}, Fact_{CRT}, Fact_{IND} and Fact_{SOC}, the tasks assignment at the begin of each block was slightly different regarding the number of questions to be answered correctly to address the corresponding RNs with their different goal condition. In particular, at the begin of Fact_{NON}, no external motivation for the participants, i.e., no numeric goal specification was given. Therefore, block Fact_{NON} is considered as an appropriate baseline for the remaining three factual search blocks which have different motivational goals according to the three RNs (cf. Sect. 4.4.1.1). Furthermore, the task difficulty in Fact_{NON} was alternated, starting with an *easy* task, to allow the users to get attuned to the search tasks and to better estimate their own performance. That was especially important for Fact_{IND}. The blocks Fact_{CRT}, Fact_{IND} and Fact_{SOC} had specific goals. In block Fact_{CRT} the instruction was to answer at least three tasks correctly (Criterion-Oriented Reference Norm). For block Fact_{IND}, the users had to choose a number of factual tasks they think they can answer correctly (Individual Reference Norm). The remaining block Fact_{SOC} was a competition with the instruction to *"be better than the other participant in the room"*, i.e., the motivational goal is to solve more search tasks correctly (Social Reference Norm). For the blocks Fact_{CRT}, Fact_{IND} and Fact_{SOC} one participant got only *easy* tasks and the other participant got only *hard* tasks. In this way, the influence of motivational goals in respect to the task difficulty can be investigated.

During the *search experiment*, each participant had at most 30 minutes to answer the two exploratory tasks Expl_1 and Expl_2 together. In contrast to *US-II*, the two exploratory tasks have been arranged in one exploratory search block, indicated by $\text{Expl}_{1\&2}$, and the time to answer both has been reduced from 40 min. to 30 min. Furthermore, the Expl_1 and Expl_2 have been presented to each participant in random order in the $\text{Expl}_{1\&2}$ block. The reason for that adaptation was threefold: first, each participant had the possibility to spend less or even more than 20 min. on the first given ES task and hence, had a bit more freedom to act; second, the number of study designs, i.e., the different orders of ES tasks and the four Fact blocks, was decreased to twelve (cf. designs below); third the time to solve the $\text{Expl}_{1\&2}$ and four Fact blocks should not exceed the time of 1 hour too much to prevent signs of fatigue. This third reason also motivated the decision to set a time frame of 10 min. for each of the four fact-findings blocks. In sum, the participants had about 70 min. to solve all search tasks in the five search blocks $\text{Expl}_{1\&2}$, Fact_{NON}, Fact_{CRT}, Fact_{IND} and Fact_{SOC}.

The maximum number of single fact-finding tasks, a Fact block could consist of, was increased to sixteen (not twelve as in *US-II*) to provide a sufficient number of tasks even if they are answered very fast. However, because of the reduced time frame of 10 min. for each Fact block it was even unlikely to reach twelve tasks. Nevertheless, reducing the number of fact-finding tasks (smaller than twelve) was considered to be critical because of the (new) used motivational goal setting. For example, it was possible, that a strongly social oriented participant "with all means" tried to be better than the other participant and hence, answered an unusual high amount of factual tasks in the block Fact_{SOC}. However, as in *US-II*, all participants have been instructed (insistently) that the correctness (Fact) resp. the quality ($\text{Expl}_{1\&2}$) of the answers is a priority over the number of tasks solved for the whole *search experiment*.

In the *search experiment*, the five search blocks $\text{Expl}_{1\&2}$, Fact_{NON}, Fact_{CRT}, Fact_{IND} and Fact_{SOC} were randomized as follows: $\text{Expl}_{1\&2}$ was always placed before or after the fact-finding blocks to minimize distracting actions between the fact-finding blocks and therefore, to enhance the comparability between the fact-finding blocks. Within the fact-finding blocks, Fact_{NON} was always placed at first to serve as baseline as already mentioned before. The remaining factual blocks were randomized in all possible permutations. That is, each pair of participants had one out of twelve possible block sequences designs within the *search experiment*:

- $Design_1$: $\text{Expl}_{1\&2} \rightarrow \text{Fact}_{NON} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{IND} \rightarrow \text{Fact}_{SOC}$
- $Design_2$: $\text{Fact}_{NON} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{IND} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Expl}_{1\&2}$
- $Design_3$: $\text{Expl}_{1\&2} \rightarrow \text{Fact}_{NON} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Fact}_{IND}$
- $Design_4$: $\text{Fact}_{NON} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Fact}_{IND} \rightarrow \text{Expl}_{1\&2}$

  ...

The complete list of all twelve designs is given in the Appendix A, Section A.1. Both participants in each pair got the same block sequence design. This guaranteed that both participants performed the same factual search blocks at approx. the same time what was important especially for Fact_{SOC} where the participants may have eye-contact what again increases the chances for the immersion into the SOC condition[18]. Fig. 4.7 illustrates the whole study procedure of *US-III* exemplified on $Design_1$. The full study took about 3 hours. Each participant received a reimbursement of 20 Euro.

4.4.3 *Search Tasks: Factual & Exploratory Tasks*

In the following, the used search tasks are described in detail. In specific, the number of fact-finding multitask assignments for the four Fact block had to be extended in contrast to *US-II*.

18 Due to technical reasons, the participants had no synchronized, i.e., no "real" competition in the Fact_{SOC} block. However, simply the pure presence of the other participant and the working space arrangement that allowed to see the other participant was considered to be sufficient to induce effects regarding the SOC.

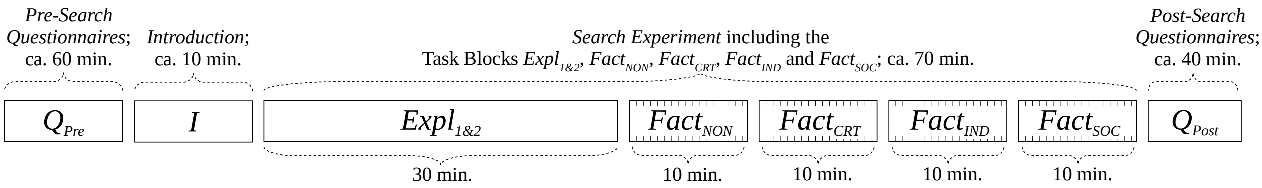


Figure 4.7: Illustration of the user study procedure of *US-III* exemplary incorporating $Design_1$.

4.4.3.1 *Factual Search Tasks*

For the *search experiment*, 117 factual search tasks have been created. Since no benchmark of fact-finding search tasks could be found that was large enough and in German language, the tasks of *US-II* have been further extended. During the creation, care was taken to ensure that tasks are from different domains. At the end, domains such as sport, natural science, geography, technology, literature and history as well as movies and music are covered. To generate the factual tasks, the constrains as in *US-II* have been used. Again, for the fact-finding tasks two difficulty levels have been implemented resulting in 57 *easy* and 60 *hard* tasks. In addition, the whole collection of factual tasks have been tested in a small pre-study with 19 students. In Tab. 4.5, examples of four *easy* and four *hard* tasks are listed. To be able to measure the correctness of the answers of the fact-finding tasks during the experiment, for each task five possible answers (one right and four wrong answers) have been generated. The pool of 117 fact-finding search tasks was used to selected the tasks randomly for the factual search blocks.

4.4.3.2 *Exploratory Search Tasks*

As mentioned in the study procedure, in *US-III* the same two exploratory search tasks $Expl_1$ and $Expl_2$ as in *US-II* (cf. Sect. 4.3) have been used but combined in one exploratory search block $Expl_{1\&2}$ and in random order for each participant.

4.4.4 *Search User Interface*

As in *US-II*, the *search experiment* was conducted using a vanilla installation of the *Mozilla Firefox* web browser and the *Google* search engine. The users' interactions were recorded using the same logger add-on with the same interface in general. All instructions and task assignments were shown in the first browser tab (also called *quiz tab*) what was fixed. However, regarding the possible interactions to answer the tasks, some adaptations were made regarding the different study procedures. While the answers to ES tasks could be given in a free text box (as in *US-II*), to answer a single factual task, the users

Table 4.5: Selection of eight factual search tasks from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers, translated from German. The * indicates the correct answer.

Nr.	Level	Task description	Answers
1	Easy	*What is the name of the third longest river in the world?*	*Yangtze*, Nile, Amazon River, Volga River, Yellow River*
2	Easy	*When was Julius Caesars birthday celebrated?*	*July 13th*, March 15th, Juliy 7th, October 13th, November 15th*
3	Easy	*What was the name of the Spacecraft were the first women was on board?*	*Wostok 6*, Wostok 5, Wostok 4 , Wostok 2, Sojus*
4	Easy	*Which element is the best electrical conductor?*	*Silver*, Copper, Gold, Platinum, Mercury*
5	Hard	*Which moon in our solar system is the closest to its planet?*	*Phobos*, Earth-Moon, Charon, Deimos, Mab*
6	Hard	*Which Summer Olympic Games host-country had not won a single gold medal during those games?*	*Canada*, Germany, France, Greece, Sweden*
7	Hard	*How are the weights to regulate the flying altitude in a hot-air balloon called?*	*ballast*, fender, filling-weight, gravitors, over-weight*
8	Hard	*Which actor won in the same year the Golden Raspberry Award for the worst actor, the worst supporting actor and the worst supporting actress?*	*Eddie Murphy*, Adam Sandler, Tom Cruise, Kevin Costner, Michael Douglas*
... 109 further fact-finding tasks (cf. Appendix A, Section A.2)			

first had to indicate this via an "Answer"-button and afterwards they had to choose one answer out of five. If users were not sure what the correct answers for a given fact-finding task was, they could select "I don't know". To select the (correct) answer, users had only 30 seconds to avoid situations where users proceed the search after clicking the "Answer"-button and check which of the given five answers is correct. After the answer selection, the next factual task was shown. At the end of each fact-finding block, a short feedback regarding the search performance, i.e., the number of correct submitted answers, was shown. In addition to that, in the RN related blocks Fact_{CRT}, Fact_{IND}

Figure 4.8: Set-up of user study *US-III*. The two working places allow the participants to search simultaneously. To see the other participant's face, the chairs are arranged in an arc of 120°. This was relevant for the factual search task block Fact_{SOC} that addressed the Social Reference Norm (SOC).

and Fact_{SOC}, a message was shown, whether the goal of the current factual block was achieved or not. After the feedback at the end of each block, a short questionnaire regarding the user's current search performance and the perceived tasks complexity have been asked.

4.4.5 *Spatial & Technical Setting*

User study *US-III* was also conducted at the Otto von Guericke University, Magdeburg, Germany. To provide a study set-up where two participants can perform search activities for about 70 min. without any distractions, a dedicated study room without any through-going traffic in the surrounding environment was chosen. This facilitated a quiet study atmosphere and allowed good circumstances for it's recorded data. Furthermore, two light panels to provide a balanced room illumination were used. Figure 4.8 depicts the study set-up. The search systems were equipped with the following recording devices:

- Eye-Tracker: One Tobii X2-60 and one Tobii T60 Eye-Tracker[19]
- Camera: two USB-CAM-152H from Phytec with 1280x960px for each user

19 Both eye-tracking devices have comparable recording properties, i.e. a sample rate of 60Hz, gaze accuracy 0.4° to 0.5°, a latency <35ms, operation distance of 40 to 90cm and both devices run on the eye-tracking software Tobii-Studio version 3.4.2.

Table 4.6: List of separate factual search tasks for each RN related block (Factual Tasks) and the number of fact-finding search sessions (as multitasking assignments) for each RN related block (Fact blocks).

	Fact_{NON}	Fact_{CRT}	Fact_{IND}	Fact_{SOC}	total
Factual Tasks	717	1076	1056	1108	3957
Fact blocks	113	113	111	110	447

That is, the recorded data for *US-III* consist of, eye-tracking data, screen records and (high resolution) user videos for troubleshooting but also for further investigations regarding face-based user variables. The participant's search, i.e., the interactions with the search engine but also with the web browser, again have been recorded using the developed logger add-on.

4.4.6 *Participants & Data Characteristics*

In the study *US-III*, 115 (76 women and 39 men) participated. The mean age is 26.78 (min = 17, max = 63) and the majority of 73 participants are students. 16 participants reported to have jobs in a variety of fields, 6 are still in schooling, 2 are in retirement, 6 are unemployed and 12 refused to give information about their current status. To achieve a broad spectrum of users, volunteers via social networks, bulletin boards and at supermarkets were acquired. As already mentioned in the study procedure, each participant received a small allowance of 20 Euro as compensation.

In sum, the data comprises 447 fact-finding search sessions (as multitasking assignments) from the four Fact blocks: Fact_{NON}, Fact_{CRT}, Fact_{IND} or Fact_{SOC}, cf. Tab. 4.6. Since some participants unintentionally closed the web browser, some fact finding block have been lost and therefore, the number of available Fact blocks is less than the maximum of 115. Tab. 4.6 also lists the number of separate fact-finding tasks for each block. The numbers already imply that users answered more questions in blocks with motivational goal but this is investigated later in Section 5.4.2.1. Since some users spend a long time on one of the ES task or just skipped the other, not $2x115 = 230$ but 226 exploratory tasks (Expl_1 or Expl_2) are available in the data set.

4.5 CHAPTER SUMMARY

In this chapter, the set of all own user studies to generate the necessary data for this thesis, i.e., to investigate the hypothesis *H*2, *H*3 and *H*4 have been described. Beginning with a two-parted user study with young users (*US-I*), where children performed a free voice-controlled search (*US-Ia*) and an exploratory voice-controlled search (*US-Ib*), the

Cf. user studies summary on p. 78

chapter proceeded with a detailed description of the two main user studies *US-II* and *US-III*. In both (*US-II* and *US-III*), participants had to perform two ES tasks Expl and at least one sequence of fact-finding search tasks Fact (also called multitasking). Goal of the studies here was to acquire data for the investigation of user models (1) that are able to differentiate between this two search activities, (2) to distinguish and better understand exploratory search as class of information seeking but also (3) to get more insights regarding the relation between factual and exploratory search activities. While *US-II* served as an important precursor for user study *US-III* regarding model generation, the data of *US-II* was also used to create an annotation ground truth for user's reading behavior for future work, cf. Sect. 6.3. *US-III* incorporated additional user variables (e.g., *personality* or *sensation seeking*), relevant for exploratory (but also factual) search behavior. Furthermore, *US-III* could extend the object of investigation because of a greater data base in terms of participants, search tasks and user variables.

For each user study, several data streams and interactions have been recorded, respectively annotated afterwards. For the sake of data re-usability, in some studies, even more data streams have been recorded than analyzed in detail for this thesis. This allows to hypothesize and investigate new research questions in context of EIS but also facilitates future research on other areas, such as emotion recognition (during search) using the video records or voice related analyzes using the acoustic utterances of the young users. Furthermore, it shall be mentioned that basically all of the user studies have been conducted in context of interdisciplinary projects. That is, the studies of course pursue the purposes as described in this chapter, but also can and have

Table 4.7: Overview of the recorded data of all user studies described in Chapter 4. Used abbreviations are User- & Screen Video (USV), Audio (A), Eye-Tracking (ET), Interaction (I), Reading Behavior (R).

Study	Participants	Recorded/ Annotated Streams	approx. $\sum$ of Session Time in h	approx. $\sum$ of Interactions	approx. Data-size in GB	Questionnaire Items	Publication
US-Ia	10	USV, A, ET, I	5	480	16	12	[76]
US-Ib	5	USV, A, ET, I	6	320	100	20	[116]
US-II	19	USV, ET, I, R	19	4,020	39	20	[114]
US-III	115	USV, ET, I	345	35,340	10,000	411	[118]

been used for investigations of other research areas[20]. In particular, user study *US-Ia* was also used to reveal general voice-controlled interactions of young users with the *Knowledge Journey* to develop appropriated, child centered SUIs [68]. In *US-III* even more (not here mentioned) personal characteristics of the users have been obtained to analyze motivational aspects in relation to micro expressions using the so-called *Facial Action Coding System* [57]. Finally, in Tab. 4.7 all user studies with their recorded data parameters are listed to provide an overview. Furthermore, the details regarding the data streams and their corresponding feature spaces are given in the Appendix A, Section A.3.

20 This is a further reason why several data streams have been recorded as well.

"Humans are explorers by nature, we seek to extend our knowledge by journeying beyond visible horizons."
— White & Roth [189], p. 1.

5 MODELING, CHARACTERIZING AND SUPPORTING EXPLORATORY INFORMATION SEEKING BEHAVIOR

In this chapter, the data acquired by the user studies is utilized to generate and analyze user models regarding specific properties of Exploratory Information Seeking (EIS). Furthermore, several user characteristics and their influence to the search behavior are investigated. Both, the models and user characteristics, contribute to reveal relevant aspects to develop search systems that are able to provide appropriate and user friendly support. If the users' current seeking behavior can be identified, an advanced information system becomes able to adapt the search interface respectively. In particular if exploration becomes necessary for the user, such adaptive features are promising and can improve the search experience for everyone in the future. Therefore, several approaches to provide support for Exploratory Search (ES) are discussed as well in the last part of this chapter.

The following sections investigate the users' Information Seeking Behavior (ISB) on different interaction levels, i.e., in terms of explicit and implicit interactions. User's exploratory (and factual) search activities are modeled, analyzed and classified (*H*3). In particular, Markovian models are defined and trained based on the corresponding study data (Sect. 5.1). Afterwards, the models are used in a classification setting to investigate their ability to identify exploratory search (Sect. 5.2). In a next step, the models are used in a clustering setting (Sect. 5.3), what allows to validate the study design but also to identify latent behavior clusters in user's search. The proposed approaches are not limited to the thesis' data but provide a methodology to model and analyze any kind of (seeking) behavior. The chapter continues with the investigation of additional user variables (*H*2) and their influence to EIS but also differences regarding fact-based multitasking search sessions are considered (Sect. 5.4). Last but not least, the revealed findings are used to discuss approaches to support exploratory search (*H*4) in terms of front-end aspects, i.e., the user interface, but also back-end aspects, i.e., approaches for the search algorithms (Sect. 5.5). The content of this chapter has already partially appeared in the following own publications: [33, 34, 70, 71, 115, 116, 118, 167, 168, 180].

5.1 MODELING SEARCH BEHAVIOR

To differentiate between user's search activities (e.g., exploratory and factual multitasking search), user models, which represent the search behavior, are a promising approach. However, the possibilities to model human information behavior are nearly unlimited and therefore, usually have to be restricted to certain, relevant aspects of interest. This restrictions, in contrast, carry a source of incompleteness if some relevant behavioral characteristics are not modeled (sufficiently).

The rather theoretical models in Chap. 2 describe user's information (seeking) behavior from a holistic point of view. Among other central model properties, such as the inclusion of the information need or the general procedure to satisfy the need, the models often have in common that aspects of exploration (on the level of information seeking behavior) are also involved. In particular if the user during the search process is confronted with situations of uncertainty, a component of exploration (sometimes in terms of browsing) is the models' approach to advance the seeking. A further common property of the theoretical models is to describe the user's information (seeking) behavior in terms of (abstract) states or phases, which are traversed over time. This approach is also utilized in the application-oriented models in Chap. 3 but more related to the actual search system interactions and hence, more restricted and less holistic. Nevertheless, the findings by utilizing these application-oriented models showed how important and descriptive detailed individual interactions are. For example, the transition between or the duration in the individual concrete (system or user) states can be used to distinguish search behavior in general or even identify more abstract states or phases within the seeking process. However, an extensive characterization and reliable identification of exploratory search, as abstract but also concrete (and therefore, connecting) search activity, is still missing in the literature.

Following the fundamentals from Chap. 2 and Chap. 3 and utilizing the data obtained in Chap. 4 allows to investigate user behavior models for such exploratory information seeking. The model based approaches, as utilized in this thesis, also interpret the user behavior as sequence of interactions (i.e., as time series) and allow to estimate how good an (unknown) user behavior fits to a given model, respectively which user model best represents a given (unknown) user behavior. That is, the task of differentiation can be interpreted as classification problem that in turn can be solved and appropriately treated by different means, such as model and feature selection or parameter optimization. In order to model, analyze and classify user's (exploratory) search activities (*H*3), in this and the following sections, Markovian models[1] are trained and used incorporating different settings and properties. This facilitates to compare the search activities but also to reveal an adequate parameter setting for the classification task at hand. This section comprises the model definition in preparation for the subsequent search activity classification and clustering afterwards.

1 As a reminder, a motivation for and the necessary fundamentals of Markovian models have been given in Sect. 3.3.2, p. 62.

5.1.1 *Explicit and Implicit Interactions*

Sequences of conscious and unconscious interactions

As already pointed out, user seeking behavior can be interpreted as a sequence of search related actions performed in a search session. This actions can be explicit, i.e., directly executed conscious "interactions" with the search system. The actions can also be implicit, i.e., indirect "exhibited" unconscious "(re-)actions" of the user. From the perspective of Markov models (or graphical models in general), explicit interactions can be interpreted and represented as discrete[2] states. Thus, in this work, user's exploratory (Expl) and factual (Fact) search activities are modeled as sequences of discrete states. In addition, users' implicit (re-)actions are often considered as attributes of the corresponding states and are modeled as emissions in the Hidden Markov Models (HMMs). In the following, at first Markov models, restricted to users' explicit interactions, and afterwards HMMs, also considering the user's implicit interaction, are defined and compared by their parameter set.

5.1.2 *Markov Model Definition*

Representing search interactions by simple states and transitions

After the preparation of the search interaction data, logged by the web browser add-on in user study *US-II*, the following four user states have been derived:

- *Query*: A user is formulating a search query. In case of *US-II*, a user could formulate an own query by themselves but could also use *Google's* feature of query auto-complete suggestions if the first letters of a query have been entered.
- *SERP*: A user is examining a Search Engine Result Page (SERP) after a query was entered.
- *Page*: A user is examining a web page as result of a clicked URL in a SERP or reached by a clicked URL from an other web page.
- *Main*: A user is viewing the first tab in the web browser (i.e., the *quiz tab*) to read, answer or make notes regarding the current search task.

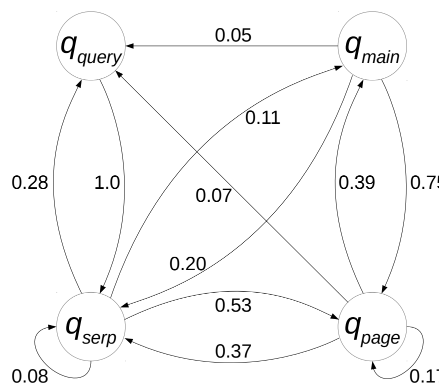
Hence, the data set $\underline{S}$ of *US-II* consists of K search sessions indexed by S^k with the length L^k and each session comprises a sequence of (visited) states from the state space $Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$. That is, $\underline{S} = (S^1, ..., S^K)$ where each search session $S^k = (S^k_1, ..., S^k_{L^k})$ and each

2 The general framework of Markovian models also allows to use continuous states resp. a continuous state space. However, the user behavior investigated in this thesis incorporates only discrete states since the interaction with the used search systems can only lead to situations which are inherently separable. Hence, the investigations here are restricted to discrete states and a discrete state space.

state $S_l^k = q_i^k \in Q$. Since the *search experiment* in *US-II* was conducted in the similar way for each participant and each participant performed the search independent of the other participants, for the data set $\underline{S}$, an independent and identical distribution for all search sessions is assumed to model the search behavior. Furthermore, the sequence of states of each individual search session forms the following *joint distribution* which will be transformed without loss of generality using the *product rule*:

$$\begin{aligned} P(\underline{S}) &= \prod_{k=1}^{K} P(S^k) = \prod_{k=1}^{K} P(S_1^k, ..., S_{L_k}^k) \\ &= \prod_{k=1}^{K} P(S_1^k) \times \prod_{l=2}^{L^k} P(S_l^k | S_1^k, ..., S_{l-1}^k) \end{aligned} \tag{5.1}$$

Applying the *Markov Assumption* for the data set results exactly into Eq. (3.9) (cf. Sect. 3.3.1, p. 64) for each S^k. Disregarding the information of the two search activities Fact and Expl, i.e., ignoring the labels of the three sub-search sessions, and training a 1st-order Markov model on the data of *US-II* leads the graphical resp. formal model in Fig. 5.1. The model shows that after entering the query, every user ($a_{query\ serp} = 1.0$) is forwarded to a SERP. About half of all interactions on SERPs result in a click on an URL ($a_{serp\ page} = 0.53$) and thereby, to a web page. Alternatively, users can click on a next (or previous) SERP, what happens relatively seldom ($a_{serp\ serp} = 0.08$); users can enter a new query resp. reformulate it ($a_{serp\ query} = 0.28$); or users can directly go (back) to the quiz tab (q_{main}). From a web page, each of the other states can be reached, whereby going (back) to the SERP or the quiz tab is similar (0.37 and 0.39 resp.) and more likely than performing a click on a URL ($a_{page\ page} = 0.17$) or enter a query on the search



$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.28 & 0.08 & 0.53 & 0.11 \\ 0.07 & 0.37 & 0.17 & 0.39 \\ 0.05 & 0.20 & 0.75 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure 5.1: Illustration of the 1st-order Markov model as graph (left) and it's corresponding components Q, A and π (right) trained on the data set from *US-II* ignoring the search activity labels.

engine's start page ($a_{page\ query} = 0.07$). The so called *hub and spoke* search pattern, mentioned by Cutrell and Guan [48, 49] (cf. Sect. 3.2.1), can also be observed in the 1st-order Markov model by the relatively strong interaction between the states q_{serp} and q_{page}. In the quiz tab users (re-)read or answered the search tasks. The quiz tab was also the starting point for each participant ($\pi_{main} = 1.0$).

The Markov model in Fig. 5.1 already revealed some insights of the user's search behavior, such as their (expected) dominant click behavior on URLs to open a web page being on a SERP. Another observed behavior is the user's apparently preference of returning from a web page to a SERP to subsequently enter a query instead of directly visit the search engine's start page and enter a query[3]. However, to differentiate the search activities Fact and Expl by analyzing the Markov model's components A and π (as exemplary done for the model in Fig. 5.1), but also to apply an automated classification, for each search activity a separate model has to be trained.

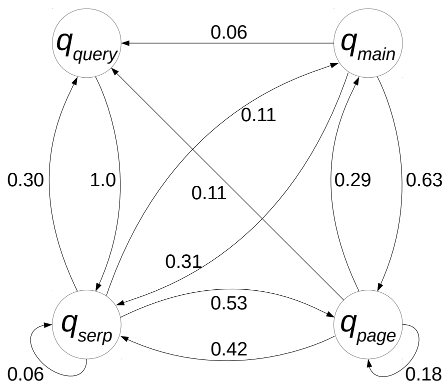
5.1.3 *Markov Model Parameter Comparison*

Differences in search activities considering simple states and transitions

Splitting each full search session (of approx. one hour) of each participant from *US-II* into three sub-sessions according to the three search blocks Fact, Expl_1 and Expl_2 (each approx. 20 min.) and train a 1st-order Markov model for each of the two corresp. search activities, results into the models θ^{Fact} and θ^{Expl} as illustrated in Fig. 5.2 and Fig. 5.3. Comparing the transition probabilities in the component A of θ^{Fact} and θ^{Expl} shows that users tend to click twice as much on a further (or previously) SERP in Expl. This may be an indication of the more complex ES tasks which require the users to examine the result pages longer and more exhaustive to estimate relevant results.

Continuing the comparison shows that the probability to go back from a web page to the *SERP* state and further to the *Query* state is higher in Fact than in Expl. At first glance, it could be assumed that this is caused by the study design and it's time restrictions on each task block (max. 20 min.): Users needed more time on web pages in Expl (cf. analysis in Sect. 5.1.5) and therefore, had less time resp. chance to go back and initiate a new search. However, the analysis of the lengths of the search blocks showed that the majority of the users spent less time than the provided 20 min. to solve the ES tasks and hence, this initial assumption does not hold (cf. Sect. 4.3.5, Tab. 4.4). A more plausible explanation is the (unintended) user's ambition to perform best: Although in the *introduction* to the *search experiment* (cf. Sect. 4.3.1) participants have been instructed (insistently) that the correctness of the answers is a priority over the number of solved tasks, it is possible

3 In fact, it is more likely that users perform the two steps $q_{page} - q_{serp} - q_{query}$: $P(q_{serp}|q_{page}) \times P(q_{query}|q_{serp}) = 0.53 \times 0.28 = 0.1484$ than typing a query on the search engine's start page after being on a page: $q_{page} - q_{query}$: $P(q_{query}|q_{page}) = 0.07$.



$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.30 & 0.06 & 0.53 & 0.11 \\ 0.11 & 0.42 & 0.18 & 0.29 \\ 0.06 & 0.31 & 0.63 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure 5.2: Illustration of the 1st-order Markov model θ^{Fact} as graph (left) and it's corresponding components Q, A and π (right) trained on the data set from *US-II* using all interactions of Fact.

that users nevertheless attempted to (correctly) solve as much factual search tasks as possible, what finally caused relatively more query formulations in Fact[4]. Furthermore, the higher probability in Fact to return to a SERP from a web page indicate a search behavior where users shortly visit more web pages to check whether they contain the current task's answer before returning to the SERP if the answer was not found. The absolute number of transitions[5] and the state durations (cf. analysis in Sect. 5.1.5) fortify that observation.

A further difference between Fact and Expl is the strong interaction between a web page and the quiz tab during the exploration. In Expl, users switch more likely to the quiz tab to collect resp. note found information to proceed and solve the ES task whereby in Fact, users mostly visited the quiz tab to submit the answer for the current task and read the next question. This behavior is also reflected by a selection of typical example sequences in the following:

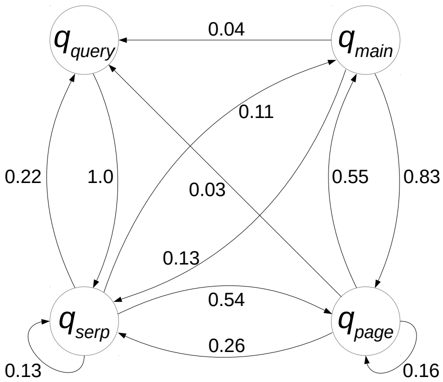
- *Fact* (example): $S^k = q_{main} - q_{qery} - q_{serp} - q_{page} - q_{serp} - q_{query} - q_{serp} - q_{page} - q_{main}\ (Answer) - q_{serp} - q_{query} - q_{serp} - q_{query} - q_{serp} - q_{page} - q_{main}\ (Answer) -\ \ldots$

- *Expl* (example): $S^k = q_{main} - q_{qery} - q_{serp} - q_{page} - q_{main} - q_{page} - q_{main} - q_{page} - q_{main} - q_{page} - q_{serp} - q_{query} - q_{serp} - q_{page} - q_{main} - q_{page} - q_{serp} - q_{query} - q_{serp} - q_{page} - q_{main} -\ \ldots$

Finally, the reason for the low transition probabilities from q_{main} to q_{query} again is caused by the observation that users tend to select

4 Also the absolute numbers show that in Fact the query state q_{query} was reached 347 times (236 times from q_{serp}) and in both ES session Expl_1 and Expl_2 together q_{query} was reached only 173 times (122 times from q_{serp}).

5 In Fact the *SERP* state q_{serp} was reached 251 from q_{page} and in both ES session Expl_1 and Expl_2 together q_{serp} was reached only 201 times from q_{page}.



$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.22 & 0.13 & 0.54 & 0.11 \\ 0.03 & 0.26 & 0.16 & 0.55 \\ 0.04 & 0.13 & 0.83 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure 5.3: Illustration of the 1st-order Markov model θ^{Expl} as graph (left) and it's corresponding components Q, A and π (right) trained on the data set from *US-II* using all interactions of Expl_1 and Expl_2.

previously used SERPs and enter a (new) query instead of first opening a new (blank) start page to type the query. This behavior can be observed in Fact and Expl. The *hub and spoke* search pattern can also be found in both search activities but is less dominant in Expl. This can be an artifact of the quiz tab where users in Expl are rather encouraged to collect content for the search task's answer.

5.1.4 *Hidden Markov Model Definition*

Representing search interactions by more complex states

The interaction with a search system represented by a Markov model, as applied in the former two sub-section, is considered as explicit user behavior because users usually are aware of the single interaction steps which also mostly require a thought-out action. In addition, the search system can also be able to identify further behavioral user aspects if the interactions are logged accordingly, resp. if the system is equipped with corresp. devices. For example, a search session record can additionally contain information about the duration users spend on certain states, the times of scrolling but also physical parameters such as gaze behavior can be used to enrich the session data and hence, to refine the modeling. This kind of (physical) user (re-)actions are less tangible for the users and are considered as implicit user behavior. The contribution of such implicit user interactions in terms of potential search parameters can facilitate the improvement for the search activity classification and can allow to gain more insight into ES. Since several of this parameters can occur at the same time and while the user is in a certain (search) state, this implicit behavior usually is modeled via emissions corresponding to the states. That makes a Markov model a Hidden Markov Model (HMM) which are defined for the given data sets and compared in the following.

Since HMMs basically extend Markov models, HMMs have the same states space Q as well as the same transition- and start probabilities. Therefore, corresponding models, such as θ^{Fact} and θ^{Expl}, can be trained. However, the difference is that states are also able to emit feature values. The features have been introduced as observations[6] in Section 3.3.2 and can represent implicit seeking behavior, e.g., the duration a user spend on a certain state. This extends the (state) sequence S (from the Markov models) by the (feature) sequence X and forms the combined data set $\underline{Z} = (Z^1, ..., Z^K)$ containing K paired sequences $Z^k = (X^k, S^k)$. That is, for each search sequence $S^k = (S_1^k, ..., S_{L^k}^k)$, an associated sequence of features $X^k = (X_1^k, ..., X_{L^k}^k)$ is given. That is, $Z^k = (Z_1^k, ..., Z_{L^k}^k)$ where each $Z_i^k = (X_i^k, S_i^k)$. All sequences S^k, X^k and hence, Z^k have the same length L_k. Analogue to the Markov models, the *joint distribution* for $\underline{Z}$ can be formed:

$$
\begin{aligned}
P(\underline{Z}) &= \prod_{k=1}^{K} P(Z^k) = \prod_{k=1}^{K} P(Z_1^k, ..., Z_{L_k}^k) \\
&= \prod_{k=1}^{K} P(Z_1^k) \times \prod_{l=2}^{L^k} P(Z_l^k | Z_1^k, ..., Z_{l-1}^k)
\end{aligned}
\tag{5.2}
$$

Applying the assumption that each feature (set) only depends on the state it is emitted, results into Eq. (3.18) (see Sect. 3.3.1, p. 68) for each $P(X_l^k | S_l^k)$ and to Eq. (3.19) (see Sect. 3.3.1, p. 68) for each $P(Z^k)$. Finally, applying the *Markov Assumption* for the data set $\underline{Z}$ results exactly into Eq. (3.20) (see Sect. 3.3.1, p. 68) for each paired search session sequence Z^k.

To model the feature's emission, different approaches can be applied. A common solution is to approximate the emission by the Probability Density Function (PDF) of the best fitting distribution[7] regarding the feature that should be modeled. An alternative solution is to estimate the probability of an emission by a HMM itself. That is, HMMs can be nested into each other. However, in the following, the first approach using a PDF will be used to illustrate the feature embedding into the HMMs of $\underline{Z}$. From the interaction data, logged by the web browser add-on in *US-II*, the following features for each state of the state space Q could be derived[8]:

- *Q.Duration*: A user is visiting a state for a specific time interval that is measured in seconds.

6 The difference in the naming of "features" resp. "observations" results from manner HMMs are applied here. Originally, in a HMM the states are hidden, i.e., not observable whereas the emissions can be observed. The way HMMs are applied in this thesis does not have this difference because the states here are known as well and therefore, the naming of states as "features" is more appropriate.

7 The process to identify the most matching distribution for a given set of feature values is called *goodness-of-fit* test.

8 For the analysis of further (implicit) features see Sect. 5.2.4, p. 134.

- *Q.Scrolling*: Accumulation of the user's scrolling time within a state measured in milliseconds.

Exemplary, the feature of state duration *Q.Duration* shall be modeled. Generating a histogram for durations regarding each state in Q including all search sessions, i.e., ignoring the labels of the search activities Fact and Expl, leads to the diagrams illustrated in Fig. 5.4. The diagrams already allow to derive some insights into the search behavior in general. For example, it can be seen that users spend often only a very short time on the query state. The probability that users enter a query in less than about five seconds is relatively high and drops down afterwards. The mean dwell time in q_{query} is 5.5 sec. According to the Cumulative Density Function (CDF) of the approximated exponential distribution, only 16.2% of all query formulation have been longer than ten sec. With a mean of 6.5 sec., users spend more time on SERPs but 78.5% spend less than ten sec. and only 4.6% of the users spend more than 20 sec. on SERPs according to the CDF. In contrast to that and as expected, web pages and the quiz tab are visited longer by the user. The duration feature for web pages reveals a wide range. Several web pages are examined really shortly (about 15.5% for three or less sec.), but the remaining visits take longer until

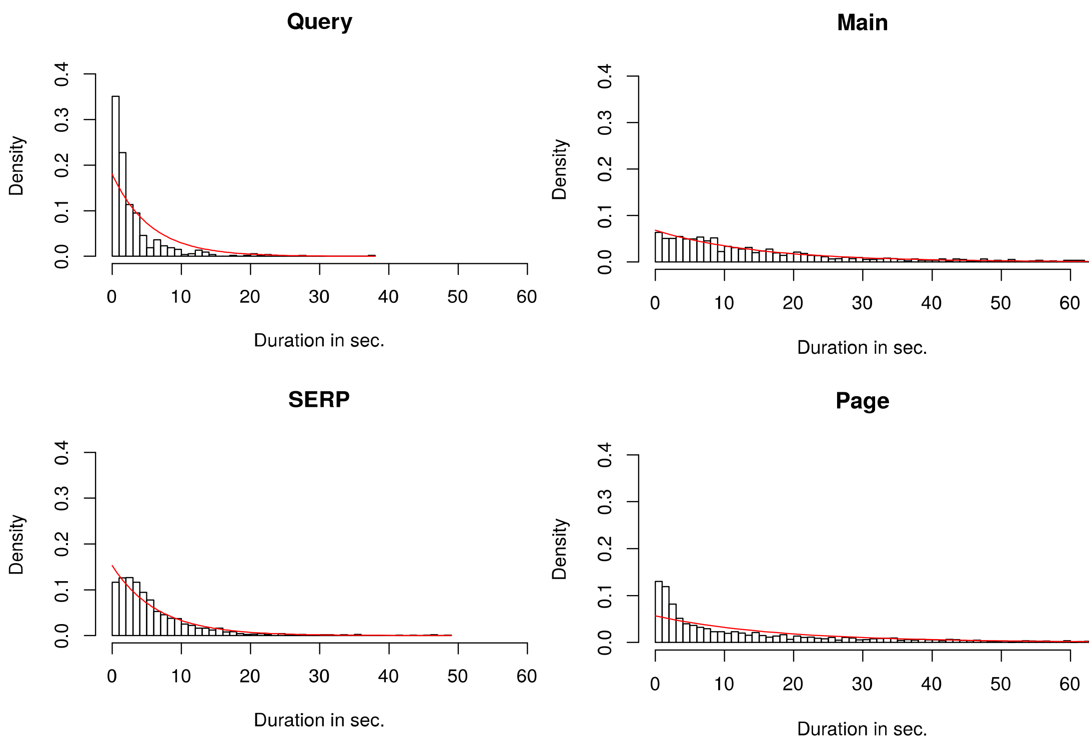


Figure 5.4: Histograms of the empirical state durations with the estimated exponential distribution in red for all search activities from *US-II* together.

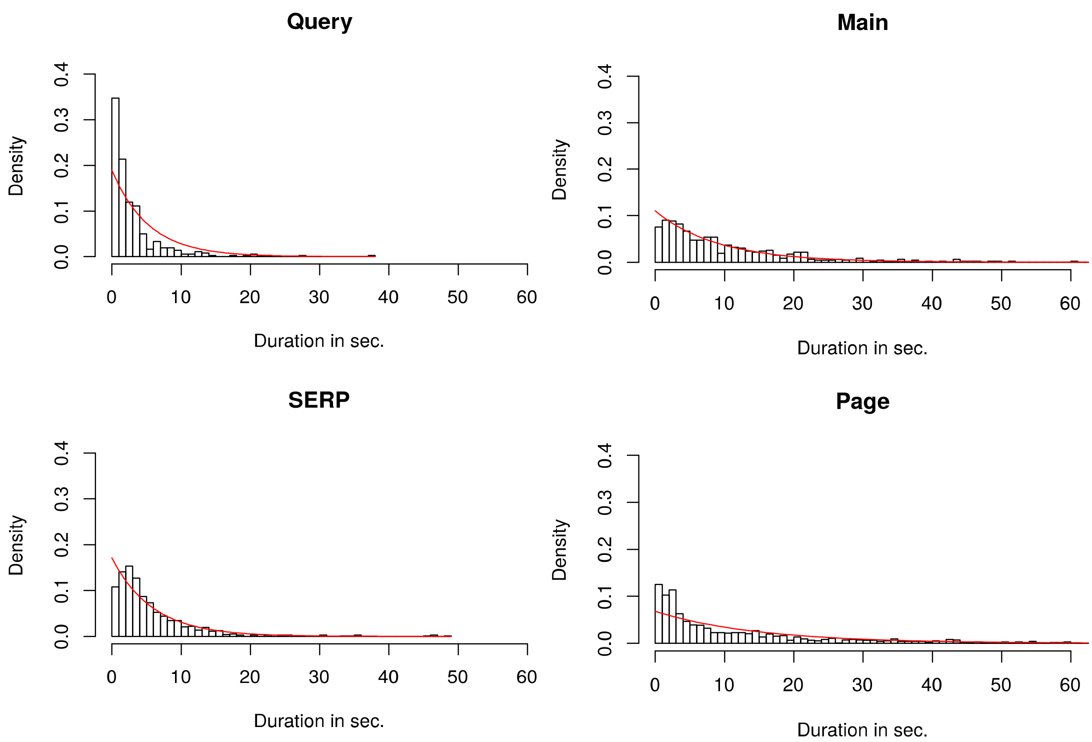


Figure 5.5: Histograms of the empirical state durations with the estimated exponential distribution in red for the search activities from Fact.

a maximum of 5.8 min., measured in *US-II*. The mean dwell time on q_{page} is 17.7 sec. and on q_{main} is 14.7 sec. Although the mean on the quiz tab is lower than for web pages, there is no set of short time visits on q_{main}. That can be explained by the content, provided in this state because in both search activities, the current search tasks had to be read (and understood) and even after a quick answer submission (in Fact), the new task was presented immediately and the state was not changed in the meanwhile. However, a differentiation between the durations in the states regarding the two search activities can reveal more details and is a necessary element anyway for the modeling (and classification afterwards).

