RESEARCH

Tatiana Gossen

Search Engines for Children

Search User Interfaces and Information-Seeking Behaviour

Springer Vieweg

Search Engines for Children

Tatiana Gossen

Search Engines for Children

Search User Interfaces and Information-Seeking Behaviour

Springer Vieweg

Tatiana Gossen
Magdeburg, Germany

Dissertation at Otto-von-Guericke University Magdeburg, Germany, 2015

ISBN 978-3-658-12068-9 ISBN 978-3-658-12069-6 (eBook)
DOI 10.1007/978-3-658-12069-6

Library of Congress Control Number: 2015954962

Springer Vieweg

Printed on acid-free paper

Springer Vieweg is a brand of Springer Fachmedien Wiesbaden
Springer Fachmedien Wiesbaden is part of Springer Science+Business Media
(www.springer.com)

"The best way to make children good is to make them happy."

— *Oscar Wilde.*

Acknowledgments

I would like to thank all the people that have supported me during these past years and in particular during the development of this thesis. First at all, I want to thank my supervisor, Andreas Nürnberger, who gave me the opportunity of writing this thesis and supported all my research endeavours. I am especially grateful for the freedom that I had being able to choose the direction of my research for my PhD and to carry out my research activities. I also want to thank Birger Larsen and Ian Ruthven for kindly agreeing to review my thesis. Thank you for taking time to read it and to give feedback.

I furthermore thank my colleagues and the whole staff of the faculty of computer science at the *Otto-von-Guericke-University* for the pleasant working atmosphere. I would further like to give special thanks to all the colleagues and students that – in one way or another – were involved in some parts of this work. In particular, I value Thomas Low and Marcus Nitsche. It was a great pleasure to have Thomas Low as my research assistant and later colleague during the logfile analysis project with the German Youth Institute. Marcus Nitsche shared his expertise in user interface design. He inspired me to conduct research in the area of human computer interaction with regards to the retrieval of information for children. A special thank you goes to Sebastian Stober whose comments, corrections and criticism have made this thesis more understandable. I would also like to thank the following students, which I had the pleasure of supervising and working with, for their contribution to this thesis: Juliane Höbel, Julia Hempel, Rene Müller, Jana Vos and Ina Bosse.

I would also like to thank the many participants of the user studies for their valuable time and effort. I especially want to thank the staff of the primary school in Biederitz and the trilingual international primary school

in Magdeburg for their kind support in the organisation of the conducted user studies with the children.

Last but not least, I want to thank my family. I thank my husband Gerhard Gossen for all his love, for his emotional and professional support. I owe my thanks to Gerhard for his patience, valuable feedback, proofreading of the manuscript and showing me useful LaTeX hacks.

Abstract

Children are a fast growing user group on the Internet. Among different online activities, children use web search engines in order to gather information related to their personal interests and school activities. Children's knowledge, cognitive abilities and fine motor skills are different from those of adults. Therefore, they may experience difficulties with search engines that are built using standard information retrieval algorithms and search interfaces for adults. Special or targeted search engines for children are essential in order to better support children in their search tasks. Therefore, the goal of this thesis is to design appropriate search engines for children with a focus on the search user interface. However, this is not an easy task to accomplish. Not only are children's abilities different from the abilities of adults, children also undergo relatively fast changes in their abilities.

The specific and dynamically changing characteristics of young users pose a great challenge. In order to address this challenge, first, the specifics of information retrieval for young users are analysed. Second, open issues are identified in user studies with children using logfile analysis and eye-tracking. The conceptual challenges in the design of user interfaces regarding search engines for children are derived based on the findings of one's own and previous user studies as well as theories of human development. Third, user interfaces of search engines that address these conceptual challenges are designed, prototypically implemented and evaluated in user studies with children following a user-centered design. Specifically, the proposed user interfaces of the search engine address the changing characteristics of the users by providing a means of adaptation. Furthermore, a novel type of search result visualisation for children with cartoon style characters is developed which takes the children's preference for visual information into account. Both approaches were very positively received

by children during evaluation. Children rated different user interface aspects of the search user interface prototypes as good, e.g. the adaptation of the search user interface towards user wishes and the helpfulness of the cartoon style characters during search. Finally, this thesis provides criteria and guidelines on how to design the user interfaces of the search engine for children.

