Towards Accessible Search Systems

Workshop of the 33rd Annual International
ACM SIGIR Conference
*on Research and Development
in Information Retrieval*

Organised by
Pavel Serdyukov
Djoerd Hiemstra
Ian Ruthven

**Association for
Computing Machinery**

SIGIR Geneva 2010

Preface

Current search systems are not adequate for individuals with specific needs: children, older adults, people with visual or motor impairments, and people with intellectual disabilities or low literacy. Search services are typically created for average users (young or middle-aged adults without physical or mental disabilities) and information retrieval methods are based on their perception of relevance as well. The workshop is the first ever to raise the discussion on how to make search engines accessible for different types of users, including those with problems in reading, writing or comprehension of complex content. Search accessibility means that people whose abilities are considerably different from those that average users have will be able to use search systems with the same success.

The objective of the workshop is to provide a forum and initiate collaborations between academics and industrial practitioners interested in making search more usable for users in general and for users with specific needs in particular.

The papers selected for this workshop are a mixture of research, discussion and position papers. We have deliberately selected a broad range of papers for this workshop to reflect the diverse research areas that contribute to the discipline of Accessible Search.

We would like to thank our panellists for providing a stimulating start to our workshop and the programme committee for generously providing comments and guidance to the submitting authors. We would particularly like to thank our keynote speakers Allison Druin and T. V. Raman.

Programme committee
Leif Azzopardi, University of Glasgow
Jamshid Beheshti, McGill University, Montreal
Dania Bilal, University of Tennessee
Carlos Castillo, Yahoo Research
Kevin Collins-Thompson, Microsoft Research
Arjen de Vries, CWI, Amsterdam
Diane Kelly, University of North Carolina
Mounia Lalmas, University of Glasgow
Barbara Leporini, CNR, Pisa
Andrew MacFarlane, City University, London
Marie-Francine Moens, University of Leuven
Meng Wang, Microsoft Research Asia

Website http://www.cis.strath.ac.uk/~ir/accessiblesearch/

Programme

9.00-9.15 Registration and welcome

9.15-10.15 Keynote (chair: Djoerd Hiemstra)
Searching for the Future: Understanding Children's Challenges, Actions, and Roles in Searching
Allison Druin

10.15-10.45 Coffee break

10.45-12.05 Session 1 Childrens' Information Retrieval (chair: Ian Ruthven)
A Closer Look at Children's Information Retrieval Usage: Towards Child-Centered Relevance
Frans Van der Sluis and Betsy Van Dijk

Assessing Fun: Young Children as Evaluators of Interactive Systems
Yusrita Mohd Yusoff, Ian Ruthven and Monica Landoni

Text Simplification for Children
Jan De Belder and Marie-Francine Moens

Children searching information on the Internet: Performance on children's interfaces compared to Google
Hanna Jochmann-Mannak and Leo Lentz

12.05 – 14.00 - Lunch

14.00-15.00 **Keynote (chair: Pavel Serdyukov)**
Toward More Accessible Search
T.V. Raman

15.00-15.40 Session 2 Information Retrieval for Users with Special Requirements (chair: Ian Ruthven)
WebMark: A Rapid Internet Browsing Technique for Visually Impaired Web Users
Hesham M. Kamel

The Mediated Information World of Children on the Autism Spectrum Disorder (ASD)
Dania Bilal

15.40-16.10 Coffee break

16.10-17.00 Panel (chair: Dania Bilal)

Table of contents

Searching for the Future: Understanding Children's Challenges, Actions, and Roles in Searching

Allison Druin
University of Maryland, College Park, USA
allisond@umiacs.umd.edu

ABSTRACT

"I don't know where it is!"

"I never find the stuff I'm looking for..."

"Maybe I can find the Vice President's birthday in the SpongeBob Square-Pants website?"