5.1.5 *Hidden Markov Model Parameter Comparison*

Differences in search activities considering more complex states

Splitting each full search session (of approx. one hour) of each participant from user study *US-II* into sessions according to Fact, Expl_1 and Expl_2 (approx. 20 min.), and model the feature *Q.Duration* for each state of the two corresp. search activities, results into the diagrams illustrated in Fig. 5.5 and Fig. 5.6. Furthermore, the mean state durations regarding the search activities (and the combination) are listed in Tab.5.1. Comparing the mean durations between Fact and Expl shows that their is no huge difference in the time users formulate queries and

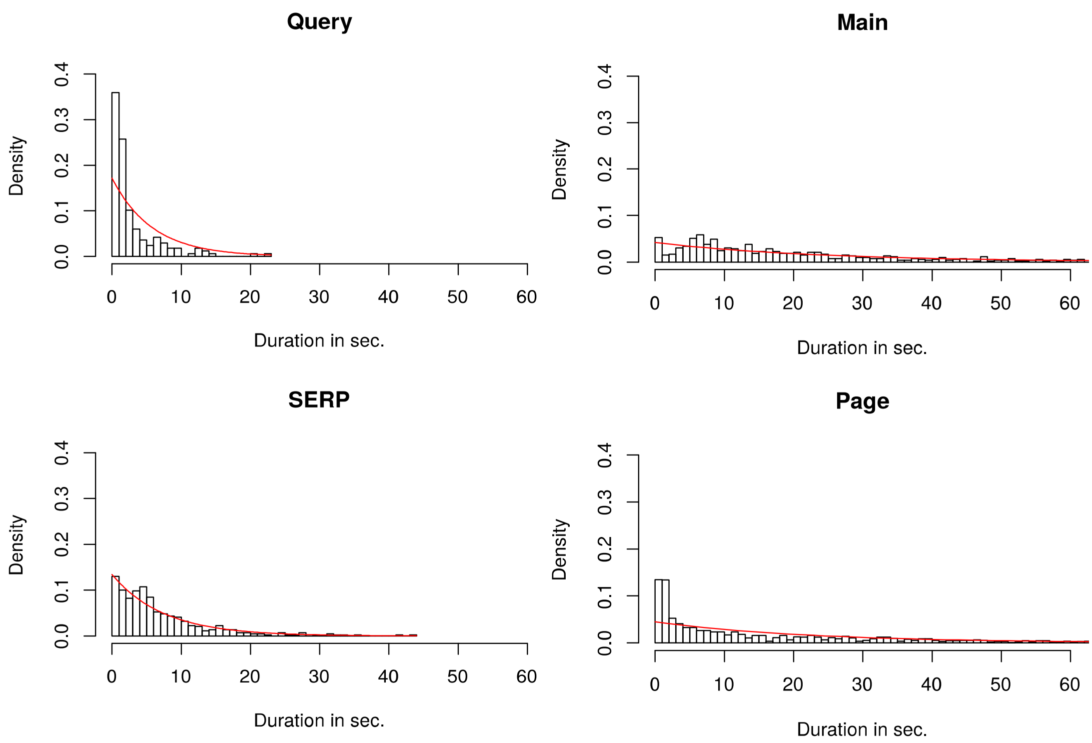


Figure 5.6: Histograms of the empirical state durations with the estimated exponential distribution in red for the search activities from Expl.

also in SERPs users spend on average only about two seconds longer in ES sessions. However, while in Fact only 19.4% of the users spend longer than ten sec. in SERPs, in Expl this number is 29% according to the CDF. As discussed in the Markov model parameter comparison above, this can be an indicator for the more complex ES tasks which require the users to examine the result pages longer and more exhaustive to estimate relevant results.

Continuing the comparison, it shows that the mean duration to spend on q_{page} and q_{main} largely differ. In the quiz tab (q_{main}), 71.7% of the visits in Fact have been ten or less sec. Hence, even answering the current factual tasks (e.g. by copy paste) and reading the next task was done often within this short time frame. In contrast to that,

Table 5.1: Mean dwell times on each state from Q in sec. according to the search activities Fact and Expl as well as the combination Fact + Expl, i.e., all search sessions together, from *US-II*.

	Query	*SERP*	*Page*	*Main*
Fact	4.6	6.1	16.1	7.9
Expl	5.9	8.1	25.1	24.9
Fact + Expl	5.5	6.5	17.7	14.7

on Expl only about one third (33.0%) of the visits of the quiz tab was shorter than ten sec. Furthermore, even 44.7% of the visits of q_{main} have been longer than 20 sec. This can be explained by the longer search task to be read in Expl on the one hand and the users more extensive actions of writing and (reordering) notes to solve the ES tasks on the other hand. Web pages in Expl are examined about nine sec. longer in average than in Fact. As implied by the analysis of the transition probabilities (cf. Sect. 5.1.3), a reason for this difference can be that users who are searching for a specific, well-formulated fact only need to fast-scan a web page to estimate, whether the sought information is present or not. For ES tasks, it can be unclear what information is useful and therefore, more time is necessary to examine a web page and/or actually read the content on the page. This observations also confirms the results of Athukorala et al. [10] that users spend more time reading documents, respectively web pages in ES.

Analyzing the feature of durations gave interesting insight into the differences between Fact and Expl. In average, users spend more time in every state of Q in Expl. In the next Section 5.2, the here defined models are used to classify the search activities Fact and Expl. In addition to the duration, also further features as implicit interactions are used and their contribution to increase the classification accuracy is evaluated.

5.2 SEARCH ACTIVITY CLASSIFICATION

In the former Section 5.1 it was shown, how a given data set $\underline{S}$ (resp. $\underline{Z}$) comprising of samples representing different search activities, can be used, to learn a generative (Hidden) Markov model θ^c with $c \in C = \{Fact, Expl\}$ for each search activity. Using the *Bayes Theorem* in combination with the sequence model from Eq. (5.1), respectively Eq. (5.2), a probabilistic classifier can be defined by:

$$P(\theta^c|S) = \frac{P(S|\theta^c) \times P(\theta^c)}{P(S)} = \frac{P(S|\theta^c) \times P(\theta^c)}{\sum_{c' \in C} P(S|\theta^{c'}) \times P(\theta^{c'})} \tag{5.3}$$

Here, $P(S|\theta^c)$ is the *likelihood* of the search session S being generated during a given search activity c, respectively by trained model θ^c. Furthermore, $P(S|\theta^c)$ corresponds to Eq. (3.9) (cf. Sect. 3.3.1, p. 64) for each S^k, respectively to Eq. (3.20) (cf. Sect. 3.3.1, p. 68) for each Z^k. The $P(\theta^c)$ represents the *prior* associated to the model θ^c and is used to handle possible class imbalances, which is given, e.g., in *US-II*, by one third of Fact and two third of Expl search sessions. The $P(S)$ in Eq. (5.3) normalizes the classification to obtain proper probabilities between zero and one for $P(\theta^c|S)$.

Since the calculation of $P(S|\theta^c)$ and hence, also $P(S)$ for the normalization, requires a multiplication of many probabilities[9], the cal-

9 To calculate $P(S|\theta^c)$, the number of multiplied probabilities at least corresponds to the number of states in the sequence but can be more if emissions in HMM are included.

culations can run quickly into inaccuracies and numerical problems. A common workaround is to transform each probability into it's logarithm (called *log-space*) beforehand, what (1) transforms the multiplication into the simpler computable addition and (2) preserves the accuracy. Unfortunately, this procedure can not be applied for the denominator in Eq. (5.3) because of the sum. However, the probabilistic classifier can be implemented nevertheless because a comparison between the nominators according to the two search activities is sufficient. That is, $P(\theta^c|S)$ is implemented by $P(S|\theta^c) \times P(\theta^c)$ with:

$$class(S) = \begin{cases} Expl & \text{if } P(S|\theta^{Expl}) \times P(\theta^{Expl}) \geq P(S|\theta^{Fact}) \times P(\theta^{Fact}) \\ Fact & \text{otherwise} \end{cases} \tag{5.4}$$

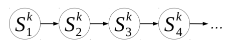
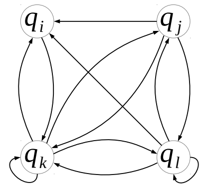
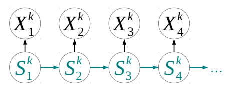
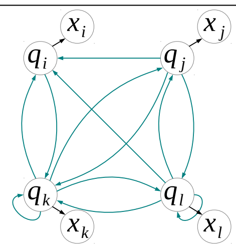
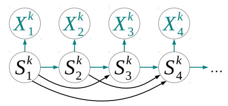
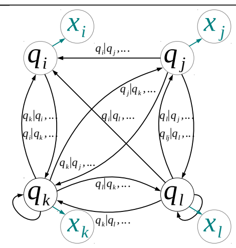
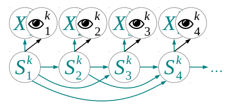
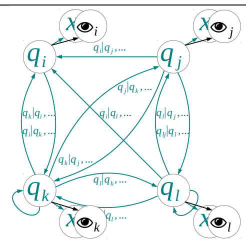
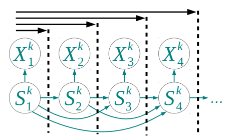
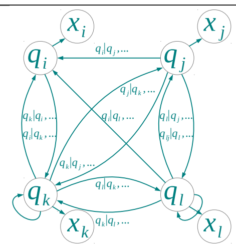
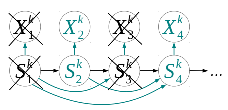
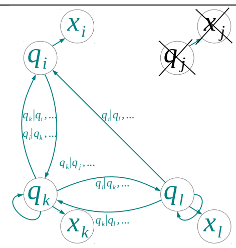
In the following six sub-sections (Sect. 5.2.1 to 5.2.6), this classification approach will be used in a successive procedure to improve the models step by step and to describe the corresponding methodology in detail. In Tab. 5.2 the successive procedure is illustrated, whereby executed (and retained) model adaptions of the previous steps are drawn in teal. The first column lists the corresp. sub-section. Column two represents the user's interactions S_i^k from the recordings. The corresp. models with their states from Q are illustrated in the last column. At first, the trained Markov models θ^{Fact} and θ^{Expl} are used in a cross validation setting to identify their classification accuracy (Sect. 5.2.1). Afterwards, incorporated additional features in terms of HMMs are used to examine their influence to the classification rate (Sect. 5.2.2). In addition to the implicit interactions, different model complexities will be used as well, i.e., the models are trained on different orders and the optimal settings are selected to proceed (Sect. 5.2.3). The set of considered feature can also be extended by recordings of additional devices, what is exemplary shown on the users' gaze behavior (Sect. 5.2.4) utilizing a methodology to select promising features. To advance the goal of providing appropriated but also timely user support, the minimal length of interactions sequences, that still provides a sufficient classification accuracy, is investigated as well (Sect. 5.2.5). Finally, to increase the generalizability of the findings, the artifacts associated to the experimental design, namely the state q_{main} (the *quiz tab*) is removed (Sect. 5.2.6).

5.2.1 *1st-order Markov Models*

Classifying search activities using Markov models with simple states

To analyze the ability to classify search activities, at first the defined Markov models θ^{Fact} and θ^{Expl} as in Fig. 5.2 and Fig. 5.3 are trained and used in a 5-fold cross validation with 2000 repetitions. That is, for each of the search activities Fact and Expl, the data set $\underline{S}$ from *US-II* is

Table 5.2: Illustration of the successive classification procedure.

Sub-Section	Interactions	Model
1st-order Markov Models 5.2.1		
1st-order HMMs 5.2.2		
Higher order HMMs 5.2.3		
Selected Gaze Features 5.2.4		
Minimal Interaction Sequence Length 5.2.5		
Ignoring Artificial States 5.2.6		

randomly divided into five sub sets and four sub sets are building the training and the remaining sub set is the test set. The size of the training set is approximately 4/5 of the data set, i.e., it comprises 15 of 19 fact-based multitasking search sessions from Fact and 30 of 38 exploratory search sessions from Expl. The training set is used to learn the parameters A and π for the two models θ^{Fact} and θ^{Exlp}. Afterwards, the remaining search sessions from the test set and the trained models are used in the classifier of Eq. (5.3). Afterwards, the next of the five sub sets is declared as test and the remaining sub sets are building the training set until each permutations was used. According to the two classes in the data set, the prior $P(\theta^C)$ was set to the relative class sizes respectively. The whole procedure was repeated 2000 times and afterwards, the median for the classification results is calculated, cf. Tab. 5.3. According to the table, the total classification rate is 73.6%.

5.2.2 *1st-order Hidden Markov Models*

Classifying search activities using Markov models with more complex states

As defined in the previous Section 5.1.4, HMMs extend the Markov models by including additional features modeled as emissions. This approach was exemplary shown on the dwell time users spend on the several states of Q. In the following, HMMs are used and the feature selection is motivated by the results of Athukorala et al. [10] which state that the parameters state duration and scrolling time might help to distinguish between lookup and exploratory tasks. The models are trained on the data set $\underline{Z}$ from *US-II* and the same cross validation procedure as in the former sub-section is applied. The classification results using the HMMs with state duration or scrolling time are depicted in Tab. 5.4 and Tab. 5.5. Using the feature scrolling, the classification rate for Fact is nearly identical to the Markov models before and for Expl only one more search session is classified correctly and leads to a total classification rate of 75.4%. Using the feature state duration, the total classification rate is increased to 89.4% with 16 correctly classified Fact and 35 correctly classified Expl sessions. If both features are used simultaneously, the classification rate drops down to a value comparable to using only scrolling as feature. A reason for this is the fact that in HMMs, as used here, all features are considered to be independent. However, the time for scrolling is always indirectly

Table 5.3: Confusion Matrix for the classification of the 1st-order Markov models regarding factual and exploratory search activities trained on the data from *US-II*.

	Real Class					
	Absolute		Relative		Relative (to class)	
	Fact	*Expl*	*Fact*	*Expl*	*Fact*	*Expl*
Classified as *Fact*	13	9	.23	.16	.68	.24
Classified as *Expl*	6	29	.10	.51	.32	.76
	19	38	1		1	1

Table 5.4: Confusion Matrix for classification using 1st-order HMMs regarding factual and exploratory search activities trained on the data from *US-II*. As emission the state duration feature was used.

	Real Class					
	Absolute		Relative		Relative (to class)	
	Fact	*Expl*	*Fact*	*Expl*	*Fact*	*Expl*
Classified as *Fact*	16	3	.28	.05	.84	.08
Classified as *Expl*	3	35	.05	.62	.16	.92
	19	38	1		1	1

contained in the state duration time and therefore the features are correlated and actually not independent. Therefore, the violation of this independence assumption might lead to increased classification errors, i.e., lowers the discriminative power of the duration.

5.2.3 *Higher order Hidden Markov Models*

Increasing the considered interaction history

The two previous classification approaches considered (Hidden) Markov Models on the first order. That is, the probability to reach a state only depends on it's direct predecessor state, cf. Eq. (3.8) et seq. To improve the classification, respectively prediction accuracy of (Hidden) Markov Models, a further possibility is to increase the model's considered state history up to a specific length, i.e., to increase the model's order. The choice of the optimal order depends on several aspects, such as the model's performance, it's complexity but also on the size of data set because the higher the order, the longer the necessary interaction sequences and the lower the probability that these interaction sequences are present in a sufficient number in the data set. In the following, two complementary approaches for model selection will be described and utilized to identify the optimal models (in terms of model's order) in order to further increase the accuracy in classifying the two search activities Fact and Expl regarding the data of *US-II*.

Table 5.5: Confusion Matrix for classification using 1st-order HMMs regarding factual and exploratory search activities trained on the data from *US-II*. As emission the state scrolling feature was used.

	Real Class					
	Absolute		Relative		Relative (to class)	
	Fact	*Expl*	*Fact*	*Expl*	*Fact*	*Expl*
Classified as *Fact*	13	8	.23	.14	.68	.21
Classified as *Expl*	6	30	.10	.53	.32	.79
	19	38	1		1	1

5.2.3.1 *Model Selection using ML and MAP Estimated Accuracy Box-Plots*

Selecting the best models: Approach 1

In Fig. 5.7 a box-plot comparison of the model's performance in terms of accuracy of different orders using a simple Maximum Likelihood (ML) estimator and a more advanced Maximum Aposteriori (MAP) estimator is illustrated. For this comparison, the HMMs from the previous sub-section with the highest accuracies, i.e., incorp. only the duration feature, are used. The ML estimator uses the Eq. (5.3) as described in the beginning of this Sect. 5.2. That is, essentially the box-plot for the ML estimator (for the 1st order) illustrates the cross validation process from the previous sub-section with a mean accuracy of 89.4%. The MAP estimator, in contrast, uses an additional dampening factor to make the classification more robust. This is achieved by extending the fraction in Eq. (5.3) by the fraction x/x, i.e., by one. Fig. 5.7 shows that the classification rate for the first order is comparable for ML and MAP, namely on 89.4%. For the higher model orders, the accuracy differs in comparison of the two estimation methods. The accuracy with the ML estimator decreases with each order. The reason for this is the lack of long interaction sequences in the data set. For the MAP estimator, a similar behavior can be observed but with a less crucial effect on the accuracy. Here the second order model has the highest accuracy with 92.1% in average with a relatively small variance. That is, according to the box-plots in the diagrams of Fig. 5.7, the order two for the HMMs incorporating (only) the duration as feature is the best choice for the data set from *US-II*.

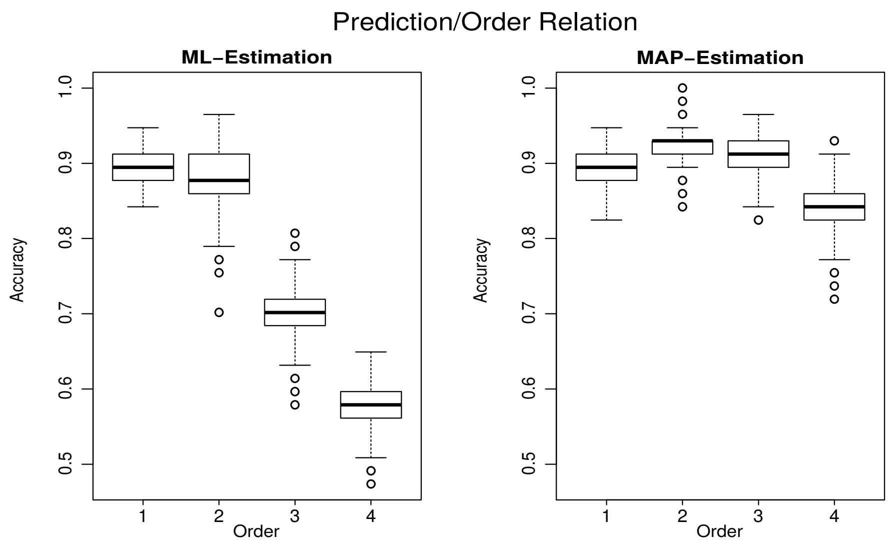


Figure 5.7: Accuracy comparison of different Markov model orders using a Maximum Likelihood (ML) and a Maximum Aposteriori (MAP) estimator.

5.2.3.2 *Model Selection using AIC and BIC*

Selecting the best models: Approach 2

Now, a further model selection approach will be applied by using the so called Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). In addition to the pure consideration of the model's performance, as applied before, this two information theoretic measures also consider the model's complexity and penalize too complex models if they lack for a corresponding performance enhancement. That is, AIC and BIC apply a trade-off approach that, e.g., prevents the model parameterisation for overfitting. Although AIC and BIC provide an objective and reliable approach of model selection, they are seldom applied, at least in the area of (search) user behavior modeling, with exception of Han et al. [83]. In the following, both measures are defined on the base of the Kullback-Leibler Information and applied to the given modeling task:

A model is selected as "the best model" if it's quality of classification, respectively prediction, is as close as possible to the empirical samples for the given set of candidate models, while penalizing it's complexity. The Kullback-Leibler Information (KLI) is a measure between a conceptual reality f and an approximating model g and is defined for continuous functions as the integral:

$$I(f,g) = \int f(x) \times ln \frac{f(x)}{g(x|\theta)} \, dx. \tag{5.5}$$

Here, f and g are n-dimensional probability distributions [32]. Akaike [5, 6] found a formal relationship between the KLI and likelihood theory [51]. This relation is defined with an estimator of expected, relative KLI based on the maximized log-likelihood function $ln(\mathcal{L})$, corrected for an asymptotic bias P. Here, P is specified as the number of independently adjusted parameters estimated by the model. The AIC is defined as:

$$AIC = -2 \times ln(\mathcal{L}) + 2 \times P \tag{5.6}$$

In contrast to the AIC, the BIC is derived within a Bayesian framework as an estimate of the Bayes factor for two competing models [1]. The BIC is defined as:

$$BIC = -2 \times ln(\mathcal{L}) + ln(K) \times P \tag{5.7}$$

The difference between both information criteria is only the coefficient multiplied with the number P of parameters. The BIC will penalize complex models harder for large K. According to Kadane [107], models chosen by BIC will be more parsimonious than those chosen by AIC in general. To apply the AIC and BIC on the given task, the

Table 5.6: Model Selection using AIC and BIC applied to Markov models trained on the factual and exploratory search activities from the data of *Us-II*.

Markov Model Order	P	AIC AIC_{Fact}	AIC AIC_{Expl}	BIC BIC_{Fact}	BIC BIC_{Expl}
1	20	5042.78	3995.37	5046.95	4009.94
2	84	**4708.33**	**3920.90**	**4725.84**	**3982.07**
3	340	5033.79	4313.11	5104.62	4560.70
4	1364	6774.96	6064.45	7059.12	7057.70

following has to be noted: The emissions of HMMs are dependent to the underlying Markov model and it's states. However, the Markov model itself is independent of the emissions what decouples the selection of the best model order from the emission model. That is, for AIC and BIC, the Markov models as defined in Sect. 5.1.2 can be used. For the parameter $\mathcal{L}$, both information criteria require the maximized likelihood of $P(\underline{S})$ what is obtained by utilizing all samples (i.e., all search sequences) for the model training. The number P of parameters, estimated by the model, results from all possible transition probabilities. For example, according to the state space Q, a 1st-order Markov model has 4 start probabilities π_i and afterwards there are 4^2 state transitions from a_{ij} possible what sums up to 20. A 2nd-order Markov model additionally has the transition to one of 4 states that depends on a two state history, i.e., a 2nd-order Markov model additionally has 4^3 parameters what sums up to $4 + 4^2 + 4^3 = 84$ parameters. Finally, $P(order) = \sum_{i=0}^{order} |Q|^{(i+1)}$. In Tab. 5.6 the results for AIC and BIC on different orders are listed. Since a high (maximized) likelihood corresponds to values close to zero and because a high model complexity (high number of parameters) is penalized, the minimal values for AIC and BIC indicate the best model(s). Table 5.6 shows that both criteria have their minimal values on the second order for Fact and Expl. In all settings, the second best model has a noticeable difference to the best model, what implies only little support for the alternatives. Only the difference between the BIC_{Expl} models on order 1 and 2 is relatively small, indicating the 1st-order also as an eligible alternative solution.

Although the optimal order can be estimated without the emissions (see argumentation above), for the sake of completeness, the AIC and BIC shall also be applied for the HMMs incorporating only the duration feature. The results are listed in Tab. 5.7. If HMMs are used, the number of estimated parameters for the emissions also have to be taken into account for the calculation of P. Since the emissions are given by the corresponding probability distribution, the parameters of the distribution have to be added. That is, for the exponential distribution, used for the HMMs in this section, P is increased by one (the λ parameter)

Table 5.7: Model Selection using AIC and BIC applied to HMMs trained on the factual and exploratory search activities from the data of *Us-II*.

Hidden Markov Model Order	P	AIC AIC_{Fact}	AIC_{Expl}	BIC BIC_{Fact}	BIC_{Expl}
1	24	16989.24	16847.84	17011.91	**16887.15**
2	88	**16832.11**	**16832.11**	**16778.10**	16976.22
3	344	17114.26	17312.51	17474.48	17875.84
4	1368	19054.20	19247.74	20346.19	21487.95

for each state and therefore, $P(order) = \sum_{i=0}^{order} |Q|^{(i+1)} + |Q|$. Table 5.7 shows a picture similar to Tab. 5.6 with one exception. While the AIC identifies the second order as best choice again, the BIC prefers the first order for Expl. This tendency was already shown in Tab. 5.6. Nevertheless, according to the analysis with AIC and BIC and considering the results from the MAP estimated accuracies, the second order is identified as best for the (Hidden) Markov models with an mean classification accuracy of 92.1%.

5.2.4 *Selected Gaze Features*

Selecting the best features

In addition to the interaction data, logged by the web browser add-on in *US-II*, also the user's gaze behavior was recorded by an eye-tracking device (cf. Sect. 4.3.4). This data provides additional implicit user interaction which can be used as emitted state features. However, the selection process for the features should not be done randomly, since (1) features that have no discriminative power just increase the models' complexity without increasing the classification rate and (2) strongly correlated features violate the independence assumption and can cause classification errors, as shown in the case of 1st-order HMMs with the features *Q.Duration* vs. *Q.Scrolling* (cf. Sect. 5.2.2, p. 129).

In the following, the first aspect (1) is addressed by identifying features with high discriminative power regarding the classification of the two search activities Fact and Expl. Afterwards, the identified features are used in the familiar classification setting to address the second aspect (2) and to reveal the best feature combination for the given task. For the feature selection procedure in this sub-section, the following three gaze features for each state of Q have been derived from the Eye-Tracker logs:

- *Q.Fixation*[10]: The number of occurrences of fixations during the visit of the state.

10 In the eye-tracking domain, eye movement is defined by the two main types *Fixations* and *Saccades*. In a fixation, the (user's) eyes stop the movement to scan a certain scene. This allows the visual system to perceive and process (i.e., extract) detailed information about the scene (e.g., a web page) what is looked at. Fixations are the most

- *Q.FixDuration*: A user's accumulated duration of fixations during a state measured in milliseconds.
- *Q.FixDurMean*: A user's average duration of fixations while being in a state.

The three features represent quantities or durations in continuous time units on the two search activity classes. Without further assumptions about the data, a two sample Kolmogorow-Smirnow Test (KST) can be applied to test for the *null hypothesis*:

> H_0: Two feature samples of *Q.Fixation*, *Q.FixDuration* or *Q.FixDurMean* from the search activities Fact and Expl respectively and on the same state $q \in Q$ are representations of the same distribution.

The purpose of this procedure is to identify features with a limited discriminative power under H_0. That is, each feature can be associated to a p-value and the selection of potential relevant resp. useful features can be done using a threshold on a certain significance level α. In other words, if H_0 has to be rejected (because of a low p-value smaller than α), their is no proof that the two samples are not from different distributions and hence, are potentially useful for the classification. This procedure is used by many variable selection algorithms that include variable ranking as a principal or auxiliary selection mechanism because of its simplicity, scalability, and good empirical success [80]. Applying a two sampled KST to derive a sub set of potential relevant features by utilizing p-values less than $\alpha = 0.1$ results into Tab. 5.8. The table shows that features associated with fixation can achieve a high discriminative power. In particular on the states *Main* and *Page*, a high significance is reached while this is missing in the states *SERP* and *Query*. This can be an indicator that users receive only limited

used feature in eye-tracking research because they enable to infer cognitive processes. The duration of a fixation is between 50-600 ms. Saccades describe the movement between points of interest, i.e., between fixations. Hence, the eye movement basically is an alternating sequence of fixations and saccades. The duration of a saccade is between 20-40 ms. In this thesis, only fixations are utilized because of their high(er) information content in general.

Table 5.8: Feature Selection based on p-values and an α of 0.1. The p denotes the p-Value. Relevant features are marked bold.

Feature	p	Feature	p	Feature	p
Page.Fixation	**0.0575**	*Page*.FixDuration	**1.7e-06**	*Page*.FixDurMean	0.6262
Main.Fixation	**3.2e-05**	*Main*.FixDuration	**3.1e-06**	*Main*.FixDurMean	0.9054
SERP.Fixation	0.9956	*SERP*.FixDuration	0.5787	*SERP*.FixDurMean	0.1281
Query.Fixation	0.9973	*Query*.FixDuration	0.3647	*Query*.FixDurMean	0.6664

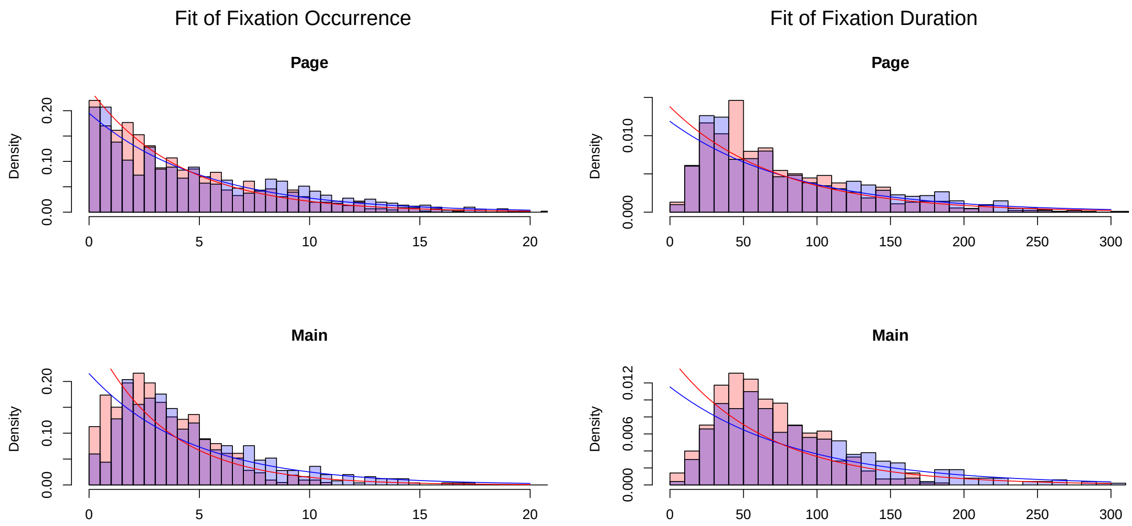


Figure 5.8: Histograms of the empirical state distribution of the four potential relevant features from *US-II*. Lines indicate the estimated exponential distribution in red for Fact and in blue for Expl.

support from search engines (regarding the engine's associated states *SERP* and *Query*) in case of different search activity.

All four potential relevant gaze features are illustrated in Fig. 5.8 by their histograms. Furthermore, for each histogram the fitted emission resp. observation model in terms of the exponential distribution is plotted. For the feature *Q.Fixation*, (the fixation occurrences) a transformation using the square root was done for two reasons: First, a better fit for the distribution was obtained what actually makes the feature *Q.Fixation*: $\sqrt[2]{Q.Fixation}$. Second, the visual inspection of the data was easier because more extreme values shrunk tighter in scale. Similar to the duration feature (*Q.Duration*), in the state *Page*, few fixations are more often represented in Fact. In contrast, Expl has more often a higher number of fixations in *Page*, i.e., Expl is heavier tailed. This implies that users might need less fixations in Fact to find relevant information on a web pages and/or users might simply read more on web pages in Expl. The shift between these two modi are centered around $\sqrt[2]{36} = 6$ fixations. In the *Main* state, users have comparably less fixations in Fact, which might indicate a lower complexity of the required tasks and answer submission but certainly is intensified by the shorter search tasks in contrast to Expl. Fixation durations (*Q.FixDuration*) were transformed the same way as the fixation occurrences for the same reasons. The fit of the exponential distribution is obviously not as appropriate as the previous approaches and it is arguable that other fits (e.g. by the gamma distribution) would capture better results with that feature. However, the exponential distribution has the advantage of only one parameter keeping a lower model complexity and therefore, will be further used here. In *Page*, the lower number of fixation durations has no clear difference in the profile but Expl is again heavier tailed. For the *Main* state, Fact is clearly overrepresented in the lower numbers of fixation durations and Expl is more dominant in more durations. The interpretation is equivalent for the former cases.

To complete the feature selection, an additional analysis with the KST on the implicit interaction features *Q.Duration* and *Q.Scrolling* have been tested on H_0 for each state. The feature *Q.Scrolling* turned out to be not relevant because all p-vales have been higher than α (similar to *Q.FixDurMean*). This was not surprising after the analysis on the 1st-order HMMs (cf. Sect. 5.2.2, p. 129). However, *Q.Durations* was confirmed as (highly) relevant because of low p-values in the states *Page* and *Main* (similar to *Q.Fixation* and *Q.FixDuration*) while *Q.Durations* on the states *SERP* and *Query* had a low discriminative power (also similar to *Q.Fixation* and *Q.FixDuration*). As final step of the feature selection, all possible combinations of features that have been identified as relevant, i.e., *Q.Duration*, *Q.Fixation* and *Q.FixDuration*, have been used in the familiar 5-fold cross validation setting with 2000 repetitions on HMMs. The results however showed that using the state durations alone as features still leads to the best classification accuracy. Nevertheless, the other two gaze features *Q.Fixation* and *Q.FixDuration* have a positive contribution in contrast to simple Markov models but never reach the values of *Q.Duration*. This can be explained by the fact that *Q.Fixation* and *Q.FixDuration* still are correlated to the state duration what violates the independence assumption.

5.2.5 *Minimal Interaction Sequence Length*

Classification in time

Providing appropriated user support for EIS also requires being able to classify the current search activity in time and not if the search session is over. While the previous investigations always applied the classification on a full (unknown) search session, here different lengths of interaction sequences will be investigated. That is, the end of each test sequence is cut and only a certain number of interactions from the beginning is taken and classified. This allows to make an estimation of the minimal interaction length where the classifier still reaches relatively high accuracies. As before, the model optimization process is continued. The best models of the previous sub-sections are used, i.e., HMMs with the duration feature on the second order. Again, for the sake of completeness also a comparison for order one, three and four will be given. In Fig. 5.9 the box plots for different interaction sequence lengths on different orders are illustrated. For the sequence length of one, each model makes basically classifications equal to the rate of the prior, cf. Eq. (5.7). The reason is, that each search activity begins on the state *Main* and therefore, only the emission, i.e., duration in that state holds aspects for classification. While the HMMs of the

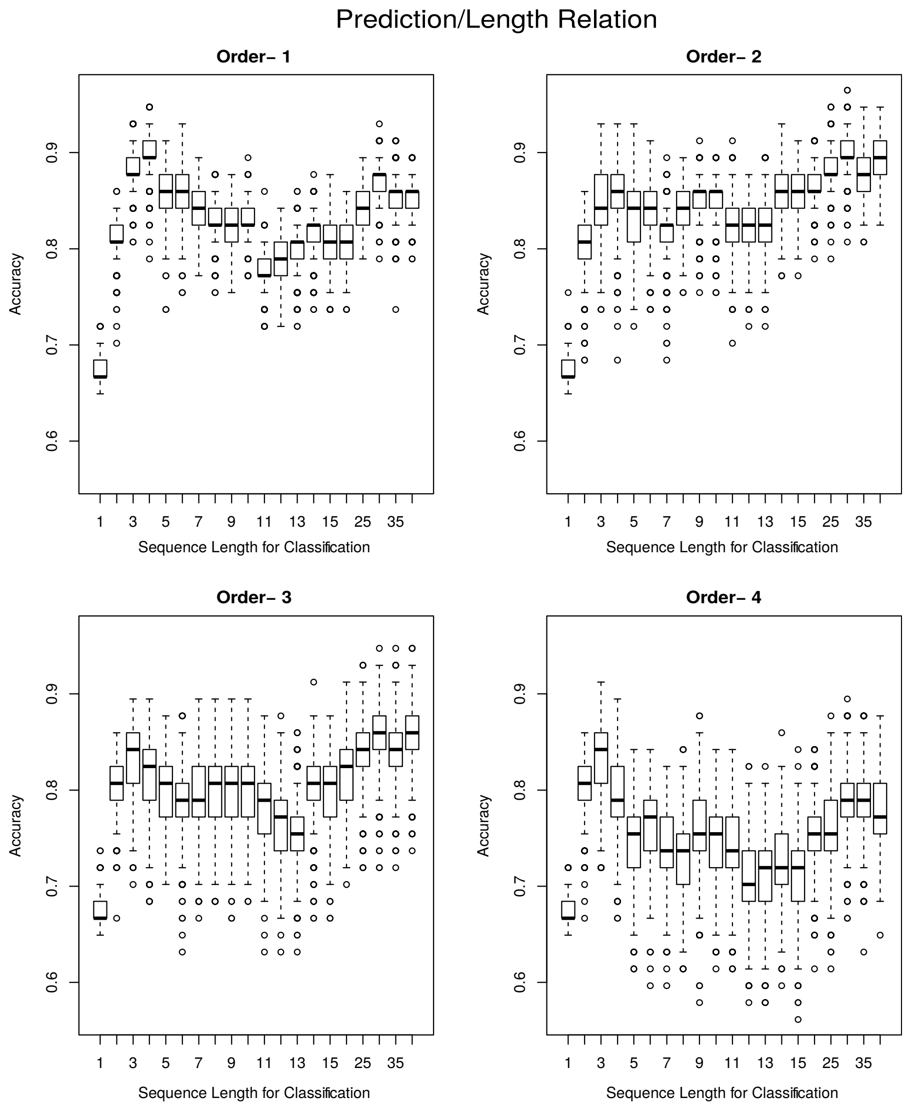


Figure 5.9: Accuracy comparison for the HMMs of order one to four using different interaction sequence lengths. Note that the scale switches at the 15th interaction.

1st-order reach their maximum mean accuracy of 89.9% within the first four interactions, there is a clear drop off under 80% within longer interactions sequence. In contrast to that, the HMMs of the 2nd-order reaches an accuracy of 85.6% within the first 4 interactions but without such a crucial drop off afterwards for longer interaction sequences. A convergence to a stable classification is reached on 30 interactions with a rate of about 89%. The HMM's of the higher orders three and four suffer from the overfitting problem and generate less accurate results. The drop off is observable even though less crucial than in the case of the 1st-order HMMs. Although the predictive power of the models is slightly reduced in the beginning of the search sessions, the diagrams in Fig. 5.9 indicate that the models are capable to discriminate the search activities Expl and Fact early within the search session beginning with acceptable rates after only four interactions.

5.2.6 *Ignoring Artificial States*

Making the classification more realistic

As described in Sect. 4.3.3, the user study design of *US-II* included: the activation of the several search blocks Fact, Expl_1, Expl_2; the presentation of the search tasks and time limits; but also the answer submission. These functions have been implemented by the *quiz tab*, corresponding to the *Main* state q_{Main} in the web browser. This implementation decision had several advantages: First, the users do not need to switch between different media, e.g., a sheet of paper and the screen. Second, users can copy paste question fragments directly to the query field but also can copy paste information from web pages directly to the answer field in the quiz tab, what saves time, can reduce distraction and is a common interaction with computer systems. Third, from the research perspective, the gaze behavior recorded by Eye-Tracker does not have gaps because the users can focus their attention for the whole search experiment on the screen. Nevertheless the quiz tab and the resulting *Main* state represent an artifact of the user study that is not present in a natural interaction with search systems. Hence, to increase the generalizability of the findings, the *Main* state (corresp. to the *quiz tab*) was removed in this sub-section[11]. That is, before the 2nd-order HMMs (using the duration feature) have been trained, in all search session of the data set of *US-II* the *Main* state was removed from the interactions sequences in the following procedure:

- If the *Main* state was at the beginning of the search session (true for all sessions), the state and it's associated feature value (duration) could be erased without further considerations. For example, $S^k = q_{main} - q_{qery} - q_{serp} - q_{page} - \ ...$ became $S^k = q_{qery} - q_{serp} - q_{page} - \ ....$
- If the *Main* state was between two different other states, the *Main* state and it's associated feature value (duration) could also be erased without further considerations. For example, $S^k = \ ... \ - q_{page} - q_{main} - q_{serp} - \ ...$ became $S^k = \ ... \ - q_{page} - q_{serp} - \ ....$
- If *Main* was between the same states, the *main* state and it's associated feature values were erased and the features for the remaining state were accumulated. For example, $S^k = \ ... \ - q_{page} - q_{main} - q_{page} - \ ...$ became $S^k = \ ... \ - q_{page} - \ ...$ and the duration in the remaining state q_{page} became equal to the sum of the two states q_{page} before and after the original *Main* state.

11 Although the classification rate is expected to be reduced, because of the missing information regarding a good discriminative state (and it's duration), the results are more realistic.

The results of the classification process are listed in Tab. 5.9. While the mean classification accuracy for Fact is further increased to 17 (89%), the accuracy for Expl is decreased to 33 (87%). This can be explained by the strong interaction between *Page* and *Main* in Expl that is now reduced to web page states *Page* with longer durations. That is, the search sessions for Expl become more similar to Fact and cause miss classifications. The total classification rate is 87.7%.

Table 5.9: Confusion Matrix for the classification using 2nd-order HMMs regarding factual and exploratory search activities trained on the data from *US-II* but reduced by the *Main* state.

	Real Class					
	Absolute		Relative		Relative (to class)	
	Fact	*Expl*	*Fact*	*Expl*	*Fact*	*Expl*
Classified as *Fact*	17	5	.30	.09	.89	.13
Classified as *Expl*	2	33	.03	.58	.11	.87
	19	38	1		1	1

5.3 IDENTIFYING LATENT SEARCH BEHAVIOR CLUSTERS

The former Section 5.2 focused on the identification of the two search activities Fact and Expl in terms of a classification setting. That is, the class information, namely the labels, have been used to train models which afterwards are able to (correctly) categorize a given (unknown) search session. This procedure implies that a given interaction sequence is either from the one or from the other search activity. Therefore, it do not allow to identify further (not label conform) seeking behavior that may also be shown by the users. In this section, the classification setting is transformed into a cluster setting. That is, the class information is not used during the model training what (1) allows to validate pre-defined search classes from study data and (2) allows to reveal additional search patterns, i.e., latent search behavior.

In the following, the data set of *US-III* is used to train classification models, as implemented in the previous Section 5.2. Afterwards, the models are extended to Mixture Models and used to cluster the users' search behavior. This facilitates to reveal (strong) associations to the already revealed (classified) search behavior (1) and thereby, to validate the experimental setting, respectively the pre-defined search classes. The number of interactions from *US-III* is a multiple in contrast to the number of interactions from *US-II* and is also necessary because the clustering setting requires a sufficiently sized data set. Finally, Mixture Models are used to investigate previous "unknown" search behavior (2).