Zusammenfassung

Kinder sind eine schnell wachsende Nutzergruppe im Internet. Sie führen unterschiedliche Aktivitäten durch, unter anderem verwenden die Kinder Suchmaschinen, um Informationen über ihre Lieblingsthemen oder Schulaufgaben zu sammeln. Der Wissensstand der Kinder, ihre motorischen und kognitiven Fähigkeiten sind anders als bei Erwachsenen. Demzufolge können Kinder Schwierigkeiten bei der Verwendung von Suchmaschinen haben, die vorrangig für Erwachsene mit entsprechenden Retrieval-Algorithmen und Suchschnittstellen entwickelt wurden. Spezielle oder nutzerorientierte Suchmaschinen für Kinder sind wichtig, um eine bessere Unterstützung der Kinder bei ihren Suchaufgaben zu gewährleisten. Deswegen ist das Ziel dieser Arbeit Suchmaschinen zu konzipieren, die auf die speziellen Bedürfnisse von Kindern eingehen. Der Schwerpunkt liegt dabei auf der Entwicklung einer ergonomischen Suchschnittstelle für Kinder. Diese Aufgabe ist herausfordernd, da sich die Fähigkeiten von Kindern nicht nur von denen der erwachsenen Nutzer unterscheiden, sondern auch relativ schnell verändern.

Die spezifischen und sich dynamisch ändernden Eigenschaften der jungen Nutzer stellen eine große Herausforderung dar. Diese Herausforderung wird in dieser Arbeit zuerst adressiert, indem die Besonderheiten der Informationssuche der Kinder analysiert werden. Im zweiten Schritt werden fehlende Informationen über das Suchverhalten in Benutzerstudien mit Logfiles und Eye-Tracking gesammelt. Aus den Erkenntnissen der eigenen und früheren Benutzerstudien sowie Theorien der menschlichen Entwicklung werden in dieser Arbeit die konzeptionellen Anforderungen an die Gestaltung von Benutzerschnittstellen von Suchmaschinen für Kinder abgeleitet. Im dritten Schritt werden die Benutzeroberflächen von Suchmaschinen, die diese konzeptionellen Herausforderungen angehen,

in einem nutzerorientierten Gestaltungsprozess konzipiert, prototypisch implementiert und in Benutzerstudien mit Kindern evaluiert. Es wird eine adaptive Suchmaschinen-Benutzerschnittstelle vorgeschlagen, die sich flexibel an die Veränderungen der Nutzerfähigkeiten anpassen lässt. Darüber hinaus wird ein neuer Ansatz zur Visualisierung von Suchergebnissen mit Charakteren im Cartoon-Stil vorgeschlagen. Dieser Ansatz berücksichtigt die Vorliebe von Kindern für visuelle Informationen. Beide Ansätze wurden in Benutzerstudien sehr positiv von den Kindern aufgenommen. Unter anderem wurde die Möglichkeit der Anpassung der Suchoberfläche an die Benutzerwünsche und die Nützlichkeit der Charaktere bei der Suche untersucht. Die Kinder vergaben eine hohe Bewertung für die Benutzerfreundlichkeit der Prototypen. Abschließend stellt diese Arbeit Kriterien und Richtlinien zusammen, wie die Benutzerschnittstellen zur Suche an Kinder angepasst werden können.