These are all responses we have received from 7, 9, or 11 year old children that have been searching online at home. In this talk, I will present seven search roles children display as information seekers using Internet keyword interfaces, based on a home study of 83 children. These roles are defined not only by the children's search actions, but also by who influences their searching, their perceived success, and trends in age and gender. These roles suggest a need for new interfaces that expand the notion of keywords, scaffold results, and develop a search culture among children. Future interfaces for mobile phones, netbooks, and more will be discussed.

BIOGRAPHY

Allison Druin is associate professor in the College of Information Studies at the University of Maryland, and director of the Human-Computer Interaction Lab. Druin's research is dedicated to designing technology for children. She believes that children should have a voice in making new technology for kids. Children's ideas need to be heard through-out the entire technology design process. In 1998, the lab began a unique technology design team. Seven children, aged seven to eleven, joined with researchers from computer science, education, art, robotics, and other disciplines, twice a week, to form an intergenerational, interdisciplinary design team. The team pursues projects, writes papers and creates new technologies. Druin's team created for instance the International Children's Digital Library, a multilingual free digital library of children's books, consisting of more than 4,000 books in over 50 languages, with more than three million users from over 160 countries worldwide. Druin also founded CHIKids at ACM SIGCHI, a program where children were CHI conference reporters, software testers, multi-media storytellers and more.

Druin received her Ph.D from the University of New Mexico's College of Education in 1997 and a master degree from MIT Media Lab in 1987. She is the author/editor of three books on the design of children's technology. She received several awards for work. ACM SIGCHI awarded Druin the SIGCHI Social Impact Award together with Ben Bederson in 2010, and the SIGCHI Distinguished Service Award in 1998. Druin received a prestigious National Science Foundation Career Award, a five year research grant for promising junior faculty research, which she used to focus on the development of the classroom of the future.

Toward More Accessible Search

T.V. Raman

Google Research, Mountain View, USA

raman@google.com

ABSTRACT

The core value of search lies in providing effective access to the right piece of information in a timely manner. Traditionally, information retrieval systems have focused on retrieving the correct piece of information. In an effective search system, the following user-centric items play a key role in determining what *right information* and *effective access* mean:

1. User context,

2. Device used to access search,

3. User interface being used.

Information retrieval systems such as Google Search perform extremely well when addressing results in the *long tail*. Addressing *accessibility* is about recognizing that user needs and abilities vary over time - said differently, accessibility is about serving users in the long tail. As we move toward an increasingly mobile world with users accessing the Web from a variety of devices and usage contexts, overall effectiveness of search systems is determined by the user's ability to complete a given task in a timely manner. In reaching this goal, information retrieval needs to match the result set with the user along a variety of axes. For example, consider the query: `Lufthansa 455`, a query for a specific flight:

1. On a desktop, one might serve up a detailed Web page showing flight tracking, status and, available future travel.

2. On a mobile device, one might show a light-weight version of the above.

3. When using a voice-only interface, one might only speak the current flight status.

In the age of information overload, the band-width between man and machine gets increasingly overloaded. Thus, the importance of search goes up directly as:

1. User's attention span goes down.

2. User's display size goes down.

3. User's network band-width goes down.

Accessible Search is about building information retrieval systems that take all of the above into account. I'd like us as a field to formally define the various axes along which we determine the *right result* to serve users appropriately.

BIOGRAPHY

T.V. Raman is Research Scientist at Google Research, Mountain View with almost 20 years of industry experience in advanced technology development, working at Xerox, Intel, Digital, Adobe, and IBM before joining Google. Raman authored 3 books and filed over 25 patents. His work on auditory interfaces was profiled in the September 1996 issue of Scientific American. Raman has leading edge expertise in Web standards, auditory interfaces and scripting languages. He participates in numerous W3C working groups and authored Aural CSS (ACSS). Raman has led the definition of XML specifications for the next generation World Wide Web, including XForms, XML Events, and Compound Document Formats such as X+V. Raman summarizes his objectives as follows:

> Develop technologies that drive the future of the Web toward eyes-free, ubiquitous information access. Speech is the next natural dimension in user interfaces, and I am developing application frameworks that combine speech technologies with the power of the Web to deliver innovative multimodal solutions that are available anytime, anywhere.