5.3.1 *Baseline Classification Model Definition*

To validate the pre-defined factual and exploratory search classes, at first baseline models, similar to the known classification setting (cf. Sect. 5.2), have been trained on the data of user study *US-III*. In the following, the baseline model specifications are given while the choices for several model parameters are based on the methodology and the revealed knowledge from the former sections:

Since the *quiz tab*, where users can read and answer the current search task, is considered as an experimental artifact, the corresponding state q_{Main} was not considered in the modeling process. Hence, the state space used here comprises $Q = \{q_{query}, q_{serp}, q_{page}\}$. Since HMMs always have been showed to be superior to the Markov models, HMMs have been used. An analysis of the implicit browser related features (*Q.Duration*, *Q.Scrolling*) and on the eye movement related features (*Q.Fixation*, *Q.FixDuration*, *Q.FixDurMean*) again identified the state duration (alone) as feature with the most discriminative power. That is, according to the feature selection, the HMMs here only incorporate the feature *Q.Duration*. In contrast to the HMMs, as defined in Section 5.1.4, the emissions for the duration here are not modeled by the fitted exponential distribution but by a multinomial distribution over 17 discrete bins for 0, 1, 2, ..., 10, 20, 30, ..., 60 and $>$ 60 seconds. This decision had to be made in advance and in favor of the cluster setting approach described afterwards[12]. The usage of 17 bins for the emissions causes the model to have 16 more state parameters in contrast to a model approach with only one exponential distribution function and it's parameter λ. Therefore, to reduce the model's complexity overall, some adaptations can be made. In particular, two adaptations have been implemented: First, the HMMs here exclusively model the state duration of the *Page* state. That is, the emission on *Query.Duration* and *SERP.Duration* remains "empty" or "silent" (corresp. to a multiplication with 1). According to the KST on the data of *US-III*, the duration feature on q_{query} and q_{serp} had (again) a low driscriminative power anyway. The second adaptation made was the model reduction to the 1st-order. This decreases the number of model parameters as well and on the data of *US-III*, the 1st-order was a supported alternative choice regarding the AIC and BIC anyway. This result is not surprising considering the BIC's tendency for the 1st-order on the data of user study *US-II* at least for Expl (cf. Sect. 5.2.3.2).

12 In the following, an explanation for the usage of discrete bin is given: PDFs provide a probability density in the range from 0 to ∞. To apply emission in HMMs, a probability mass in the range from 0 to 1 is necessary. This can be achieved by using the corresp. CDF of the distribution but requires to specify a δ to calculate the area under the curve in the CDF what can be interpreted as probability. However, in the Expectation Maximization (EM) algorithm, necessary for the cluster setting, the δ can cause mathematical inaccuracies at the distribution border what would cause the EM algorithm to not monotonically increasing likelihoods.

Cf. Data Characteristics in Sect. 4.4.6, p. 111

To reach the goal of this section, the interaction data of *US-III* was used as follows: From the 115 participants, 226 exploratory search sessions could be extracted. This exploratory search sessions result from the two exploratory search tasks Expl_1 and Expl_2[13]. To train the model regarding the exploratory search, the pool of these 226 exploratory search sessions, respectively sequences, have been used. To train the model regarding the factual search, only the interactions from the search block Fact_{NON} have been taken into account. The reason is to avoid biases by the motivational goal conditions in the other blocks. Furthermore, the interactions sequence of each fact-finding task was used separately, not as multitasking search activity. The reason for this decision was threefold: First, this increased the number of (unbiased) fact-finding search instances for the classification and clustering. Second, the ability to differentiate between Fact (as multitasking assignment) and Expl could already be shown in Sect. 5.2 while the focus of this Sect. 5.3 is to identify implicit behavior that may also be induced by individual factual tasks. Third, from the model training perspective, the separation has no huge influence to the model parameters because the state q_{Main} as linking element between two fact-finding searches was not modeled; and the training (of a 1st-order model) is based on two consecutive states which are likewise contained in separated fact-finding tasks. This finally results into a pool of 717 factual search sessions, respectively sequence for the training (and test set).

According to the model and data specifications above, baseline classification HMMs of the 1st-order have been build using the data set $\underline{Z}$ of user study *US-III*. In particular, data set $\underline{Z}$ (for this Sect. 5.3) consists of $K = 717 + 226$ search sequences, indexed by Z^k and associated to the corresponding factual and exploratory search tasks with the individual length L^k. Again, each interaction at index l in a search session Z^k is defined as $Z_l^k = (S_l^k, X_l^k)$, while S_l^k is from the state space $Q = \{q_{query}, q_{serp}, q_{page}\}$ and X_l^k is representing the emission. The only special characteristic here is that the duration for q_{page} is modeled using a multinomial distribution over 17 bins while the emissions on q_{query} and q_{serp} are not modeled, i.e., they remain empty.

5.3.2 *Cluster Model Definition*

To transform the classification setting into a cluster setting, the former used classification approach can be extended, respectively generalized, to form a Finite Mixture Model (FMM) with a finite number of components. In this thesis, each component of a FMM will be a HMM that represents the search behavior induced by the processing of an factual or exploratory search task. The advantage of FMMs is their ability to cluster data when labels are missing. This benefit will be used later

13 As a reminder: Since some users spend a long time on one ES task or just skipped the other, not $2x115 = 230$ but 226 exploratory tasks are available in the data set.

to reveal new (former "unknown") latent search behavior. Nevertheless, FMMs are also able to classify (interaction) sequences if the class information, i.e., the labels, are given what basically results into the classification approach described and already used before. The FMMs used here will be defined in the form of a Mixture of Hidden Markov Models (MHMM) to model the Likelihood of the given data set $\underline{Z}$. That is, for the definition of the MHMMs, the Eq. (5.2) for the calculation of $P(\underline{Z})$ will be reformulated incorporating the several components:

$$\begin{aligned}
P(\underline{Z}) &= \prod_{k=1}^{K} P(Z^k) \\
&= \prod_{k=1}^{K} \sum_{c \in C} P(Z^k|\theta^c) \times P(\theta^c) \\
&= \prod_{k=1}^{K} \sum_{c \in C} P(X^k, S^k|\theta^c) \times P(\theta^c) \\
&= \prod_{i=1}^{K} \sum_{c \in C} P(X_1^k|S_1, \theta^c) \times P(S_1^k|\theta^c) \times \prod_{l=2}^{L^k} P(X_l^k|S_l^k, \theta^c) \times P(S_l^k|S_{l-1}^k, \theta^c)
\end{aligned} \tag{5.8}$$

If the class information is available, corresponding HMMs can be learned respectively trained on the given search behavior samples, as done for the baseline models. That is, the θ^c represents the HMM component according to the given factual or exploratory search. The prior $P(\theta^c)$ denotes the mixture coefficients regarding the search session and $P(Z^k|\theta^c)$ denotes the probability of generating a certain sequence Z^k by the given model θ^c. This exactly leads to the probabilistic classifier given in Eq. (5.3) on p. 126 and allows to apply the classification implementation according to Eq. (5.4) as already used multiple times. However, if the class information for the individual search sequence is not available or not used, Eq. (5.8) can be utilized in the $\mathcal{Q}$ formula of the Expectation Maximization (EM) algorithm to derive the estimates within a clustering setting [27]. In particular, the EM algorithm here can be implemented by it's *E step*:

$$\mathcal{Q}(\theta, \theta_{old}) = \sum_{c \in C} P(\theta_{old}^c|\underline{Z}) \times ln(P(\underline{Z}, \theta^c)) \tag{5.9}$$

and it's corresp. *M step* (cf. Bishop [27]):

$$\theta_{new} = \underset{\theta}{\text{argmax}}(\mathcal{Q}(\theta, \theta_{old})) \tag{5.10}$$

Summing up, Eq. (5.8) can now be used to investigate the user's seeking behavior in terms of search behavior classification and parameter analysis as previously done on the data of *US-II*. Furthermore, the

Table 5.10: Confusion Matrix for the classification using 2nd-order HMMs on the data from *US-III*.

	Real Class					
	Absolute		Relative		Relative (to class)	
	Fact	*Expl*	*Fact*	*Expl*	*Fact*	*Expl*
Classified as *Fact*	670	61	.71	.06	.93	.27
Classified as *Expl*	47	165	.05	.18	.07	.73
	717	226	1		1	1

formulas in Eq. (5.9) and (5.10) can be used to derive two (or more) clusters of search activities and compare the clusters to the pre-defined factual and exploratory classes to validate the experimental design.

5.3.3 *Search Class Validation*

If Eq. (5.8) is applied in a classification setting on a 5-fold cross-validation averaged over 2000 repetitions to identify the factual and exploratory search behavior the model achieves an accuracy of 88.58%. The corresponding confusion matrix is given in Tab. 5.10. The here reached accuracy is similar to the 89.4% of the 1st-order HMMs with duration on the full state space using the data of *US-II* (cf. Sect. 5.2.2) and also similar to the 87.7% of the 2nd-order HMMs with duration on the state space without q_{Main} using the data of *US-II* (cf. Sect. 5.2.6). Considering the class imbalance of 717 fact-finding and 226 exploratory search sequences, the relative class accuracies with about 93% for factual and about 73% for exploratory tasks show an inverse result comparing the models trained on the data of *US-II*. The 1st-order HMMs with duration on the full state space reached about 84% for Fact and about 92% for Expl. That is, the classification in Sect. 5.2.2 achieved higher values for exploration and lower values for fact-finding. Here it is the opposite case. Comparing the results to the 2nd-order HMMs with duration on the state space without q_{Main} using the data of *US-II*, which reached about 89% for Fact and about 87% for Expl, the relative class accuracies are higher for factual but lower for exploratory search tasks. However, the results here allow to argue that the models trained on the data of *US-III* still recognizes the two search classes sufficiently what allows to draw further conclusions from it.

Fig. 5.10 plots several characteristics of the classification models trained here. Considering only the states of the underlying Markov model, Fig. 5.10 illustrates the stationary distribution of the relative proportion for each state in a long-run behavior. It can be seen that users predominantly interact with web pages during exploratory search sessions and interactions on the *Query* state happen relatively seldom. In contrast to that, users in fact-finding search interact more

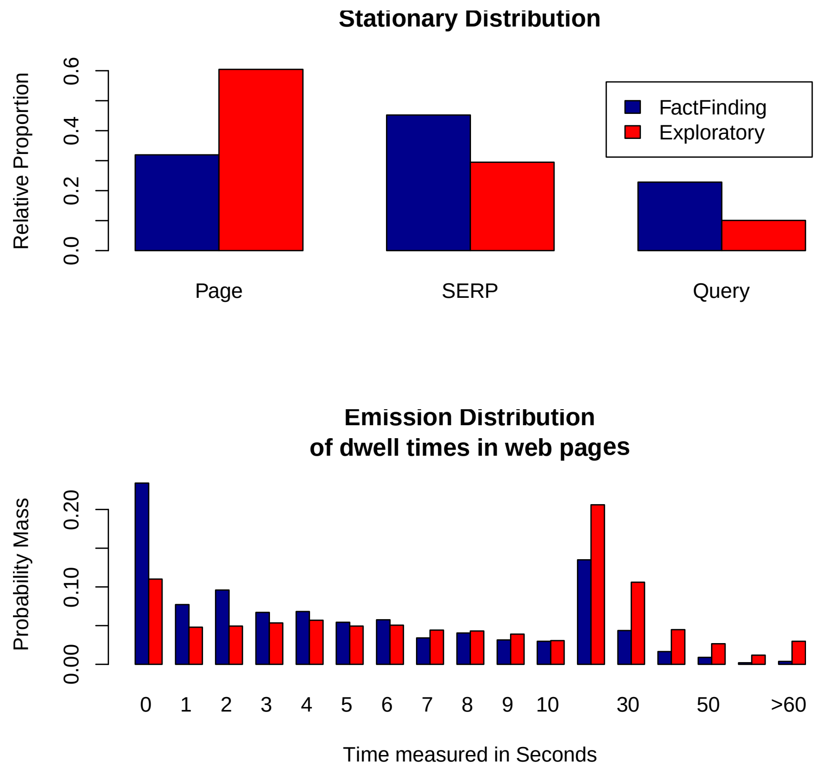


Figure 5.10: Distribution diagrams regarding the classification of fact-finding search in blue and ES in red. The upper plot illustrates the stationary distributions over the state space Q while the lower plot represents the emission distribution for the duration feature on web pages.

homogeneous with all three states but with a preference towards the *SERP* state. This is conform with the analysis in Sect. 5.1, where user in Fact click twice as much on a further (or previously) SERP. Considering the duration distributions of the observation models, Fig. 5.10 shows that users who spend more than seven sec. on web pages have been predominantly in the exploratory search sessions and users who spend less than seven sec. on web pages have been predominantly in fact-finding search. This again is conform with the analysis in Sect. 5.1.1[14] and as well with the results of Athukorala et al. [10].

Now Eq. (5.8) and the corresponding models will be used in a cluster setting with exactly two components to investigate the user's search behavior without any class information. That is, the label information for the corresponding search interaction sequences is omitted leading to the usage of Eq. (5.9) and (5.10) with the underlying EM algorithm. This procedure enables to capture evidence-based characteristics of the data even when the experimental design marginally violates the

14 According to the CDF of the approximated exponential distribution, 75.6 percent of all web site visits in Expl last longer than seven sec. while in Fact only 64.7 percent of all web site visits last longer than seven sec.

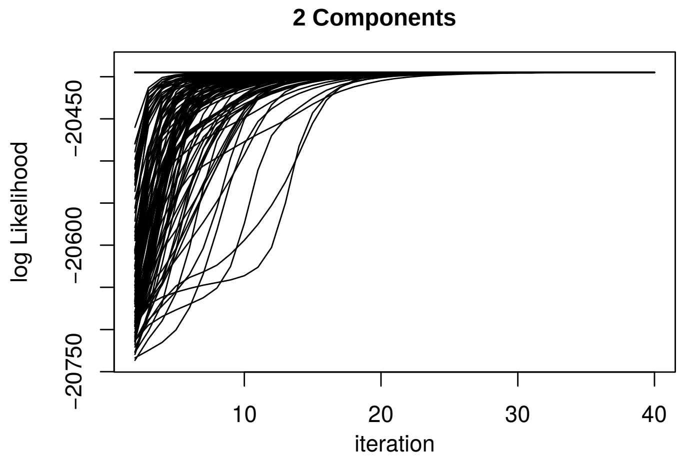


Figure 5.11: Plot of the EM progression for the MHMMs with two components on 100 random initializations until convergence on the data of user study *US-III*.

"true" underlying process, i.e., the pre-defined and assumed factual and exploratory search classes.

To illustrates the iterative clustering process of the EM algorithm, the progression on 100 random initializations is shown in Fig. 5.11. The final result of the algorithm depends on model initializations. Hence, 100 MHMMs have been initialized randomly to use this models in repeated EM runs. The likelihood maximization is applied on the full data set until the incremental increase dropped under a predefined termination threshold. As Fig. 5.11 shows, all runs converge to the same log likelihood plateau of $-20,394.71$[15]. This allows to conclude that there is few evidence for other cluster separations on two components.

15 In addition, the MAP version of the EM was used implementing Dirichlet Priors with hyperparameters set to $\beta = 1.1$. Hence, the estimated model θ can be selected by $ln(P(\underline{Z}|\theta)) + ln(Prior(\theta|\beta)) = -20,394.71$.

Table 5.11: Contingency table that represents the cluster assignments to the search sessions according to the experimental design, i.e. the "Real" classes identified by 1st-order HMMs on the data set from *US-III*.

	"Real" Class					
	Absolute		Relative		Relative (to class)	
	Fact	*Expl*	*Fact*	*Expl*	*Fact*	*Expl*
$Cluster_1$	585	65	.62	.07	.82	.29
$Cluster_2$	132	161	.14	.17	.18	.71
	717	226	1		1	1

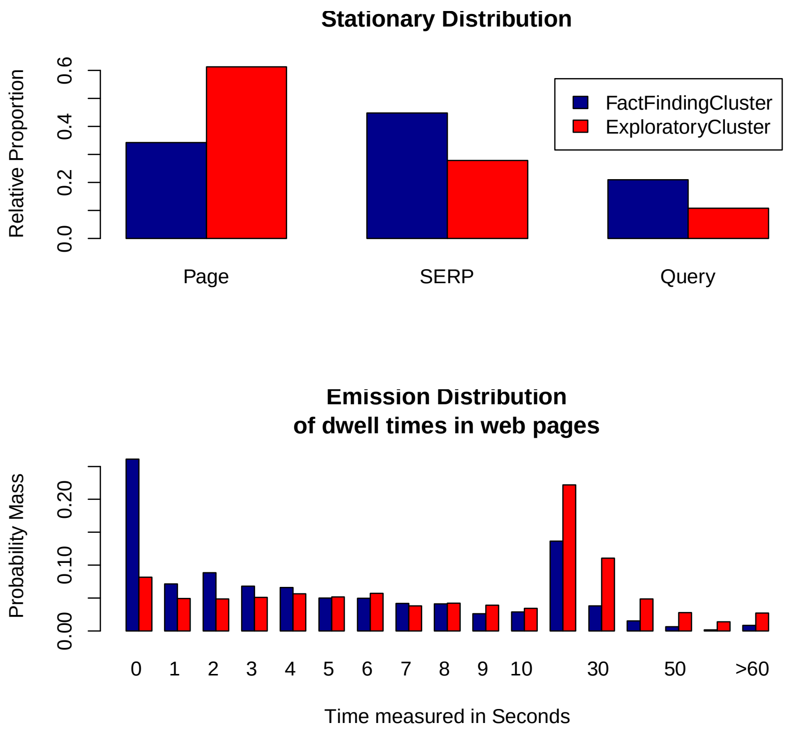


Figure 5.12: Distribution diagrams regarding the search activity clusters for fact-finding search in blue and for ES in red. The upper plot illustrates the stationary distributions over the state space Q while the lower plot represents the emission distribution for the duration on web pages.

For each search sequence, the maximal latent expectation as the assignment to it's hypothesized search behavior is used. The contingency Tab. 5.11 shows the cluster assignments to the search classes according to the experimental design. The agreement of 79% (62% + 17%) allows to infer implicit semantics from the proposed behavior clusters. In particular, $Cluster_1$ consists of 90% fact-finding assigned search sequences, making the cluster an arguable representation of the fact-finding behavior as postulated in the experimental design. The size of $Cluster_2$ is less than a half of $Cluster_1$. $Cluster_2$ comprises of 45% fact-finding assigned search sequences, and 55% exploratory assigned searches respectively. Although, their is a lack of significant evidence, $Cluster_2$ can be postulated as the representation of exploratory search behavior because of the majority of present Expl.

Similar to the analysis above, Fig. 5.12 plots several characteristics of the cluster models. Considering the stationary distribution of the relative proportion for each state, the search behavior in the two clusters indicates a high level of similarity to the models from the classification setting. In $Cluster_2$, what now will be considered as "ExploratoryCluster", users have an increased orientation towards web sites. In $Cluster_1$, what now will be considered "FactFindingCluster", user interact more homogeneous but with a preference towards the SERPs. Almost the same similarity holds for the duration distribution. Users in the "ExploratoryCluster" have a tendency to spend more than eight sec. on web pages and users in the "FactFindingCluster" predominantly spend less than five sec. on web pages.

The analysis showed that the factual and exploratory search can be identified by a clustering approach. That is, based on the interactions sequences, user's seeking behavior can be differentiated according to the experimental design even when task assignments are missing. This allows the conclusion that fact-finding related search behavior is a reaction to, respectively is induced by, the here used factual search tasks. The same conclusion can be made for exploratory related search behavior. However, correlations in the clusters and the experimental assignments indicate some easily recognizable but also some borderline cases of search behavior. Therefore, it can be argued that based on the proposed model, some vague cases or outliers should be considered in more detail because the used search tasks might also have induced further unexpected search behavior. This investigation will be implemented in the following sub-section.

5.3.4 *Identification of Latent Search Activities*

To validate the experimental setup of *US-III* (and implicitly of *US-II*), the number of components in the cluster setting had to be equal to the number of proposed search behavior classes, namely two. However, the analysis indicated that there may be further search behavior induced by the search tasks what causes the mixture and the narrow majority of exploratory search sessions in $Cluster_2$. In general, the results of the cluster approach above but also results from the literature indicate that a straight separation of search activities might be difficult and/or artificial (caused by the experimental design) since the number of components, which represent different groups of search behavior, can be different. For example, Athukorala et al. [10] proposed that there are three search modes, namely fact-finding search, exploratory search and a borderline cases. Consequently, the results here but also the results from the literature provide a sound motivation to investigate the number of different behavior types in the data of *US-III*.

The task to identify an appropriate number of components can be interpreted as a model selection task. Therefore, the information criteria AIC and BIC, cf. Eq. (5.6) and (5.7), can be used. Since the data of *US-III* comprises a relatively high number K of search samples, it is reasonable to include this parameter in the model selection what leads to the choice of BIC instead of AIC that does not considers K. In contrast to the model selections regarding the best order of the (Hidden) Markov model before, the calculation of the estimated parameter

Table 5.12: Model selection using BIC for different components and using the data from *US-III*.

Components	P	log-Likelihood of $P(\underline{Z})$	BIC
1	30	-20667.509	41540.491
2	60	-20343.942	41098.829
3	**90**	**-20223.716**	**41063.849**
4	120	-20159.884	41141.656
5	150	-20114.927	41257.215
6	180	-20073.690	41380.213

P here is dominated by the emission features and not by the state structure of the Markovian models. That is, the HMMs used here, have the same order one what results in an increasing of P only by 3 for each new component. The number of discrete bins for the modeling of the emission feature in turn has a higher impact. For example, if the MHMM has only one component, P is the sum of 3 start probabilities, 3^2 state transitions from order one, 17 bins on the *Page* state and 1 further parameter for the prior, what equals to 30. For a MHMM with two components, P is two times the number of a single component, i.e., 60. That is, the 17 bins on each of the two components capture already $(2 \times 17)/60 = 56.6\%$ of the parameters. Finally, if $comp$ denotes the number of components for a given MHMM and bin denotes the number of discrete bins to model the emission for the duration feature on web pages, $P(comp, bin) = comp \times (|Q| + |Q|^2 + bin + 1)$. In the next step, MHMMs with components from 1 to 6 have been trained analog to the previous cluster approach. Afterwards, the model selection applying the BIC was executed to infer the optimal number components, i.e., the appropriate number of search behavior groups. The results are listed in Tab. 5.12. According to the information criterion, a model with three components has the smallest values and therefore, has the most support.

Similar to the cluster analysis on two components, Fig. 5.13 illustrates the iterative clustering process of the EM algorithm, the progression on 100 random initializations. Again, the likelihood maximization is applied until the incremental increase dropped under a predefined termination threshold. All runs converge to the same log likelihood plateau of $-20,293.1$, cf. Fig. 5.13. This allows to conclude that there is few evidence for other cluster separations on three components.

The corresp. contingency table illustrates the three resulting components, cf. Tab. 5.13. According to the table, $Cluster_1$ is the biggest cluster comprising about 50% of all search sessions. Furthermore, $Cluster_1$ consists of about 96% fact-finding assigned search sessions

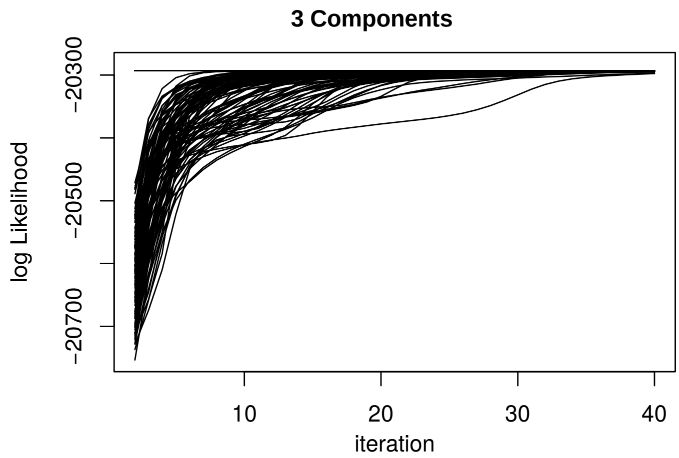


Figure 5.13: Plot of the EM progression for the MHMMs with three components on 100 random initializations until convergence on the data of user study *US-III*.

and hence, is relatively homogeneous. $Cluster_2$ is less than half of $Cluster_1$'s size with about 21% of all search sessions. It contains 45.5% fact-finding assigned and 54.5% exploratory assigned searches respectively. $Cluster_3$ has more of half the size of $Cluster_1$ with about 28% of all search sessions. It consists of 64.8% fact-finding assigned sessions. That is, $Cluster_3$ is bigger than $Cluster_2$, has a bigger portion of fact-finding assigned sessions but is still relatively heterogeneous. Based on the evidence in the contingency table, it is postulated that $Cluster_1$ is a representation of fact-finding, $Cluster_2$ represents exploratory and $Cluster_3$ borderline search behavior. This postulation can also be confirmed by stationary distribution of the relative proportion for the states in each cluster and the corresp. duration distributions, cf. Fig. 5.14. The diagrams show that the relative state proportion for the "FactFindingCluster" and the "ExploratoryCluster" are similar to the proportions in the two cluster setting. That is, users in the "FactFindingCluster" interact more homogeneous but with a prefer-

Table 5.13: The contingency table that represents the cluster assignments to the search activities according to the experimental design, identified by 1st-order HMMs on the data from *US-III*.

	Experimental Design Class					
	Absolute		Relative		Relative (to class)	
	Fact	*Expl*	*Fact*	*Expl*	*Fact*	*Expl*
$Cluster_1$	448	20	.47	.02	.62	.09
$Cluster_2$	92	110	.10	.12	.13	.49
$Cluster_3$	177	96	.19	.10	.25	.42
	717	226	1		1	1

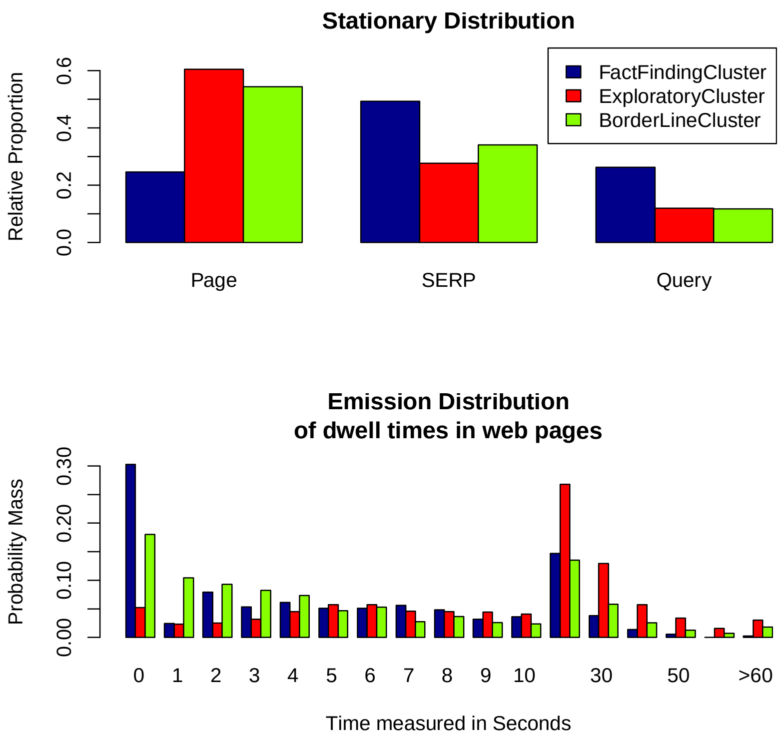


Figure 5.14: Distribution diagrams regarding the search activity clusters for fact-finding search in blue, for ES in red and for borderline behavior in green. The upper plot illustrates the stationary distributions over the state space Q while the lower plot represents the emission distribution for the duration on web pages.

ence towards the *SERP* state and users in the "ExploratoryCluster" have an increased orientation towards web sites. One difference is that in the "FactFindingCluster" users preferences towards *SERP* and *Query* state is more developed in comparison to the two cluster setting. Investigating the "BorderLineCluster" reveals several similarities to the "ExploratoryCluster" regarding the state proportions. A difference is that the borderline behavior has a small decrease on web site proportions but a small increase in the *SERP* state. Considering the state durations, similar to the two cluster setting, in the "ExploratoryCluster" users have the tendency to spend more than ten sec. on web pages while in the "FactFindingCluster" users spend less than five sec. on web pages. In the borderline behavior the durations are more similar to the fact-finding search behavior. That is, the notation as "BorderLineCluster" is appropriate in so far as it is more similar to "ExploratoryCluster" regarding the relative state proportion but more similar to "FactFindingCluster" regarding the duration distributions.

For user study *US-III* (but also in *US-II*), participants could submit their experience regarding the current ES task with it's answer on a five point Likert-Scale. In particular, the users had to answer how much they agree with the statement *"I am familiar with the topic:"* on the scale *"Yes, I am an expert"* (1) to *"Not at all"* (5). Considering the user's experience on the "ExploratoryCluster", 83% of the users submitted values of 4 or 5 (i.e., almost no experience), and only 6% submitted values of 1 or 2 (i.e., familiar with the topic). For the "FactFindingCluster" and the "BorderLineCluster" the stated experience regarding the two ES tasks were on a similar level and both were higher in contrast to the "ExploratoryCluster" with: 72% of submitted experience values of 4 or 5 resp. 11% of submitted experience values of 1 or 2 for the "FactFindingCluster" and 66% of submitted experience values of 4 or 5 resp. 10% of submitted experience values of 1 or 2 for the "BorderLineCluster". That is, the "ExploratoryCluster" is characterized by a lower user experience regarding the search topic. However, the number of ES tasks in the "FactFindingCluster" was quite low and therefore, the relative numbers have to be taken carefully, whereas the number of ES tasks in the "BorderLineCluster" are almost as high as in the "ExploratoryCluster". This in turn, characterizes the "BorderLineCluster" by a higher user experience regarding the exploratory search topic(s).

Based on the analysis above, it can be argued that different search sessions, respectively activities, can be identified by the proposed search behavior model even if the amount of "actual" sub models is not known. The correlation between the resulting clusters components and the original experimental search assignments indicate that exploratory and fact-finding behavior are highly conserved task behaviors. Some participants show a kind of borderline search behavior by applying aspects of exploratory search behavior regarding the state proportions but with an short duration on web pages what usually is associated to factual search behavior. This implies and allows to conclude that fact-finding and exploratory search behavior should not necessarily be considered as two concepts on a dual continuum. Rather they incorporate a multitude of aspects what may result in seeking behavior that instantiates patterns or characteristics which seems to be contradictory but are apparently complementary. This again, is in accordance with the Marchionini's and White's (et al.) proposition regarding exploratory search (cf. Sect. 2.6, p. 33 et seq.) but also with the conclusions in Section 2.6.3 (p. 38 et seq.) to address $H1$. All in all, it can be concluded that information seeking behavior is triggered by a given search task type but the subjective reaction to a given task might vary from user to user and may also depend on personal characteristics (cf. the following Sect. 5.4). Furthermore, time pressure or individual motivation might be additionally factors for adapting a certain behavior patterns.

To finalize this section, the following shall be noted. The methodology described in this section allows to identify new, former unknown, search behavior solely on search parameters independent of the known search class. Therefore, this statistical approach can be used to reveal seeking behavior from interaction data in general. Even though the method was applied to the data set of *US-III*, it is suitable to detect new behavior as generic patterns in a latent parameter space. This enables the identification of user's search activities that might be triggered contradictory to the originally expected search behavior and/or can reveal behavioral sub patterns.

5.4 INFLUENCING VARIABLES ON EXPLORATORY SEARCH

As pointed out in Section 2.3, Wilson proposed a list of several intervening variables which can influence the information (seeking) behavior [195]. Therefore, some of these variables are also important for this thesis to examine whether indications of influences can be found in the data and if so, to discuss to what extent they are relevant for EIS. The underlying motivation and long term goal remains, namely to further understand the paradigm of ES and thereby to reveal potentially means which enable (adaptive) information systems to provide an appropriate user support. In particular, *psychological and demographic* as well as aspects of *role-related* variables will be considered by utilizing the obtained user interactions as well as personal characteristics from the studies *US-I* and *US-III* (*H*2). At first, the demographical variable "age" will be addressed by analyzing the data of *US-I* and it's sub studies (Sect. 5.4.1). Afterwards, the psychological variables, obtained by user study *US-III*, will be investigated (Sect. 5.4.2).

5.4.1 *User Variable Age: Young and Adult Users*

The target group of Internet users continuously becomes wider and more heterogeneous. This makes the user variable "age" more and more important because the user group for the Internet comprises a huge proportion of young, but also elderly users[16] which have different needs and capabilities in contrast to adults. Young users have less experience in general. That is why, these users are a promising user group for less-biased investigations regarding web search activities, such as ES. While the focus of investigation in the related own works [76] and [114] was more on the development of voice-controlled Search User Interfaces (SUIs) for children, in the following, a more analytical and model oriented perspective on the interactions of the young users from *US-I* will be given.

16 Although the consideration of elderly users is not less important than young users, given the user studies and data in context of this thesis, only the search behavior of young users will be analyzed in the following.

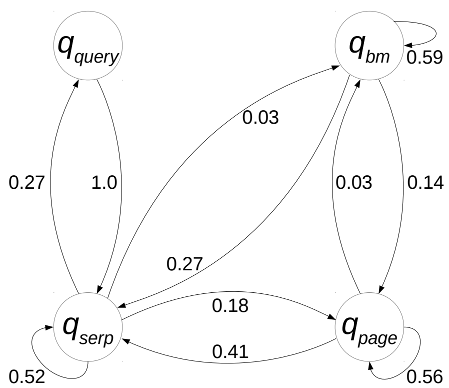
5.4.1.1 *Modeling Young Users Search Behavior*

After the preparation of the search interaction data with the SUI *Knowledge Journey* using the Eye-Tracker software *Tobii-Studio* and it's annotation function, for both, *US-Ia* and *US-Ib*, the following four user states could be derived:

- *Query*: A young user expressed the desire to enter a search query or (in case of *US-Ia*) to select an item on the navigation menu.
- *SERP*: A user is examining a SERP after a query was entered.
- *Page*: A user is examining a web page as result of an opened SERP item or reached by a clicked URL from an other web page.
- *BM*: A user is (re-)viewing a bookmarked (BM) web page(s) in the *treasure chest* (*US-Ia*) or the *logbook* (*US-Ib*).

Recalling the SUI *Knowledge Journey* and in particular the interaction graph in Fig. 4.3 (Sect. 4.2.3, p. 82), there are some differences in the user states listed above in contrast to the user states derived from *US-II* or *US-III*. First, the query input (resp. the selection of an item in the navigation menu) was performed by the investigator (the wizard) after receiving the children's corresponding control demand. That is, the young user first had to express the command, than the wizard had to interpret the command and finally had to act correspondingly. Therefore, the query process comprises a mixture of explicit and implicit interactions of the child and the wizard. As a result, the annotation of the *Query* state was set exactly at the end of the children's utterance but before the wizard's query input. The time to type the query, respectively to click in the navigation menu, was usually short since the wizard was mostly able to anticipate the next desired interaction in general and hence, the duration of the wizard's interactions will be neglected. Consequently, the *Query* state is allocated relatively exactly (in time) but does not have emissions, such as duration or fixations. However, for the sake of a user friendly interaction, achieved by the voice-control, this missing emissions on the *Query* state was considered as justifiable. Second, the SERP in the *Knowledge Journey* and it's coverflow visualization always shows only one item of the result list. Therefore, the user had to express whether the next or the previously result shall be shown, what potentially leads to high self references for the *SERP* state (i.e., $a_{serp\ serp}$). Third, the *BM* state also allows several internal interactions, such as illustrate a preview of the bookmarked items, make some notes for the items or delete them, what results into corresponding self references (i.e., $a_{bm\ bm}$). The usage of and possible interactions in the *BM* state makes this state similar to the *Main* state in the user studies *US-II* or *US-III*.

To further formalize the states above and the corresponding sequences, the data set $\underline{S}$ of *US-I* consists of K search sessions indexed by S^k with the length L^k and each session comprises a sequence of (visited) states from the state space $Q = \{q_{query}, q_{serp}, q_{page}, q_{bm}\}$. That is, the data set $\underline{S} = (S^1, ..., S^K)$ where each search session $S^k = (S^k_1, ..., S^k_{L^k})$ and each state $S^k_l = q^k_i \in Q$. Similar to the factual and exploratory search in *US-II* or *US-III*, the young users' search activities will be distinguished in two different search types. One type is the free search from *US-Ia* (Free). The other type is the exploratory search –regarding the zoo animal housing topic– from *US-Ib* (Expl). Although the number of interactions from *US-I*, given the small user base, is strongly restricted and does not allow to claim for generalizability, modeling and analyzing the young users' interactions can still reveal insights for ES that can not (yet) be found in the literature. Training a 1st-order Markov model for each of the two corresponding search activities Free and Expl results into the models θ^{Free} and θ^{Expl} as illustrated in Fig. 5.15 and Fig. 5.16. In Free and Expl, users always started their search on an empty SERP ($\pi_{serp} = 1.0$). From the *SERP* state, every other state could be reached. As mention above, the SERP always provided only one result item in the center of the screen. The resulting high amount of self references for the *SERP* state (i.e., $a_{serp\ serp}$) can be observed in both models whereat the portion in θ^{Free} is even higher than in θ^{Expl}. The query usage in both search activities is relatively equal. The probability to open a web page in contrast differs strongly and is more prominent in Expl ($a_{serp\ page} = 0.35$) than in Free ($a_{serp\ page} = 0.18$). Furthermore, the self references on web pages in Expl ($a_{page\ page}$) are higher and are mostly caused by "clicks" on the page. This can be an indicator that users in a free search try to estimate the relevance of a web page (i.e., how interesting the page's content is) by the preview in the SERP. In Expl however, this approach appears to be less successful (to solve the ES task) and therefore, more web pages are opened more likely and

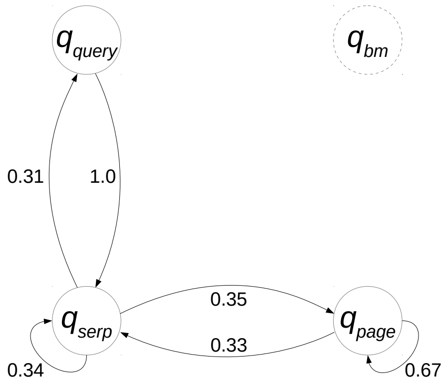


$$Q = \{q_{query}, q_{serp}, q_{page}, q_{bm}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.27 & 0.52 & 0.18 & 0.03 \\ 0.00 & 0.41 & 0.56 & 0.03 \\ 0.00 & 0.27 & 0.14 & 0.59 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{bm} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 1.0 \\ 0.0 \\ 0.0 \end{pmatrix}^T$$

Figure 5.15: Illustration of the 1st-order Markov model θ^{Free} as graph (left) and it's corresponding components Q, A and π (right) trained on the data set from *US-Ia* using all interactions of Free.



$$Q = \{q_{query}, q_{serp}, q_{page}, q_{bm}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.31 & 0.34 & 0.35 & 0.00 \\ 0.00 & 0.33 & 0.67 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{bm} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 1.0 \\ 0.0 \\ 0.0 \end{pmatrix}^T$$

Figure 5.16: Illustration of the 1st-order Markov model θ^{Expl} as graph (left) and it's corresponding components Q, A and π (right) trained on the data set from *US-Ib* using all interactions of Expl.

the web pages themselves are used more intense. That is, the presence of the ES tasks indeed can cause a different seeking behavior even for young users. In *US-Ia*, many bookmarks of web pages have been made by the users but the corresponding state (q_{bm}) was seldom visited. If users opened the *treasure chest*, they mostly managed the bookmarks (e.g., by make notes or delete items) and did not used them to open the corresponding web page. That is, the *treasure chest* literally was utilized rather as a collection of "valuable" items instead of include the items in the search process. In *US-Ib* many bookmarks of web pages have been made as well but not a single user opened the *logbook*.

5.4.1.2 *Comparing Relative State Proportion*

Next, the stationary distribution of the relative proportion for each of the four states regarding Free and Expl will be considered. The corresponding diagram is illustrated in Fig. 5.17. Comparing the

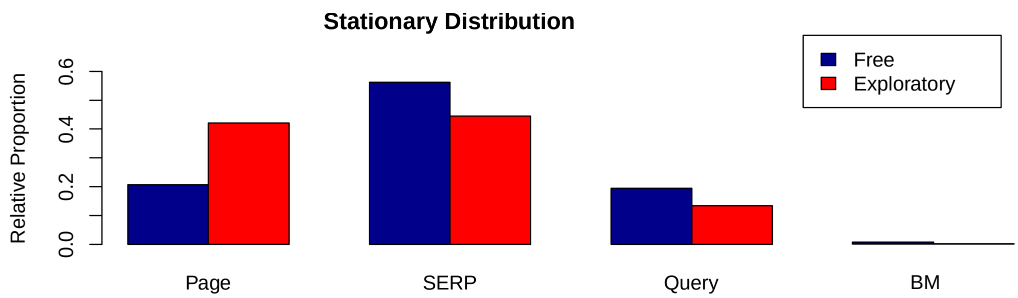


Figure 5.17: Distribution diagrams regarding the search activities Free in blue and Expl in red from *US-I*. The plot illustrates the stationary distributions over the state space Q.

Table 5.14: Mean dwell times on the states from *Q* in sec. according to the search activities Free and Expl as well as the combination Free + Expl, i.e., all search sessions together, from *US-I*.

	Query	*SERP*	*Page*	*Main*
Free	-	23.5	34.8	25.3
Expl	-	23.9	34.0	-
Free + Expl	-	23.6	34.3	-

proportions of the young users' search behavior with the stationary distribution of the factual and exploratory search sessions of adults (cf. Sect. 5.3.3, Fig. 5.10), reveals remarkably similarities. Also the young users predominantly interact with web pages and the SERP during exploratory search and interactions regarding the query state happen relatively seldom. In contrast, young users in the free search use more query interactions, less web page interactions (making both more equal distributed) but mostly interact with the SERP. For completeness, the visits, respectively proportion, of the state q_{bm} are illustrated but because of the low values on both search activities, the influence to the other state proportions is marginal. That is, free search of young users has similarities to the adults' fact-finding search. Furthermore, in both user groups, the induced exploratory search behavior is similar as well. The only noticeable divergence here is that children use web pages and the SERP almost equally during their ES while adults clearly are more often on web pages. This difference however can also be an artifact of the *Knowledge Journey*' SUI where the young users need more often to interact (and stay) on the SERP to view the results because of the coverflow visualization.