Contents

List of Figures

List of Tables

List of Acronyms

API	Application Programming Interface
BK	Blinde-Kuh.de
ESUI	Evolving Search User Interface
GL	Google-like
GUI	Graphical User Interface
HCIR	Human-Computer Interaction in Information Retrieval
ID	Document Identifier
IR	Information Retrieval
KJ	Knowledge Journey
KJE	Knowledge Journey Exhibit
OPAC	Online Public Access Catalog
SERP	Search Engine Result Page
SUI	Search User Interface
UI	User Interface
URL	Uniform Resource Locator

Chapter 1

Introduction

Nowadays, Internet usage knows no age limits. Since an increasing number of households all over the world own a computer and have Internet access, many children have access to the Internet and explore the web from a young age. The German study KIM[1] continuously examines the media usage of children aged between six and thirteen years old. According to this study [15], on average 62 % of German children use the Internet. This number for children increases with age from 21% by six years old to 93% by thirteen years old. 75% of the children use a search engine (e.g. Google) at least once a week.

Children use the Internet not only for entertainment, but it also plays an increasing role in education. On average half of the children use a computer to find information for school at least once per week [15]. They look for facts about historical events, mathematical formulas, the latest news and much more. To do so, children use computers at school or at home. Teachers or parents are not always around to support them: 60% of children search the Internet predominantly alone [15]. In order to better support children at their search tasks, several websites that provide special search services for children have been launched. Meanwhile, their main purpose is helping children to find child-suitable, in particular child-safe, content on the Internet.

However, this is only one important aspect of such search engines for children. Another important aspect is the usability of these search engines. Children should be able to successfully use a search engine without the help of adults. But unfortunately, not all children succeed in information inquiry, and especially the younger children can experience strong difficulties [15]. A possible explanation of failure is the children's difficulty in being able to use the search engines due to their insufficient usability.

The usability of web search engines is of special importance for children since their cognitive abilities and motor skills are not fully formed. Young

[1] KIM is a German acronym for Children and Media ("Kinder + Medien, Computer + Internet").

T. Gossen, *Search Engines for Children,*
DOI 10.1007/978-3-658-12069-6_1, © Springer Fachmedien Wiesbaden 2015

children have difficulties with abstract concepts, can process less information and their performance in pointing movements, e.g. using a mouse, are lower than that of adults [102]. It is not only desirable that children are supported according to their skills during search sessions, so that they are able to find good results, but the success in searching also plays a major role in the development of children. Erikson [62] found that primary school-aged children want to learn and to show what they can produce. They want to achieve the skills that seem to be important to their cultural environment and win the recognition of parents, teachers and peers by doing so [62]. Finding information on the Internet is an important skill that children need to develop and it is important to provide them with the necessary tools to succeed. If children succeed in finding information, they will feel competent and develop their self confidence. In contrast, if they are not able to find good results, children may develop a feeling of incompetence that could even lead to a feeling of inferiority [62]. In order to avoid those consequences, a search engine for children has to support children in finding good results, that is, it has to be adapted to the special needs — the motor skills, cognitive abilities and knowledge — of children in the respective age groups.

Currently, there are many good techniques in Information Retrieval (IR) for adults, but not much insight on how to design search engines for children in terms of both user interfaces and underlying algorithms. Most of the current IR systems are designed for adults. However, previous user studies indicate that there are significant differences between a child's and an adult's search behaviour. For instance, children can get easily frustrated if they do not find relevant results, do not understand the search engine output or if a failure emerges [17]. The fact that children also have difficulties when trying to evaluate the relevance of retrieved documents to their information needs aggravates this [92]. Furthermore, most children have difficulties with typing [25]. They are not able to type commands without looking at the keyboard (touch-typing). Instead they typically hunt-and-peck on the keyboard for correct keys. By looking at the keyboard while typing, children often do not spot spelling mistakes. In addition, some interaction techniques like scrolling or drag-and-drop are difficult for young users [25]. Therefore, young users would benefit from search user interfaces and algorithms that would take the special requirements of children into account.

1.1 Research Questions

Based on the motivation, this thesis addresses the following research questions (*RQ*):

1. What aspects of child development are important for information retrieval tasks?

2. What components of an IR system can be adapted to the targeted user group?

3. To what extent are the existing search engines for children appropriate for their motor and cognitive skills?

4. What are the characteristics of children's information seeking behaviour? Specifically, what are the differences of a child's and an adult's web information seeking behaviour:

   a) with regard to queries and search interactions?
   b) with regard to search performance?
   c) with regard to perception of search engine result pages?
   d) with regard to search strategies?