T.V. Raman was born and raised in Pune, India. He was partially sighted (sufficient to be able to read and write) until he was 14. Raman received his B.A. in Mathematics at Nowrosjee Wadia College in Pune and his Masters in Math and Computer Science at the Indian Institute of Technology, Bombay. For his final-year project, he developed CONGRATS, a program that allowed the user to visualize curves by listening to them. Many of the ideas on audio formatting mathematics come from his experiences in having math read to him, in dictating math exams and having them written by a writer, and in listening to RFB (Recordings for the Blind) books on tape. Raman was introduced to computing in 1987 with an introductory course on programming in Fortran77. He did his computing with someone behind him to read the display. He joined the PhD program in Applied Mathematics at Cornell University in Fall 1989 supervised by prof. David Gries. In 1994, he received the ACM Doctoral Dissertation Award for his Ph.D. thesis: *Audio System For Technical Readings*. Raman received several other awards, including in 1999 the Computerworld Award from the Smithsonian Institution for Emacspeak: The complete audio desktop.

A Closer Look at Children's Information Retrieval Usage

Towards Child-Centered Relevance

Frans van der Sluis
Human Media Interaction, University of Twente
P.O. Box 217, 7500AE
Enschede, The Netherlands
f.vandersluis@utwente.nl

Betsy van Dijk
Human Media Interaction, University of Twente
P.O. Box 217, 7500AE
Enschede, The Netherlands
bvdijk@ewi.utwente.nl

ABSTRACT

Access to information suitable and understandable for children is key to their development. Regrettably, current state-of-the-art Information Retrieval (IR) is mainly made with adults in mind, resulting in IR systems that do not suit children well: they require complicated queries and often retrieve inapprioriate results in a format unsuitable for children. To confirm this, this paper presents four groups of salient problems children have with IR. To explain these problems, a comprehensive review of children's use of IR systems is given, defining relevant aspects of the user, system, interaction, and context, and relating these to the search performance of children. Based on this framework, an integrative perspective on relevance is proposed, specificially geared at children's needs. It is proposed that complexity, interestingness, and affective value are key relevance criteria for children, and should be incorporated in an information system for children, if to arrive at an optimal search result and experience.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Human Factors, Algorithms

Keywords

Children, Information Retrieval, Relevance

1. INTRODUCTION

Several studies indicate that children, compared to adults, use IR systems in a different way. However, most current IR systems are focussed on adults. Accordingly, they require the typing of complex queries and the use of a correct spelling and efficient vocabulary. Moreover, the search results are often aimed at people who are average literate, requiring sufficient reading skills and a large enough knowledge base, including abstract concepts [12, 5].

Children are still developing their skills and knowledge, and, consequently, have different needs for an IR system.

One of the most notable developmental theories illustrates this. Jean Piaget [29] indicated four developmental stages: sensorimotor (age 0-2), preoperational (age 2-7), concrete operational (age 7-11), and formal operational (age 11 - ...). Some achievements in these stages, relevant to IR systems, are: in the preoperational stage children learn to use language and can classify objects according to one feature; in the concrete operational stage children learn to think logically about objects and events, achieve a notion of the conservation of number, mass, and weight, and can classify objects according to several features; and, only in the last stage children learn to think logically about abstract concepts. Some simple implications of these stages can already be derived, for example the use of simple language and the use of only concrete concepts until age 11.

Current research has found several shortcomings in the use of IR systems by children. Considering the vast amount of studies showing many different problems, an overview of them is given. For ease of comprehension, the problems will be grouped in four categories.

The first group of problems is related to *an insufficient mental model*. Numerous researchers found that children to some extent miss an understanding of the workings of the IR system. This manifestates itself while making search queries by the use of natural language, repetitive keywords, and the limited use of Boolean logic operators [5, 21, 6]. This lack of system understanding is not a sure thing, as most studies show large interpersonal differences.