5.4.1.3 *Comparing State Durations*

Literature and the results of this thesis so far identified the duration, as implicit interaction, to be a valuable and discriminative feature. Therefore, the dwell times on the states of *Q* will be examined also for young users. Tab. 5.14 lists the mean dwell times for the visited states. As already pointed out above, for the *Query* state no features could be extracted in *US-I* and the *BM* state in *US-Ib* (i.e., the *logbook*) was never opened. For the two states *SERP* and *Page* it can be seen that the durations in Free and Expl are virtually equal respectively. Independent of the search activity, the young users spend about 24 sec. on SERPs. It has to be noted here that each time, the users instructed the system (the wizard) to show the next item in the SERP, there was a reset on the dwell time counter. That is, the time young users spend on a SERP regarding one given query was even longer. Comparing the dwell times to those of the adults (the avg. was 6.5 sec.), young users spend about 3.5 times longer on SERPs in Free and in Expl. The histograms of

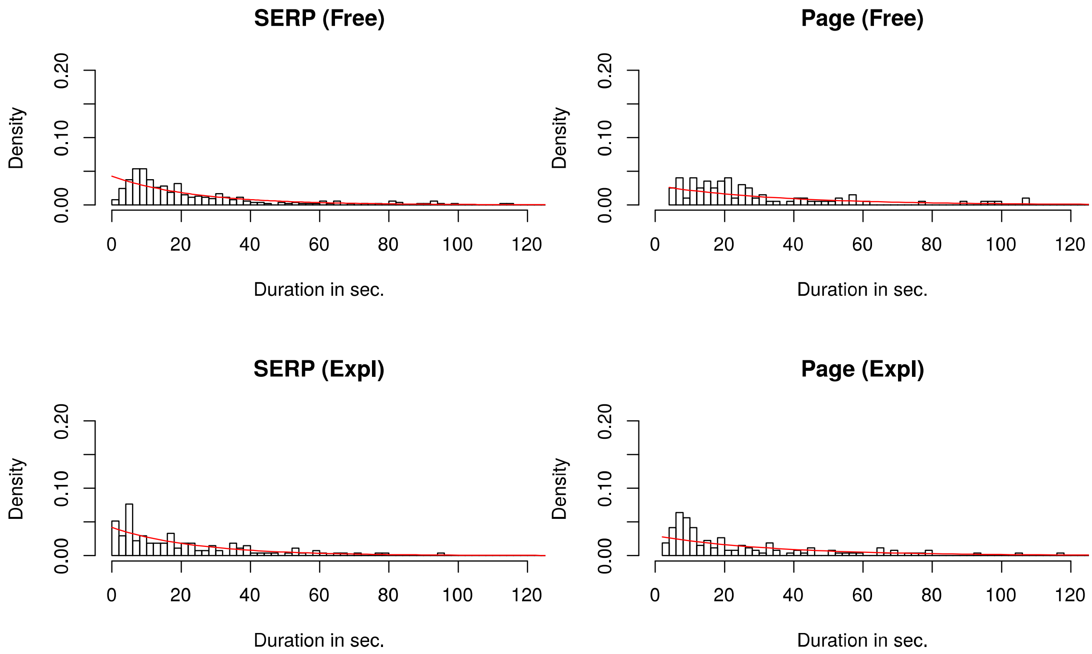


Figure 5.18: Histograms of the empirical state durations with the estimated exponential distribution in red for the search activities Free and Expl from *US-I* together.

the empirical state durations for q_{SERP} in Fig. 5.18 confirm the similarities between Free and Expl[17]. On web pages, children again spend a multiple time than adults, here about twice the time. Furthermore, the dwell time on q_{page} is also virtually equal for Free and Expl. That is, in contrast to the analysis of the relative proportion above, the picture on dwell times is a different one. First, the differences between Free and Expl of young users are not comparable with the differences between fact-finding and exploratory search of adults because the durations on the states (*SERP* and *Page*) are almost equal in the search of young users. Fact-finding and exploratory search of adults turned out to have different dwell times on the states (*SERP* and *Page*). Second, the (mean) dwell times are by far longer for young users than in adult user's search activities.

5.4.1.4 *Young Users Gaze Features*

Next, a brief examination of the young users' eye-tracking features will be given. As shown in Section 5.2.4, user's fixations as implicit interactions (like dwell times on states) can help to distinguish between search activities. A KST on the data of *US-I* and in particular on the

17 Remark: In contrast to the histograms of the empirical state durations of adult users in Fig. 5.4, 5.5 and 5.6, the density axis is scaled to the half and the duration axis is doubled for a more convenient illustration of the young users' longer dwell times.

Table 5.15: Feature Selection based on p-values and an α of 0.1. The p denotes the p-Value. Relevant features are marked bold.

Feature	p	Feature	p	Feature	p
Page.Fixation	**4.1e-07**	*Page*.FixDuration	**1e-07**	*Page*.FixDurMean	**1.0e-05**
SERP.Fixation	**3.3e-06**	*SERP*.FixDuration	**2.3e-07**	*SERP*.FixDurMean	**4.4e-12**

features *Q.Fixation*, *Q.FixDuration* and *Q.FixDurMean* was made and the results are listed in Tab. 5.15[18]. The table shows that each of the features (on the states *SERP* and *Page*) are potentially relevant resp. useful (because of p-values lower than $\alpha = 0.1$) to distinguish between Free and Expl in theory. However, the utilization of these gaze related features as emissions on HMMs for automatic search activity classification will not be further investigated here because of the relatively small number of interactions from *US-I*, respectively the small given user base. In addition to that, the young users moved much in front of the screen and also often searched eye-contact to the investigator. This leads to a low(er) recognition of the gaze behavior by the Eye-Tracker. In particular, the mean detection rate in *US-I* was 48.1% with a relatively high variance, cf. Fig. 5.19. Nevertheless, the tests on gaze related features indicated them again as potential useful. Interestingly, even for the virtually equal dwell times, the KST revealed that the corresponding feature *Q.Duration* (barley) could be used as well with a p-value of $0.001988 < \alpha$ for the *Page* state and a p-value of $0.001913 < \alpha$ for the *SERP* state.

5.4.1.5 *Conclusions on User Variable Age*

Summing up, the analysis of the user variable "age" regarding the data sets of this thesis revealed that there are similarities between young users and adults in the search behavior but also differences could be found. While the exploratory search behavior of the two user groups in terms of state proportions was surprisingly similar, the dwell times on the several states are by far longer in the children's search activities. Furthermore, the free search of young users, performed

18 A benefit of the KST is that it can also be applied on smaller samples.

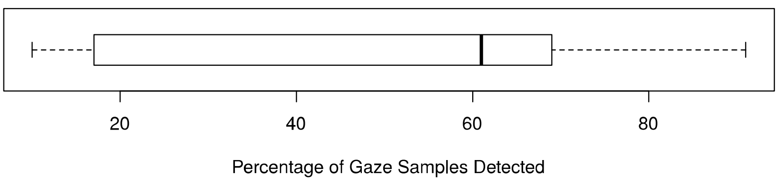


Figure 5.19: Boxplot (with median) illustrating the percentage of gaze samples that could be detected by the Eye-Tracker in *US-I*.

in user study *US-Ia*, is similar to the fact-finding search of adults (in *US-III*) from the perspective of state proportions but regarding the dwell times could not show differences between the free and the exploratory search of the young users. This can be an indicator of the young user's ability to (already) apply different search behavior regarding the given information need (induced by the search tasks) but also indicate the young user's not (yet) well developed ability to identify relevant information on the specific states, such as *Page* and *SERP*. Recalling Bates discussion of strategies and tactics (cf. Sect. 2.2, footnote on p.19), this can be an indication that children already have an overall plan (a strategy) how to seek for information and solve the given (search) task but the specific moves to advance the search (the tactic) may lack for experience with the provided information system and sources.

5.4.2 *Analyzing Personal User Characteristics*

To continue the investigation of intervening variables, the data set of user study *US-III* is utilized to get insights into the degree of influence of the several recorded personal user characteristics. At first, the users motives are addressed. As described in Sect. 4.4.1.2, the Motivation Theory of Rheinberg (et al.) [53, 157, 158] and the related Reference Norms (RNs) associate motivation to goals or results. Therefore, in this thesis, goals will be measured by the number of (correctly) answered factual tasks and are implemented by three corresponding factual task blocks, Fact_{CRT}, Fact_{IND}, Fact_{SOC} and a fourth one Fact_{NON} as baseline (cf. Sect.Sect. 4.4.2). Consequently, the analysis for the user's motives have to be limited to factual search behavior (Sect. 5.4.2.1). In contrast to that, the obtained remaining psychological variables can and will be investigated in context of ES. In particular, relations to the user's personality traits (Sect. 5.4.2.2), aspects of intelligence (Sect. 5.4.2.3) as well as user's sensation seeking (Sect. 5.4.2.4) will be examined. This includes the investigation of ES but also fact-finding search will be considered.

5.4.2.1 *User's Motives*

To investigate the influence of motivational goals, in a first step, it will be analyzed, whether there are differences between the two difficulty levels *hard* and *easy* in the factual search tasks. That allows to confirm whether the extended task set[19] and the corresponding difficulty levels induce a natural and expected search behavior. Therefore, the following three sub-hypothesis $H2_a$, $H2_b$ and $H2_c$ will be examined:

$H2_a$: *The two variables task difficulty*

19 As a reminder: the task set of original twelve factual search tasks from *US-II* has been extended to 117 tasks for *US-III*.

and correctness are not independent. *Hypothesis $H2_a$*

$H2_b$: Hard questions are answered more often wrong than easy questions. *Hypothesis $H2_b$*

$H2_c$: It takes more time to answer hard questions than easy questions (correctly). *Hypothesis $H2_c$*

To test two variables for independence and hence, to examine $H2_a$, the χ^2-Test can be used. Applying the χ^2-Test reveals that the variables *task difficulty* and *correctness* are significantly not independent for all factual search tasks together: $\chi^2_{df=1} = 338.34$, $p < .000$. In addition to that, for each of the four factual search task blocks separately, the χ^2-Test reveals the same result: Fact_{NON} (79.14), Fact_{CRT} (72.46), Fact_{IND} (89.65), Fact_{SOC} (103.28) with $p < .000$. Therefore, hypothesis $H2_a$ can be confirmed. Furthermore, a correlation (Phi-coefficient) of $\phi = 0.307$ has been calculated between the two variables *task difficulty* and *correctness*. This result indicates that it is more likely that participants have answered hard tasks wrong than easy tasks. This in turn supports $H2_b$ and thereby, the procedure of the factual task generation (cf. Sect. 4.3.2.1, resp. 4.4.3.1) is supported as well. The corresponding cross tables regarding the two variables *task difficulty* and *correctness* for all tasks but also for the individual search blocks are given in Tab. 5.16 and Tab. 5.17 respectively[20]. Next, the average time to answer the tasks was calculated. If the variable *correctness* is not considered, users needed 86.72±61.29 sec. to answer *hard* questions and 53.24±38.42 sec. to answer *easy* questions. Considering the *correctness*, users needed 79.74±55.62 sec. to answer *hard* questions correctly and 51.55±37.35 sec. to answer *easy* questions correctly. That result confirms $H2_c$. Nevertheless, it has to be noted that the difference between the times to answer *hard* and *easy* tasks is not huge, what was already indicated by the relatively low correlation of 0.307. After testing for the three sub-hypotheses, the following has to be mentioned: Of course, it is not surprising that $H2_a$, $H2_b$ and $H2_c$ can be confirmed, but this

20 As a remark, the sum of correct and wrong answers in Fact_{NON} (619) in Tab. 5.17 is not equal to the number of fact-finding search sequences in Sect. 5.3.1 (717) because users could also choose "I don't know" as answer. Furthermore, the task processing of the current factual search tasks was aborted if the time limit of a block was reached. Both cases were counted neither as correct nor as wrong.

Table 5.16: Cross table for the variables *task difficulty* and *correctness* regarding the total number of answered factual tasks from *US-III*.

	correct	wrong	total
easy	1708	519	2227
hard	626	721	1347
total	2334	1240	3574

Table 5.17: Cross table for the variables *task difficulty* and *correctness* regarding the total number of answered factual tasks from user study *US-III* but separated to the corresponding search tasks blocks. The arrow symbol indicates an increasing or decreasing error rate for Fact_{CRT}, Fact_{IND} or Fact_{SOC} regarding the baseline block Fact_{NON}. E.g. errorRate($\text{Fact}_{NON-hard}$) = 144/(130+144) = 52.5% and errorRate($\text{Fact}_{SOC-hard}$) = 212/(153+212) = 59.5% and therefore, $\text{Fact}_{SOC-hard}$ is marked with ↑.

	Fact_{NON}		Fact_{CRT}		Fact_{IND}		Fact_{SOC}	
	correct	wrong	correct	wrong	correct	wrong	correct	wrong
easy	281	64	453	155 ↑	489	130 ↑	485	170 ↑
hard	130	144	169	188 ↑	174	177 ↓	153	212 ↑
total	411	208	622	343	663	307	638	382

tests are important because the results confirm that the data set of *US-III* comprises natural (seeking) behavior and therefore, affirm the applicability for any future investigation. In Fig. 5.20 the average durations to answer factual search tasks are illustrated for "easy" (green) and "hard" (red) factual search tasks. The diagram shows that most of the "easy" tasks have been answered faster by the users than "hard" tasks, what again conforms $H2_c$.

In a second step to investigate the influence of motivational goals, the four several fact-finding search blocks will be examined. Finally, it will be shown that the user's motives indeed have a significant impact on specific aspects of the search behavior while other aspects are unaffected. In Tab. 5.18, the average number of questions N and the average time t to answer a single question within a block are listed. All N in search blocks with goal, i.e., Fact_{CRT}, Fact_{IND} and Fact_{SOC}, are increased and all t are decreased in contrast to the baseline block Fact_{NON} without goal. This is an indicator that the users have been motivated and/or in a hurry to solve questions in blocks with goal.

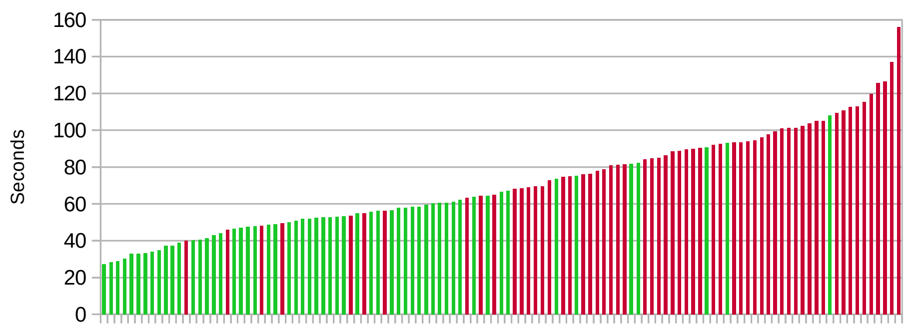


Figure 5.20: Average durations to answer factual search tasks from *US-III* (considering all search blocks) ordered increasingly. Green bars represent "easy" and red bars represent "hard" search tasks.

Table 5.18: Block statistics (mean value and standard deviation). Number of: questions answered (N_q), questions answered correctly ($N_{correct}$), hard questions answered correctly (N_{chard}), easy questions answered correctly (N_{ceasy}); Time to answer: one question (t_q), one question correctly ($t_{correct}$), one hard question correctly (t_{chard}), one easy question correctly (t_{ceasy}).

	N_q	$N_{correct}$	N_{chard}	N_{ceasy}	t_q	$t_{correct}$	t_{chard}	t_{ceasy}
Fact_{NON}	5.3±2.2	3.9±2.3	2.7±1.6	5.0±2.3	92.0±67.5	81.0±59.1	99.3±61.4	72.6±56.1
Fact_{CRT}	8.2±3.6	5.8±3.6	3.3±2.2	8.0±3.1	61.7±43.7	56.1±39.1	72.1±46.9	50.2±33.9
Fact_{IND}	8.2±3.4	5.9±3.8	3.1±1.9	8.5±3.3	61.1±45.7	55.6±41.5	78.2±55.7	47.6±31.4
Fact_{SOC}	8.7±3.7	5.7±4.0	2.8±1.7	8.5±3.6	58.3±45.3	51.4±37.3	73.1±54.9	44.5±26.2

Examining the results in Tab. 5.18 further, reveals that the total number of correct answered easy questions was relatively more increased than the number of correctly answered hard questions. Nevertheless, the relative task correctness in general is decreased, respectively the error rate is increased, cf. Tab. 5.17 indicated by the ↑ symbol. For example, the error rate for easy questions in Fact_{NON} is 18.5% and for blocks with goal in average 5.6% higher. Furthermore, the error rate for hard questions in Fact_{NON} is 52.5%, and the remaining block is Fact_{CRT} 52.6%, Fact_{IND} 50.4% and Fact_{SOC} 59.5%. That is, if user's motives are addressed in terms of a goal setting, this additional condition causes to decrease the time users spent on a task but also increases the error rates with one exception. The error rate for hard questions in the individual search block Fact_{IND} is decreased. This is an indicator that the Individual Reference Norm (IND) can causes a more attentive task processing. Furthermore, the error rates in Fact_{SOC} were increased by 7.0%, which is an indicator that the Social Reference Norm (SOC) can cause a more careless task processing in contrast. Finally, a statistical analysis with the Tukey's range test regarding N (cf. Tab. 5.18) reveals that there is a significant difference between the block without goal (Fact_{NON}) and blocks with goal (Fact_{CRT}, Fact_{IND} and Fact_{SOC}). However, in between the blocks with goal, no significant differences could be found. Analyzing the state proportions also no significant differences for the fact-finding search blocks could be found. That is, even if the motivational goals influences the task processing speed, the executed search strategies are not influenced.

5.4.2.2 *User's Personality*

As described in the introduction and motivation to include the user's personality in *US-III* (cf. Sect. 4.4.1.2), personal traits have been obtained by the *NEO Five Factor Inventory*. Comparing the five factor's mean (and standard deviation) of the participants with the values of Körner et al. [113], shows that the user group in *US-III* has a relative

Table 5.19: Comparison of the NEO five factors of personality of *US-III*'s participants with the values of the population of the Federal Republic of Germany (FRD) in terms of mean (and standard deviation). Used abbreviations are *Neuroticism* (N), *Extraversion* (E), *Openness to experience* (O), *Agreeableness* (A) and *Conscientiousness* (C).

	N	E	O	A	C
US-III	1.73±.60	2.35±.55	2.59±.54	2.56±.48	2.62±.62
FRD [113]	1.62±.62	2.20±.50	2.05±.46	2.54±.47	2.71±.55

strong agreement with the population of the Federal Republic of Germany in general, cf. Tab. 5.19. Each of the five factors is measured by twelve corresponding items in the questionnaire on a five point Likert-Scale from "strong refusal" (0) to "strong agreement" (4). The biggest difference to [113] considering the means is the .54 higher value on openness to experience in *US-III*.

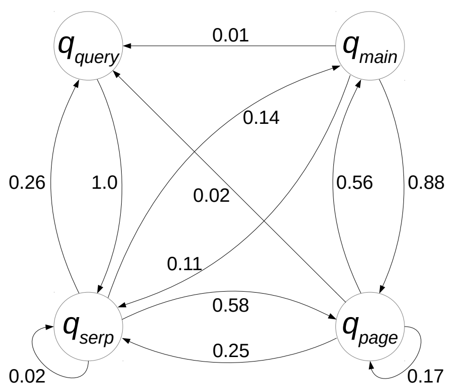
Applying the One-sample KST shows that all five factors are normal distributed[21]. That is, the values are distributed around one center point and therefore, the user groups can and will be split into two to train models on the corresponding sub groups. For the analysis of the five personality factors in context of ES, the user group of *US-III* is divided in two groups regarding each of the five factors. However, the split was not done on the mean but the median for the corresponding factor. For instance, the median for *openness to experience* is 2.58 in *US-III* what resulted into 64 participants with higher (H) and 51 participants with lower (L) values on *openness to experience* than the median. The reason for that choice is that the mean usually is vulnerable to outliers what can have negative influences to the model training. In contrast, the median is more robust at this point. To train (Hidden) Markov models, in this section, the full state space $Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$ (as in Sect. 5.1.2) is used. Including the *Main* state (the *quiz tab*) is motivated by the capability to compare the user's five personality factors on all available states. Since the focus of investigation here is to analyze the influence of personality factors on EIS, the models (at first) are trained on the interactions induced by the ES tasks, not the factual tasks. That is, one model $\theta^H_{<f>}$ for the exploratory search behavior from users with higher (H) values than the median of the corresponding personality factor $< f >$ have been trained and one model $\theta^L_{<f>}$ for the users with lower (L) values have been trained. Comparing the (1-st order) Markov model's components A (and π) however revealed only marginally differences on all five factors N, E, O, A, C. Taking into account the state dwell times under application of the KST also reveals only weak

21 Applying the One-sample KST to the remaining user variables such as intelligence and sensation seeking (with a significance level of $\alpha = 0.01$) confirms the normal distribution as well.

Table 5.20: Mean dwell times on each state from Q in sec. according to the models θ_O^H and θ_O^L from *US-III* using all Expl of users with higher (H) and lower (L) values on *Openness to experience* (O). Furthermore, the p-Values of the corresp. KST are given.

	Query	*SERP*	*Page*	*Main*
θ_O^H	3.8	9.3	21.9	24.9
θ_O^L	4.4	8.5	19.1	24.0
p-Value	0.497	0.0552	2.9e-06	0.0023

indications of different behavior. For example, the models θ_O^H and θ_O^L are illustrated in Fig. 5.21 and Fig. 5.22. The state durations of θ_O^H and θ_O^L are given in Tab. 5.20. The biggest (but still small) differences can be noted regarding the *SERP* and *Page* states but only allow to formulate the following relative week statements: Being on a SERP, users with higher values on *Openness to experience* more likely click on a result to visit a web page ($a_{serp\ page} = 0.58$) than users with low values on *Openness to experience* who rather switch to the *Main* state ($a_{serp\ main} = 0.19$). Furthermore, being on a web pages, users with higher values on *Openness to experience* more likely switch to the *Main* state ($a_{page\ main} = 0.56$) and return ($a_{main\ page} = 0.88$) than users with low values on *Openness to experience* who rather click on links on the web pages and hence, stay on *Page* ($a_{page\ page} = 0.25$). Considering the dwell times, user with higher values on *Openness to experience* spend also slightly more time on SERPs, Web Pages and the *quiz tab* than users with lower values on *Openness to experience*. However, as mentioned, the indications for these three statements are not strong. Considering the other personality factors, basically for the traits on *Neuroticism*

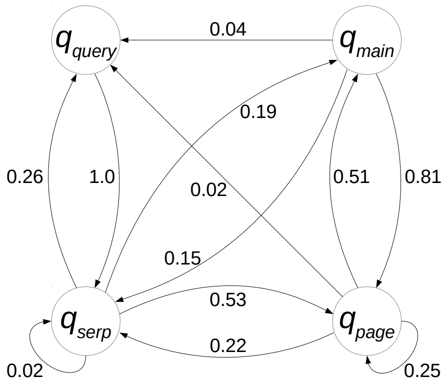


$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.26 & 0.02 & 0.58 & 0.14 \\ 0.02 & 0.25 & 0.17 & 0.56 \\ 0.01 & 0.11 & 0.88 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure 5.21: Illustration of the 1st-order Markov model θ_O^H as graph (left) and it's corresponding components Q, A and π (right) trained on the data set from *US-III* using all Expl of users with higher (H) values on *Openness to experience* (O).



$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.26 & 0.02 & 0.53 & 0.19 \\ 0.02 & 0.22 & 0.25 & 0.51 \\ 0.04 & 0.15 & 0.81 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure 5.22: Illustration of the 1st-order Markov model θ_L^H as graph (left) and it's corresponding components Q, A and π (right) trained on the data set from *US-III* using all Expl of users with lower (L) values on *Openness to experience* (O).

and *Conscientiousness* similar (not strong supported) but cognizable indications and corresponding statements can be found, e.g.: Being on a SERP, for users with higher values on *Neuroticism* it is more likely to enter a (new) query than for users with lower values on *Neuroticism* who more likely click on a result to visit a web page. Furthermore, users with higher values on *Conscientiousness* spend more time on the *Main* state and less time on SERPs than users with lower values on *Conscientiousness*. The complete set of Markov model components and corresponding state dwell times can be found in the Appendix A, Section A.4.

Even though the differences are marginal, it has been tested, whether the (Hidden) Markov models are able to classify users with high, respectively low, values on the five personality factors with the motivation to provide appropriate support for users in the future. Applying the methodology and gained knowledge from Section. 5.2, results to the following model specification: To include the state durations,

Table 5.21: Confusion Matrix for the classification of the 2nd-order HMMs regarding ES using a median split of users with higher (H) and lower (L) values regrading the corresp. personality factor.

	Real Class									
	N		E		O		A		C	
	H	L	H	L	H	L	H	L	H	L
Classified as H	32	29	28	30	44	33	38	29	26	25
Classified as L	29	25	30	27	20	18	26	22	35	29
Accuracy	.50		.48		.54		.52		.48	

HMMs with an fitted exponential distribution on each of the four states are used. Applying the AIC and BIC revealed the 2nd-order for the split on all five factors as a supported choice (similar to *US-II*) but again with a tendency to the 1st-order. For the classification here, HMMs of the 2nd-order are used. In a next step, the familiar 5-fold cross validation setting with 2000 repetitions for the HMMs was applied. Tab. 5.21 lists the results. As expected, the classification accuracies are low. Only for *Openness to experience* there are slightly positive classifications tendencies with .54 but this includes a high false positive rate for the "higher" class (33 of L are classified as H). The classification of other factors results basically in guessing the classes with a balanced true and false positive resp. negative assignment. Disregarding the dwell times (i.e., using a 2nd-order Markov model) leads to similar results. Only *Openness to experience* is experiencing a small boost to an accuracy to .56. Applying the same procedure (i.e., 2nd-order HMMs and Markov models) on the four factual search blocks leads to similar values. The highest accuracy values on all five factors have been received by 2nd-order Markov models on the Fact_{IND} block again with an maximum for *Openness to experience* with .60, cf. Tab. 5.22. Summing up, after the analysis of user's variable personality it has to be concluded that the user's (exploratory) seeking behavior basically does not differ considering the user's the traits at least regarding interactions (modeled as transitions) and state dwell times. Furthermore, the identification of users with higher or lower values on personality factors using search interactions is basically not reliable. There is only one small indication that users with higher values on *Openness to experience* can be estimated if they perform a fact-finding search activity and if they are motivated to find a self chosen number on information (in reference to the own performance, cf. Fact_{IND}) during their search.

5.4.2.3 *Aspects of Intelligence*

In addition to the user's traits, also aspects of intelligence have been obtained in *US-III*, measured by three sub-tests of the WAIS. These

Table 5.22: Confusion Matrix for the classification of the 2nd-order Markov models regarding Fact_{IND} using a median split of users with higher (H) and lower (L) values regrading the corresp. personality factor.

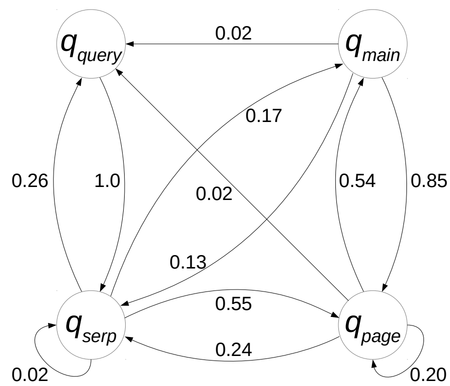
	Real Class									
	N		E		O		A		C	
	H	L	H	L	H	L	H	L	H	L
Classified as H	37	27	27	28	45	26	37	26	31	23
Classified as L	22	25	28	28	18	22	25	23	27	30
Accuracy	.56		.50		.60		.54		.55	

Table 5.23: User's mean (and standard deviation) values for the intelligence aspects *Similarities* (Si), *Symbol Search* (Sy) and *Letter-Number Sequencing* (LN) obtained by the WAIS in *US-III*.

Si	Sy	LN
24.25±4.45	41.77±8.53	11.10±2.90

sub-tests with their corresponding scales are: *Similarities* (Si), on a point scale from 0 to 33; *Symbol Search* (Sy), on a point scale from 0 to 60; and *Letter-Number Sequencing* (LN), on a point scale from 0 to 21. Tab. 5.23 lists the mean values for each of the three aspects.

Similar to the analysis of the five personality factors, (first) the ES sessions of *US-III* are used and the users are divided in two groups by a median split for the corresponding aspect of intelligence leading to a group of users with higher (H) and a group with lower (L) values. Again, the full state space $Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$ (as in Sect. 5.1.2) is used to train the (Hidden) Markov models. Comparing the (1-st order) Markov model's components A (and π) however revealed also marginally differences on all three intelligence aspects Si, Sy and LN. The same holds for the state dwell times under application of the KST. Exemplary, the models θ^H_{LN} and θ^L_{LN} are illustrated in Fig. 5.23 and Fig. 5.24. The state durations of θ^H_{LN} and θ^L_{LN} are given in Tab. 5.24. The biggest (but still small) differences can be noted regarding the *SERP* states where users with higher values on LN more likely switch to the *Main* state ($a_{serp\ main} = 0.17$) than users with low values on LN who rather click on result to visit a web page ($a_{serp\ page} = 0.58$). However, as mentioned, the indications for these three statements are marginal. The complete set of Markov model components and state



$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.26 & 0.02 & 0.55 & 0.17 \\ 0.02 & 0.24 & 0.20 & 0.54 \\ 0.02 & 0.13 & 0.85 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure 5.23: Illustration of the 1st-order Markov model θ^H_{LN} as graph (left) and it's corresponding components Q, A and π (right) trained on the data set from *US-III* using all Expl of users with higher (H) values on *Letter-Number Sequencing* (LN).

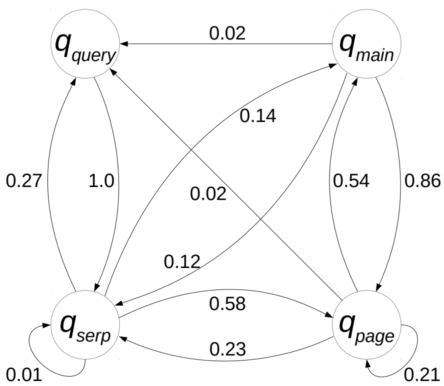
Table 5.24: Mean dwell times on each state from Q in sec. according to the models θ^{H}_{LN} and θ^{L}_{LN} from *US-III* using all Expl of users with higher (H) and lower (L) values on *Letter-Number Sequencing* (LN). Furthermore, the p-Values of the corresp. KST are given.

	Query	*SERP*	*Page*	*Main*
θ^{H}_{LN}	4.0	8.5	18.7	24.2
θ^{L}_{LN}	4.1	9.6	23.3	25.0
p-Value	0.2991	0.0292	0.0434	0.3921

dwell times can be found in the Appendix A, Section A.5. An analysis of the possibility to identify users with high, respectively low, values on the three aspects of intelligence in terms of classification have been implemented as well but similar to the personality factors, the results revealed that an identification is basically not reliable. Summing up, after the analysis of user variable intelligence it has to be concluded that the user's (exploratory and fact-finding) seeking behavior basically does not differ considering the user's verbal comprehension, processing speed and working memory (as measured by the WAIS) at least regarding interactions (modeled as transitions) and state dwell times.

5.4.2.4 *User's Sensation Seeking*

The last user variable considered here is the sensation seeking. As described in the Section 4.4.1.2, the Sensation Seeking Scale (SSS) measures the four sub-factors *Thrill and Adventure Seeking* (TAS); *Experience Seeking* (EXS); *Disinhibition* (DIS); and *Boredom Susceptibility* (BOS);



$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.27 & 0.01 & 0.58 & 0.14 \\ 0.02 & 0.23 & 0.21 & 0.54 \\ 0.02 & 0.12 & 0.86 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure 5.24: Illustration of the 1st-order Markov model θ^{L}_{LN} as graph (left) and it's corresponding components Q, A and π (right) trained on the data set from *US-III* using all Expl of users with lower (L) values on *Letter-Number Sequencing* (LN).

Table 5.25: User's mean (and standard deviation) values for the sensation seeking values *Thrill and Adventure Seeking* (TAS), *Experience Seeking*(EXS), *Disinhibition* (DIS), and *Boredom Susceptibility* (BOS) obtained by the SSS in *US-III*.

TAS	EXS	DIS	BOS
.465±.24	.63±19	.62±.28	.38±.19

whereby in particular, the second sub-factor EXS appears to be relevant for investigations on ES. Each of the four sub-factors is measured on a scale from 0 to 1. Tab. 5.25 lists the user's mean vales (from *US-III*) for each of the factor. Worth mentioning are the relative high values for EXS with a relative low variation.

However, similar to the analysis of personality and intelligence aspects before, the differences in the search behavior between two user groups with high (H) and low (L) values after a median split in terms of interactions sequences and dwell times are minimal. For completeness, the set of all Markov model components and state dwell times regarding each of the four sub-factors can be found in the Appendix A, Section A.6. Applying the classification setting with different model parameters, i.e., (Hidden) Markov models on different orders trained on the available tasks blocks, could, as expected, not clearly distinguish the user groups H and L. Nevertheless, the highest values on the EXS factor could be reached with 1st-order Markov models in the tasks block without goal Fact_{NON}, cf. Tab. 5.26.

Table 5.26: Confusion Matrix for the classification of the 1st-order HMMs regarding Fact_{NON} using a median split of users with higher (H) and lower (L) values regrading the four sensation seeking sub-factors.

	Real Class							
	TAS		EXS		DIS		BOS	
	H	L	H	L	H	L	H	L
Classified as H	39	26	42	21	37	19	24	24
Classified as L	31	17	27	23	35	22	35	30
Accuracy	.50		.56		.52		.48	

5.5 SYSTEMS TO SUPPORT EXPLORATORY INFORMATION SEEKING

As outlined in the thesis' motivation (Sect. 1.1), current Information Retrieval (IR) systems are already quite successful in providing relevant results if users are able to specify their information need precisely. However, if users are not able to specify their request, e.g., because they are not familiar with the (search) domain, the demand to explore increases. Thereby, the search behavior does not have an explicit fact-finding character anymore. Unfortunately, current search systems still lack for sufficient approaches to support the users if exploration becomes necessary. In their report, White et al. [188] confirmed this with their statement:

> *"However, [search engines] do not work well in situations where users lack the knowledge or contextual awareness to formulate queries or navigate complex information spaces."* (p. 1),

what in turn is based on the observations of Bates [15]. To provide appropriated user support (e.g., for ES) and thereby, to increase the usefulness of IR systems, it is reasonable to adapt the provided support to each search individually. This first requires to understand and identify the (search) mode that should be supported before the actual supporting means can be designed and provided. For the search paradigm of ES, these first challenges have been addressed in the former parts of this thesis. In the present section, the findings of the thesis will be used to discuss selected supporting approaches for ES (H4). On the one hand, this includes approaches regarding the information systems' front-end, i.e., the visualization of the search (result) items, available settings and supporting means in the SUI (Sect. 5.5.1). On the other hand, this also includes approaches regarding the back-end aspects, i.e., the underlying search algorithms and utilized data bases, considered in context of ES (Sect. 5.5.2). This procedure is also conform with the suggestions of Athukorala et al. [10], who name:

> *"three aspects of IR systems that can be tailored: interface design, retrieval algorithm design, and user model design." (p. 16).*

While the modeling was already addressed in the former part(s) of the thesis as main contribution, the interface design (front-end) and algorithm design (back-end) are discussed in the following. Nevertheless, it is self-evident that the division in front- and back-end is rather a theoretical one. In practice, of course, the back-end algorithms mostly restrict their calculation to aspects that are relevant for the front-end and the means in the front-end control and are dependent on the calculation's internal (document) representations in the back-end.

5.5.1 *Front-End Aspects*

The presentation of information in the SUI, in particular on the SERP, has a strong influence to the user's seeking behavior and therefore, is a crucial aspect. The study of Cutrell and Guan [48, 49] on eye-tracking of users performing a web search showed that providing additional information to the snippets in the SERP improved the user performance for so-called *informational tasks*. In contrast to that, using the same additional information in so-called *navigational tasks* did not increase but decreased the user performance. That is, the same feature or supporting mechanism can have positive but also negative effects on the users' search performance, depends on the current search tasks, and thus has an influence to the applied search activity. This (further) confirms the need for techniques to distinguish between user's search activities to provide appropriated information or supporting means in the SUI. The *navigational tasks* (utilized in [48, 49]) aim to locate specific web pages, e.g., to buy a product. Therefore, the *navigational tasks* are clearly related to lookup (as investigated in this thesis). Furthermore, "Navigation" is also listed as related activity for lookup in Marchionini's definition (cf. Sect. 2.6.1). The *informational tasks* (utilized in [48, 49]) aim to find a piece of information on a web page and therefore, can in theory induce both, lookup and ES depending on the certain task (complexity) and the knowledge of the user. However, the type of informational tasks in the study of Cutrell and Guan [48, 49] are strongly related to the type of fact-finding tasks as used in this thesis (cf. Sect. 4.3.2.1 and 4.4.3.1) and hence, apparently tend to induce lookup related search activities. As a matter of fact, this actually raises the demand to distinguish even between lookup related search activities, i.e., navigational vs. informal (as used in Cutrell and Guan's study), for future work.

Similar to the investigation above, Lorigo et al. [127] showed that informational searches take more time on average then navigational searches. Hypothesizing that (more complex) informal tasks require more elaborated interactions, more information gaining regarding the topic and more effort in general than navigational tasks, the results of Lorigo et al. [127] are in accordance with the results of this thesis in so far that: (1) the dwell time on each state in ES tasks was longer than the state dwell times on fact-finding tasks, cf. Tab. 5.1; and (2) the time to answer hard factual tasks was longer than the time to answer easy factual tasks, cf. Section 5.4.2.1, $H2_c$. Therefore, the time users spend with the SUI (in addition to the interaction patterns in general) provides a valuable instrument for differentiation to finally adapt single elements of the SUI or even change the whole SUI.

In order to support ES task processing, several dedicated SUIs already have been proposed. For instance, Nitsche and Nürnberger [138] described an interface, called *Trailblazer*, that utilizes a domino metaphor for the SERP visualization. That is, the results for a query are illustrated as horizontal thumbnails and clicking on a result generates a specification of the query regarding the clicked result leading to a new vertical (i.e., orthogonal) result visualization. This process can be continued and allows the user to keep track of the search process as an iterative specification during the ES. In designing the SUI for the *Trailblazer*, the three-dimensional axes character from Noël et al. (cf. Sect. 2.6.2) has been considered by Nitsche and Nürnberger [138] as well. A feature that allows the user to traverse trails to the document set was also suggested by White and Roth [189]. Ruotsalo et al. [162, 163] implemented a radial result visualization in the so-called *Internet Radar*, and utilize the interactions of the users to estimate their (search) intents. The more relevant the (intended) result, the more it is placed in the center of the visualization while the users also have the option to provide feedback for the system. The results of the implemented user study suggest that the interface can be useful for directing an ES and improve the user's performance. Furthermore, Huurdeman and Kamps [95] investigated different exploratory features of a user interface, e.g., query refinement or histories, workspaces and progress (cf. *Trailblazer*). They suggest that there are differences in the interaction flow with SUI features depending on the stage of ES.

5.5.1.1 *Ontology supported Exploratory Search*

In designing approaches to support ES, the (search) task domain and the available means are central points. Since users want to investigate and learn about the domain during their exploration, the number and quality of the provided resources are important aspects as well. This includes the individual result "documents" but also the type of information source in general. To address this issue and to contribute to the aspect of manifold sources, in Kotzyba et al. [115], an ontology supported SUI for ES was proposed. During their exploration, users have access to a multitude of information sources simultaneously. According to the given (search) tasks domain, in addition to a "common" search engine, also further domain specific information sources can be provided for the user. In particular, the SUI in [115] was exemplary implemented for the search and exploration in context of the fitness domain, i.e., users here are seeking for physical training exercises.

To illustrate the system's components, the exemplary considered scenario is briefly described in the following: To achieve and maintain a healthier life, a user wants to create a fitness plan and therefore, uses an interactive system designed to provide personal training assistance. The system has access to ontology based knowledge about the fitness domain that builds the basis for suggestions of training plans. This includes a list of physical training exercises that can be combined. Furthermore, the fitness plan is open for updates from external information sources, e.g., available in the Internet, because a "One size fits all" approach for individual training would not be appropriated. That is, users can add a new exercise to the system's knowledge base. Such an option is in particular important if a user has specific health demands. In this scenario, the user initiates an online search to explore and identify new exercises which suit to the user's physical conditions. The training system supports the user by providing a SUI that also has access to several knowledge bases.

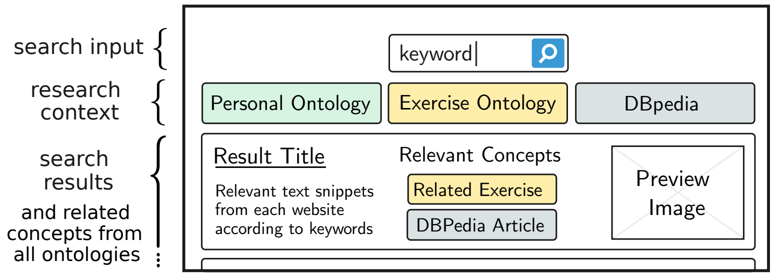


Figure 5.25: Schematic illustration the ontology supported SUI consisting of a search input in the top, the related (found) concepts from the ontologies as research context below and the results of the web search engine including relations to the ontologies and a thumbnail in the center.