5. How can the user interface better support search to fulfill a child's information needs:

   a) considering children in a concrete operational development stage (age 7-11)?
   b) designing an IR system that grows with the children?

6. What are alternative ways to visualise search results for children?

   a) What features of web documents do children consider to be important?
   b) How do children visualise a web document as a search result?

These questions aim at achieving the following overall research goal: the development of interactive systems for information search for children as a targeted user group with special focus on the user interface. The goal is furthermore to identify open issues, analyse conceptual challenges when designing web search user interfaces for children, propose feasible solutions and demonstrate their applicability.

This thesis addresses young users between seven and eleven as a target user group. This choice is motivated by the theories of human development that are described later in Section 2.2 (see also Section 2.4). Within this work, we use the terms "children" and "young users" interchangeably. Furthermore, this thesis mainly concentrates on information search in a web document collection. During the search children intend to find information relevant to their information needs. In order to search successfully, children require special search engines that should be designed considering the specific requirements and needs of a child.

1.2 Thesis Outline

This thesis is structured into four main parts and an appendix. Part I provides the reader with fundamental knowledge that is important for the understanding of this thesis: Chapter 2 introduces the fundamental concepts and approaches in Information Retrieval (IR), describes the specifics of young users from the human development perspective (that are important to consider in IR for children), and briefly discusses the basic user research methods that can be applied for information-seeking investigation and evaluation of IR systems. Chapter 3 summarises the main previous finding about the information-seeking behaviour of children and provides an overview of existing algorithms and search user interfaces that are developed for children. This chapter provides information about what components of an IR system should be adapted to the targeted user group children and what methods currently exist.

Part II describes the research that was conducted in this thesis in order to identify open issues concerning targeted search engines for children and children's usage of those search engines. In particular, in Chapter 4 a case study about the usability of existing search engines for young users is conducted. Chapter 5 presents the large-scale study of logfiles of search engines for children. Children's search queries and interactions are analysed in order to identify the differences of a child's and an adult's web information search. In order to compare children's and adults' perception and performance on targeted web search engines, an eyetracking user study was conducted. This study is described in Chapter 6.

Part III focuses on the design of search user interfaces for children in primary school age. In Chapter 7 both, the findings of this thesis and from pre-

vious research, were considered in the analysis of conceptual challenges in the design of user interfaces in search engines for children. Chapter 7 also proposes several solutions for the design of children's search user interface (SUI). In order to demonstrate these solutions, a SUI for children called *Knowledge Journey* was developed. *Knowledge Journey* was evaluated in a user study with children against a classical search user interface. This user study motivated the need for an evolving search user interface that adapts to a particular young user. A second pilot study was conducted in order to investigate the potentials of a voice-controlled version of the *Knowledge Journey*. The idea of an evolving search user interface that addresses changing user requirements is elaborated in Chapter 8. An evolving search user interface which enables a flexible adaptation of the SUI to address changing user characteristics was developed and a user study was conducted in order to find a mapping between users of different age groups and SUI elements. Based on the Evolving Knowledge Journey, *Knowledge Journey Exhibit* (KJE) was developed as a robust information terminal. Chapter 9 investigates alternative ways to visualise search results for children and support them in the processing of search results. A novel approach is described that suggests to visualise each search result as a character where a character visually provides clues about the content of the result web page. Following a user-centered design, children were involved in the design of characters and the evaluation of different search result layouts with characters.

Part IV summarises the achieved results in respect to the specified research questions. Furthermore, an overview of open research issues is given. Directions concerning search histories, ranking algorithms, collaborative IR systems and evaluation methods for children are discussed.