Having an insufficient mental model is not unique to children. Borgman [7] shows that unexperienced undergraduates also have problems in utilizing Boolean logic for an IR task, especially those from social majors. Such findings have been explained by the Mental Model Theory [28], stating that users construct a mental model of a system by continually forming and verifying hypotheses. An incorrect model decreases success when using the system, especially on complex search tasks [7]. Following its definition, the correctness of a mental model depends mainly on training and previous experience.

The second group of problems is related to *the vocabulary problem*. Numerous studies show children have difficulties in choosing the right words. Often, they misspell their keywords or use keywords that are too broad or too narrow [5]. These findings are often attributed to a lack of vocabulary, which is a known problem in IR research; i.e., the vocabulary problem.

The vocabulary problem has been a major problem in IR. A suprisingly large interpersonal variability in word choice,

Table 1: Dimensions and relations of interaction.

Concept	Interrelations
search process	perceived problem**, affective state** [22]
information seeking pattern	epistemological believes* [40], learning style** [14],
relevance judgements	epistemological believes** [40]
document search strategy	

Note. * *weak relation,* ** *strong relation*

something fundamental to language, leads to a bare 10% hit rate for a single access term [13]. Furthermore, a lack of domain knowledge reduces the change of finding a correct access term.

The third group of problems considered is the *chaotic search behavior* often employed by children using an IR system. This is witnessed by: little reading of the retrieved sources, little focus on the search goal, and many looping and backtracking actions [5, 33]. Rouet and Coutelet [31] show that document search behavior changes with age, such that older children use more cues to search through a document in a more directed way.

Partly, chaotic search behavior is normal at the start of an information search; i.e., the so-called pre-focus phase [22]. It is indicative for not having a directed search goal. However, it is also typical for children that they often don't form a focus at all, nor seem to remain concentrated for a long time [27].

The fourth group of encountered problems is in the used *relevance judgements* by children. Often they scan the text for a specific word or a ready-made answer. For example, a source was seen as irrelevant if not containing a searched-for word in its topic [27]. Furthermore, children tend to take everything as being true and correct [33].

Normally, relevance judgements change over the search process. Where in the beginning a lot of documents are considered relevant, near the end only documents very specific about the problem are considered relevant. Furthermore, the relevance of the early gathered documents is reconsidered [22].

As the different groups of problems show, current IR systems are not optimal for children. And, as several of the problems are not unique to children, the problems indicate the general need for systems focused on the special needs of (less experienced) users. In order to get more grips on the use of IR systems, the next section will give a comprehensive overview of the various aspects relevant during an information search. Using this dissection of IR use, implications are derived for the notion of relevance in Section 3. It is argued that relevance, in the strict sense of topicality, is not enough for a good search experience for children. Finally, the resulting understanding will be discussed in Section 4.

2. USAGE DIMENSIONS

This section gives a theoretical view on the problems encountered (Section 1), reviewing theories from library and information science [e.g., 22], cognitive and educational psychology [e.g, 23], Information Retrieval Interaction (IRI), and Information Search & Retrieval (IS&R) research [e.g, 37, 36]. This section will further elaborate on gaps in functionality and knowledge regarding children's IR usage.

The next sections elaborate on the process of IR use; i.e.,

the different steps users normally proceed through in their search for information. Furthermore, the dimensions (i.e., variables) influencing IR use are reviewed in four categories. These categories are constructed upon the combination of the classifications made by Lazonder and Rouet [23] and Tanni and Sormunen [36], and follow a basic HCI structure:

- Interaction. Search process [22], Information Problem Solving (IPS) activity [23], and access and interaction dimensions [36].

- Context. Contextual variables [23] and learning task dimensions [36].

- User. Individual variables [23] and learner dimensions [36].

- System. Resources variables [23].