A schematic illustration of the SUI is illustrated in Fig. 5.25 and a screenshot of the actual interface is given in Fig. 5.26. In the beginning, the user can choose from search resources and ontologies, cf. Fig. 5.26, left. The search resources can be web search APIs (e.g., from Bing[22] or Google) or local document directories, such as folders containing own documents regarding a certain topic. The ontologies can be provided by the (training) system or other external sources. After entering a search query, the SUI provides results from all chosen information sources. Found concepts from the chosen ontologies are shown directly below the search input. In the prototypical implementation, an exercise ontology (from the hypothetical training system), the user's personal ontology and the *DBpedia*[23] ontology have been selected. In the own ontology, the user can store and retrieve personal entries, e.g., exercises in context of spine therapy. The DBpedia ontology provides a general knowledge base that supports the user's ES in addition to learn and investigate the domain. For each found ontology item, related information can be displayed, such as title and the corresponding super concepts. A mouse over allows to present even more information, e.g., how to execute a physical exercise or a snippet from a DBpedia entity. Results from the web search APIs are presented in the center of the SUI. Each result comprises of a title, a summarizing text, the web pages' URL, a thumbnail and a list of so-called annotations (also called Relevant Concepts). These annotations connect the several information sources because they represent terms that appear in the content of the web page and they represent concepts found in the ontologies. The annotations are also colored in accordance with ontologies, where the concept appears. For example, the web results in Fig. 5.26 are all enriched with gray annotations, which represent concepts from the DBpedia ontology (cf. screenshot on the left: Annotations: Wikipedia). If an annotation is hovered with the mouse courser, the term and it's surrounding context on the Web page is shown. Than the user

22 https://www.bing.com/
23 https://wiki.dbpedia.org/

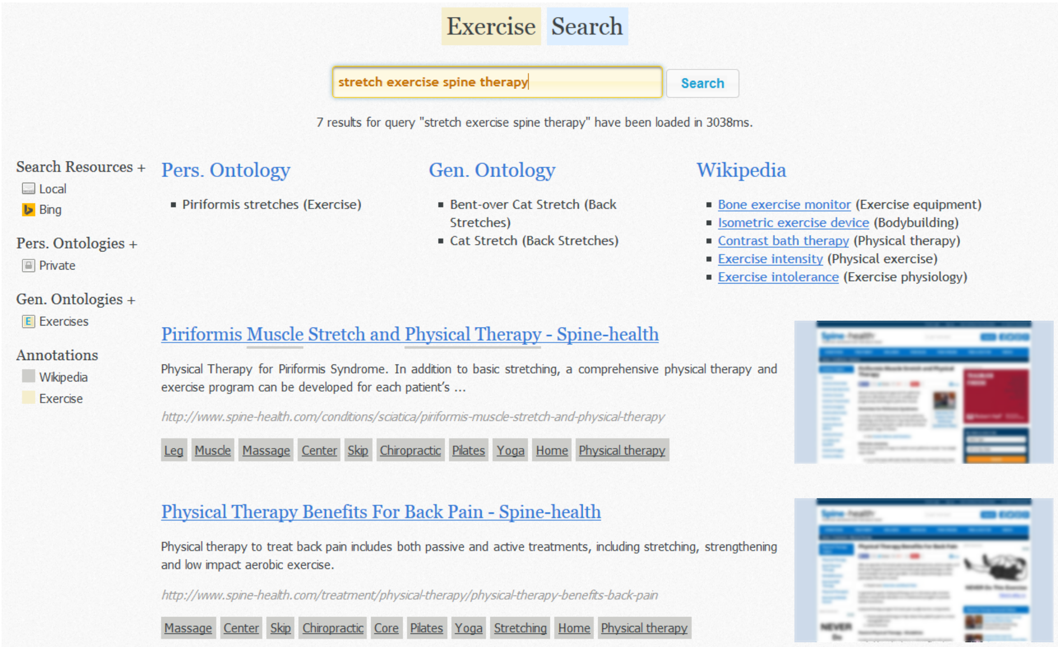


Figure 5.26: Screenshot of the ontology supported SUI consisting of a search input in the top, the related (found) concepts from the ontologies as research context below, available search settings and information sources on the left and the results of the web search engine including relations to the ontologies and a thumbnail in the center.

can decide to investigate (and learn more about) the retrieved term on the web page or in the ontology. This allows the user to acquire and discover knowledge from different sources as well as to analyze and compare the results what in turn are crucial aspects of learning and investigation in context of ES, cf. Marchionini: Section 2.6.1. The ontology supported SUI further allows to select and transfer found information (e.g., new physical exercise) to the system in order to check whether the found information (or physical exercise) is already present in one of the ontologies. This facilitates Marchionini's aggregation, respectively integration aspect. In the given example scenario, the explored content can be even directly used for the (training) planning aspects. By extending the available sources for the information (seeking) behavior, also the *Source characteristics* (described by Wilson, Sect. 2.3) are addressed in so far as the sources differ in their properties what enables the user to proceed the search in the preferred directions. Considering the *Source characteristics* in terms of document types, the usefulness of book records and documents in the *Portable Document Format* was exemplary considered by Ingwersen et al. [99]. As a result, the usefulness of book records appeared to be higher than for documents in the *Portable Document Format*. However, the smaller number of book records for the given data set may have influenced this result. Nevertheless, the fact that different document types have been perceived as differently useful (for the given tasks) underline the need to consider that *Source characteristics* as well in designing the support for user's ES[24].

Summing up, domain ontologies can be helpful to support ES and enrich a standard web based search if the specific search domain is known and the corresponding ontologies are available. This complements the spectrum of application for ontologies in context of IR because they also have been used to improve navigation in the Web. For instance, Crampes and Ranwez [46] discussed strategies and formulated principles of ontology based conceptional navigation. Furthermore, an ontology-based FAQ system has been proposed by Yang et al. [199] that processes queries to enhance ranking techniques. Yang and Ho [200] showed that even user models for interface agents can be improved by using a domain-specific ontology. Considering the example scenario above, the personal ontology here can be a promising point of entry for future investigations in that direction.

Last but not least, it has to be mentioned that the here described prototype for ontology supported SUI has not been explicitly tested in a user study to measure the benefits for ES task processing. The reason for this was twofold: First, the objective evaluation of search systems and their effect on ES is an extensive topic (and therefore, a potential thesis) in itself. This is also confirmed by White et al. [190]:

> *"Determining whether an exploratory search system is effective is a challenge in itself. No metrics exist to determine how well a system supports exploration, yet users will undoubtedly be able to tell what works well for them." (p. 2)*

Second, the goal of this section (as later part in the thesis) is rather to discuss possible supporting approaches for ES to provide a selected collection and inspiration for future work.

5.5.1.2 *Exploratory Search for Young Users*

In addition to the search task domain and the available information sources, also the user group has to be considered in designing support for ES. With the increasing number of Internet users, also the number of young users with their specific demands on the search systems grow. As the analysis of the young users' search behavior in Section 5.4.1 showed, the behavior of children differs to adult users. For instance,

24 Furthermore, the authors of [99] analyzed (amongst others) the relation between the complexity of a work task (cf. Section 3.3, footnote, p. 55) and an usefulness assessment. The results show that (for both document types) highly useful documents are associated to tasks perceived as complex what confirmed the results to a former study [98]. In the context of ES, this is coherent in so far that for complex (ES) tasks, where users have little domain knowledge, the odds for a document to increase the users' knowledge are higher in contrast to tasks where users already have a certain expertise. That in turn can increase the perceived usefulness.

young users need far more time on the states *SERP* and *Page*. Furthermore, although the state distributions in free and exploratory search of young users is similar to the factual and exploratory search of adults, the dwell times showed a different picture. With their limited domain knowledge [94], young users more likely need to explore (new) search domains and require adequate SUIs. To address this demand, special or targeted search engines for children become more necessary. In the own publication [71], an overview of seven crucial aspects and corresponding challenges to support children's seeking have been proposed. It has been shown that the SUI *Knowledge Journey* (KJ, cf. Sect. 4.2.3) addresses each of the seven aspects with it's provided functionalities. In the following, the seven aspects will be revisited from the perspective of children's ES:

Requirements for young users

- *Emotional Support:* According to Erickson's theory of psychosocial development [62], children need emotional support. This can be achieved by guidance. In case of the KJ, the present guidance figure takes this role by representing a personification of the SUI itself and by providing help, e.g., via spelling correction. Furthermore, the possibility of personalization allow the young users to adapt the SUI in a way they feel comfortable with. The emotional support is likewise or even more important for young user's ES because of the increased amount on uncertainty. A guidance figure with a sufficient functionality here represents a suitable possible solution. For example, if a young user is probably performing an ES (because of an increased state distributions on web pages), the guidance figure could suggest a related (super) category to search for, e.g., derived from an ontology.

- *Language Support:* With their difficulties with keyboard typing [175] and their smaller domain knowledge [94], children more often have problems to (correctly) formulate a queries. A spelling correction and suggestion, e.g. via a guidance figure, is a suitable option for compensation. A navigation menu (as in the KJ) is a further approach and in particular, if a provided category corresponds to the (new) domain to explore. Especially in ES and it's inherent lack of vocabulary (also for adult users), the aspect of language support becomes important.

- *Cognitive Support:* Children pass through different stage in their development [143]. Therefore, in some of the stages (in particular the so-called *concrete operational stage* during the primary school) children's reasoning and abstract thinking is still limited. To get an easy and smooth access to (new) explored knowledge, metaphors (as the treasure hunt or space travel) as a connection to the physical world can facilitate a successful EIS. In addition to that, the provided content, such as, terms and search categories should also not be too abstract.

- *Memory Support:* Young users can process less information than adults [108]. Therefore, it is important to provide means for an overview of already found information. Especially in ES this is required to learn about and investigate the (new) domain. Similar to the text field in *US-II* and *US-III*, provided for the adults in their ES task processing, a SUI for children also should allow to track the exploration. In the KJ this is achieve by the storage functionality, e.g., the *treasure chest*. But also by the coverflow visualization, where the young users easily can scroll between the results, only a limited amount of information (exactly one result) is presented, what supports the memory aspect as well. Alternative visualizations such as the *Trailblazer* (with a limited amount of search results) are also possible solutions for children since they (additionally) allow users a result tracing.

- *Interaction Support:* To support children's motor skills, the SUI should (at least) alternatively allow simple interactions, e.g., touch and swipe. This requirement is independent of the underlying search task, such as ES. The KJ provides both, interactions with mouse and keyboard as well as touch. Touch interactions are more natural for children than the use of a mouse. Furthermore a voice-controlled SUI, as imitated in *US-I*, can allow an additional and preferred input modality and at the same time can address the emotional support requirement.

- *Relevance Support:* The studies of Jochmann-Mannak et al. [105] showed the difficulties of young users to judge the relevance of retrieved results regarding the current information need. To evaluate the relevance of results in ES is an inherent task for all user groups and to support young users in this aspect is especially challenging. One approach is to consider the representation of the results in terms of the related information. For example, in addition to a text based summary of the web page, the SUI can provide a thumbnail (cf. KJ) and further meta information. In the ontology supported SUI, also information regarding the concepts (from the knowledge base) have been shown as annotations what provides more context about the results[25]. A second approach to provide more context and hence, to easier estimate the relevance is to visualize the relation between the results. A standard result list represents only the relevance of the results to the query by a rank but does not considers the relation between

25 Also for children, more context, e.g., from ontology knowledge, can be helpful: For example, in *US-Ib*, one child searched for the animal lion (in German: "Loewe") but the search engine retrieved television devices from the brand "Loewe" in the first results. It took several iterations and trails for the child to recognize that the results do not have a relation to the animal. An ontology and disambiguation functionality here could help to solve this kind of issues faster.

the results. A two dimensional visualization such as the *Internet Radar* (explained in Sect. 5.5.1) overcomes this issue.

- *Diversity Support:* The inter-individual cognitive differences between children can be huge. Therefore, an individual support by the SUI becomes even more important. In Gossen et al. [74], so-called *evolving* SUIs for young users are introduced to address this challenge. However, for ES the unfamiliarity of the user with the search system (resp. SUI), was identified as an crucial point (in addition to the unfamiliar search domain and the increased task's complexity). Therefore, the adaption (or evolution) of the SUI is important on the one side to satisfy the young users' needs but also has to be applied carefully because adaptations in the interfaces while ES sessions can complicate the already more demanding seeking. Nevertheless, *evolving* SUIs are anyway designed to be long-term accompanied systems which do not change the SUI elements often and quickly.

Knowledge Journey Exhibit

In a final step, the *Knowledge Journey* SUI was implemented in terms of an information terminal for an exhibition [70]. A photo of the so-called *Knowledge Journey Exhibit* is given in Fig. 5.27. To demonstrate it's functionality, an age-adaptable version of the *Knowledge Journey* SUI was developed. The terminal enables a flexible adaptation of the SUI by a slider element to address changing requirements of users at different age groups. At the same time, the flexible SUI can be utilized to adapt to the demands of (young) users who are performing an ES because the adaptation also allows to change the number of presented results, the activation of thumbnails, length of results snippet, font size, etc. In the proposed *"three aspects of IR systems that can be tailored"* (cf. the begin of this Sect. 5.5), Athukorala et al. [10] further suggest

Figure 5.27: Photo of the *Knowledge Journey Exhibit* device [70] at the exhibition in Munich, Germany.

"Adjusting the number of result items shown per SERP." and *"Adjusting the length of results snippet according to task type."* for the interface design. The *Knowledge Journey Exhibit* as well fulfill this suggestion by the possible SUI adaptations described above.

5.5.2 *Back-End Aspects*

Not only the SUIs in the front-end, but also retrieval algorithms in the back-end of the information system can and should provide mechanisms to support ES. Since ES requires aspects of learning and investigation, ensuring a diversity of search results to cover many different perspectives regarding the (query) topic and thereby to get an overview of the domain is a promising approach. In particular, White and Roth [189] stated:

> *"To this end, exploratory search systems should offer collection overviews (glimpses), the ability to traverse trails through the collection (exploratory browsing), and document examination/retention." (p. 27)*

Possible options to traverse trails have been shown in the former SUI related section, e.g., in the *Trailblazer*. The examination of documents could be enriched, e.g., by the annotations of the ontology supported SUI (cf. Sect. 5.5.1.1). However, to support the overview and glimpsing aspect of browsing (also cf. Sect. 2.5.1), the *ESSENCE* system from Homoceanu et al. [92] should be mentioned. The system supports researcher in getting familiar with scientific literature by considering how paper's keywords change over time in the corresp. domain. In particular, the novelty of certain keywords is approximated by their usage in the documents on a yearly basis. An alternative approach to address an overview was proposed by Pratt and Fagan [149] with their dynamic categorization of search results. Their approach utilizes search result grouping, category selection and a hierarchic generation of categories what turned out to be more useful (in terms of answer finding and user satisfaction) than traditional clustering and relevance ranking techniques. Like the ontology supported SUI above (cf. Sect. 5.5.1.1), the information system in [149] uses a knowledge base in the background for the processing of the result presentation. In their studies on the enrichment of search results by semantic category information, Dumais et al. [55] could confirm that providing categories are more effective that list based result pages. That is, providing context information of the results for the users is a promising approach also for ES.

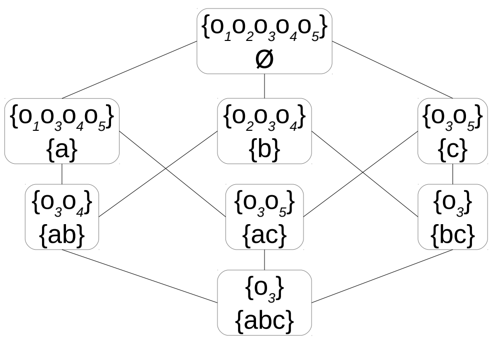
Graphs to provide search context

An alternative to provide context is to represent the search results as a graph. Here the nodes can be the results and links between the nodes represent a certain relation in between. One example for such an interactive exploration of graph representations is the so-called

Creative Exploration Toolkit (CET) from Haun et al. [86]. Although the CET is dedicated to explore and analyze complex graph structures, it can also be used to explore search results (e.g., from the Bing API) presupposed an additional graph structure of the results has been generated in the background. This underlying graph structure can be created from the scratch based only on the content of the nodes (e.g., web pages) or additional knowledge (bases), such as ontologies. However, such a search result graph structure can hold the additional value that is necessary for ES since the relations between the graph nodes (the links) may reveal similarities not easy to recognize in a common search engine the hence, these approaches can be used to provide new perspectives to the explored domain.

5.5.2.1 *Formal Concept Analysis to support Exploratory Search*

In the own publication [33], the graph visualization of the CET was used as well to present the results of a search system back-end process implemented for ES. In particular, the work in [33] describes a conceptual framework that is able to structure search results utilizing a so-called concept lattice. This lattice is created by methods from the Formal Concept Analysis (FCA) [65]. To apply methods of FCA in the area of IR however is not new. A survey of related work can be found in [148]. The underlying idea of FCA is to arrange collections of objects (e.g., result web pages) with their properties or attributes (e.g., size, topic, etc.) in an object-attribute model, the lattice. Fig. 5.28 exemplary illustrates a concept lattice of five arbitrary objects and their three binary attributes. The concept lattice itself is a graph. Each node of the concept lattice graph consists of a set of objects and a set of the corresponding attributes the objects have. The top node contains all objects disregarding the attributes and the node at the bottom contains only the objects, which have all attributes, what is true in the example only for object o_3. The nodes between represent all possible sub-set combinations regarding the attributes and including the corresponding objects. Therefore, the lattice represents a hierar-



	a	b	c
o_1	x		
o_2		x	
o_3	x	x	x
o_4	x	x	
o_5	x		x

Figure 5.28: Concept lattice on the left according to the five objects o_1 to o_5 and their corresp. attributes a, b or c in the table on the right.

chical structure where the edges describe a relation of generalization, respectively specification, within the nodes according to the attributes. Consequently, the nodes in the lattice represent clusters of objects with the same attributes.

The approach of concept lattices and FCA can be transferred to the IR domain, as exemplary implemented in the system described in [33]. The result is a search engine that provides a set of query related web documents (the search results) which are organized by a hierarchy of clusters (the lattice nodes). The clusters in turn comprise of documents with equal, respectively similar, attributes what is beneficial in particular for ES. To generate the concept lattice, a mapping from the search results (the web pages) to the object-attribute model is necessary. The derivation of web pages attributes here can be designed in manifold ways. Obviously, the more the attributes represent the results, the better. From the perspective of IR, for example, the attributes can describe the presence of terms within the web document, the term's weight in the underlying document representation model but also meta-data can be a representative property. However, in the origin of FCA, attributes are binary but in a IR setting this feature may be limiting. Hence, in [33], a fuzzy approach on the attributes (e.g., to represent the term weights) was used to allow and place also similar (not only equal) objects in one cluster. As mentioned above, the visualization of the CET was used as one possible way to present the resulting lattice graph. In contrast to the original visualization in the CET, the lattice graph here represents clusters (the concepts) of search results in each node. With the available generalization, respectively specification, within the nodes on the edges of the lattice, such a framework provides users means to get an overview of a search domain what is a desired functionality for systems to support ES. Finally, there are different approaches to use the concept lattice in this context. The lattice can be used directly to allow the user an exploration of the full lattice and all of it's concepts. Alternatively, the lattice can be reduced (e.g., by generating a hierarchy) before visualizing the clusters with the goal to minimize the amount of information and therefore, not to overwhelm the user. Methods form FCA itself can also be applied to the lattice, e.g., to further analyze relations between the objects and concepts or to order the result sets according to the structure of underlying concept lattice and it's properties. Even though the system described in [33] was not intended to propose a search system with an innovative new way to interact with document results, it could highlight the value of concept lattices and FCA also for ES. In the publication of Butka et al. [34], the implementation of the FCA based system was further extended by an alternative, interactive graph visualization. The visualization involves different types of attributes regarding the search results with the goal to better understand the visualized concept lattice and hence, to further support ES.

5.5.2.2 *Deriving an Exploration Graph*

In addition to the representation of the search results as (concept lattice) graph in the back-end, also the exploratory seeking behavior of the users themselves can be represented as a graph. To model and interpret the behavior of users' exploratory search, in the own publication [180] a so called *exploration graph* is used. The *exploration graph* is the result of logged user interactions with a web browser during an exploratory search. Basically, the nodes of the graph are visited web pages and the edges represent user actions which created or traversed through the web pages. The presented data model in the back-end facilitates the storage and interpretation of users' ES and, e.g., can help to improve result ranking and recommendation given the current exploration state. However, there are several and essential differences to the graph based Markovian models described and used in the section 5.1 to 5.4 because the perspectives to the user search behavior is different: First, the Markovian models have a fixed number of nodes (the states) to calculate the transition probabilities in between and in particular focus on the interplay between the SERPs and the other stats. The number of nodes in the *exploration graph* is not fixed, correspond to the number of visited web pages (i.e., the graph is different for each ES). Furthermore, the SERPs in constructing the *exploration graph* take the role of connecting elements in the graph[26]. Second, the structure of the *exploration graph* is more related to explicit user interactions with the web browser but this in a high degree of detail: Up to eight different interactions (represented as the graph edges) are considered using an own defined browser history. The Markovian models in contrast do not differentiate between the transition "types" but are able to include different kinds of implicit user interactions as emissions. Third, the Markovian models have a strong focus on the probabilistic background while the *exploration graph* can rather benefits from graph-theoretical methods, such as diameter, density, number of connected components, etc. Nevertheless, both approaches have in common that they are based on cyclic, directed graphs.

In addition to the generation of a straight web browser based interaction graph (as default), the work in [180] also proposes an additional *interpretation* step. The interpretation function allows to include assumptions while generating the *exploration graph*. For example, Fig. 5.29 illustrates the differences between two exploration graphs. Both graphs have been created on browser interactions of nine users solving the "Home Heating" topic related exploratory search task Expl1[27]. The exploration graphs represent the sum of all iterations of the users, not individual search sessions. White nodes are web pages visited by exactly one user, black nodes represent web pages which

26 The possible way to explicitly differentiate the SERPs is also described in [180].

27 This additional small scale user study basically served as a prove of concept and therefore, was not mentioned or described in detail in Chapter 4.

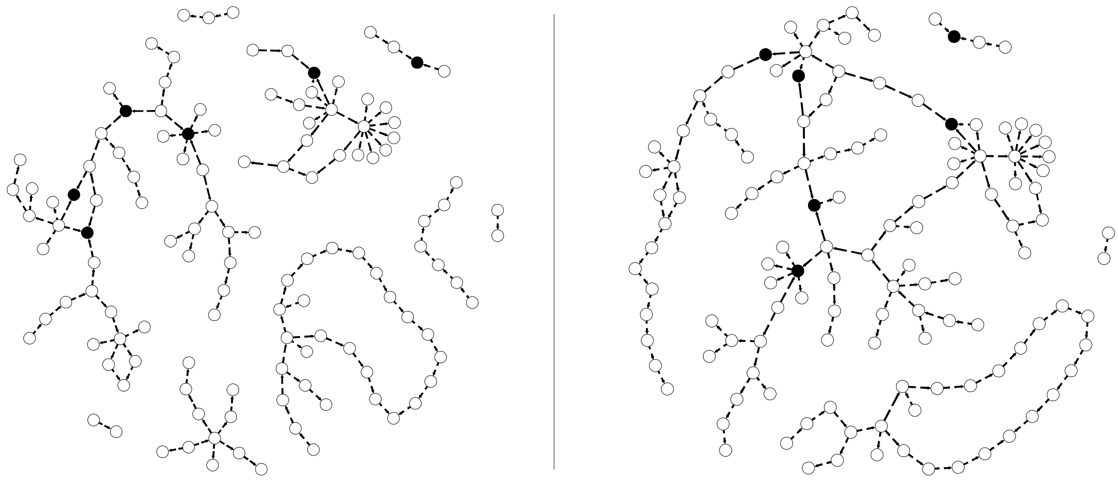


Figure 5.29: Users' search interaction as exploration graphs without assumptions in the interpretation step on the left; and with SERP related assumptions in the interpretation step on the right. The right graph is more connected. White nodes represent web pages which have been visited by only one user, black nodes are web pages which have been visited by at least two different users.

have been visited by at least two different users. For the graph on the left, no assumptions have been made in the interpretation step to generate the graph, leading to a graph with several single components. For the graph on the right, the following assumption have been made in the interpretation step: If a users finds a certain (helpful) piece of information on a web page and uses this as (new) search query, the resulting SERP is not interpreted as a new search path (i.e., a new component) but is connected to the web page where the piece of information was found. This leads to a more connected *exploration graph* that represents how users explore the (new) search domain (implicitly) over time.

5.6 CHAPTER SUMMARY

This chapters comprises the main contribution of this thesis. The integrated search paradigm of ES (cf. Chap. 2) was analyzed and investigated from different perspectives utilizing sequential model approaches (cf. Chap. 3). A comparison of ES to it's most related counter part, namely factual search, was done in parallel as well. To define and train the models for exploratory and factual search, explicit and implicit interaction data from several user studies (cf. Chap. 4) have been used. ES sessions could be described and classified utilizing different model parameters, feature selection and model selection approaches which facilitated accuracies up to 92.1%. In particular the duration being in a certain state was revealed (and confirmed) to be an helpful feature in user seeking behavior modeling but also eye-tracking data as implicit interaction could show it's value. Transforming the classification setting into a cluster scenario described a methodology for (search) behavior analysis in general; confirmed the search tasks design of the user studies; and also identified a borderline search behavior in the interactions data set. Analyzing (further) user variables revealed similarities and differences between young and adult users. Furthermore, the effect of motivation to factual search activities of users was shown. Significant differences considering psychological user characteristics (of adults) however could not be found. In the end, several approaches to support ES have been proposed in terms of search system front-end and back-end aspects.

Part IV

DISCUSSION AND OUTLOOK

"Only in its seeking, the spirit of man finds the secret which it is looking for." (paraphrased)

— Friedrich Schlegel, Lucinde

6 DISCUSSION, FUTURE WORK AND CONCLUSION

User's Exploratory Information Seeking (EIS) has been revealed as a challenging but also all too natural and especially necessary behavior. EIS is challenging for the users who have to get familiar with new search domains if a perceived information need can not be satisfied and/or specified by (single) factual query and search approaches, or if the underlying (complex) search tasks is open-ended. Nevertheless, to discover new environments, gain new experience and be confronted with open (search) tasks is indisputable necessary and natural for everybody (cf. Sect. 2.1). EIS is also challenging from the perspective of the information system (providers) because sophisticated interfaces and algorithms become required to support the exploring users appropriately. Fortunately, this twofold challenge can be approached by the modeling of user's information behavior since the models allow to analyze and better understand exploratory seeking on the on hand but also enables to identify users who are performing Exploratory Searchs (ESs) with systems to provide adequate means. This chapter discusses the revealed findings of the thesis. In the beginning, a discussion of ES in a collaborative information seeking scenario (H5) is given (Sect. 6.1). Afterwards, the results and contribution of the thesis are summarized and discussed in context of the proposed hypotheses and research questions (Sect. 6.2). The chapter continues with an outline of possible future work (Sect. 6.3) and finalizes with a conclusion (Sect. 6.4).

6.1 COLLABORATIVE EXPLORATORY INFORMATION SEEKING

In Section 2.6, the paradigm of ES has been described, characterized and integrated into the theoretical framework of Information Seeking Behavior (ISB). The need to perform an ES, respectively the circumstances where exploratory seeking behavior emerges, depend on various aspects. One key point is the user's experience with the search domain and the available means (i.e., the search tools). Another essential point is the given (search) task, respectively the underlying information need, the associated complexity and uncertainty to solve the task (cf. Sect. 4.3.2.2). The more complex a (exploratory) search task is, the more challenging it is for a solitary user to solve the task sufficiently. In the work of Aula and Russell [12], is has been shown under which circumstances ES tasks become complex and (among other criteria) the authors confirm that complex search often requires exploration. In case of open-ended and multifaceted tasks and information needs also time constrains come into account, e.g., if a proper solution would take days, weeks or even month over several search sessions. Therefore, a suitable strategy to address this issue is to share the task processing. In the context of competitive (business) environments, usually a group of (domain) experts work together to find satisfying answers for the given task(s). A representative example setting is the so-called area of Technology Scouting (TS) [197][1]. The goal of TS is to stay up to date regarding the (company) related technologies, learn about and investigate state-of-the-art methods and thereby, to be competitive. In contrast to the statement that users are (also) unfamiliar with the search domain (often formulated for ES), in TS the opposite is the case. The users here are domain experts and can precisely formulate corresponding questions. That is, the emphasis is not on overviewing the given domain but rather on the discovery and exploration of new developments in the domain what immediately becomes an open-ended and complex (search) task and certainly demands a team of experts. However, to investigate and develop information systems for such a collaborative EIS is challenging for several reasons: First (1), the models for the theoretical background, i.e., ISB but also Information Behavior (IB) in general, are usually oriented and empirically investigated towards single user scenarios. Second (2), to support collaborative EIS even more aspects (in addition to the ones discussed in Sect. 5.5) have to be considered, since a corresponding search system also has to support the coordination between the individual (expert) users. In the following, the two challenges (1) and (2) are exemplified in more detail. The content of this section has already partially appeared in the following own publication [178].

6.1.1 *Models for Collaborative Exploration*

The IB and ISB models, described in Chap. 2, illustrate the process of users' seeking, elaborated relevant components and underlined the (possible) complexity of the search process. Furthermore, the paradigm of ES, as instance of ISB, has been embedded by identifying the relations to the theoretical models. However, the majority of the theoretical models rather consider user's information seeking as a process of an individual user not a group. To overcome this drawback, new models can be developed and investigated or the existing models can be adapted, respectively extended. At first, the *C5* model of Shah (et al.) [169, 174] should be mentioned. The development of the model was motivated by the minor consideration of collaboration in the domain of Information Retrieval (IR) but also by the demand to better understand the concept of collaboration. The resulting *C5* model

Shah's C5 model

1 As a reminder, the area of Technology Scouting was already mentioned as example regarding the integration of ES into Ellis' model of ISB, namely for the *Monitoring* feature, in Sect. 2.6.3.

comprises of five layers with: *Communication* between the individual users in terms of exchanging information as model core; followed by *Contribution* where the group members help each other regarding the individual information need; *Coordination* of the available information sources and resources but also sharing responsibilities and common (search) goals; *Cooperation* what in addition to coordination includes facets of actions, such as planning and negotiation; and finally the last layer *Collaboration*. For Shah [169], collaboration is a concept that incorporates the four other layers above, engages all seeking parties and furthermore deploys a (search) process with the potential to go *"beyond their own individual expertise and vision by constructively exploring their differences and searching for common solutions"*. That (again) confirms the involvement of ES in collaborative search scenarios with complex search tasks and additionally includes the consideration of group members' expertise and resources. Actually, this is not surprising recalling the characteristics to ES and the fact that for (complex) search tasks in a team the amount of uncertainty at least in the initiation phase is higher what in turn increases demand for the "inner" *C5* model layers, such as *communication*. In their empirical investigation, Shah et al. [174] could confirm the complexity of collaborative group work that includes multiple factors. Nevertheless, the authors showed that collaborative seeking can result in successful but also unsuccessful solutions. Furthermore, the individual seeking behavior also has influencing effects to the final solution. Shah (et al.) provided a substantial contribution to the area of collaborative information seeking. A comprehensive and recommendable overview is given in [171].

Considering the demands of a collaborative exploration discloses and confirms that, e.g., Wilson's first model of IB (cf. Sect. 2.3), is focused on individual seeking settings: The user's information behavior is motivated by a perceived individual information need and consequently, individual information seeking is applied. The model provides means to exchange or transfer information with other people too but merely to satisfy the individual need. None of the search related actions performed with the information system are associated or synchronized with the search activities of others. Finally the process of information use remains an individual tasks as well. Of course, the generality of Wilsons' model allows to assume the illustrated IB is applied by each group member and therefore, runs in parallel. This is without any question possible but in the end, the interplay between the individual users is not represented.

Collaborative extension of Wilson's IB model

In the own publication [178], Wilson's IB model has been reviewed under the perspective of collaborative seeking. As indicated above, also the demands for collaborative EIS are related and can be included. In Fig. 6.1, an illustration of user's collaborative information seeking according to [178] and adapted from of Wilson's first model of

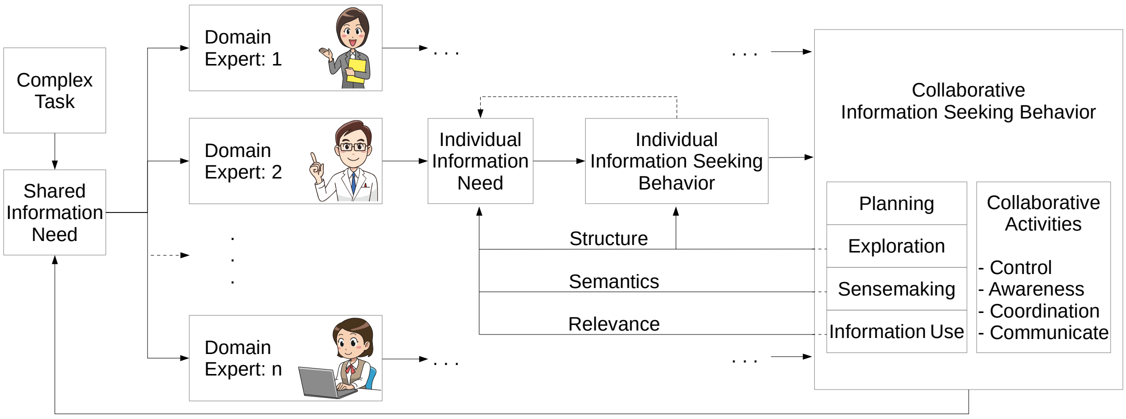


Figure 6.1: Illustration of collaborative information seeking with an emphasis on a group of domain experts according to [178] and based on Wilson's first model of IB [193], cf. Sect.2.3.

IB [193] is shown[2]. Under consideration of the TS scenario, described above, the diagram has an emphasis on a group of domain experts. Given a complex exploratory search task, the group of domain experts first is confronted with the challenge to analyze and identify the corresponding, shared information need and possible aspects of uncertainty. Assuming the identified information need(s) can be confirmed to be too complex for a single user, the need(s) have to be divided into sub-needs which can (at least initially) be processed by each user individually. At this point, Wilson's IB model can be applied (in parallel) as mentioned above. That is, each domain expert has the sub-task to satisfy the identified sub-need individually by performing individual information seeking. Actually, the present scenario and the so far described steps already allow to interpret the "Information Exchange" with "Other People" of Wilson's IB model as possible intersection for collaboration. Furthermore, the individual IB, as in Wilson's model, can pass through several iterations (indicated by the dashed arrow in Fig. 6.1) before collaborative related interactions with others become necessary. However, as pointed out and confirmed by Shah et al. [174], collaboration includes more than the pure "Information Exchange" with "Other People". Fig. 6.1 continues with the collaborative information seeking behavior component and exemplifies seeking and collaboration related activities. For a successful tasks processing, the group members should be able to participate the overall seeking process, organize and understand the revealed information. Similar to Shah's *C5* model, the organization of the collaborative ISB can be illustrated on different layers (which not necessarily need to be encapsulated). The collaborative activities, illustrated on the right, exemplify the "core" aspects which enable a complex task processing in a group

2 For the sake of attribution, the images of the domain experts in Fig. 6.1 are from `https://publicdomainq.net/` and have been found on `https://creazilla.com`.

at all. The control component comprises all activities to promote and regulate the seeking process in the group. The component of awareness enables each user to review the current state and progress of the search task and assess individual contributions (cf. Shah's $C5$ model) of the group. All means to synchronize the collaborative seeking activities between the users are covered by the coordination component. Last but not least, the communication component refers to exchange the so far found information among the group members. The four components planning, exploration, sensemaking and information use require a cooperation of the users and facilitate the collective act to advance the seeking process. Therefore, the four components are similar to the cooperation layer of Shah's $C5$ model. In particular, by exploration, the group of domain experts reveal structural information that can influence the individual information need or the information seeking behavior directly[3]. In context of TS, this structural information, e.g., could be derived from a (together collected) overview of new technologies. The components of sensemaking and information use allow the experts to reveal new findings and key aspects that influence the information need. According to Russell et al. [164] and their rather economic perspective, sensemaking can be understood as the *"process of searching for a representation and encoding data in that representation to answer task-specific questions"*. Therefore, sensemaking on the so far collected information, respectively knowledge, fosters the semantic level. In addition to that, questioning the collected (new) information, resp. knowledge, in terms of information use enables to perform a relevance evaluation. The sum of the components: structure, semantics and relevance again represent the contribution of the group to the seeking process. Furthermore, the experts' individual information need and therefore, the resulting individual ISB, is influenced by the contribution. Finally, the cooperative planning involves to (re-)consider (shared) information need(s) as in the beginning and establish the next necessary actions.

6.1.2 *Support for Collaborative Exploration*

Already in 1996, Twidale and Nichols [183] argued that more support in and the propagation of collaborative IR settings could lead to more effectively used search systems and would facilitate user's learning and working. As mentioned in the beginning of this section, in addition to the support of individual users, the coordination and exchange of revealed information, resp. knowledge, within the group of (expert) users becomes a crucial issue that needs to be supported. This leads to high requirements for the search system, e.g., to enable users to trace and interpret the search related activities of other group members. In

3 The term exploration as used in Fig. 6.1 is rather to be understood as exploration in Kuhlthau's model: cf. Sect.2.4.2. Also cf. the relation to ES discussed in Sect. 2.6.3.

the following, several requirements and challenges for search systems to support collaborative EIS will be exemplified.

Aula and Russell [12] argue that complex exploratory search requires the feature to take notes since the users' ability to hold gathered information in memory is limited. The interactions of the participants in the user studies *US-II* and *US-III* while answering the ES tasks can confirm this demand by an extensively usage of the text field (to make notes) in the *quiz tab*. Therefore, in the TS scenario, a feature for own notes but also to provide notes for the others group members becomes important to address the coordination and awareness requirements. A further important challenge is that collaborative EIS not necessarily is performed synchronous. That is, it is possible that some (even new) experts join the group while others have already started to seek for new, relevant technologies. This further[4] increases the demand on traceability of the seeking process. The ability to exchange the found information by communication with the other group members bilaterally or in conference has also been revealed as a core aspect, cf. Shah's *C5* model. Such a feature can be provided by a text based chat system but also video and voice communications are nowadays potentially options. In case of asynchronous seeking, the ability to leave some notes, messages, etc. for the other members also can be suitable means. In the ES setting, the (expert) users update their domain knowledge during the seeking. Therefore, means to integrate the new relevant information into the (own) storage system can be helpful for future tasks. Finally, the requirement of a tool to support sensemaking is a crucial but still neglected feature. As integral part of several information seeking models (cf. Sect. 2.4.1 and 2.4.2) this quite creative process represents a key point for exploratory (collaborative) ISB but how this aspect exactly can be supported is an open research issue.

In the context of user's search activities, the experimental *Coagmento* [170] web browser plugin from Shah and González–Ibáñez was mentioned (cf. Sect. 3.3.1). The plugin was utilized to analyze the six stages of the Information Search Process (ISP) over two search sessions in a collaborative setting with two people [172]. *Coagmento* allows the user to make annotations and to observe several web page and task specific statistics. Furthermore, the search history of the individual users can be reviewed by the other members. In the own publication [178], the conceptual design of the so called *Search Maps* in the use case of TS is described to support the aspect of traceability in collaborative search. One benefit of search maps [179] is their ability to visualizing the explored "landscape" of a seeking group, e.g., in terms of visited web pages, utilized queries or extracted notes. Last but not least, in Shah [171] further approaches and systems to support

4 "Further" in the sense that the ability for the users to keep track of the search process was already highlighted as supporting mean for individual ES in Sect. 5.5.

collaborative information seeking are presented which also involve possible features to support exploratory (collaborative) search.

6.2 DISCUSSION OF RESULTS

In this section, the results and contributions of the thesis are summarized and discussed. In particular, each of the following sub sections will review one of the thesis' hypotheses, will refer the related chapters and sections and outline in how far the corresponding hypothesis can be confirmed. Furthermore, the contribution of the individual hypothesis processing to answer the thesis' research questions will be discussed in the corresponding sub-sections.

6.2.1 *Reviewing the Integration of Exploratory Search into Information Seeking Behavior Models (H1)*

The contribution by addressing Hypothesis $H1$ is to elaborate crucial aspects of human's information (-seeking) behavior on the one hand and to enable the integration of the paradigm of ES on the other hand. In a first step, several related user behavior models for the different levels of abstraction have been described and discussed (cf. Sect. 2.3 to 2.5). Furthermore, the often presupposed need for information as motivation and initial step of any further informational actions has been introduced and examined (cf. Sect. 2.2). The consideration of the information need's origin from the perspective of exploration is one contribution; the discussion of the information need as a construct that runs in parallel to (Maslow's hierarchy of) human needs is a second contribution. Nevertheless, the integration, i.e., to explicate the relation of ES to the models of information (-seeking) behavior, is the main contribution in context of $H1$ (cf. Sect. 2.6.3). While the models of IB rather describe user's information acquisition in terms of major steps, possible resources, individual motivation and influencing characteristics; "models" of search behavior consider the user's behavior in high resolution but from a limited point of view, e.g., in term of click behavior. Both model types for their own are not able to capture the multifaceted demands and interactions of users who are performing ES. Furthermore, both provide a suitable framework to approach the integration from a deductive and an inductive perspective. Hence, the relation of ES to the abstract models of IB and the detailed search models was exemplified as well. Nevertheless, models of ISB have been revealed as the most appropriate foundation although the relation to ES also had to be explicated in more detail. Reconsidering Kuhlthau's and Ellis' models (as two representative examples for ISB) revealed that their states, respectively features, describe user seeking on a behavior level that is more related to the individual steps or "moves"[5] which

Contributions regarding the information need

Contribution by integrating ES into ISB

5 According to Choo et al. [39].

can appear during the processing of a search tasks. The representation of ES in contrast requires the interplay between the individual steps. In other words, ISB models are rather made to describe the *"what"* while ES rather requires the *"how"*. This in the end leads to the conclusion that ES should be considered as an instance or a class of the ISB process that can cover all states, respectively features of the corresponding ISB models. However, the operationalization of the single ISB model components in a first step would include more inaccuracies for the ES identification process. Therefore, in this thesis, the more important prior step, namely to identify ES sessions on the user interaction level was implemented. The challenge to split, respectively to map the ES process to the single ISB model components, remains for future work but have been already addressed more or less successful, e.g., by Shah and González–Ibáñez [172].

Confirmation of $H1$

In the end, the integration of ES can be considered as successful what allows to state that $H1$ can be confirmed. As a result, the concept of ES turns out to be an approved construct to investigate and establish a conjunction between the theoretical information behavior models and the analytical models at the same time. This in turn advances a seamless deductive resp. inductive view on human seeking behavior performed on specific search systems and hence, a further step to close the gap could be made. Now, with the consideration of ES as class of ISB and the resulting so-called EIS, the research question $Q1$ can be answered on different levels of abstraction. On the most abstract level, the need to satisfy a recognized unknown matter initiates actions of information behavior. To resolve this unknown matter can of course have the goal to enable the satisfaction of a further, maybe much more relevant need. However, if the way to answer the initial matter includes uncertainties, known alternative paths have to be chosen or unknown alternatives have to be identified. Certainly, a known alternative (quasi strategy) can itself be to identify unknown alternatives, what represents a recursive process and can also be interpreted as learning. Learning and to reduce uncertainty are strongly linked to the concept of exploration as a potentially approach to achieve both.

Answering $Q1$

Therefore, on that level of abstraction, the answer to $Q1$, how users behave during EIS, is: Users utilize the available knowledge, in terms of lookup, learn and investigate new information, to derive new knowledge[6] and thereby, reduce uncertainty. Decreasing the level of abstraction results in user's seeking behavior. According to the related (theoretical) models, during exploration, users browse, identify, extract, monitor, gather and verifying different information sources in an iterative processes with the aim to acquire, compare, aggregate and integrate the (new) information and analyze, synthesize, evaluate and interpret (new) knowledge. On the most concrete level (least abstrac-

6 This conclusion basically is conform with Marchionini's framework of ES, cf. Sect. 2.6.1.

tion), users interact with a certain information-, respectively search system. According to the results of the investigations in this thesis, during exploration, users: spend more time on web pages and Search Engine Result Pages (SERPs) in contrast to factual search; they use means to make notes (e.g., the *quiz tab*) more extensively and longer; click more often on further (or previously) SERPs; show a strong interaction between a web pages and the *quiz tab*; visit web pages more often and SERPs less often and finally, have a significantly different number and duration in their fixations on the web pages and SERPs than in factual search.