Part I
Fundamentals

Chapter 2

Information Retrieval for Young Users

This chapter introduces the fundamental concepts and approaches that are prerequisites for the work described in this thesis. In Section 2.1 an introduction to Information Retrieval (IR) is given by defining the research field and describing IR system architecture. Furthermore, this section provides explanation of relevance ranking and introduction to search user interfaces. An introduction to targeted search engines is given in Section 2.1.4. The main aspects of children's development relevant to IR tasks are described in Section 2.2. In Section 2.3, basic user research methods and user evaluation types are briefly discussed.

2.1 Basics of Information Retrieval

The research field of Information Retrieval was defined by Salton [167, p. v] as:

Definition

"*Information retrieval* is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information."

Information retrieval is an activity that a user is engaged in. The user has a perceived gap in his or her knowledge, also called *information need*. This information is assumed to be present in a collection. In order to find this information, the user interacts with an IR system. The most common scenario is a search in a collection of text documents. In this thesis, we focus on textual information retrieval that is important in order to make the search in a web document collection work. Multimedia retrieval that deals with visual and sound data, e.g. images, music, videos [10, Chapter 14], is out of the scope of this thesis.

T. Gossen, *Search Engines for Children,*
DOI 10.1007/978-3-658-12069-6_2, © Springer Fachmedien Wiesbaden 2015

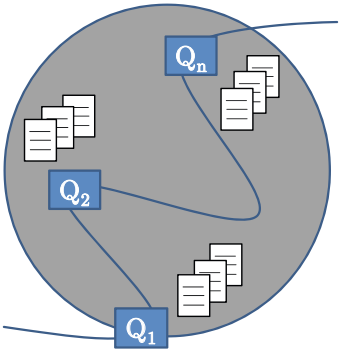

Fig. 2.1: Berrypicking search consists of a series of user interaction with an IR system. User learn bits of information one step at a time. Starting with a query Q_1, the user learns new information from the retrieved documents. This may lead to new queries until the user's information need is satisfied (this figure is modified from [13]).

The human interaction process with an IR system may be complex [13, 128]. Starting with one information need that may be vague in nature, a user submits a query and views the search results. They learn new information from these results, such as new facts or new vocabulary, that might lead to a drift in their information need or the understanding that rephrasing the query will lead to better search results. Therefore, they continue to query the search engine until the information need is satisfied. In other words, not every information need can be resolved with a single query and a single set of search results, but a series of queries and user examining the results might be necessary where a user is learning bits of information one step at a time. This bit-at-a-time retrieval is called berrypicking [13] by the analogy of picking berries from a bush (see Fig. 2.1).

Manning et al. [125, p. 1] defines Information Retrieval pointing out the unstructured nature of documents:

Definition

"*Information retrieval (IR)* is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)."

This means that in IR we deal with data which does not have a clear semantic structure. This leads to differences between data and information retrieval. The main differences are summarized by van Rijsbergen [164] as in Table 2.1. In data retrieval our goal is to find an exact match that is conformant with our query. An example of data retrieval is searching in a relational database [32]. A database has an underlying schema. Therefore, it is possible to find the exact database entries that match the query. In information retrieval we are interested in best matches, even if these matches are only partial. In some cases the answer to a user's information need is spread across several documents and these two partial matches reveal the desired information (see Berrypicking model above – Fig. 2.1). The query language for data retrieval is usually artificial with a restricted syntax and vocabulary. This language allows a user to give an exact and complete specification of what is wanted. In IR an exact specification is not possible because of the unstructured nature of text documents. To specify their information need, users in IR mainly use the vocabulary of natural languages. The retrieved results in IR indicate the likelihood of their relevance to user's information need and are usually sorted according to the relevance. Relevance is a measure of how closely a given document matches a user's information need. This judgment is done by the user and depends on different factors, e.g. his domain knowledge, the context of the search or previously seen results. As we deal with probabilities, in IR small errors in matching generally are not critical, whereas errors in data retrieval imply a total failure of the system.

2.1.1 Architecture of an IR System

Fig. 2.2 shows the main components of an IR system according to Baeza-Yates and Ribeiro-Neto [10]. An IR system has a frontend, also known as a