2.1 Interaction Dimensions

Table 1 gives an overview of the dimensions characterizing an IR interaction, and their relation to other dimensions. The *search process* is the first interaction dimension reviewed. It involves the cooperation between the user and the system in solving the (perceived) information problem, and thus, in performing the information search activity throughout the *search process*. Hence, the achieved *search performance* is one of the dimensions. First, we will describe various models of the search process, after which other models of the interaction are given.

IR use is viewed as a process. Table 2 summarizes different views on this process. As can be seen, there is a considerable overlap between the views. The process starts with some shortage of information (e.g., a learning task, an information problem); i.e., initiation and selection, pre-focus, or define problem. Then, through searching, the user refines the problem and forms a focus on the problem at hand; i.e., exploration or search. Finally, the user gathers relevant information solving the problem; i.e., collection and presentation, post-focus, or integration [23].

The processes in Table 2 are iterative in nature, thus different steps are repeated. Furthermore, the dimensions (i.e., context, interaction, user) change when proceeding through the processes.

Pattern of information seeking is the second interaction dimension reviewed. Three patterns have been discerned: fast surfing, broad scanning, and deep diving. The search goal is characteristic for the pattern, respectively: finishing the search as quickly as possible, finding as much information as possible, trying to understand. The fast surfing pattern is often found with children (see Section 1). These patterns have been shown to be (partly) determined by study approach and personality traits [14]. Moreover, the study approach is directly related to motivation and interest, discussed in, respectively, Section 2.3.2 and 3.

Table 2: Overview of three search process models.

Information Search Process [22]	IR Process [37]	Information Problem Solving Activity [23]
Initiation	Pre-focus	Define problem
Selection		
Exploration		Search (scanning/studying)
Formulation	Formulation	
Collection	Post-focus	Integration
Presentation		

Related to the pattern of information seeking are the performed *relevance judgements*. Fast surfers tend to only accept documents which contain the whole answer, as often done by children as well (see Section 1). Broad scanners tend to select new, unique, documents related to the search topic. Deep divers select high quality, detailed, highly topical documents [36].

Within a document children perform different *document search strategies* as well. As Moore [27] showed, children tend to linearly scan documents searching an exact match with their keyword. Rouet and Coutelet [31] differentiate between a top-down strategy, where students utilize structural cues of a document, and a linear strategy, similar to what [27] showed. Children develop their search strategies with age; 9 year olds mostly use a linear approach, whereas 13 year olds often use a top-down strategy. It is hypothesized that working memory, metatextual knowledge, and functional comprehension strategies are key factors in efficient search strategies [31].

The interaction dimensions already cover part of the identified problems. In particular, the pattern of information seeking is directly related to the chaotic search behavior and the performed relevance judgements, and can be explained by latent user variables (personality, Section 2.3.1) and actualized user dimensions (motivation, Section 2.3.2).

2.2 Context Dimensions

The second category refers to the context of the user [36]. Meaning, *"all relevant characteristics of the situation (place, time, equipment, people and messages) that pre-exist to the search activity"* [23, p. 756]. The context dimensions are summarized in Table 3.

Inspired by the learning sciences (e.g., Vygotski), *collaboration* and *scaffolding* (i.e., instructional support) are possibly important factors. There is some evidence for a non-beneficial effect of collaboration on *information seeking pattern*: pupils are more concerned with getting pieces of information from their peers rather than creating a shared understanding [2]. On the contrary, group learning has been shown to foster the *study approach* of deep learning [cf. 19]. Considering scaffolding, there is ample empirical evidence of its role on information search behavior. Numerous suggestions have been made; e.g., the tutor could influence thoughts, feelings, and actions within the learning process, for example by encouraging the student [22], and the tutor can keep students in their zone of proximal development, meaning an area of improvement only possible with sufficient scaffolding [22]. Scaffolding and collaboration can be done virtually as well; i.e., resp., by a virtual tutor or virtual peers.

Jonassen [20] identifies key characteristics of a problem.