Finally, it can be concluded that users show a certain behavior if they explore. This behavior can be described on different levels of abstraction and partially can even be automatically identified (as shown in Chap. 5). Though, the reasons leading to an exploratory behavior and how the exploration in particular is conducted depends on multifarious aspects such as search task, search system or user characteristics.

6.2.2 *Influencing User Characteristics (H2)*

To address several intervening variables, relevant for ISB and hence, also ES, user's characteristics, additionally collected in the user studies have been analyzed (cf. Sect. 5.4). At first, the user variable "age" was identified as relevant. In Section 4.2 the design of user study *US-I* to analyze a free (*US-Ia*) but also exploratory (*US-Ib*) voice-controlled search of young users was described[7]. Reconsidering *US-Ia* and summarizing the results of the related own publication [76], it could be revealed that voice controlled Search User Interfaces (SUIs) can increase the usability of information systems for children in general but to develop systems that provide fully voice-controlled search dialogues, further conceptual adaptations and investigations are necessary. Reviewing the own publication [114], which is related to *US-Ib*, several insights into the children's search behavior in context of ES have been gained. For instance, the discussed patterns of interaction but also the analyzed acoustic characteristics can help to improve support for the young user's exploration. Furthermore, the analysis of the speech patterns suggests that it is possible to build a speech recognition program. The results of both, [76] and [114], can serve as fundamentals to develop voice-controlled search dialogues for young users. A more general perspective to apply a user-centered design on search engines for children is given in [71]. An overview about young user's requirements for information search is given in [69]. However, in this thesis, a more analytical and model oriented perspective on the interactions of the

Contributions regarding the user variable "age"

7 As a reminder, voice-controlled interfaces have the advantage that children do not need to have good spelling skills. Therefore, the interaction can be more intuitive, motivating and hence, was taken as means of choice for the user studies in *US-I*.

young users from *US-I* was given. User models of the children's free and exploratory search activities have been generated[8]. The differences in the interactions between free and exploratory search of the young users could be restricted to the interplay between SERP and web pages. In particular, the probability to open a web page from a SERP differs strongly and is twice as high in exploratory than in the free search. Considering the stationary distribution of the relative proportion for each of the search engine states revealed that also young users (in comparison with adults) predominantly interact with web pages and SERPs in exploratory search and interactions regarding the query state happen relatively seldom. In contrast, young users utilize more query interactions during the free search, show less web page interactions but mostly interact with the SERP. Investigating the state durations showed that children spend about 3.5 times longer on SERPs than adults in both activities, exploratory and free search. To complement this results, a further analysis of the differences between the young and adult users' web search can be found in [72]

Contributions regarding the user variable "motivation"

In a next step, several user variables (of adults) have been analyzed by using the data obtained in user study *US-III*. This includes the study procedure to address users' motivational goals but also the several psychological user variables gathered via pertinent questionnaires. Similar to the user *age* (as demographical variable), Willson proposed a possible intervening effect to the seeking behavior also by psychological variables, cf. Sect. 2.3. At first, users motivation was considered. Since motivational goals, as defined and used in this thesis, require an assessment of their achievement, this aspect of search was investigated regarding factual search tasks. After confirming the study procedure by testing hypotheses regarding the search task *difficulty* and *correctness*, it has been shown that motivational goals indeed cause differences in the search behavior. In particular, given a motivational goal, users answered more task and spend less time on each tasks. This indicates that the users have been motivated and/or in a hurry to solve questions only by providing a corresponding search tasks assignment. As expected, the total number of correct answered easy questions was higher than the number of correctly answered hard questions and the relative tasks correctness in general was decreased. One exception was the individual search block where the correctness for hard tasks was increased. This indicated that the Individual Reference Norm (IND) can causes a more attentive task processing. However, between the search tasks blocks with motivational goal, no significant differences could be found regarding the total number of answered questions and used time per task. Analyzing the state proportions revealed no significant differences in the blocks with motivational goal as well. This is an indicator that even if the motivational goals influences the

8 As a reminder, the user base in *US-I* was relatively small. Therefore, the results are limited to the certain user group of the studies and can not necessarily be generalized.

task processing speed, the executed search behavior, i.e., the applied strategies, are not changed. In addition to the user's motivation, also aspects of personality, intelligence and sensation seeking have been investigated using the exploratory and factual search activities of users in *US-III*. Splitting the users regarding the individual psychological variables however showed only slight indications for differences in exploratory or factual search behavior. Considering the personality factor *Openness to experience* (*O*), for example: being on a SERP, users with higher values on *O* more likely click on a result to visit a web page than users with low values on *O*, who rather switch to the Main state. Analyzing the user's search behavior regarding their skill in the *Letter-Number Sequencing* (LN) task, as aspect of intelligence, also revealed only slightly indications, e.g.: being on a SERP, users with higher values on LN more likely switch to the Main state than users with low values on LN who rather click on result to visit a web page. The user variable *sensation seeking* and it's four factors again did not indicated strong differences for users with high or low values on the corresponding factors. That is, the investigations on user characteristics let conclude that exploratory and factual seeking behavior basically does not differ considering the user 's the traits, aspects of intelligence or sensation seeking at least regarding inter-actions (modeled as transitions) and state dwell times as implemented in this thesis.

Contributions regarding users' psychological variables

This new finding has several implications. First, the relation between users' psychological characteristics and their performed search behavior as proposed by Wilson (cf. Sect. 2.3) could not be clearly identified. However, that is no proof that such a relation does not exist because it is possible that the models and model parameters as used here just do not capture the influence of the users psychological characteristics. Further parameters or modeling the user interactions on different levels of abstraction (cf. previous sub section) may reveal the proposed relations. Second, if differences in the exploratory and factual seeking behavior regarding the users' psychological characteristics would be identifiable by the models, a classification of certain user groups would be possible. To support the corresponding user groups according their demands is a desirable goal but comes along with huge responsibility to not utilize this personal data. From that point of view, it can be interpreted as a fortunate case that such characteristics can not be derived "simply" by analyzing the search related interactions. In contrast the psychological characteristics, the analysis of the search behavior of young and adult users revealed differences. That is, there is a relation between the user's age (as demographical variable) and the applied interactions on search. Hence, the Hypothesis *H*2, can be confirmed regarding the users' age but not for the user's psychological characteristics (at least in context of this thesis). Consequently, the question *Q*1 from the perspective of *H*2 and in context of the user's age, can be answered with: Young users need more time for their ES

Differentiated confirmation of H2

Answering Q1

than adults, but behave similar in term of state proportions. Given an task with an open-end character induces exploratory search behavior for both, young and adult users, who more probably (have to) open web pages to advance the search.

6.2.3 *Analyzing, Modeling and Classifying User's Exploratory Information Seeking Behavior (H3)*

The ability to analyze and model user's seeking behavior benefits from a phenomenological theoretical background. This background was given in Chapter 2. In addition to that, relevant behavioral aspects have been identified and methods which are able to capture the corresponding behavior were discussed. In context of user's search behavior, several approaches have been presented in Chapter 3. In particular, the users' click behavior (on SERP) was identified by research as a promising source to reflect user's search process for several reasons. However, to approach the paradigm of ES, also the behavior regarding web pages, queries and possible supporting means (e.g., to make notes) becomes relevant. Furthermore, the time, users spend on SERPs, web pages or queries but also the interplay between the individual search engine components have been revealed as additional valuable source. For that reason(s), in this thesis, a sequential approach that captures user's (exploratory) search behavior in terms of individual states has been selected as the most promising method to reveal the specific characteristics of ES (cf. Chap. 5). In particular, the framework of Markovian models were utilized. In contrast to other approaches, these models have the benefit to represent time series (i.e., sequences), allow to interpret their parameters and provide the possibility to extend the states with emissions. Other methods lack for this combination of properties. For example, classic, static methods like SVMs, Random Forests, etc. are simple, (partially) descriptive but are originally not designed to represent time series. Neuronal networks and in particular LSTMs are able to handle sequential data but require a huge amount of data and provide (as black boxes) only limited support in the understanding the trained models.

The choice for Markovian Models

Contribution by analyzing ES on model parameters

In a first step, with the chosen Markovian models, the user's search sessions could be analyzed by the model parameters. Considering factual and exploratory search sessions, the model parameters could reveal that users more likely (re-)formulate queries and more often return to the SERP in fact-finding search. In exploratory search, users show a stronger interaction between a web page and the *quiz tab* (to make notes and answer the ES tasks). That is, only by analyzing this explicit user interactions (in terms of state transition), the models could show and confirm that users with more specified information needs (induced by the fact-finding tasks) also more likely use query related interactions whereas less specified information needs (induced by the ES tasks) cause more web page related interactions. Utilizing additional search parameters as implicit interactions further completed the picture[9]. Analyzing the state durations showed that the mean dwell times especially for web pages and the *quiz tab* largely differ between the two investigated search activities, i.e., are longer in ES. To model this implicit user behavior as emissions in the Hidden Markov Models (HMMs), a fitted exponential distribution was used. As an alternative, e.g., Hassan et al. [84] suggests to use a gamma distribution since it is frequently used to model waiting times. However, the exponential distribution has the advantage of only one parameter what results into a lower model complexity. Furthermore, the exponential distribution is basically a special case of the gamma distribution namely if the shape parameter is equal to 1.

Contribution by classifying ES and improving the corresp. models

In a next step, the Markovian models have been successfully used in a classification setting. Search sessions with ES tasks could be identified with accuracies up to 92.1% utilizing 2nd-order HMMs with state duration as emission. To achieve sufficient accuracies was one goal but to reveal the limits of the models and the corresponding data was also part of investigation. Therefore, the models have been analyzed on different parameters and in different conditions to investigate the influence to the classification rates. The results showed that further emission, such as scrolling on web pages or the number and duration of user's gaze fixations, are indeed valuable implicit features but did not reach the discriminative power of the dwell time on states[10]. Furthermore, HMMs have been showed to be always superior to the Markov models (without any modeled emissions as implicit interactions) in terms of classification. To chose an appropriate model order was implemented as a model selection task utilizing accuracy Box-Plots and the two information theoretic measures Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Given the interaction data of the user studies *US-II* and *US-III* showed the 2nd order as preferable but a tendency to the 1st order (especially for ES) could also be noted. To identify the user's current search activity as soon as possible, e.g., to provide supporting means for users, the minimal number of interactions was analyzed. After just four interactions, HMMs of the 2nd-order reaches an accuracy of 85.6% but with a drop off afterwards. The resulting recommendation here is to set the interactions to 30 because of the corresponding convergence to a stable classification. A further experiment to approach real life search scenarios was implemented by deleting the study related *quiz tab* (as artificial state) what resulted into still acceptable accuracy rates between 87% and 89%.

9 Also Lorigo et al. [127] highlighted the importance of implicit and explicit user feedback indicators for the modeling and the development of adaptive search engines.

10 Many investigations could show, that time, resp. the duration users spend in several states, is a crucial factor to distinguish different search and seeking behavior and finally, this conclusion can also be made here.

Contribution by clustering ES to validate the study design and identify latent search behavior

In a third step, the classification setting was transformed into a cluster setting. This approach had two main goals. The first was to validate the design of the user studies with the developed factual and exploratory search tasks. If the cluster setting is used to identify two components, the resulting clusters have a high agreement with factual and exploratory search behavior. That is, the own created search tasks indeed induce a certain, corresponding search behavior and this behavior can even been recognized, respectively represented, by the user models. The second goal was to identify latent search behavior. After revealing three components as the best choice (according to the model selection), a fact-finding, an exploratory and a borderline cluster could be identified. Such a borderline case was also proposed by the literature and the fact that this case could also be identified in the study data confirms the natural search behavior on the data on the one hand and confirms an appropriated representation of the user's search behavior by the utilized user models on the other hand.

Confirmation of H3

The sum of these results, as contribution, allow to conclude that the thesis' (main) hypothesis *H*3 can be confirmed: The chosen Markovian models allow to appropriately model user's exploratory search behavior, analyze and reveal it's characteristics and allows to differentiate ES from other search activities such as factual search. In addition to that, the models can be used to identify even further (sub) search patterns if corresponding (experimental) search (task) assignments are not given or used. This also allows to answer the question *Q*2. User's EIS can be modeled by sequence based models (such as Markovian models) which are able to represent users explicit interactions but also allow to incorporate implicit interactions.

Answering Q2

6.2.4 *Approaches to Support Exploratory Search (H4)*

The results of the former sub sections and their related hypothesis clearly depict that ES represents a search and seeking behavior with certain characteristics and therefore, demands for appropriate support on different levels. In Section 5.5, supporting means for front- and back-end components of the corresp. information systems have been proposed and discussed. A first potential means is to let the user traverse trails to the document (result) sets. Example implementations from the literature for such a feature have been described with the *Trailblazer* or the *Internet Radar*. A second mean is to provide different information source for the user. As a possible solution, an own prototypical implementation in terms of an ontology supported SUI for ES was presented. If domain specific ontologies are available they can be complement the sources utilized during ES. For instance, the entities and relations in the ontology can enrich the common web document based search results by annotations in the SERP snippet or on the web documents themselves and hence, support users in

Contribution regarding possible front-end support

Support by ontologies

getting an overview of the current (search) domain. In addition to possible supporting means (for adults) also requirements for young users have been discussed in context of ES and it was exemplified in how far this requirements have been addressed by the SUI *Knowledge Journey*, respective it's prototypical implementation, the *Knowledge Journey Exhibit*.

Requirements for young users

To support ES not only the SUI in the front-end but also components in the back-end can and should be considered. In particular means to provide an overview regarding the search domain are promising. The literature here could show that the utilization of search result grouping, category selection or a hierarchic generation of categories are helpful approaches. In addition to that, providing clues regarding the results (e.g., web documents), graphs can be used (in front- and back-end). Combining result objects and their (derived) attributes by methods of Formal Concept Analysis (FCA) also results into a graph structure. This graph represents a hierarchical structure where the graph edges describe a relation of generalization, respectively specification, within the corresponding concepts (i.e., similar result documents). An prototypical implementation with the Creative Exploration Toolkit (CET) as front-end was described. Considering the hierarchical structure of the results actually allows to address the three-dimensional character of ES, as described in Section 2.6.2: The horizontal axis would be conform to concepts (i.e., clusters) on the same level, the vertical axis with concepts on higher or lower levels and the transversal axis by the same concepts retrieved by different search queries (preferably in related domains). A further back-end approach was presented by the generation of a so-called *Exploration Graph* using browser based interactions. The Exploration Graph allows to represent the user's (exploratory) search over several search sessions which can be combined. An additional interpretation step was proposed as well that leads to a more connected version of the Exploration Graph and therefore, allows better to trace and interpret the ES over time.

Contribution regarding possible back-end support

Support by FCA

Support by exploration graphs

Reviewing the proposed approaches to support ES on the front-, resp. back-end components but also for different users (young and adult), allows to state that the fourth hypothesis *H*4 can be confirmed. Consequently the question *Q*3 in an individual search scenario can be answered by: providing means that allow the user to get an overview of the search domain, means to trace the search process itself by corresp. visualizations but also by available tools to make notes. Furthermore, means to compare results, set the results in relations to each other but also means that consider the user group specifically requirements and do not overcharge the user are crucial.

Confirmation of H4

Answering Q3

6.2.5 *Exploratory Information Seeking in Collaborative Settings (H5)*

The last topic of interest was ES in collaborative settings. In Section 6.1

Contribution regarding exploratory collaborative ISB it was pointed out under which circumstances ES is performed by more than a solitary user. The chosen representative scenario for the exemplification was the Technology Scouting (TS). Often applied in business settings, a TS quickly turns out to be to challenging for just on user. Main reasons are the associated complexity and uncertainty to solve the underlying task to stay up to date regarding the (company) related technologies, learn about and investigate state-of-the-art methods and thereby, to be competitive. *Confirmation of H5* This actually enable to confirm the thesis hypothesis *H*5 but to answer the question *Q*3 regarding a collaborative setting appropriately, Section 6.1 continued by providing possible model aspects and means of support. In particular, models to capture collaborative information seeking were discussed and an own extension of Wilson's IB model was proposed. For this thesis, an emphasizes on exploratory collaborative information seeking for the model have been provided and relation to a famous model from the literature was pointed out. For the support of collaborative ES, requirements and challenges were exemplified and the usage of *Search Maps* was described. *Answering Q3* Consequently, *Q*3 in a collaborative search scenario can be answered by: providing means that enable to coordinate and exchange the revealed information, resp. knowledge, within the group of users; means that enable users to trace and interpret the search related activities of other group members (e.g., by *Search Maps*); but also features to take notes; means that provide synchronous but also asynchronous search have been revealed as promising approaches.

6.3 FUTURE WORK

Given the theoretical (Chap. 2) and analytical (Chap. 3) background of the thesis' topic of user behavior models for EIS, the methodology (Chap. 5) explicated and utilized for the investigation regarding the recorded and prepared data sets (Chap. 4) of course facilitates a multitude of further possible research questions. In the following, several entry points and inspirations for future research will be outlined.

Identifying ISB components To further close the gap between the theoretical and analytical perspectives on humans ISB, the identification of individual components of user's seeking is a promising direction. For example, the recognition of states and features during the seeking, as proposes in Kuhlthau's and Ellis' model, provide an adequate framework. As highlighted in this thesis, the ES and it's related diverse characteristics covers (and partially requires) a broad spectrum of seeking related interactions of users. Therefor, ES, as an instance of ISB, is an excellent concept to advance research in that direction. *Identifying ES activities* Furthermore, a differentiation of ES related several activities (according to Marchionini, c.f. Fig. 2.8) would advance the understanding of ES but also could reveal the relation to the ISB components mentioned before. *Considering Learning* As described in Sect. 2.6.1, the concept of learning is also relevant for ES and can help to further understand the (exploratory) search process in more detail. In [24], Bhattacharya and Gwizdka addressed this issue by the attempt to measure learning during search. Furthermore, the relevance of user's domain knowledge for learning while searching was considered by Roy et al. [161] and O'Brien et al. [141]. In addition to the demographical and psychological intervening variables, also environmental and further role-related variables, respectively source characteristics (as proposed by Wilson) and the variables' influence to users' EIS behavior remains an research topic with open questions. In particular, role-related variables become also relevant in collaborative (exploratory) settings.

Further search related variables influencing EIS

Considering the framework of Markovian models and the corresponding literature, there are manifold extensions to overcome possible drawback of the original structure. For example, McCallum and Pereira [133] extend HMMs to Maximum Entropy Markov Models where observations do not only depend on the emitting state but also on the previous observations what allows a richer representation of the features. A further extension, respectively abstraction, are Conditional Random Fields (CRFs), e.g., used by Ageev et al. [2]. However, under consideration of the HMMs as used in this thesis (where the user states are actually not hidden), a richer representation of the features not necessarily improves the user modeling. For instance, using a HMM with (only) state duration as feature would imply that the dwell time not only depends on the current state (e.g., a web page) but also on the duration of the previous state (e.g., a SERP), what is questionable. However, the ability of Markovian models to be nested into each other (e.g., as emission generating sub processes) makes a richer representation of the features relevant: Assuming the user's gaze behavior is not utilized directly as feature but is first used to derive the user's reading state (classified by a HMM) and the output of the reading classification in turn is used as feature for a state (e.g., a web page). In that case, the reading state detecting HMM can benefit from a richer feature representation because it is plausible that, e.g., the duration of a former fixation is related to a later fixation duration in the same reading class.

Extensions and variants of Markovian approaches

To motivate the inclusion of user's reading behavior (as mentioned before) for future research, in the following, the annotated reading behavior on the data of user study *US-II* is briefly described and illustrated. Literature shows, humans exhibit (at least) three different types of reading activities if they acquire information, namely (pure) *Reading*, *Skimming* and *Scanning*. While pure (silent) reading *"involves a sequence of eye movements that typically move from left-to-right across the page and then down the page"* (p. 3) [155] with the goal to understand the read text, *Scanning* and *Skimming* are rapid forms of text reading with different goals and corresponding eye movements. According to Rodeghero and McMillan [160], *Skimming* does not intent to read and

Include user's reading behavior and different reading activity types understand the full text but rather aims to get the general meaning, respectively a summary, of the presented information. Hence, *Skimming* is characterized by less and shorter eye fixations and rather vertical than horizontal [40] eye movements. *Scanning* is similar to *Skimming*, as it is also a rapid reading. The aim of *Scanning* in contrast is to *"gain insight about a particular piece of information"* (p. 2) [160], for example if a user attempts to find a certain keyword on a web page or SERP.

Assuming that users during ES have to read and understand the text more than during factual search where (sometimes) only a piece of information has to be identified. This leads to the motivation to include the user's reading behavior into the search process as a valuable feature to differentiate the two search activities. However, reliable methods to recognize whether a user is reading or not (given eye movement data) are currently missing and to identify even different reading activities complicates the tasks further. Nevertheless, under a further review of the literature regarding user's reading, guidelines to annotate the different reading activities have been derived and finally applied to the data of user study *US-II*. While the tasks to implement a reading activity classifiers remains as one possible future work, the data annotation itself can be used as ground truth to extend the models for a further future work, namely as a proof of concept. The indications that reading behavior can be beneficial to classify ES are clearly noticeable. In Fig. 6.2 the reading annotation is exemplary illustrated as sequence plots for the factual and exploratory search sessions. In addition to the annotation of *Reading*, *Skimming* and *Scanning* it was also marked, if users view pictures or videos during the search, marked as *MediaView*. In case, users performed a behavior that was not in accordance with the four activities describes above, the annotation used a unclassified label *Unknown*. A visual comparison of the sequences shows that a user in fact-finding more frequently switch

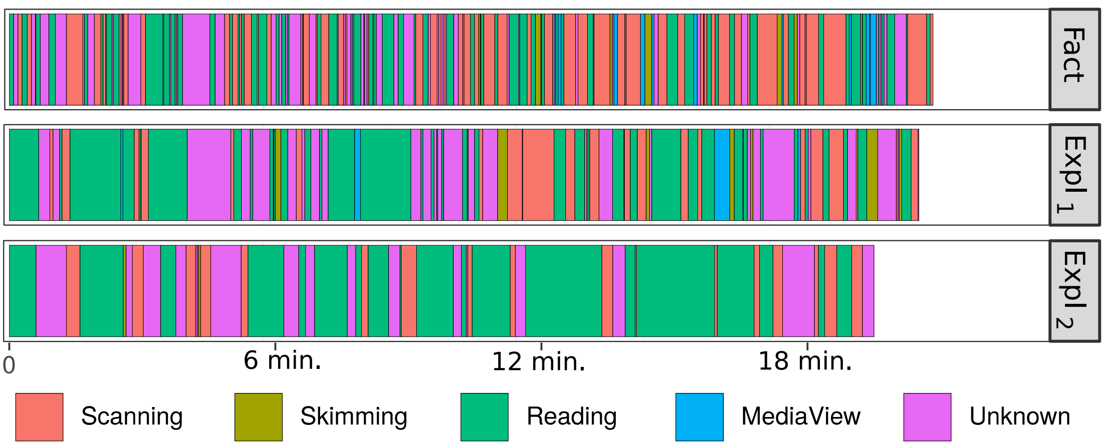


Figure 6.2: Exemplary illustration of the reading annotation as sequence plots on *US-II* for factual and exploratory search sessions.

between the reading activities and that the proportion of amount of *Scanning* is relatively high. During both ES tasks, there are less reading activity switches, the time for the reading activities are longer and (pure) *Reading* has the highest proportion. This indications are promising for future research.

Differentiating exploring and struggling users

In addition to the user's reading behavior, the recorded data from the user studies *US-II* and *US-III* allow to extract further valuable features: As pointed out by Hassan et al. [85], several characteristics of exploratory search are similar to characteristics in seeking behavior if users are struggling. For example, the number of queries or the dwell time on web pages may increase (cf. Sect. 3.3.1). This makes it difficult (especially from the perspective of usual search engines log files) to differentiate between both. However, the ability to distinguish between a struggling or an exploring user is important for the analysis of search behavior and understanding of seeking in general but also for search engines to provide adequate means of support. The methodology presented in this thesis provide possible approaches for these purposes. In particular, recorded user video data from user study *US-II* and *US-III* can serve as rich source for the detection of negative (or positive) facial expression to further incorporate emotional features in the search activity classification process.

Designing and evaluating supporting means for ES

A further possible direction of investigation is the evaluation of means to support ES. Here the described aspects of Section 5.5 can serve as an entry point. However, also more simple adaptations of common list based result visualizations are thinkable, for example, to adapt the diversity of results provided in the SERP in case of detected exploratory search behavior.

Extending the explored search domain to image data bases

To extend the perspective on ES also on other (non-textual) search domains, the concept, methodology but also data of the own work [128] can be used. The goal of the corresp. study was to record and investigate users who are exploring a multimedia collection of images. In particular, a group of 15 participants navigated through a set of images (in particular polygons) while their gaze behavior was recorded. Exploring large multimedia collections can be challenging and exhausting for users because it is not always clear which direction (or paths, cf. tasks complexity in Sect. 3.3, footnote on p. 55) should be followed. That is, the search can become a highly dynamic process where users have to investigate and learning about the structure and/or content of the collection. Therefore, a search system that is able to estimate the user's search intent and goals, to provide support during an ongoing ES, is desirable. However, it is not much known about the exploratory seeking behavior of users in large multimedia collections and thus needs to be investigated more to develop such systems.

Including semantic content in the modeling

Last but not least, the inclusion of semantic information into the seeking process can further advance the understanding of ES but also users ISB in general. The methodology presented and applied in this these is rather focused on the interaction level without consideration of the provided or acquired content. For example, an analysis of the utilized queries, visited (or read) web pages and viewed SERPs during (exploratory) search may provide additional insights of the search process on the one side but also can further closes the gap between theoretical and analytical perspective on the other side. However, a rather interaction based methodology (as proposed in this work) in contrast has the advantage to reveal seeking related findings independent of semantic aspects which therefore, can be better generalized.

6.4 CONCLUSION

This work was dedicated to the topic of Exploratory Information Seeking (EIS) and related user behavior models. To approach the topic, the thesis reviewed the literature of human information (seeking) behavior and exemplified selected theoretical user models by their inherently given hierarchical structure. In contrast to that, the thesis also reviewed several application-oriented, analytical user models, which are (partially) the result of the still growing amount of interaction related data with search system in the Internet. In between this two research areas, the literature identified a long-time existing gap. Therefore, the contribution of this thesis was to further bridge the gap, provide a more elaborated foundation for the research area and advance the methodology to satisfy the user's information need sufficiently. To accomplishes this goal, the paradigm of Exploratory Search (ES) was selected and served as the major objective of investigation. This kind of search comprises a multitude of search related activities simultaneously and hence, can encompass a high complexity of the user's search process. Furthermore, ES represents a connecting concept between theoretic information behavior models and analytical interaction models. Therefore, the thesis explicated the integration of ES into the framework of user's information seeking behavior. Due to the lack of appropriate data sets, the generation and conductance of several user studies to create a proper basis for the thesis' research questions was implemented. To study EIS from different perspectives, on the one hand, the influence of users' personal characteristics are examined and on the other hand, the (exploratory) search process itself, including the analysis, modeling and classification on different interaction levels was investigated. The provided and applied methodology facilitates new research regarding the recognition of certain search activities, identification of latent search behavior patterns and appropriate model selection. Considering the results of the conducted investigations, means to support several aspects of exploratory search behavior via search engine interfaces have been exemplified. In addition, a discussion regarding collaborative (exploratory) information seeking was given. Summing up, the contribution of the thesis' investigations is a deeper understanding of users' exploratory information seeking to establish a stronger link between the theoretical and application-oriented, analytical behavior models.

Part V

APPENDIX

A APPENDIX

This appendix provides several content which has been outsourced for the sake of comprehensibility.

A.1 LIST OF ALL DESIGNS IN USER STUDY III

This section lists all twelve designs containing the five search blocks $\text{Expl}_{1\&2}$, Fact_{NON}, Fact_{CRT}, Fact_{IND} and Fact_{SOC} from *US-III*. Each pair of participants had one out of the twelve possible block sequences designs within the *search experiment*. The twelve designs have been equally distributed among the participants:

- *Design*$_1$: $\text{Expl}_{1\&2} \rightarrow \text{Fact}_{NON} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{IND} \rightarrow \text{Fact}_{SOC}$
- *Design*$_2$: $\text{Fact}_{NON} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{IND} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Expl}_{1\&2}$
- *Design*$_3$: $\text{Expl}_{1\&2} \rightarrow \text{Fact}_{NON} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Fact}_{IND}$
- *Design*$_4$: $\text{Fact}_{NON} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Fact}_{IND} \rightarrow \text{Expl}_{1\&2}$
- *Design*$_5$: $\text{Expl}_{1\&2} \rightarrow \text{Fact}_{NON} \rightarrow \text{Fact}_{IND} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{SOC}$
- *Design*$_6$: $\text{Fact}_{NON} \rightarrow \text{Fact}_{IND} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Expl}_{1\&2}$
- *Design*$_7$: $\text{Expl}_{1\&2} \rightarrow \text{Fact}_{NON} \rightarrow \text{Fact}_{IND} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Fact}_{CRT}$
- *Design*$_8$: $\text{Fact}_{NON} \rightarrow \text{Fact}_{IND} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Expl}_{1\&2}$
- *Design*$_9$: $\text{Expl}_{1\&2} \rightarrow \text{Fact}_{NON} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{IND}$
- *Design*$_{10}$: $\text{Fact}_{NON} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Fact}_{IND} \rightarrow \text{Expl}_{1\&2}$
- *Design*$_{11}$: $\text{Expl}_{1\&2} \rightarrow \text{Fact}_{NON} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Fact}_{IND} \rightarrow \text{Fact}_{CRT}$
- *Design*$_{12}$: $\text{Fact}_{NON} \rightarrow \text{Fact}_{SOC} \rightarrow \text{Fact}_{IND} \rightarrow \text{Fact}_{CRT} \rightarrow \text{Expl}_{1\&2}$

A.2 LIST OF ALL FACTUAL SEARCH TASKS IN USER STUDY III

This section lists all 117 factual search tasks designed for user study *US-III* (cf. Sect. 4.4.3.1) in the Tables A.1 to A.10. To minimize possible translation issues and/or to reduce the scope for interpretation, the tasks and answer options are given in their origin language German.

Table A.1: List of factual search tasks 1 to 15 from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers. The * indicates the correct answer.

Nr.	Level	Task description	Answers
1	Easy	*In welcher Sportart wird der Euler gesprungen?*	*Eiskunstlauf*, Hochsprung, Weitsprung, Reitsport, Dreisprung*
2	Easy	*Welches Land ist mehr als zehnmal so lang wie breit?*	*Chile*, China, Algerien, Russland, Mosambik*
3	Easy	*Wie heißt der zweitlängste Fluss der Welt?*	*Amazonas*, Nil, Wolga, Donau, Yangtsekiang*
4	Easy	*Wie wurde die Pferderennbahn in einer antiken Millionenstadt genannt?*	*circus*, orbis, nomus, coetus, cyclus*
5	Easy	*Wann wurde Weissenburg/Bayern erstmals urkundlich erwähnt?*	*867*, 1331, 90, 1581, 270*
6	Easy	*Welcher Planet unseres Sonnensystems hat den kürzesten Tag?*	*Jupiter*, Merkur, Saturn, Mars, Neptun*
7	Easy	*Welches ist das leichteste Metall unter normalen Bedingungen?*	*Lithium*, Magnesium, Aluminium, Titan, Barium*
8	Easy	*In welchem Jahr ging die Suchmaschine Google das erste Mal online?*	*1998*, 1996, 2000, 2004, 1992*
9	Easy	*Wie nennt man das Ausgeben von fremden Geistesgut als das Eigene?*	*Plagiat*, Raubkopie, Gedankenraub, Sampling, Origanat*
10	Easy	*Wie bezeichnet man das japanische Theater für das gemeine Volk?*	*Kabuki*, Wakizashi, Takayama, Seppuku, Daimyo*
11	Easy	*Die Protagonisten welches Films sind Namensgeber für einen Pub in Magdeburg?*	*Blues Brothers*, Spione wie wir, American Werewolf, Bloody Mary – Eine Frau mit Biss, Die Glücksritter*
12	Easy	*Was ist ein Mufti?*	*Ein islamischer Rechtsgelehrter*, Ein indischer Prinz, Ein korsisches Wildschaf, Ein afghanischer Bettler, Ein Angehöriger eines afrikanischen Stammes*
13	Easy	*In Bad Hersfeld stehen Denkmäler zwei berühmter Persönlichkeiten, die sich den Vornamen teilen. Wie ist deren Vorname?*	*Konrad*, Wilhelm, Karl, Helmut, Holger*
14	Easy	*Wie breit muss ein Fußballfeld sein?*	*45 bis 90 m*, 40 bis 60 m, 45 bis 75 m, 50 bis 80 m, 40 bis 50 m*
15	Easy	*Wie heißt der drittlängste Fluss der Welt?*	*Jangtsekiang*, Nil, Amazonas, Wolga, Gelber Fluss*

Table A.2: List of factual search tasks 16 to 29 from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers. The * indicates the correct answer.

Nr.	Level	Task description	Answers
16	Easy	*Nach wem ist der höchste Berg Grönlands benannt?*	*Gunnbjörn Úlfsson (auch Gunnbjörn Ulf-Krakuson)*, Lars Styggehorn, Paul-Emile-Victor Mound, Pierre-Emil Hojberg, Qaqqaaq Johnson*
17	Easy	*Wie viele Berge befinden sich im Himalaya, die höher als 8000 Meter sind?*	*10*, 7, 8, 9, 11*
18	Easy	*Wann wurde Julius Caesars Geburtstag gefeiert?*	*13. Juli*, 15. März, 7. Juli, 13. Oktober, 15. November*
19	Easy	*Wie nannte man vor 1970 die sog. eidesstattliche Versicherung?*	*Offenbarungseid*, Staatliche Absicherung, Bescheinigung zur Offenlegung, Verzichtseid, Verzichtserklärung*
20	Easy	*Wie heißt der größte natürliche Mond unseres Sonnensystems?*	*Ganymed*, Kallisto, Europa, Io, Erdmond*
21	Easy	*Welche Base wird während der Translation ausgetauscht?*	*Thymin*, Guanin, Cytosin, Adenin, Purin*
22	Easy	*Wie heißt das größte Passagierflugzeug?*	*A380*, Boeing 747, Boeing 777, A340, A390*
23	Easy	*Wie heißt der vierte Teil des „Rings der Nibelungen"?*	*Götterdämmerung*, Zweiter Tag, Das Rheingold, Die Walküre, Siegfried*
24	Easy	*Was ist ein „Weggla"?*	*Brötchen*, Ball, Einweckglas, Wecker, Wegweiser*
25	Easy	*Jäger haben ihre eigene Sprache. Wie nennen sie beispielsweise eine junge Gämse?*	*Kitz*, Welpe, Fohlen, Frischling, Ricke*
26	Easy	*Wie heißt die Hauptstadt des Landes, in dem 2010 die Fußballweltmeisterschaft ausgetragen wurde?*	*Pretoria*, Johannesburg, Durban, Germiston, Pietersburg*
27	Easy	*Wie heißt die aus 3 Torbögen bestehende Brücke der südspanischen Kleinstadt, die den traditionellen Stierkampf maßgeblich beeinflusste?*	*Puente Nuevo*, Pont de Llierca, Puente de Barrancales, Zubizuri, Puente de Itero*
28	Easy	*In welchem Staat/Gebiet leben die meisten Mitglieder des Volkes der Mari?*	*Russland/Republik Mari EL.*, Russland/Komi-Republik, Ukraine, Senegal, Tansania/Mari-Region*
29	Easy	*Wie viele offizielle Sprachen gibt es in der Republik Südafrika?*	*11*, 12, 10, 5, 3*

Table A.3: List of factual search tasks 30 to 43 from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers. The * indicates the correct answer.

Nr.	Level	Task description	Answers
30	Easy	*In welchem Jahr wurde die Erste der heute überall bekannten Säulen zur Plakatierung aufgestellt?*	*1855*, 1854, 1873, 1874, 1890*
31	Easy	*Welche Unabhängigkeitsbewegung hieß wie ein Kartenspiel?*	*Mau Mau*, Tichu, Hanafuda, Kuku, Piquet*
32	Easy	*Woher kommen kurzperiodische Kometen?*	*Kuiper-Gürtel*, Oort'sche Wolke, Schwarzes Loch, Starburstgalaxie, „Antennen"-Galaxie*
33	Easy	*Bei welchem Syndrom glaubt man, dass nahestehende Personen ausgetauscht wurden?*	*Capgras-Syndrom*, Klüver-Bucy-Snydrom, Cotard-Syndrom, Couvades-Syndrom, Dorian-Grey-Syndrom*
34	Easy	*Wie nennt man die Vorrichtung, welche bei der Eisenbahn das Überfahren eines Gleisendes verhindern soll?*	*Prellbock*, Gleisstopper, Gleisbremse, Abstoppbock, Kopfbremse*
35	Easy	*Wie hieß das Raumschiff auf dem die erste Frau im All war?*	*Wostok 6*, Wostok 5, Wostok 4, Wostok 2, Sojus*
36	Easy	*Wie alt ist Mickey Mouse heute?*	*88*, 72, 85, 93, 78*
37	Easy	*Nach aktuellen Angaben, wie viele Zimmer gibt es im Buckingham Palace in London?*	*755*, 346, 188, 822, 574*
38	Easy	*Der Quetzalcoatl ist eine Gottheit, die aus einem Vogel und einem weiteren Tier besteht. Welches weitere Tier ist das?*	*Schlange*, Quetzal, Krokodil, Jaguar, Geier*
39	Easy	*Welches Land produziert am meisten Pistazien?*	*Iran*, USA, Syrien, Türkei, Russland*
40	Easy	*Welches Land produziert am meisten Safran?*	*Iran*, Spanien, Südfrankreich, Griechenland, Marokko*
41	Easy	*Welcher Bundeskanzler war länger im Amt als Konrad Adenauer?*	*Helmut Kohl*, Willy Brandt, Helmut Schmidt, Gerhard Schröder, Kurt Georg*
42	Easy	*Was zeichnet eine Hochkultur nicht aus?*	*Bergbau*, Religion, Handel, Schrift, Wissenschaft*
43	Easy	*Mit welcher Abkürzung sind Kraftfahrzeuge von Diplomaten gekennzeichnet?*	*CD*, CC, Dipl., DD, 0*

Table A.4: List of factual search tasks 44 to 57 from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers. The * indicates the correct answer.

Nr.	Level	Task description	Answers
44	Easy	*Welches Element ist der beste Leiter für Elektrizität?*	*Silber*, Kupfer, Gold, Platin, Quecksilber*
45	Easy	*Wovor haben Sciaphoben Angst?*	*Schatten*, Alpträumen, Großen Höhen, Pflanzen, lauten Geräuschen*
46	Easy	*Um das Unternehmen Apple gründen zu können, verkauften Steve Jobs und Steve Woszniak welche Dinge?*	*Einen VW-Bus und einen Taschenrechner*, Eine Uhr und einen Computer, Ein Haus und einen Cadillac, Eine Comicsammlung und einen Fernseher, Ein Skateboard und eine Stereoanlage*
47	Easy	*Wie viele Menschen waren bis Juni 2004 auf dem Mond?*	*12*, 11, 10, 7, 8*
48	Easy	*Wie wird das vierte Buch des Alten Testamentes auf Griechisch genannt?*	*Numeri*, Genesis, Exodos, Levitikon, Deuteronomion*
49	Easy	*Der Kurzfilm The Big Shave bezieht sich auf welchen Krieg?*	*Den Vietnamkrieg*, Den dreißigjährigen Krieg, Den zweiten Weltkrieg, Den ersten Weltkrieg, Den Irakkrieg*
50	Easy	*Welcher Speisefisch wandert zwischen Meer und Süßwasser?*	*Lachs*, Flunder, Kabeljau, Scholle, Makrele*
51	Easy	*Wie heißt der Ort in Georgien, mit dem im Land bedeutsamsten Fund menschlicher Vorfahren?*	*Dmanissi*, Tiflis, Batumi, Uplisziche, Rabat*
52	Easy	*Welche deutsche Band bezeichnet sich selbst ganz gerne als „die drei Götter in schwarz"?*	*Die Ärzte*, Die Toten Hosen, Fettes Brot, Die Prinzen, Kraftwerk*
53	Easy	*Welchen Namen geben die Einheimischen Neuseelands der Süßkartoffel?*	*Kumara*, Melanzani, Sweet Potato, Cucurbita, Batate*
54	Easy	*In welchem Land der Erde wurden die größten Temperaturunterschiede gemessen?*	*Russland*, Finnland, Kanada, Mongolei, Dänemark*
55	Easy	*Von wem wird Miho im ersten Film zu Frank Millers Comicreihe über eine moralisch verdorbene Stadt gespielt?*	*Devon Aoki*, Lucy Liu, Michelle Yeoh, Zhang Ziyi, Jamie Chung*
56	Easy	*Welches philosophische Gedankenexperiment beschäftigt sich mit ungewollter Schwangerschaft?*	*Der Geiger*, Das Schiff des Theseus, Das Kind im Teich, Der Schleier des Nichtwissens, Russels Teekanne*
57	Easy	*Der Edelstein Jade welcher Farbe kostet am meisten?*	*Grün*, Orange, Weiß, Lavendel, Rot*

Table A.5: List of factual search tasks 58 to 66 from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers. The * indicates the correct answer.

Nr.	Level	Task description	Answers
58	Hard	*Welches Land erlangte in dem Jahr, in dem es selbst Gastgeber der Olympischen Sommerspiele war, keine einzige Goldmedaille?*	*Kanada*, Deutschland, Frankreich, Griechenland, Schweden*
59	Hard	*Welche europäische Halbinsel ist eigentlich nur durch eine Brücke mit der restlichen Landmasse verbunden?*	*Peloponnes*, Iberische Halbinsel, Gibraltar, Nordstrand, Hornstrandir*
60	Hard	*Welche Stadt/Region gilt traditionell als die Stadt mit dem besten Hopfen der Welt?*	*Saaz/Zatec*, In der Hallertau, In Spalt, In Hersbruck, Im Mühlviertel*
61	Hard	*In welchem System des Paläozoikums traten die ersten Reptilien auf?*	*Karbon*, Perm, Devon, Jura, Kreide*
62	Hard	*Wie oft musste das sich in der 11-größten Stadt Deutschlands (nach Einwohnerzahl) befindliche Opernhaus im Laufe seiner Geschichte wiederaufgebaut werden?*	*dreimal*, einmal, zweimal, viermal, fünfmal*
63	Hard	*Welches der sieben Weltwunder der Antike hat es auch auf die Liste der sieben Weltwunder der Neuzeit geschafft?*	*keines*, Die Pyramiden von Gizeh, Kolosseum, Der Koloss von Rhodos, Chinesische Mauer*
64	Hard	*Welches Wort ist im Morse-Code als „–. – . – – – – – – . – . .“ codiert?*	*COOL*, BOOT, HALT, KAHN, CHAM*
66	Hard	*Wie wird der Überrest bezeichnet, der bei der Herstellung von Zuckerrübensaft entsteht?*	*Melasse*, Rübenkaramell, Teig, Mole, Rum*
66	Hard	*Im vierten Teil der Buchreihe Harry Potter gibt es am Ende des Buches einen Fehler, der einem im fünften Buch der Reihe bewusst wird. Welcher ist das?*	*Harry kann die Thestrale am Ende des vierten Teils nicht sehen. obwohl er bereits jemanden sterben gesehen hat.*, Cedric kann nach seinem Tod mit Harry sprechen. im fünften Teil allerdings nicht., Rubeus Hagrid verlässt Hogwarts nie. im fünften Teil ist er aber verschwunden., Die maulende Myrte beschimpft Harry wegen Cedrics Tod. im fünften Teil ist sie glücklich. dass Cedric verstorben ist., Harry erhält das Preisgeld des Trimagischen Turniers. beklagt sich aber am Anfang des fünften Teils über Geldnot.*

Table A.6: List of factual search tasks 67 to 78 from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers. The * indicates the correct answer.

Nr.	Level	Task description	Answers
67	Hard	*Wie hießen die beiden Hauptcharaktere des Romans „50 Shades of Grey“ ursprünglich?*	*Edward und Bella*, Darcy und Elizabeth, Ron und Hermione, Anthony und Pepper, Aragorn und Arwen*
68	Hard	*Mit wem war Lisa Marie Presley am längsten verheiratet?*	*Michael Lockwood*, Danny Keough, Michael Jackson, Nicolas Cage, Tom Cruise*
69	Hard	*Welcher Schauspieler gewann im gleichen Jahr die Goldene Himbeere, unter anderem als schlechtester Schauspieler, schlechtester Nebendarsteller und schlechteste Nebendarstellerin?*	*Eddie Murphy*, Adam Sandler, Tom Cruise, Kevin Costner, Michael Douglas*
70	Hard	*Welche deutsche Band spielt mit einem Song auf das Buch „Uhrwerk Orange“ an?*	*Die Toten Hosen*, Nena, Matzen, Bakkushan, Kraftwerk*
71	Hard	*Wie nennt man auch den Anteil eines einzelnen Gesellschafters bei einer Gesellschaft mit beschränkter Haftung?*	*Stammeinlage*, Bürgschaft, Anteilseignung, Stammkapital, Kapitalumsatz*
72	Hard	*Unter welchen Umständen ist es in Paraguay erlaubt, sich zu duellieren? Wenn beide Teilnehmer …*	*... Blutspender sind*, ... männlich sind, ... blonde Haare haben, ... einen Impfausweis mit sich führen, ... Knochenmarkspender sind*
73	Hard	*Welcher Mond in unserem Sonnensystem ist am nächsten gelegen zu seinem Planeten?*	*Phobos*, Charon, Deimos, Erdmond, Mab*
74	Hard	*Wie wird im Deutschen das Organ zur Eiablage bei Insekten genannt?*	*Legebohrer*, Eileger, Legerüssel, Positionsorgan, Eistecher*
75	Hard	*Wie wird im Fachjargon das Hilfsmittel genannt, das Raubfischen die Drehbewegung kranker Fische vortäuschen soll?*	*Blinker*, Schwimmer, Spinnrute, Spinnangel, Einholer*
76	Hard	*Wie heißt die Vorrichtung an einem Segelboot, welche dem Steuermann die Windrichtung verrät?*	*Verklicker*, Ventumeter, Barometer, Kompass, Einsteller*
77	Hard	*Wo ist das kleinste Brauhaus der Welt?*	*Saaz/Zatec*, In Köln, Bacharach, Bayreuth, Dresden*
78	Hard	*Wie weit liegt der Ort mit dem ersten Radonbad der Welt von Magdeburg entfernt (mit dem Auto in km)?*	*285 km*, 350 km, 200 km, 500 km, 1500 km*

Table A.7: List of factual search tasks 79 to 89 from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers. The * indicates the correct answer.

Nr.	Level	Task description	Answers
79	Hard	*Wieviel Watt hat ein Gigawatt?*	*1 Milliarde*, 1 Million, 1 Billiarde, 1 Billion, 10 Million*
80	Hard	*Wie nennt man die Gewichte eines Heißluftballons zur Regelung der Flughöhe?*	*Ballast*, Übergewicht, Fender, Gravitoren, Auffülllast*
81	Hard	*Welcher osteuropäische Maler war für seine alptraumhaften Bilder bekannt?*	*Zdzisław Beksiński*, Iwan Konstantinowitsch Aiwasowski, Juri Pawlowitsch Annenkow, Marc Chagall, Michail Wassiljewitsch Matjuschin*
82	Hard	*Wie heißt der Sänger, Gitarrist und Pianist der in John Nivens erstem Buch behandelten Band?*	*Robbie Robertson*, Glenn Frey, Bob Dylan, Ronnie Hawkins, Paul Simon*
83	Hard	*Eine Band, deren Mitglieder als Pioniere des Heavy Metals gelten, schrieb ein Lied, was denselben Namen trägt, wie ein beliebter Comicheld. Wovon handelt das Lied?*	*Einen Mann. der Gutes tun will und Böses tut*, Den Comichelden, Die Schönheit der Schmiedekunst, Eigentlich hat das Lied keinen tieferen Sinn, Prothesen von Kriegsopfern*
84	Hard	*Von welchem Lied sagt die Band Muse, dass sie selbst nicht wisse, was der Text bedeute?*	*Plug in Baby*, MK Ultra, Blackout, Hysteria, Resistance*
85	Hard	*Sie sind auf einer akademischen Veranstaltung und werden Herrn Professor Dr. Dr. h. c. mult. Wilhelm Carstens vorgestellt. Wie lautet die korrekte Anrede?*	*Herr Professor Carstens*, Herr Carstens, Herr Professor Doktor Carstens, Herr Doktor Carstens, Wilhelm*
86	Hard	*Krawatten kann man auf vielerlei Arten binden. Wie viele Krawattenknoten gibt es ungefähr?*	*85*, 15, 30, 55, 100*
87	Hard	*Manchmal braucht man Wochen, in schlimmeren Fällen sogar Monate für die Genesung nach einer Krankheit. Wie nennt man diese Zeit?*	*Rekonvaleszenz*, Rehabilitation, Rekonstruktion, Remission, Reevaluation*
88	Hard	*Wie ist das ungefähre Verhältnis zwischen Erdumfang und Entfernung von der Erde zum Mond?*	*etwa 1:10*, etwa 1:20, etwa 1:5, etwa 1:7, etwa 1:15*
89	Hard	*Weltweit gibt es lediglich eine Hunderasse, deren Zunge nicht die Farbe pink hat. Welche ist gemeint?*	*Chow-Chow*, Pekingese, Havaneser, Eurasier, Irish Setter*

Table A.8: List of factual search tasks 90 to 99 from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers. The * indicates the correct answer.

Nr.	Level	Task description	Answers
90	Hard	*Wie viele ganze Runden auf der Laufbahn in einem Stadion müsste man ungefähr laufen, um die Strecke von Rom nach Paris (Luftlinie) zu überbrücken?*	*2762*, 5423, 1063, 4698, 9754*
91	Hard	*In welcher olympischen Sportart werden bei Meisterschaften in der Regel zwei Bronzemedaillen vergeben?*	*Tischtennis*, Tennis, Fussball, Turnen, Boxen*
92	Hard	*Was ist die östlichste Millionenstadt Europas (Stichtag 1. Januar 2015)?*	*Perm*, Jekaterinburg, Moskau, Wolgograd, Kasan*
93	Hard	*Wie nennt man Vorsprünge der Mauern von neuzeitlichen Festungen, die tote Winkel verhindern?*	*Bastionen*, Kasematten, Speerspitzen, Glacis, Courtine*
94	Hard	*Wie nennt man die Strecke, die ein Fahrzeug ohne erneute Treibstoffzuführung zurücklegen kann?*	*Aktionsradius*, Reststrecke, Ausrollweg, Stehbereich, Ausrollstrecke*
95	Hard	*Was ist das optimale Verhältnis zwischen Luft und Brennstoff (Benzin), um den Kraftstoff eines PKWs im Katalysator vollständig zu verbrennen?*	*14.7:1*, 10.5:1, 1:5, 1:22.3, 1:1*
96	Hard	*Wie heißt das erste vom Begründer des Gonzo-Journalismus geschriebene Buch?*	*Hell's Angels*, Kill Your Friends, Der Electric Kool-Aid Acid Test, Skagboys, On the Road*
97	Hard	*In „Die Elexiere des Teufels“ von E.T.A. Hoffmann werden ähnliche Motive aufgegriffen, wie in ...*	*Der seltsame Fall des Dr. Jekyll und Mr. Hyde von Robert Louis Stevenson*, Die Verwirrungen des Zöglings Törleß von Robert Musli, Doktor Faustus von Thomas Mann, Schuld und Sühne von Fjodor Dostojewski, Der große Gatsby von Francis Scott Key Fitzgerald*
98	Hard	*Wer führte bei dem Film, der bei den Oscarverleihungen im Jahr 2014 als „bester Film“ gekürt wurde, die Regie?*	*Steve McQueen*, Tom Hooper, Peter Jackson, George Lucas, Roman Polanski*
99	Hard	*Mit welchem Instrument wollte man ursprünglich einen ganz anderen Klang erzeugen, als es nach dem Bau der Fall war?*	*Hammondorgel*, Klavier, Geige, Cello, Gitarre*

Table A.9: List of factual search tasks 100 to 109 from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers. The * indicates the correct answer.

Nr.	Level	Task description	Answers
100	Hard	*Für welche beiden Pokémon der ersten Spielgeneration lässt sich durch das Rückwärtslesen des Namens auf das Aussehen der Pokémons schließen?*	*Rettan und Arbok*, Leschum und Retsua, Suam und Ettar, Lesma und Relda, Emulb und Muab*
101	Hard	*Wie heißt der kleine Kampftrupp, der einer sich zurückziehenden Truppe folgt?*	*Nachhut*, Sturmtrupp, Späher, Vorheer, Folgertrupp*
102	Hard	*Magersucht (Anorexia nervosa) und die soziale Phobie sind ernstzunehmende psychische Störungen. Was ist beiden Erkrankungen gemein?*	*Beide Störungen beginnen typischerweise in der Jugend*, Beide Störungen treten ungefähr gleich häufig auf, Für beide Störungen gilt Untergewicht als Diagnosekriterium, Der Krankheitsverlauf beider Störungen ist meist kurzweilig, Beide Störungen findet man eher bei Männern. als bei Frauen*
103	Hard	*Nach welchem Gemüse ist die Dekoration der bekannten feinkeramischen Erzeugnisse aus Meißen benannt?*	*Zwiebel*, Gurke, Aubergine, Tomate, Blaukraut*
104	Hard	*In welchem nicht-europäischen Land wurde der einzige nicht-europäische Fußballer geboren, der für drei Nationalteams gespielt hat?*	*Argentinien*, Brasilien, Kolumbien, Venezuela, Chile*
105	Hard	*Wie heißt eine Verlängerung des Kiels eines Kite-Boards, das zur Stabilisierung dient?*	*Foil*, Finne, Trapez, Bar, Kite*
106	Hard	*Nach was ist Namibia benannt?*	*Wüste Namib*, Volk der Nama, Einen früheren Präsidenten des Landes, Kopfschmuck der Frauen in Namibia, Der Stadt Namib*
107	Hard	*An welchem süddeutschen Wallfahrtsort steht das Gnadenbild der schwarzen Muttergottes?*	*Altötting*, Benrath, Unterfauenhaid, Insel Mainau, Ramsdorf*
108	Hard	*Wer schrieb es dem Rauchen, Trinken und wenig Schlafen zu, dass er gut in Form sei?*	*Winston Churchill*, Helmut Schmidt, Sigmund Freud, Carlo Thränhardt, Ludwig Erhard*
109	Hard	*Die deutsche Predigt begann mit dem öffentlichen Auftreten welcher Mönche?*	*Dominikaner*, Johanniter, Templer, Pauliner, Kopten*

Table A.10: List of factual search tasks 110 to 117 from user study *US-III* including difficulty levels, corresponding fact-finding search tasks and possible answers. The * indicates the correct answer.

Nr.	Level	Task description	Answers
110	Hard	*Bei Henry Ford wurde das Fließband für Autobauer erfunden. Wer hat es in Europa zuerst genutzt?*	*Citroën*, Opel, Volkswagen, Fiat, Seat*
111	Hard	*Wie viele Jahre liegen zwischen dem ersten Flug des Kitty Hawk Flyers und Neil Armstrongs Mondlandung?*	*66*, 64, 72, 69, 70*
112	Hard	*Welcher französische Maler des 19en Jahrhunderts hat erst mit 40 Jahren angefangen zu malen, nach seinem Dienst beim Zoll?*	*Henri Rousseau*, Henri-Julien-Félix, Paul Gauguin, Vincent Van Gogh, Paul Cezanne*
113	Hard	*Welches Beatles Album kam im selben Jahr wie Martin Scorseses zweiter Kurzfilm raus?*	*Sgt. Pepper's Lonely Hearts Club Band*, Abbey Road, Help!, Yellow Submarine, With the Beatles*
114	Hard	*Wie heißt das Diagramm aus rein geometrischen Formen, welches vor allem im japanischen Buddhismus bei der Meditation helfen soll?*	*Yantra*, Mandala, Mantra, Huda, Boshra*
115	Hard	*Wie heißt der größte bekannte Planet des Doppelsternsystems Kepler-47?*	*Kepler-47-c*, Kepler-47-a, Kepler-47-b, Kepler-16-a, Kepler-47-d*
116	Hard	*Wie viel Prozent der deutschen Männer im Alter von 70 bis 79 leiden an Diabetes (Angaben für das Jahr 2011)?*	*22%*, 20%, 18%, 35%, 12%*
117	Hard	*Welches Tier gibt die Milch mit dem höchsten Fettgehalt?*	*Robbe*, Rind, Ziege, Kamel, Schaf*

A.3 FEATURE SPACE OF THE USER STUDIES' DATA

This section provides details about the feature spaces of the data (streams) recorded for the user studies *US-I*, *US-II* and *US-III*. This includes the User- & Screen Video, Audio and Eye-Tracking if available:

A.3.1 *User- & Screen Video*

Tab. A.11 lists the resolutions regarding the User- & Screen Videos. In *US-Ib* user videos from two different perspectives have been recorded, from the top of the screen and from the side.

Table A.11: Resolutions regarding the User- & Screen Videos.

	US	Resolution	approx. length
	US-Ia	640x480 px	10-15 min.
User	*US-Ib*	2x 1388x1038 px	20 min.
Video	*US-II*	1280x960 px	60 min.
	US-III	1280x960 px	70 min.
	US-Ia	1280x1024 px	10-15 min.
Screen	*US-Ib*	1280x1024 px	20 min.
Video	*US-II*	1920x1200 px	60 min.
	US-III	1280x1024 px	70 min.

A.3.2 *Audio*

Audio was recorded for the two studies *US-Ia* and *US-Ib*. In *US-Ia*, the built in Eye-Tracker microphone was used with a record quality of 22 kHz for the young users. In *US-Ib*, a Sennheiser headset microphone HSP 2-EW-3 was used with a record quality of 44.100 kHz for the young users and the investigator.

A.3.3 *Eye-Tracking*

Both Eye-Trackers, the Tobii X2-60 and the Tobii T60, have comparable recording properties, i.e. a sample rate of 60Hz, gaze accuracy 0.4° to 0.5°, a latency <35ms, operation distance of 40 to 90cm and both devices have been used with the eye-tracking software Tobii-Studio version 3.4.2. The software allows to extract several features. Screenshots of the features, a description and the corresp. format are given in the Figures A.1 to A.8.

General data

Contains data about export data, studio version, project and test name, participant characteristics, recording information and current fixation filter

12 of 12 selected

Name	Description	Format
ExportDate	Date when the file was exported (date of export).	YYYYMMDD
StudioVersionRec	Tobii Studio release version used to perform the recording.	
StudioProjectName	Name of the Tobii Studio Project.	
StudioTestName	Name of the Tobii Studio Test.	
ParticipantName	Name of the Participant associated with the Tobii Studio recording.	
[VariableName]Value	Displays the independent variable values associated with the participant. Enabling this column generates one column per independent variable.	
RecordingName	Name of the Recording.	
RecordingDate	Date when the recording was performed.	YYYYMMDD
RecordingDuration	Total duration of the recording.	Milliseconds
RecordingResolution	The resolution of the screen or video capture device used during the recording.	Pixels (horizontal x vertical)
PresentationSequence	The name of the presentation sequence used during the recording.	
FixationFilter	The fixation filter applied when the data was exported (Raw data; I-VT Fixation Filter; ClearView Fixation Filter; Tobii Fixation Filter).	

Figure A.1: Screenshot of the Eye-Tracker's Features: General data

Media data

Contains data about media name, position and size

5 of 5 selected

Name	Description	Format
MediaName	Name of the media/stimuli element from the Tobii Studio test timeline.	
MediaPosX (ADCSpx)	Recorded horizontal position of the media on the screen. The value represents the horizontal position of the left edge of the media in relation to the left edge of the screen.	Pixels
MediaPosY (ADCSpx)	Recorded vertical position of the media on the screen. The value represents the vertical position of the top edge of the media to the top edge of the screen.	Pixels
MediaWidth	Recorded horizontal size of the media element - width.	Pixels
MediaHeight	Recorded vertical size of the media element - height.	Pixels

Figure A.2: Screenshot of the Eye-Tracker's Features: Media data

Segment and scene data

Contains data about segments and scenes

8 of 8 selected

Name	Description	Format
SegmentName	Name of the segment in the recordings selected for export in the Select Segments for Export option. This column will only contain values when the recordings are selected in theSelect Segments for Export type.	
SegmentStart	The start time for each segment in each recording. The time value is calculated relative to the recording timeline. This column will only contain values when the recordings are selected in theSelect Segments for Export type.	Milliseconds
SegmentEnd	The end time for each segment in each recording. The time value is calculated relative to the recording timeline. This column will only contain values when the recordings are selected in theSelect Segments for Export type.	Milliseconds
SegmentDuration	Duration of each segment in the recordings. This column will only contain values when the recordings are selected in theSelect Segments for Export type.	Milliseconds
SceneName	Name of the Scene selected for export in the Select Media for Export option. This column will only contain values if the scene is selected in the Select Media fo Export type.	
SceneSegmentStart	The start time for each segment belonging to a scene. The time value is calculated relative to the recording timeline. This column will only contain values if the scene is selected in the Select Media fo Export type.	Milliseconds
SceneSegmentEnd	The end time for each segment belonging to a scene. The time value is calculated relative to the recording timeline. This column will only contain values if the scene is selected in the Select Media fo Export type.	Milliseconds
SceneSegmentDuration	Duration of each segment belonging to a scene. This column will only contain values if the scene is selected in the Select Media fo Export type.	Milliseconds

Figure A.3: Screenshot of the Eye-Tracker's Features: Segment and scene data

Timestamp data

Contains data about the time stamps available for export

3 of 3 selected

☑	Name	Description	Format
☑	RecordingTimestamp	Timestamp counted from the start of the recording (t0=0 ms).	Milliseconds
☑	LocalTimeStamp	Recording computer local "date time value" timestamp.	HHMMSSmmm
☑	EyeTrackerTimestamp	Timestamp obtained from the eye tracker firmware (TET server) clock.	Microseconds

Figure A.4: Screenshot of the Eye-Tracker's Features: Timestamp data

Recording event data

Contains data about events recorded during a recording such as mouse clicks, key presses, studio events and external events

15 of 15 selected

☑	Name	Description	Format
☑	MouseEventIndex	Represents the order in which a mouse click event was recorded. The index is an auto-increment number starting with 1 (first mouse event detected).	Count
☑	MouseEvent	Mouse click type: Left button and Right button.	
☑	MouseEventX (ADCSpx)	Horizontal coordinate of the mouse event location on the screen.	Pixels
☑	MouseEventY (ADCSpx)	Vertical coordinate of the mouse event location on the screen.	Pixels
☑	MouseEventX (MCSpx)	Horizontal coordinate of the mouse event location on the media element.	Pixels
☑	MouseEventY (MCSpx)	Vertical coordinate of the mouse event location on the media element.	Pixels
☑	KeyPressEventIndex	Represents the order in which a key press event was recorded. The index is an auto-increment number starting with 1 (first key press event detected).	Count
☑	KeyPressEvent	Displays the information of which key was pressed on the keyboard.	
☑	StudioEventIndex	Represents the order in which a Tobii Studio recording event was registered. The index is an auto-increment number starting with 1 (first event detected).	Count
☑	StudioEvent	Type of media element or manual logging event (START, END and manual logging description).	
☑	StudioEventData	Displays the name of the media element, webpage URL or PDF page. For Manual logging it displays the text entered in the log message.	
☑	ExternalEventIndex	Represents the order in which external events were logged during the recording. The index is an auto-increment number starting with 1 (first event detected).	Count
☑	ExternalEvent	Type of event logged by an external device.	
☑	ExternalEventValue	Value of external event.	
☑	EventMarkerValue	Reports whether the signal on the TX300's sync port is off or on: 0 = off, 1 = on.	

Figure A.5: Screenshot of the Eye-Tracker's Features: Recording event data

Gaze event data and AOI activity information

Contains data about gaze event data such as fixations and saccades as well as about AOI activity information

10 of 10 selected

Name	Description	Format
FixationIndex	Represents the order in which a fixation (i.e. fixations and saccades). The index is an auto-increment number starting with 1 (first fixation detected).	Count
SaccadeIndex	Represents the order in which a saccade was recorded. The index is an auto-increment number starting with 1 (first saccade detected).	Count
GazeEventType	Type of eye movement event classified by the fixation filter settings applied during the gaze data export: Fixation, Saccade and Unclassified.	Fixation; Saccade; Unclassified
GazeEventDuration	Duration of an eye movement event.	Milliseconds
FixationPointX (MCSpx)	Horizontal coordinate of the fixation point on the media. This column is affected by the settings in the Fixation Filter Tab (Global Settings).	Pixels
FixationPointY (MCSpx)	Vertical coordinate of the fixation point in the media. This column is affected by the settings in the Fixation Filter Tab (Global Settings).	Pixels
SaccadicAmplitude	Distance in visual degrees between the previous fixation location and the current fixation location as defined by the fixation filter.	Degrees
AbsoluteSaccadicDirection	Offset in degrees between the horizontal axis and the current fixation location where the previous fixation location is set as origo.	Degrees
RelativeSaccadicDirection	The difference between the absolute saccadic direction of the current and previous saccade where the current saccade is between the current and previous fixation.	Degrees
AOI[Name of AOI]Hit	Reports whether the AOI is active and whether the the fixation is located inside of the AOI: -1 = AOI Not active; 0 = AOI active, the fixation is not located in the AOI, 1 = AOI active and the fixation is located inside of the AOI. Enabling this column generates one column per AOI.	

Figure A.6: Screenshot of the Eye-Tracker's Features: Gaze event data and AOI (Area of Interest) activity information

Gaze tracking data

Contains the raw gaze point of each data sample expressed in pixels or millimeters for each eye individually or as an average

15 of 15 selected

Name	Description	Format
GazePointIndex	Represents the order in which the gaze sample was acquired by Tobii Studio from an eye tracker. The index is an auto-increment number starting with 1 (first gaze sample).	Count
GazePointLeftX (ADCSpx)	Horizontal coordinate of the unprocessed gaze point for the left eye on the screen.	Pixels
GazePointLeftY (ADCSpx)	Vertical coordinate of the unprocessed gaze point for the left eye on the screen.	Pixels
GazePointRightX (ADCSpx)	Horizontal coordinate of the unprocessed gaze point for the right eye on the screen.	Pixels
GazePointRightY (ADCSpx)	Vertical coordinate of the unprocessed gaze point for the right eye on the screen.	Pixels
GazePointX (ADCSpx)	Horizontal coordinate of the averaged left and right eye gaze point on the screen.	Pixels
GazePointY (ADCSpx)	Vertical coordinate of the averaged left and right eye gaze point on the screen.	Pixels
GazePointX (MCSpx)	Horizontal coordinate of the averaged left and right eye gaze point on the media element.	Pixels
GazePointY (MCSpx)	Vertical coordinate of the averaged left and right eye gaze point on the media element.	Pixels
GazePointLeftX (ADCSmm)	Horizontal coordinate of the unprocessed gaze point for the left eye on the screen.	Millimeters
GazePointLeftY (ADCSmm)	Vertical coordinate of the unprocessed gaze point for the left eye on the screen.	Millimeters
GazePointRightX (ADCSmm)	Horizontal coordinate of the unprocessed gaze point for the right eye on the screen.	Millimeters
GazePointRightY (ADCSmm)	Vertical coordinate of the unprocessed gaze point for the right eye on the screen.	Millimeters
StrictAverageGazePointX (ADCSmm)	Horizontal coordinate of the averaged gaze point for both eyes on the screen. "average" function similar to the one used for Eye selection.	Millimeters
StrictAverageGazePointY (ADCSmm)	Vertical coordinate of the averaged gaze point for both eyes on the screen. "average" function similar to the one used for Eye selection.	Millimeters

Figure A.7: Screenshot of the Eye-Tracker's Features: Gaze tracking data

Eye tracking data and validity codes

Contains data such as eye position, pupil size and validity codes

19 of 19 selected

Name	Description	Format
EyePosLeftX (ADCSmm)	Horizontal coordinate of the 3D position of the left eye.	Millimeters
EyePosLeftY (ADCSmm)	Vertical coordinate of the 3D position of the left eye.	Millimeters
EyePosLeftZ (ADCSmm)	Distance/depth coordinate of the 3D position of the left eye.	Millimeters
EyePosRightX (ADCSmm)	Horizontal coordinate of the 3D position of the right eye.	Millimeters
EyePosRightY (ADCSmm)	Vertical coordinate of the 3D position of the right eye.	Millimeters
EyePosRightZ (ADCSmm)	Distance/depth coordinate of the 3D position of the right eye.	Millimeters
CamLeftX	Horizontal location of the left eye in the eye tracker sensor image (0 is the left edge of the image, 1 is the right edge) - Only available for recordings made with Tobii Studio versions prior to Tobii Studio 2.3.	Normalized Coordinate
CamLeftY	Vertical location of the left eye in the eye tracker sensor image (0 is the top edge of the image, 1 is the bottom edge) - Only available for recordings made with Tobii Studio versions prior to Tobii Studio 2.3.	Normalized Coordinate
CamRightX	Horizontal location of the right eye in the eye tracker sensor image (0 is the left edge of the image, 1 is the right edge) - Only available for recordings made with Tobii Studio versions prior to Tobii Studio 2.3.	Normalized Coordinate
CamRightY	Vertical location of the right eye in the eye tracker sensor image (0 is the top edge of the image, 1 is the bottom edge) - Only available for recordings made with Tobii Studio versions prior to Tobii Studio 2.3.	Normalized Coordinate
DistanceLeft	Distance between the left eye and the eye tracker. The distance is calculated based on the length of the vector from the center of the eye to the UCS origin point (User Coordinate System) on the eye tracker.	Millimeters
DistanceRight	Distance between the right eye and the eye tracker. The distance is calculated based on the length of the vector from the center of the eye to the UCS origin point (User Coordinate System) on the eye tracker.	Millimeters
PupilLeft	Estimated size of the left eye pupil.	Millimeters
PupilRight	Estimated size of the right eye pupil.	Millimeters
ValidityLeft	Indicates the confidence level that the left eye has been correctly identified. The values range from 0 (high confidence) to 4 (eye not found).	
ValidityRight	Indicates the confidence level that the right eye has been correctly identified. The values range from 0 (high confidence) to 4 (eye not found).	
IRMarkerCount	The number of IR-Markers detected in the current sample.	Count
IRMarkerID	The ID numbers of the IR-Markers detected in the current sample. The information is presented as a comma separated list, ordered from the lowest ID to the highest ID.	
PupilGlassesRight	Estimated value of the right pupil size in relation to the average value of the pupil during the calibration. The pupil size is measured during the recording and the calibration as the largest value of the diameter of the pupil as seen in the eye image.	%

Figure A.8: Screenshot of the Eye-Tracker's Features: Eye tracking data and validity codes

A.4 ALL PERSONALITY RELATED MODELS & DURATIONS

This section lists all Markov model components and state dwell times for the models regarding the five personality factors obtained in the user study *US-III*. That is, one model $\theta^H_{<f>}$ for the exploratory search behavior from users with higher (H) values than the median of the corresponding personality factor $< f >$ have been trained and one model $\theta^L_{<f>}$ for the users with lower (L) values have been be trained. In particular, Tab. A.12 lists all dwell times and figures afterwards contain the model components regarding: *Neuroticism* (N) A.9, *Extraversion* (E) A.10, *Openness to experience* (O) A.11, *Agreeableness* (A) A.12 and *Conscientiousness* (C) A.13.

Table A.12: Mean dwell times on each state from Q in sec. according to the models $\theta^H_{<f>}$ and $\theta^L_{<f>}$ from *US-III* using all Expl of users with higher (H) and lower (L) values than the median on the personality factor $< f >$. Furthermore, the p-Values of the corresp. KST are given.

		Query	*SERP*	*Page*	*Main*
N	θ^H_N	4.1	9.2	20.0	23.8
	θ^L_N	4.0	8.6	21.4	25.4
	p-Value	0.6675	0.0272	0.7748	0.0773
E	θ^H_E	3.8	9.2	19.2	24.6
	θ^L_E	4.4	8.7	22.4	24.4
	p-Value	0.8285	0.206	0.2291	0.002
O	θ^H_O	3.8	9.3	21.9	24.9
	θ^L_O	4.4	8.5	19.1	24.0
	p-Value	0.497	0.0552	2.9e-06	0.0023
A	θ^H_A	3.5	8.6	20.0	22.4
	θ^L_A	4.8	9.5	21.7	27.7
	p-Value	0.0194	0.1204	0.1471	0.0882
C	θ^H_C	4.1	8.4	20.1	26.4
	θ^L_C	4.1	9.7	21.4	22.2
	p-Value	0.9419	0.0205	0.2763	0.0008

$$\theta_N^H : \qquad\qquad \theta_N^L :$$

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\} \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.29 & 0.01 & 0.54 & 0.16 \\ 0.02 & 0.22 & 0.22 & 0.54 \\ 0.02 & 0.13 & 0.85 & 0.00 \end{bmatrix} \qquad A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.22 & 0.03 & 0.59 & 0.16 \\ 0.02 & 0.24 & 0.20 & 0.54 \\ 0.02 & 0.11 & 0.87 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T \qquad \pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.9: Illustration of the components Q, A and π of the 1st-order Markov models θ_N^H (left) and θ_N^L (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Neuroticism* (N).

$$\theta_E^H : \qquad\qquad \theta_E^L :$$

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\} \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.25 & 0.02 & 0.57 & 0.16 \\ 0.03 & 0.24 & 0.20 & 0.53 \\ 0.02 & 0.13 & 0.85 & 0.00 \end{bmatrix} \qquad A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.28 & 0.02 & 0.55 & 0.15 \\ 0.02 & 0.23 & 0.20 & 0.55 \\ 0.02 & 0.13 & 0.85 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T \qquad \pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.10: Illustration of the components Q, A and π of the 1st-order Markov models θ_E^H (left) and θ_E^L (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Extraversion* (E).

$$\theta_O^H : \qquad\qquad \theta_O^L :$$

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\} \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.26 & 0.02 & 0.58 & 0.14 \\ 0.02 & 0.25 & 0.17 & 0.56 \\ 0.01 & 0.11 & 0.88 & 0.00 \end{bmatrix} \qquad A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.26 & 0.02 & 0.53 & 0.19 \\ 0.02 & 0.22 & 0.25 & 0.51 \\ 0.04 & 0.15 & 0.81 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T \qquad \pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.11: Illustration of the components Q, A and π of the 1st-order Markov models θ_O^H (left) and θ_O^L (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Openness to experience* (O).

$$\theta_A^H : \qquad\qquad \theta_A^L :$$

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\} \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.26 & 0.02 & 0.56 & 0.16 \\ 0.02 & 0.24 & 0.20 & 0.54 \\ 0.02 & 0.12 & 0.86 & 0.00 \end{bmatrix} \qquad A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.27 & 0.01 & 0.57 & 0.15 \\ 0.03 & 0.22 & 0.21 & 0.54 \\ 0.03 & 0.13 & 0.84 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T \qquad \pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.12: Illustration of the components Q, A and π of the 1st-order Markov models θ_O^H (left) and θ_O^L (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Agreeableness* (A).

$$\theta_C^H:\qquad\qquad \theta_C^L:$$

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\} \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.27 & 0.01 & 0.56 & 0.16 \\ 0.03 & 0.23 & 0.21 & 0.53 \\ 0.02 & 0.13 & 0.85 & 0.00 \end{bmatrix} \qquad A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.26 & 0.02 & 0.56 & 0.16 \\ 0.02 & 0.23 & 0.20 & 0.55 \\ 0.03 & 0.12 & 0.85 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T \qquad \pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.13: Illustration of the components Q, A and π of the 1st-order Markov models θ_O^H (left) and θ_O^L (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Conscientiousness* (C).

A.5 ALL INTELLIGENCE RELATED MODELS & DURATIONS

This section lists all Markov model components and state dwell times for the models regarding the three aspects of intelligence obtained in the user study *US-III*. That is, one model $\theta_{<i>}^H$ for the exploratory search behavior from users with higher (H) values than the median of the corresponding aspects of intelligence $< i >$ have been trained and one model $\theta_{<i>}^L$ for the users with lower (L) values have been be trained. In particular, Tab. A.13 lists all dwell times and the figures afterwards contain the model components regarding: *Similarities* (Si) A.14, *Symbol Search* (Sy) A.15 and *Letter-Number Sequencing* (LN) A.16.

Table A.13: Mean dwell times on each state from Q in sec. according to the models $\theta^H_{<i>}$ and $\theta^L_{<i>}$ from *US-III* using all Expl of users with higher (H) and lower (L) values than the median on the aspects of intelligence $< i >$. Furthermore, the p-Values of the corresp. KST are given.

		Query	*SERP*	*Page*	*Main*
Si	θ^H_{Si}	3.9	8.8	20.0	24.7
	θ^L_{Si}	4.2	9.2	21.6	24.2
	p-Value	0.7831	0.0337	0.1934	0.4494
Sy	θ^H_{Sy}	3.7	8.5	19.3	24.6
	θ^L_{Sy}	4.4	9.4	22.2	24.4
	p-Value	0.541	0.2445	0.0061	0.3285
LN	θ^H_{LN}	4.0	8.5	18.7	24.2
	θ^L_{LN}	4.1	9.6	23.3	25.0
	p-Value	0.2991	0.0292	0.0434	0.3921

θ^H_{Si} :

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.25 & 0.02 & 0.56 & 0.17 \\ 0.02 & 0.23 & 0.20 & 0.55 \\ 0.02 & 0.12 & 0.86 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

θ^L_{Si} :

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.28 & 0.01 & 0.56 & 0.15 \\ 0.02 & 0.24 & 0.21 & 0.53 \\ 0.03 & 0.13 & 0.84 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.14: Illustration of the components Q, A and π of the 1st-order Markov models θ^H_{Si} (left) and θ^L_{Si} (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Similarities* (Si).

$$\theta^H_{Sy}: \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.22 & 0.02 & 0.58 & 0.18 \\ 0.02 & 0.25 & 0.19 & 0.55 \\ 0.02 & 0.12 & 0.86 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

$$\theta^L_{Sy}: \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.30 & 0.02 & 0.54 & 0.14 \\ 0.02 & 0.22 & 0.23 & 0.53 \\ 0.03 & 0.12 & 0.85 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.15: Illustration of the components Q, A and π of the 1st-order Markov models θ^H_{Sy} (left) and θ^L_{Sy} (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Symbol Search* (Sy).

$$\theta^H_{LN}: \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.26 & 0.02 & 0.55 & 0.17 \\ 0.02 & 0.24 & 0.20 & 0.54 \\ 0.02 & 0.13 & 0.85 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

$$\theta^L_{LN}: \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.27 & 0.01 & 0.58 & 0.14 \\ 0.02 & 0.23 & 0.21 & 0.54 \\ 0.02 & 0.12 & 0.86 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.16: Illustration of the components Q, A and π of the 1st-order Markov models θ^H_{LN} (left) and θ^L_{LN} (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Letter-Number Sequencing* (LN).

A.6 ALL SENSATION SEEKING RELATED MODELS & DURATIONS

This section lists all Markov model components and state dwell times for the models regarding the four sensation seeking sub-factors three obtained by the SSS in the user study *US-III*. That is, one model $\theta^{H}_{<s>}$ for the exploratory search behavior from users with higher (H) values than the median of the corresponding sub-factor $< s >$ have been trained and one model $\theta^{L}_{<s>}$ for the users with lower (L) values have been be trained. In particular, Tab. A.14 lists all dwell times and the figures afterwards contain the model components regarding: *Thrill and Adventure Seeking* (TAS) A.17, *Experience Seeking* (EXS) A.18; *Disinhibition* (DIS) A.19; and *Boredom Susceptibility* (BOS) A.20.

Table A.14: Mean dwell times on each state from Q in sec. according to the models $\theta^{H}_{<s>}$ and $\theta^{L}_{<s>}$ from *US-III* using all Expl of users with higher (H) and lower (L) values than the median on the sensation seeking sub-factors $< s >$. Furthermore, the p-Values of the corresp. KST are given.

		Query	*SERP*	*Page*	*Main*
TAS	θ^{H}_{TAS}	3.9	9.3	20.7	23.5
	θ^{L}_{TAS}	4.4	8.4	20.7	26.0
	p-Value	0.2199	0.0036	0.5515	0.1612
EXS	θ^{H}_{EXS}	3.7	8.6	20.3	24.0
	θ^{L}_{EXS}	4.9	9.6	21.2	25.4
	p-Value	0.0526	0.4533	0.1999	0.0556
DIS	θ^{H}_{DIS}	3.7	8.4	19.2	23.6
	θ^{L}_{DIS}	4.8	10.2	24.0	26.5
	p-Value	0.3422	0.0030	8.4e-05	0.1362
BOS	θ^{H}_{BOS}	4.4	9.3	22.1	25.3
	θ^{L}_{BOS}	3.7	8.7	19.3	23.8
	p-Value	0.2839	0.5541	0.0802	0.1361

$$\theta^{H}_{TAS}: \qquad \qquad \theta^{L}_{TAS}:$$

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\} \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.29 & 0.02 & 0.56 & 0.13 \\ 0.02 & 0.23 & 0.19 & 0.56 \\ 0.02 & 0.12 & 0.86 & 0.00 \end{bmatrix} \qquad A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.22 & 0.02 & 0.56 & 0.20 \\ 0.02 & 0.24 & 0.23 & 0.51 \\ 0.03 & 0.13 & 0.84 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T \qquad \pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.17: Illustration of the components Q, A and π of the 1st-order Markov models θ^{H}_{TAS} (left) and θ^{L}_{TAS} (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Thrill and Adventure Seeking* (TAS).

$$\theta^{H}_{EXS}: \qquad \qquad \theta^{L}_{EXS}:$$

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\} \qquad Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$

$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.29 & 0.02 & 0.54 & 0.15 \\ 0.02 & 0.23 & 0.19 & 0.56 \\ 0.02 & 0.13 & 0.85 & 0.00 \end{bmatrix} \qquad A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.21 & 0.02 & 0.59 & 0.18 \\ 0.02 & 0.23 & 0.25 & 0.50 \\ 0.02 & 0.13 & 0.85 & 0.00 \end{bmatrix}$$

$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T \qquad \pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.18: Illustration of the components Q, A and π of the 1st-order Markov models θ^{H}_{EXS} (left) and θ^{L}_{EXS} (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Experience Seeking* (EXS).

θ^{H}_{DIS} :

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$
$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.25 & 0.02 & 0.56 & 0.17 \\ 0.02 & 0.23 & 0.23 & 0.52 \\ 0.02 & 0.14 & 0.84 & 0.00 \end{bmatrix}$$
$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

θ^{L}_{DIS} :

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$
$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.28 & 0.01 & 0.57 & 0.14 \\ 0.02 & 0.23 & 0.17 & 0.58 \\ 0.03 & 0.10 & 0.87 & 0.00 \end{bmatrix}$$
$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.19: Illustration of the components Q, A and π of the 1st-order Markov models θ^{H}_{DIS} (left) and θ^{L}_{DIS} (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Disinhibition* (DIS).

θ^{H}_{BOS} :

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$
$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.27 & 0.02 & 0.58 & 0.13 \\ 0.03 & 0.21 & 0.22 & 0.54 \\ 0.02 & 0.12 & 0.86 & 0.00 \end{bmatrix}$$
$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

θ^{L}_{BOS} :

$$Q = \{q_{query}, q_{serp}, q_{page}, q_{main}\}$$
$$A = \begin{bmatrix} 0.00 & 1.00 & 0.00 & 0.00 \\ 0.26 & 0.02 & 0.54 & 0.18 \\ 0.01 & 0.25 & 0.20 & 0.54 \\ 0.02 & 0.13 & 0.85 & 0.00 \end{bmatrix}$$
$$\pi = \begin{pmatrix} \pi_{query} \\ \pi_{serp} \\ \pi_{page} \\ \pi_{main} \end{pmatrix}^T = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \end{pmatrix}^T$$

Figure A.20: Illustration of the components Q, A and π of the 1st-order Markov models θ^{H}_{BOS} (left) and θ^{L}_{BOS} (right) trained on the data set from *US-III* using all Expl of users with higher (H) and lower (L) values on *Boredom Susceptibility* (BOS).

[1] Henry De-Graft Acquah. "Comparison of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in Selection of an Asymmetric Price Relationship." In: *Journal of Development and Agricultural Economics* 2.1 (2010), pp. 1–6.

[2] Mikhail Ageev, Qi Guo, Dmitry Lagun, and Eugene Agichtein. "Find It If You Can: A Game for Modeling Different Types of Web Search Success Using Interaction Data." In: *Proc. of the 34th Int. ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. 2011, pp. 345–354.

[3] Eugene Agichtein, Eric Brill, Susan Dumais, and Robert Ragno. "Learning User Interaction Models for Predicting Web Search Result Preferences." In: *Proc. of the 29th Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '06. Seattle, Washington, USA: ACM, 2006, pp. 3–10.

[4] Francis Joseph Aguilar. *Scanning the Business Environment*. Studies of the modern corporation. New York, USA: Macmillan, 1967.

[5] Hirotogu Akaike. "Information Theory and an Extension of the Maximum Likelihood Principle." In: *Second Int. Symposium on Information Theory*. Ed. by B. N. Petrov and F. Csaki. Budapest: Akademiai Kiado, 1973.