Namely, *structuredness*, *complexity*, and *abstractness*. Information problems, coined rule-using problems [20], tend to be ill-structured and complex. Though, this depends heavily on the exact search task and the user. For example, problem *structuredness* and *complexity* influence *search performance* [34]. Moreover, the *search performance* of ill-defined (i.e., *structuredness*) problems is partly determined by *epistemic believes* [34] (See Section 2.3.1).

Also in an educational perspective, the *means* and *time* a student is given to solve a research assignment, together with the *educational setting* (e.g., a problem-based curriculum focused on the process of learning), are important determinants for, respectively, search performance [36] and learning outcomes [17].

The context dimensions do not explain any of the problems encountered in Section 1, but do indicate methods to improve: scaffolding and collaboration. Scaffolding can change the motivation and affective state of the child, whereas collaboration can foster a deep learning approach. Hence, indirectly, the context can solve many of the shown problems.

2.3 User Dimensions

The user dimensions can be divided in trait characteristics and state characteristics, which respectively indicates long-term, relatively stable (i.e., latent), and short-term, more unstable (i.e., actualized) user characteristics. Not mentioned but imaginable other characteristics are *gender*, *age*, and *personality* (e.g., the big five). Please note that gender differences are often easily stated (e.g., in technology use / interests), but should be interpreted with care because of often undecided accompanying evidence [11]. Table 4 and 5 give an overview of the, respectively, latent and actualized user characteristics.

2.3.1 Latent User Dimensions

Domain knowledge is a trait which specifies the knowledge about a topic. This is more than just factual knowledge, especially *structural knowledge* about the relations between facts is an important determinant in problem solving (e.g., *search performance*) [20]. Knowledge about the workings of an IR system, i.e., a *mental model*, can be seen as a type of domain knowledge which increases search performance for complex problems [7]. The use of domain knowledge is related to age. Children are less apt to use previous knowledge when using an IR system [6].

Cognitive skills and procedural knowledge is knowing how to do something [29]. For the retrieval of often written information, *literacy skills* are important; i.e., the ability to read, comprehend, and write. Around age 9 there often is a turning point, where the child goes from learning to read

Table 3: Interrelations of context dimensions.

Concept	Interrelations
collaboration	study approach** [19]
scaffolding	motivation* [22]
	affective states* [22]
problem	search performance** [34]
	search performance X epistemic beliefs** [34]
means, time	search performance* [36]
educational setting	learning outcomes* [17]

Note. * weak relation, ** strong relation

Table 4: Interrelations of latent user dimensions.

Concept	Interrelations
age	
gender	
personality	
domain knowledge	search performance X age* [6]
structural knowledge	search performance** [20]
mental model	search performance X problem** [7]
cognitive skills	
literacy	age**
information literacy	search performance** [9]
familiarity	perceived problem** [20]
learning style	information seeking pattern* [14]
metacognitive skills	age** [29, 20]
self-regulation	learning outcomes* [10]
epistemological beliefs	relevance judgements* [40]
	information seeking pattern* [40]

Note. * weak relation, ** strong relation

to reading to learn. *Information literacy* is the "ability to access, evaluate, organise and use information in order to learn, problem-solve, make decisions in formal and informal learning contexts, at work, at home and in educational settings"[9], and thus somewhat overarching the IR use. Related to the information literacy is the *familiarity* with the problem type, stating that routine problems appear more well-structured to the experienced solver [20].

Complementary to cognitive skills is the concept of *learning styles*, reflecting patterns of thinking [20]. One of them, the study approach, differentiates between individuals with a surface approach and a deep approach. Respectively, aiming for finishing a task with the least of effort or aiming for personal understanding [36]. These are related to the *information seeking pattern*; the surface approach is linked to fast surfing, whereas the deep approach is linked to deep diving [14].

The next set of traits are *metacognitive skills and knowledge*; i.e., thinking about thinking. *Self-regulation* is an important skill. It embodies skills such as goal setting, planning, self-motivation, attention control, application of learning strategies, self-monitoring, self-evaluation, and self-reflection [29]. *Epistemological beliefs*, knowing about knowing, includes beliefs about: the certainty of knowledge, the structure of knowledge, the source of knowledge, criteria for determining the truth, the speed of learning, and the nature of learning ability. In other words, people form a theory about knowledge and learning [29].