[6] Hirotogu Akaike. "A New Look at the Statistical Model Identification." In: *IEEE Transactions on Automatic Control* AC-19.6 (1974), pp. 716–723.

[7] M. N. Al-Suqri and A. S. Al-Aufi. *Information Seeking Behavior and Technology Adoption: Theories and Trends.* IGI Global, 2015.

[8] Manfred Amelang, Dieter Bartussek, Gerhard Stemmler, and Dirk Hagemann. *Differentielle psychologie und persönlichkeitsforschung*. Ed. by Herbert Heuer, Frank Rösler, and Werner H. Tack. Standards Psychologie. W. Kohlhammer Verlag, 2006.

[9] Michael von Aster, A. Neubauer, and Horn R. *Wechsler Intelligenztest für Erwachsene. WIE: Manual. Deutschsprachige Bearbeitung und Adaption des WAIS-III von David Wechsler (2. Aufl.)* Frankfurt am Main: Pearson Assessment & Information GmbH, 2009.

[10] Kumaripaba Athukorala, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. "Is Exploratory Search Different? A Comparison of Information Search Behavior for Exploratory and Lookup Tasks." In: *Journal of the Association for Information Science and Technology* 67.11 (2016), pp. 2635–2651.

[11] Anne Aula, Rehan M. Khan, and Zhiwei Guan. "How does Search Behavior Change as Search becomes more Difficult?" In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2010, pp. 35–44.

[12] Anne Aula and Daniel M Russell. "Complex and exploratory web search." In: *Information Seeking Support Systems Workshop (ISSS 2008), Chapel Hill, NC, USA*. 2008.

[13] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern information retrieval*. Vol. 463. ACM press New York, 1999.

[14] Marcia J. Bates. "Information Search Tactics." In: *Journal of the American Society for information Science* 30.4 (1979), pp. 205–214.

[15] Marcia J. Bates. "The Design of Browsing and Berrypicking Techniques for the Online Search Interface." In: *Online Information Review* 13.5 (1989), pp. 407–424.

[16] Marcia J. Bates. "What is Browsing-really? A Model drawing from Behavioural Science research." In: *Information Research* 12(4) paper 330 (2007), pp. 1–22.

[17] Marcia J Bates. "Information behavior." In: *Encyclopedia of library and information sciences* 3 (2010), pp. 2381–2391.

[18] André Beauducel and Burkhard Brocke. "Sensation Seeking Scale - Form V: Merkmale des Verfahrens und Bemerkungen zur deutschsprachigen Adaption." In: ed. by M. Roth and P. Hammelstein. Sensation Seeking - Konzeption, Diagnostik und Anwendung. Göttingen: Hogrefe, 2003, pp. 77–99.

[19] André Beauducel, Alexander Strobel, and Burkhard Brocke. "Psychometrische Eigenschaften und Normen einer deutschsprachigen Fassung der Sensation Seeking Skalen, Form V." In: *Diagnostica* 49.2 (2003), pp. 61–72.

[20] N.J. Belkin, Helen M. Brooks, and Robert N. Oddy. "ASK for information retrieval: Part I Background and theory." In: *Journal of Documentation* (1982), pp. 61–71.

[21] Nicholas J. Belkin. "Anomalous states of knowledge as a basis for information retrieval." In: *Canadian journal of information science* 5.1 (1980), pp. 133–143.

[22] Nicholas J. Belkin. "Some (what) grand Challenges for Information Retrieval." In: *ACM SIGIR Forum*. Vol. 42. 1. ACM. 2008, pp. 47–54.

[23] Nicholas J. Belkin, Colleen Cool, Adelheit Stein, and Ulrich Thiel. "Cases, Scripts, and Information-Seeking Strategies: On the Design of Interactive Information Retrieval Systems." In: *Expert Systems with Applications* 9.3 (1995), pp. 379–395.

[24] Nilavra Bhattacharya and Jacek Gwizdka. "Measuring learning during search: Differences in interactions, eye-gaze, and semantic similarity to expert knowledge." In: *Proc. of the 2019 conference on human information interaction and retrieval*. 2019, pp. 63–71.

[25] Ralf Bierig, Jacek Gwizdka, and Michael J Cole. "A user-centered experiment and logging framework for interactive information retrieval." In: *Proc. of the SIGIR 2009 Workshop on Understanding the User: Logging and interpreting user interactions in information search and retrieval*. 2009, pp. 8–11.

[26] Dania Bilal and Joe Kirby. "Differences and Similarities in Information Seeking: Children and Adults as Web Users." In: *Information Processing & Management* 38.5 (2002), pp. 649–670.

[27] Christopher M Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer-Verlag New York, 2006.

[28] Paolo Boldi, Francesco Bonchi, Carlos Castillo, Debora Donato, Aristides Gionis, and Sebastiano Vigna. "The Query-Flow Graph: Model and Applications." In: *Proc. of the 17th ACM Conference on Information and Knowledge Management*. CIKM '08. Napa Valley, California, USA: ACM, 2008, pp. 609–618.

[29] P. Borkenau and F. Ostendorf. *NEO-Fünf-Faktoren Inventar (NEO-FFI); nach Costa und McCrae (2. Aufl.)* Göttingen: Hogrefe, 2008.

[30] J. C. Brunstein and H. Heckhausen. "Leistungsmotivation." In: *Motivation und Handeln*. Ed. by Jutta Heckhausen and Heinz Heckhausen. Springer, Berlin, Heidelberg, 2018, pp. 163–221.

[31] Raluca Budiu and Jakob Nielsen. *Usability of Websites for Children: Design Guidelines for Targeting Users Aged 3–12 Years, 2nd edition*. Fremont, CA: Nielsen Norman Group Report, 2010.

[32] Kenneth P Burnham and David R Anderson. "Multimodel Inference: Understanding AIC and BIC in Model Selection." In: *Sociological Methods & Research* 33.2 (2004), pp. 261–304.

[33] Peter Butka, Thomas Low, Michael Kotzyba, Stefan Haun, and Andreas Nürnberger. "A Framework for FCA-based Exploratory Web Search." In: *Proc. of the 1st Int. Symposium on Companion-Technology, ISCT '15*. 1. COST. Ulm, Germany, 2015, pp. 131–136.

[34] Peter Butka, Thomas Low, Michael Kotzyba, Stefan Haun, and Andreas Nürnberger. "Exploration of Web search results based on the formal concept analysis." In: *Semanitic Keyword-based Search on Structured Data Sources*. Vol. 10546. Third Int. KEYSTONE Conference, IKC '17. Gdańsk, Poland: Springer, 2017, pp. 123–135.

[35] Katriina Byström. "Information and information sources in tasks of varying complexity." In: *Journal of the American Society for information Science and Technology* 53.7 (2002), pp. 581–591.

[36] Donald J Campbell. "Task complexity: A review and analysis." In: *Academy of Management Review* 13.1 (1988), pp. 40–52.

[37] Donald Owen Case and Lisa M. Given. *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Vol. 4. Studies in Information. Emerald Group Publishing Limited, 2016.

[38] Olivier Chapelle and Ya Zhang. "A dynamic bayesian network click model for web search ranking." In: *Proc. of the 18th Int. conference on World wide web*. WWW '09. Madrid, Spain: ACM, 2009, pp. 1–10.

[39] Chun Wei Choo, Brian Detlor, and Dan Turnbull. "Information Seeking on the Web: An Integrated Model of Browsing and Searching." In: *First Monday* 5.2 (2000).

[40] Malcolm Clark, Ian Ruthven, Patrik O'Brian Holt, Dawei Song, and Stuart Watt. "You have e-mail, what happens next? Tracking the eyes for genre." In: *Information Processing & Management* 50.1 (2014), pp. 175–198.

[41] Mark Claypool, David Brown, Phong Le, and Makoto Waseda. "Inferring User Interest." In: *IEEE Internet Computing* 5.6 (2001), pp. 32–39.

[42] Michael J. Cole, Chathra Hendahewa, Nicholas J. Belkin, and Chirag Shah. "Discrimination between Tasks with User Activity Patterns During Information Search." In: *Proc. of the 37th Int. ACM SIGIR Conference on Research & development in Information Retrieval*. ACM. 2014, pp. 567–576.

[43] Linda Z. Cooper. "A case study of information-seeking behavior in 7-year-old children in a semistructured situation." In: *Journal of the American society for information science and technology* 53.11 (2002), pp. 904–922.

[44] Linda Z. Cooper. "Methodology for a project examining cognitive categories for library information in young children." In: *Journal of the American Society for Information Science and Technology* 53.14 (2002), pp. 1223–1231.

[45] Paul T. Costa Jr and Robert R. McCrae. "Four ways five factors are basic." In: *Personality and individual differences* 13.6 (1992), pp. 653–665.

[46] Michel Crampes and Sylvie Ranwez. "Ontology-supported and ontology-driven conceptual navigation on the World Wide Web." In: *Proc. of the eleventh ACM on Hypertext and hypermedia*. 2000, pp. 191–199.

[47] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. "An Experimental Comparison of Click Position-Bias Models." In: *WSDM '08: Proc. of the Int. conference on Web search and web data mining*. ACM, 2008, pp. 87–94.

[48] Edward Cutrell and Zhiwei Guan. "Eye tracking in MSN Search: Investigating snippet length, target position and task types." In: *City* (2007), pp. 1–13.

[49] Edward Cutrell and Zhiwei Guan. "What are you looking for? An eye-tracking study of information usage in web search." In: *Proc. of the SIGCHI conference on Human factors in computing systems*. 2007, pp. 407–416.

[50] Richard Daft and Karl Weick. "Toward A Model of Organizations as Interpretation Systems." In: *Academy of Management Review* 9.2 (1984), pp. 284–295.

[51] Jan De Leeuw. "Introduction to Akaike (1973) Information Theory and an Extension of the Maximum Likelihood Principle." In: *Breakthroughs in Statistics*. Ed. by Samuel Kotz and Norman L. Johnson. Vol. 1. London: Springer-Verlag, 1992, pp. 599–609.

[52] Brenda Dervin. "An Overview of Sense-Making Research: Concepts, Methods, and Results to Date." In: *Int. Communications Association Annual Meeting* (1983).

[53] Oliver Dickhäuser and Falko Rheinberg. "Bezugsnormorientierung: Erfassung, Probleme, Perspektiven." In: *Diagnostik von Motivation und Selbstkonzept* (2003), pp. 41–55.

[54] Doug Downey, Susan T Dumais, and Eric Horvitz. "Models of Searching and Browsing: Languages, Studies, and Application." In: *IJCAI Int. Joint Conference on Artificial Intelligence*. Vol. 7. 2007, pp. 2740–2747.

[55] Susan Dumais, Edward Cutrell, and Hao Chen. "Optimizing Search by Showing Results in Context." In: *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '01. New York, NY, USA: ACM, 2001, pp. 277–284.

[56] Georges E. Dupret and Benjamin Piwowarski. "A User Browsing Model to Predict Search Engine Click Data from Past Observations." In: *Proc. of the 31st Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2008, pp. 331–338.

[57] Paul Ekman and Wallace V. Friesen. “Facial Action Coding System: A technique for the measurement of facial movement.” In: Palo Alto, CA; Consultion Psychologists Press, 1978.

[58] Robert J Elliott, Lakhdar Aggoun, and John B Moore. *Hidden Markov models: estimation and control*. Vol. 29. Springer Science & Business Media, 2008.

[59] David Ellis. “A Behavioural Approach to Information Retrieval System Design.” In: *Journal of Documentation* 45.3 (1989), pp. 171–212.

[60] David Ellis, Deborah Cox, and Katherine Hall. “A Comparison of the Information Seeking Patterns of Researchers in the Physical and Social Sciences.” In: *Journal of Documentation* 49.4 (1993), pp. 356–369.

[61] David Ellis and Merete Haugan. “Modelling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment.” In: *Journal of Documentation* 53.4 (1997), pp. 384–403.

[62] E.H. Erikson. “Children and society.” In: *NY: Narton* (1963).

[63] Leon Festinger. “A Theory of Social Comparison Processes.” In: *Human relations* 7.2 (1954), pp. 117–140.

[64] Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White. “Evaluating implicit measures to improve web search.” In: *ACM Transactions on Information Systems (TOIS)* 23.2 (2005), pp. 147–168.

[65] Bernhard Ganter and Rudolf Wille. *Formal concept analysis: mathematical foundations*. Springer Science & Business Media, 2012.

[66] Howard Gardner. “The Mind’s New Science: A History of the Cognitive Revolution.” In: (1985).

[67] Zoubin Ghahramani. “Learning Dynamic Bayesian Networks.” In: *Adaptive Processing of Temporal Inforamtion*. Ed. by C. L. Giles and M. Gori. Lecture notes in artificial intelligence. Springer-Verlag, 1998, pp. 168–197.

[68] Tatiana Gossen. *Search Engines for Children: Search User Interfaces and Information-Seeking Behaviour*. Springer, 2016.

[69] Tatiana Gossen, Michael Kotzyba, and Andreas Nürnberger. “Anforderungen und Verhalten junger Nutzer bei der Informationssuche - Ein Überblick.” In: *Information-Wissenschaft & Praxis* 66.1 (2015), pp. 57–64.

[70] Tatiana Gossen, Michael Kotzyba, and Andreas Nürnberger. "Knowledge Journey Exhibit: Towards Age-Adaptive Search User Interfaces." In: *Advances in Information Retrieval*. Vol. 9022. 37th European Conference on IR Research, ECIR '15. Vienna, Austria: Springer, 2015, pp. 781–784.

[71] Tatiana Gossen, Michael Kotzyba, and Andreas Nürnberger. "Search Engine for Children: User-Centered Design." In: *Datenbank-Spektrum* 17.1 (2017), pp. 61–67.

[72] Tatiana Gossen, Thomas Low, and Andreas Nürnberger. "What are the real differences of children's and adults' web search?" In: *Proc. of the 34th Int. ACM SIGIR conference on Research and development in Information*. ACM. Jan. 2011, pp. 1115–1116.

[73] Tatiana Gossen, Marcus Nitsche, and Andreas Nürnberger. "Knowledge Journey: A Web Search Interface for Young Users." In: *Proc. of the Sixth Symposium on Human-Computer Interaction and Information Retrieval*. HCIR '12. Cambridge, California, USA: ACM, 2012, pp. 1–10.

[74] Tatiana Gossen, Marcus Nitsche, and Andreas Nürnberger. "Evolving Search User Interfaces." In: *Proc. of the Workshop on euroHCIR at SIGIR Conf.* 2013, pp. 31–34.

[75] Tatiana Gossen, Marcus Nitsche, Jana Vos, and Andreas Nürnberger. "Adaptation of a Search User Interface towards User Needs: A Prototype Study with Children & Adults." In: *Proc. of the Symposium on Human-Computer Interaction and Information Retrieval*. HCIR '13. Vancouver BC, Canada: ACM, 2013, pp. 1–10.

[76] Tatiana Gossen, Michael Kotzyba, Sebastian Stober, and Andreas Nürnberger. "Voice-Controlled Search User Interfaces for Young Users." In: *7th Annual Symposium on Human-Computer Interaction and Information Retrieval*. HCIR '13. 2013, pp. 2–5.

[77] Laura A Granka, Thorsten Joachims, and Geri Gay. "Eye-tracking analysis of user behavior in WWW search." In: 2004, pp. 478–479.

[78] Fan Guo, Chao Liu, and Yi Min Wang. "Efficient Multiple-Click Models in Web Search." In: *Proc. of the second acm Int. conference on web search and data mining*. WSDM '09. Barcelona, Spain: ACM, 2009, pp. 124–131.

[79] Fan Guo, Chao Liu, Anitha Kannan, Tom Minka, Michael Taylor, Yi-Min Wang, and Christos Faloutsos. "Click Chain Model in Web Search." In: *Proc. of the 18th Int. conference on World wide web*. WWW '09. Madrid, Spain: ACM, 2009, pp. 11–20.

[80] Isabelle Guyon and André Elisseeff. "An Introduction to Variable and Feature Selection." In: *Journal of Machine Learning Research* 3.Mar (2003), pp. 1157–1182.

[81] Jacek Gwizdka and Irene Lopatovska. "The Role of Subjective Factors in the Information Search Process." In: *Journal of the American Society for Information Science and Technology* 60.12 (2009), pp. 2452–2464.

[82] Jacek Gwizdka and Ian Spence. "What Can Searching Behavior Tell Us About the Difficulty of Information Tasks? A Study of Web Navigation." In: *Proc. of the American Society for Information Science and Technology* 43.1 (2006), pp. 1–22.

[83] Shuguang Han, Zhen Yue, and Daqing He. "Automatic Detection of Search Tactic in Individual Information Seeking: A Hidden Markov Model Approach." In: *iConference 2013 Proc.* 2013, pp. 712–716.

[84] Ahmed Hassan, Rosie Jones, and Kristina Lisa Klinkner. "Beyond DCG: User Behavior as a Predictor of a Successful Search." In: *Proc. of the third ACM Int. Conference on Web Search and Data Mining*. ACM. 2010, pp. 221–230.

[85] Ahmed Hassan, Ryen W. White, Susan T. Dumais, and Yi-Min Wang. "Struggling or Exploring?: Disambiguating Long Search Sessions." In: *Proc. of the 7th ACM Int. conference on Web search and data mining*. ACM. 2014, pp. 53–62.

[86] Stefan Haun, Andreas Nürnberger, Tobias Kötter, Kilian Thiel, and Michael R. Berthold. "CET : A Tool for Creative Exploration of Graphs." In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by José Luis Balcázar, Francesco Bonchi, Aristides Gionis, and Michéle Sebag. Lecture Notes in Computer Science 6323. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 587–590.

[87] Yin He and Kuansan Wang. "Inferring Search Behaviors Using Partially Observable Markov Model with Duration (POMD)." In: *Proc. of the Fourth ACM Int. Conference on Web Search and Data Mining*. WSDM '11. Hong Kong, China: ACM, 2011, pp. 415–424.

[88] Marti A. Hearst. *Search user interfaces*. Cambridge university press, 2009.

[89] Heinz Heckhausen. *Leistung und Chancengleichheit*. Motivationsforschung - Band 2. Göttingen: Hogrefe, 1974.

[90] Chathra Hendahewa and Chirag Shah. "Segmental Analysis and Evaluation of User Focused Search Process." In: *Proc. of the 12th Int. Conference on Machine Learning and Applications*. Vol. 1. ICMLA '13. IEEE Computer Society, 2013, pp. 291–294.

[91] Saul Herner. "Browsing." In: *Encyclopedia of Library and Information Science*. Ed. by Allen Kent and Harold Lancour. Vol. 3. Marcel Dekker, New York, 1970, pp. 408–415.

[92] Silviu Homoceanu, Sascha Tönnies, Philipp Wille, and Wolf-Tilo Balke. "Time-Based Exploratory Search in Scientific Literature." In: *Research and Advanced Technologies for Digital Libraries: 17th Int. Conference on Theory and Practice of Digital Libraries (TPDL)*. Valletta, Malta, 2013.

[93] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, and Raj Reddy. *Spoken language processing: A guide to theory, algorithm, and system development*. 1st. USA: Prentice Hall PTR, 2001.

[94] Hilary Hutchinson, Allison Druin, Benjamin B Bederson, Kara Reuter, Anne Rose, and Ann Carlson Weeks. "How do I find blue books about dogs? The errors and frustrations of young digital library users." In: *Proc. of the 11th Int. Conf. on HCII 2005* (2005), pp. 22–27.

[95] Hugo C. Huurdeman and Jaap Kamps. "From Multistage Information-Seeking Models to Multistage Search Systems." In: *Proc. of the 5th Information Interaction in Context Symposium*. IIiX '14. New York, NY, USA: ACM, 2014, pp. 145–154.

[96] Peter Ingwersen. *Information Retrieval Interaction*. Taylor Graham London, 1992, p. 246.

[97] Peter Ingwersen. "Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory." In: *Journal of Documentation* 52.1 (1996), pp. 3–50.

[98] Peter Ingwersen and Peiling Wang. "Relationship between usefulness assessments and perceptions of work task complexity and search topic specificity: An exploratory study." In: *TBAS* (2012), pp. 19–23.

[99] Peter Ingwersen, Christina Lioma, Birger Larsen, and Peiling Wang. "An exploratory study into perceived task complexity, topic specificity and usefulness for integrated search." In: *Proceedings of the 4th Information Interaction in Context Symposium*. 2012, pp. 302–305.

[100] Beth St. Jean, Soo Young Rieh, Yong-Mi Kim, and Ji Yeon Yang. "An Analysis of the Information Behaviors, Goals, and Intentions of frequent Internet Users: Findings from Online Activity Diaries." In: *First Monday* 17.2 (2012).

[101] Thorsten Joachims. "Optimizing search engines using clickthrough data." In: *Proc. of the eighth ACM SIGKDD Int. conference on Knowledge discovery and data mining*. ACM, 2002, pp. 133–142.

[102] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. "Accurately Interpreting Clickthrough Data as Implicit Feedback." In: *Proc. of ACM SIGIR 2005*. ACM, 2005, pp. 154–161.

[103] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. "Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search." In: *ACM Trans. Inf. Syst.* 25.2 (2007).

[104] Hanna Jochmann-Mannak, Theo W.C. Huibers, Leo Lentz, and Ted Sanders. "Children Searching Information on the Internet: Performance on children's Interfaces compared to Google." In: *SIGIR'10 Workshop on accessible search systems* (2010), pp. 27–35.

[105] Hanna Jochmann-Mannak, Leo Lentz, Theo Huibers, and Ted Sanders. "How Interface Design and Search Strategy Influence Children's Search Performance and Evaluation." In: *Evaluating Websites and Web Services: Interdisciplinary Perspectives on User Satisfaction*. Hershey, PA, USA: IGI Global, 2014, pp. 241–287.

[106] Daniel Jurafsky and James Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Vol. 2. Prentice Hall series in artificial intelligence. Upper Saddle River, NJ: Pearson Int., 2008.

[107] Joseph B Kadane and Nicole A Lazar. "Methods and Criteria for Model Selection." In: *Journal of the American statistical Association* 99.465 (2004), pp. 279–290.

[108] Robert V Kail. *Children and their development*. Prentice-Hall, Inc, 2001.

[109] George A. Kelly. *A Theory of Personality: The Psychology of Personal Constructs*. Norton library. W. W. Norton, 1963. ISBN: 9780393001525.

[110] Jeonghyun Kim. "Task Difficulty as a Predictor and Indicator of Web Searching Interaction." In: *CHI '06 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '06. ACM, 2006, pp. 959–964.

[111] Shirlee-ann Knight and Amanda Spink. "Toward a Web Search Information Behavior Model." In: *Web Search*. Ed. by Amanda Spink and Michael Zimmer. Vol. 14. Information Science and Knowledge Management. Springer Berlin Heidelberg, 2008, pp. 209–234.

[112] Joseph A Konstan, Bradley N Miller, David Maltz, Jonathan L Herlocker, Lee R Gordon, and John Riedl. "GroupLens: Applying Collaborative Filtering to Usenet News." In: *Communications of the ACM* 40.3 (1997), pp. 77–87.

[113] Annett Körner, Michael Geyer, and Elmar Brähler. "Das NEO-Fünf-Faktoren Inventar (NEO-FFI) – Validierung anhand einer deutschen Bevölkerungsstichprobe." In: *Diagnostica* 48 (2002), pp. 19–27.

[114] Michael Kotzyba, Ingo Siegert, Tatiana Gossen, Andreas Wendemuth, and Andreas Nürnberger. "Exploratory Voice-Controlled Search for Young Users: Challenges & Potential Benefits." In: *Kognitive Systeme, DuEPublico, Duisburg-Essen Publication Online* 2015-1 (2015), pp. 1–10.

[115] Michael Kotzyba, Denis Ponomaryov, Thomas Low, Marcus Thiel, Birte Glimm, and Andreas Nürnberger. "Ontology - supported Exploratory Search for Physical Training Exercises." In: *Demonstrations Track at the 14th Int. Semantic Web Conference*. Vol. 1486. ISWC '15. 2015.

[116] Michael Kotzyba, Tatiana Gossen, Johannes Schwerdt, and Andreas Nürnberger. "Exploration or Fact-Finding: Inferring User's Search Activity Just in Time." In: *Proc. of the Conference on Conference Human Information Interaction and Retrieval*. SIGIR CHIIR '17. Oslo, Norway: Association for Computing Machinery, 2017, pp. 87–96.

[117] Michael Kotzyba, Tatiana Gossen, Sebastian Stober, and Andreas Nürnberger. "Model-Based Frameworks for User Adapted Information Exploration: An Overview." In: *Companion Technology*. Cognitive Technologies. Springer, 2017, pp. 37–56.

[118] Michael Kotzyba, Johannes Schwerdt, Tatiana Gossen, Martin Krippl, and Andreas Nürnberger. "The Effect of Motivational Goals on Information Search for Tasks of Varying Complexity Levels." In: *Int. Conference on Systems, Man, and Cybernetics*. SMC '18. Miyazaki, Japan: IEEE, 2018, pp. 2602–2607.

[119] C. C. Kuhlthau. "Meeting the information needs of children and young adults: Basing library media programs on developmental states." In: *Journal of Youth Services in Libraries* 2.1 (1988), pp. 51–57.

[120] Carol Collier Kuhlthau. "Inside the Search Process: Information Seeking from the User's Perspective." In: *Journal of the American Society for Information Science* 42.5 (1991), pp. 361–371.

[121] Carol Collier Kuhlthau. *Seeking Meaning: A Process Approach to Library and Information Services*. 1993.

[122] Carol Collier Kuhlthau. "Information Search Process." In: *Encyclopedia of Library and Information Sciences*. Ed. by Marcia Bates and Mary Niles Maack. Boca Raton: Taylor Francis, 2009.

[123] B. H. Kwasnik. "A descriptive study of the functional components of browsing." In: *Proc. of the IFIP TC2/WG2. 7 Working conference on Engineering for Human Computer Interaction*. 1992, p. 191.

[124] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data." In: *Proc. of the Eighteenth Int. Conference on Machine Learning*. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 282–289.

[125] Yuelin Li and Nicholas J Belkin. "A faceted approach to conceptualizing tasks in information seeking." In: *Information Processing & Management* 44.6 (2008), pp. 1822–1837.

[126] Jingjing Liu, Chang Liu, Jacek Gwizdka, and Nicholas J. Belkin. "Can Search Systems Detect Users' Task Difficulty?: Some Behavioral Signals." In: *Proc. of the 33rd Int. ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM. 2010, pp. 845–846.

[127] Lori Lorigo, Bing Pan, Helene Hembrooke, Thorsten Joachims, Laura Granka, and Geri Gay. "The influence of task and gender on search and evaluation behavior using Google." In: *Information Processing & Management* 42.4 (2006), pp. 1123–1131.

[128] Thomas Low, Nikola Bubalo, Tatiana Gossen, Michael Kotzyba, André Brechmann, Anke Huckauf, and Andreas Nürnberger. "Towards Identifying User Intentions in Exploratory Search using Gaze and Pupil Tracking." In: *Proc. of the Conference on Human Information Interaction and Retrieval*. SIGIR CHIIR '17. Oslo, Norway: Association for Computing Machinery, 2017, pp. 273–276.

[129] Gary Marchionini. "Information-Seeking Strategies of Novices using a Full-Text Electronic Encyclopedia." In: *Journal of the American Society for Information Science* 40.1 (1989), pp. 54–66.

[130] Gary Marchionini. *Information Seeking in Electronic Environments*. 9. Cambridge University Press, 1997.

[131] Gary Marchionini. "Exploratory Search: From Finding to Understanding." In: *Communications of the ACM* 49.4 (2006), pp. 41–46.

[132] Abraham Harold Maslow. "A theory of human motivation." In: *Psychological Review* 50.4 (1943), p. 370.

[133] Andrew McCallum, Dayne Freitag, and Fernando CN Pereira. "Maximum Entropy Markov Models for Information Extraction and Segmentation." In: *Icml*. Vol. 17. 2000. 2000, pp. 591–598.

[134] Günter Molz, Ralf Schulze, Ulrich Schroeders, and Oliver Wilhelm. "TBS-TK Rezension: Wechsler Intelligenztest für Erwachsene WIE. Deutschsprachige Bearbeitung und Adaptation des WAIS-III von David Wechsler." In: *Psychologische Rundschau* 61 (Jan. 2010), pp. 229–230.

[135] Clifford T. Morgan and Richard A. King. *Introduction to Psychology, 4th edition*. New York, USA: McGraw-Hill, 1971.

[136] Yashar Moshfeghi, Peter Triantafillou, and Frank E Pollick. "Understanding information need: An fMRI study." In: *Proc. of the 39th Int. ACM SIGIR conference on Research and Development in Information Retrieval*. 2016, pp. 335–344.

[137] Valerie Nesset. "Two representations of the research process: The preparing, searching, and using (PSU) and the beginning, acting and telling (BAT) models." In: *Library & Information Science Research* 35.2 (2013), pp. 97–106.

[138] Marcus Nitsche and Andreas Nürnberger. "Trailblazing Information: An Exploratory Search User Interface." In: *Human Interface and the Management of Information. Information and Interaction Design*. Ed. by Sakae Yamamoto. Vol. 8016. LNCS. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 230–239.

[139] Laurence Noël and Ghislaine Azémard. "From Semantic Web Data to Inform-Action: A Means to an End." In: *Workshop: Semantic Web User Interaction, CHI 2008, 5-10 April, Florence, Italy* (2008), pp. 1–7.

[140] Laurence Noël, Olivier Carloni, Nicolas Moreau, and Stéphanie Weiser. "Designing a Knowledge-Based Tourism Information System." In: *Int. Journal of Digital Culture and Electronic Tourism* 1.1 (2008), pp. 1–17.

[141] Heather L O'Brien, Andrea Kampen, Amelia W Cole, and Kathleen Brennan. "The role of domain knowledge in search as learning." In: *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. 2020, pp. 313–317.

[142] K.R. Olson and B. Chapin. "Relations of Fundamental Motives and Psychological Needs to Well-Being and Intrinsic Motivation." In: *Psychological Well-being*. New York, NY, USA: Nova Science Publishers, Inc, 2010, pp. 231–243.

[143] Jean Piaget and Bärbel Inhelder. *The psychology of the child*. Basic Books, 1969.

[144] Peter Pirolli. "Information Foraging." In: *Encyclopedia of Database Systems*. Ed. by Ling Liu and M. Tamer Özsu. New York, NY: Springer New York, 2016, pp. 1–7.

[145] Peter Pirolli and Stuart Card. "Information Foraging in Information Access Environments." In: vol. 1. Jan. 1995, pp. 51–58.

[146] Peter Pirolli and Stuart Card. "Information Foraging." In: *Psychological Review* 106 (Oct. 1999), pp. 643–675.

[147] Benjamin Piwowarski, Georges Dupret, and Rosie Jones. "Mining User Web Search Activity with Layered Bayesian Networks or How to Capture a Click in Its Context." In: *Proc. of the Second ACM Int. Conference on Web Search and Data Mining*. WSDM '09. Barcelona, Spain: ACM, 2009, pp. 162–171.

[148] Jonas Poelmans, Dmitry I Ignatov, Stijn Viaene, Guido Dedene, and Sergei O Kuznetsov. "Text Mining Scientific Papers: A Survey on FCA-Based Information Retrieval Research." In: *Advances in Data Mining. Applications and Theoretical Aspects*. Ed. by Petra Perner. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 273–287.

[149] Wanda Pratt and Lawrence Fagan. "The usefulness of dynamically categorizing search results." In: *Journal of the American Medical Informatics Association* 7.6 (2000), pp. 605–617.

[150] Yan Qu and George W Furnas. "Model-driven formative evaluation of exploratory search: A study under a sensemaking framework." In: *Information Processing & Management* 44.2 (2008), pp. 534–555.

[151] Boele De Raad. *The Big Five Personality Factors: The Psycholexical Approach to Personality.* Seattle, Toronto, Bern, Göttingen: Hogrefe & Huber Publishers, 2000.

[152] Lawrence R Rabiner. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." In: *Proc. of the IEEE* 77.2 (1989), pp. 257–286.

[153] Lawrence R Rabiner and B Juang. "An Introduction to Hidden Markov Models." In: *IEEE ASSP Magazine* 3.1 (1986), pp. 4–16.

[154] Filip Radlinski and Thorsten Joachims. "Query Chains: Learning to Rank from Implicit Feedback." In: *Proc. of the Eleventh ACM SIGKDD Int. Conference on Knowledge Discovery in Data Mining*. KDD '05. Chicago, Illinois, USA: ACM, 2005, pp. 239–248.

[155] Keith Rayner and Monica S Castelhano. "Eye movements during reading, scene perception, visual search, and while looking at print advertisements." In: (2008).

[156] Falko Rheinberg. *Leistungsbewertung und Lernmotivation*. Motivationsforschung - Band 8. Göttingen: Hogrefe, 1980.

[157] Falko Rheinberg. "Teachers Reference-Norm Orientation and Student Motivation for Learning." In: *AERA, conf., Seattle*. 2001, pp. 10–14.

[158] Falko Rheinberg and Stefan Fries. "Bezugsnormorientierung." In: *D. H. Rost (Ed.), Handwörterbuch Pädagogische Psychologie (4 ed.)* 2010, pp. 61–67.

[159] Ronald E. Rice, Maureen McCreadie, and Shan-Ju L. Chang. *Accessing and Browsing Information and Communication*. MIT Press, 2001.

[160] Paige Rodeghero and Collin McMillan. "An empirical study on the patterns of eye movement during summarization tasks." In: *2015 ACM/IEEE Int. Symposium on Empirical Software Engineering and Measurement (ESEM)*. IEEE. 2015, pp. 1–10.

[161] Nirmal Roy, Felipe Moraes, and Claudia Hauff. "Exploring Users' Learning Gains within Search Sessions." In: *Proceedings of the 2020 conference on human information interaction and retrieval*. 2020, pp. 432–436.

[162] Tuukka Ruotsalo et al. "Directing Exploratory Search with Interactive Intent Modeling." In: *Proc. of the 22nd ACM Int. Conference on Information & Knowledge Management*. CIKM '13. New York, NY, USA: ACM, 2013, pp. 1759–1764.

[163] Tuukka Ruotsalo, Jaakko Peltonen, Manuel J. A. Eugster, Dorota Głowacka, Patrik Floréen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. "Interactive Intent Modeling for Exploratory Search." In: *ACM Trans. Inf. Syst.* 36.4 (2018).

[164] Daniel M Russell, Mark J Stefik, Peter Pirolli, and Stuart K Card. "The Cost Structure of Sensemaking." In: *Proc. of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*. CHI '93. New York, NY, USA: ACM, 1993, pp. 269–276.

[165] Tefko Saracevic. "Modeling Interaction in Information Retrieval (IR): A Review and Proposal." In: *Proc. of the ASIS Annual Meeting*. Vol. 33. ERIC. 1996, pp. 3–9.

[166] Tefko Saracevic. "The stratified model of information retrieval interaction: Extension and application." In: vol. 34. 1997, pp. 313–327.

[167] Johannes Schwerdt, Michael Kotzyba, and Andreas Nürnberger. "Inferring User's Search Activity using Interaction Logs and Gaze Data." In: *Int. Conference on Companion Technology*. ICCT '17. Ulm, Germany: IEEE, 2017, pp. 1–6.

[168] Johannes Schwerdt, Michael Kotzyba, and Andreas Nürnberger. "Fact-Finding or Exploration: Identifying Latent Behavior Clusters in User's Search Activities." In: *Int. Conference on Systems, Man and Cybernetics*. SMC '19. Bari, Italy: IEEE, 2019, pp. 1465–1471.

[169] Chirag Shah. "Toward collaborative information seeking (CIS)." In: *arXiv preprint arXiv:0908.0709* (2009).

[170] Chirag Shah. "Coagmento - A Collaborative Information Seeking, Synthesis and Sense-Making Framework." In: *Integrated demo at CSCW* (2010).

[171] Chirag Shah. *Collaborative information seeking: The art and science of making the whole greater than the sum of all*. Vol. 34. Springer Science & Business Media, 2012.

[172] Chirag Shah and Roberto González-Ibáñez. "Exploring Information Seeking Processes in Collaborative Search Tasks." In: *Proc. of the American Society for Information Science and Technology* 47.1 (2010), pp. 1–7.

[173] Chirag Shah, Chathra Hendahewa, and Roberto González-Ibáñez. "Rain or Shine? Forecasting Search Process Performance in Exploratory Search Tasks." In: *Journal of the Association for Information Science and Technology* 67.7 (2015), pp. 1607–1623.

[174] Chirag Shah and Chris Leeder. "Exploring collaborative work among graduate students through the C5 model of collaboration: A diary study." In: *Journal of Information Science* 42.5 (2016), pp. 609–629.

[175] Paul Solomon. "Children's information retrieval behavior: A case analysis of an OPAC." In: *Journal of the American Society for information Science* 44.5 (1993), pp. 245–264.

[176] Amanda Spink. "Study of Interactive Feedback during Mediated Information Retrieval." In: *Journal of the American Society for Information Science* 48.5 (1997), pp. 382–394.

[177] Amanda Spink, Minsoo Park, Bernard J Jansen, and Jan Pedersen. "Multitasking during Web Search Ssessions." In: *Information Processing & Management* 42.1 (2006), pp. 264–275.

[178] Dominic Stange, Michael Kotzyba, and Andreas Nürnberger. "Professional Collaborative Information Seeking: Towards Traceable Search and Creative Sensemaking." In: *Transactions on Computational Collective Intelligence XXVI, TCCI*. Vol. 10190. Springer, 2017, pp. 1–25.

[179] Dominic Stange and Andreas Nürnberger. "Search Maps - Enhancing Traceability and Overview in Collaborative Information Seeking." In: *Advances in Information Retrieval - Proc. of 36th Europ. Conf. on IR Research, ECIR 2014*. Springer. Amsterdam, The Netherlands, 2014, pp. 763–766.

[180] Dominic Stange, Michael Kotzyba, Stefan Langer, and Andreas Nürnberger. "Interpreting and Leveraging Browser Interaction for Exploratory Search Tasks." In: *Proc. of the 19th Int. Conference on Enterprise Information Systems: ICEIS*. Vol. 3. Porto, Portugal: ScitePress, 2017, pp. 191–197.

[181] Charles Sutton and Andrew McCallum. "An introduction to conditional random fields." In: *Foundations and Trends® in Machine Learning* 4.4 (2012), pp. 267–373.

[182] Vu Tuan Tran and Norbert Fuhr. "Using Eye-Tracking with Dynamic Areas of Interest for Analyzing Interactive Information Retrieval." In: *In Proc. of the 35th Int. ACM SIGIR conference on Research and development in information retrieval*. ACM, 2012, pp. 1165–1166.

[183] Michael Twidale and David Nichols. "Collaborative browsing and visualisation of the search process." In: *Aslib Proc.* Vol. 48. 7-8. 1996, pp. 177–182.

[184] Kuansan Wang, Nikolas Gloy, and Xiaolong Li. "Inferring Search Behaviors Using Partially Observable Markov (POM) Model." In: *Proc. of the Third ACM Int. Conference on Web Search and Data Mining*. WSDM '10. New York, New York, USA: ACM, 2010, pp. 211–220.

[185] Edward S Warner et al. "Information Needs of Urban Residents. Final Report." In: (1973).

[186] Karl E. Weick and Richard L. Daft. "The Effectiveness of Interpretation Systems." In: *Organizational Effectiveness*. Ed. by Kim S. Cameron and David A. Whetten. Academic Press, 1983, pp. 71–93.

[187] Ryen W. White and Steven M. Drucker. "Investigating Behavioral Variability in Web Search." In: *Proc. of the 16th Int. World Wide Web Conference(WWW)*. 2007.

[188] Ryen W. White, Bill Kules, and Ben Bederson. "Exploratory Search Interfaces: Categorization, Clustering and beyond." In: *SIGIR Forum* 39.2 (2005), pp. 52–56. ISSN: 0163-5840.

[189] Ryen W White and Resa A Roth. "Exploratory search: Beyond the query-response paradigm." In: *Synthesis lectures on information concepts, retrieval, and services* 1.1 (2009), pp. 1–98.

[190] Ryen W. White, Bill Kules, Steven M. Drucker, and M. C. Schraefel. "Supporting Exploratory Search, Introduction, Special Issue, Communications of the ACM." In: *Communications of the ACM* 49.4 (2006), pp. 36–39.

[191] Barbara M. Wildemuth and Luanne Freund. "Assigning Search Tasks Designed to Elicit Exploratory Search Behaviors." In: *Proc. of the Symposium on Human-Computer Interaction and Information Retrieval*. HCIR '12 4. ACM, 2012.

[192] Thomas D. Wilson. "Human Information Behavior." In: *Informing Science* 3.2 (2000), pp. 49–56.

[193] Tom D. Wilson. "On User Studies and Information Needs." In: *Journal of Documentation* 37.1 (1981), pp. 3–15.

[194] Tom D Wilson. "Information needs and uses: fifty years of progress." In: *Fifty years of information progress: a Journal of Documentation review* (1994), pp. 15–51.

[195] Tom D. Wilson. "Information Behaviour: An Interdisciplinary Perspective." In: *Information processing & management* 33.4 (1997), pp. 551–572.

[196] Tom D. Wilson. "Models in Information Behaviour Research." In: *Journal of Documentation* 55.3 (1999), pp. 249–270.

[197] Michael F Wolff. "Managers at Work: Scouting for Technology." In: *Research-Technology Management* 35.2 (1992), pp. 10–12.

[198] Wan-Ching Wu, Diane Kelly, and Kun Huang. "User Evaluation of Query Quality." In: *Proc. of the 35th Int. ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '12. Portland, Oregon, USA: ACM, 2012, pp. 215–224.

[199] Sheng-Yuan Yang, Fang-Chen Chuang, and Cheng-Seen Ho. "Ontology-supported FAQ processing and ranking techniques." In: *Journal of Intelligent Information Systems* 28.3 (2007), pp. 233–251.

[200] Sheng-Yuan Yang and Cheng-Seen Ho. "Ontology-supported user models for interface agents." In: *IEEE expert* (1996).

[201] Zhen Yue, Shuguang Han, and Daqing He. "Modeling Search Processes using Hidden States in Collaborative Exploratory Web Search." In: *Proc. of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM. 2014, pp. 820–830.

[202] Marvin Zuckermann. *Sensation Seeking: Beyond the Optimal Level of Arousal*. Hillsdale: L. Erlbaum Associates, 1979.

[203] Marvin Zuckermann, Sybil Bianca Giuliett Eysenck, and Hans Jürgen. Eysenck. *Sensation seeking in England and America: Cross-cultural, age and sex comparisons.* Vol. 46. 1. Hillsdale: Journal of Consulting and Clinical Psychology, 1978, pp. 139–149.