Self-regulation is stated to be related to reading compre-

hension; i.e., *learning outcomes*[10]. Furthermore, *Epistemological beliefs*, are related to *relevance judgements* (e.g., recognizing authorative information sources) and *information seeking pattern* (e.g., handling conflicting information sources)[40]. Metacognitive skills and knowledge, among which self-regulation and epistemological beliefs, improve with *age*. For example, children become increasingly realistic about their memory capabilities and the speed of learning. They learn more learning strategies and become better in structuring facts [29, 20].

Since the problems and the described user traits are both related to age, the latent user dimensions explain all problems described in Section 1. In specific, first, domain knowledge, which increases with age, explains the problems regarding the mental model of the user. Second, the literacy skills explain part of the vocabulary problem. Third, the learning style, epistemological believes, and self-regulation skills influence the pattern of information seeking and relevance judgements.

2.3.2 Actualized User Dimensions

The *perceived problem*, as contrary to the actual problem (e.g., an assignment), is the user's view of the problem. This depends on, among others, the familiarity of a problem [20], the motivation, etc. The importance of motivation is perfectly illustrated by Tanni and Sormunen [36]: most students see a research assignment as a reporting exercise, where *"reporting means seeking other peoples' answers to someone else's questions"* (p. 901). Furthermore, the per-

6

Table 5: Interrelations of actualized user dimensions.

Concept	Interrelations
perceived problem	problem**
	familiarity**[20]
	motivation*[36]
	search process**[22]
motivation	search performance**[20]
	perceived problem**[36]
	affective state*[30]
affective state	search performance**[22]
	information seeking pattern**[22]

Note. * *weak relation,* ** *strong relation*

ceived problem changes when the search proceeds [22], from an ill-structured problem to a more structured problem; i.e., the search proceeds to a focus. This point is typical for a reduction in the difficulty of the perceived problem: the user has given structure to it, has gained control over the problem. Task complexity and difficulty have been confirmed to have a negative correlation with the experienced valence (i.e., the *affective state*, see Section 2.3.2) [3].

Motivation is an important user state, influencing the effort exerted and the persistence shown in solving a problem. It has been found to be a strong predictor of problem-solving success [20]. Motivation is often divided in intrinsic motivation; i.e., a genuine interest, and extrinsic motivation; i.e., some external incentive. Intrinsic motivation can be explained by two determinants: a task performance which leads to a sense of mastery and competence, and a novelty which leads to a sense of curiosity, attention, and interest [30]. Hence, this is closely related to latent user characteristics [35], the perceived problem (and its progress; e.g., a focus), and the affective state.

Kuhlthau [22] stresses the *affective states* a user experiences throughout an information search. She observed the following emotions from students performing an information search: uncertainty, optimism, confusion, frustration, doubt, clarity, confidence, relief, and (dis)satisfaction. Hence, it seems like searching isn't always a positive experience. The emotions can partly be explained by one of the previous parameters: task performance and sense of mastery. Simply put, advancement in a (not too easy) task will bring about a category of mental states: enjoyment, sense of mastery, sense of control, and competence [30]. Positive affect has been theorized and shown to enhance problem solving, information integration, and intrinsic motivation [18].

In sum, the actualized user dimensions are all directly related to the search outcome: the perceived problem, the motivation, and the affective state of the user throughout the search are all great predictors of search success. In relation to the problems described in Section 1, the actualized user dimensions mainly operate via the information seeking pattern and relevance judgements.

2.4 System Dimensions

Although there are several exceptions, most IR systems currently use queries for users to communicate their information need. And, this method works well; regular users know how to operate an IR system to find what they want. However, as the presented outline shows, there is more to

an information search than just one query. Except for the query, there are more facets the system influences.

First, the *interface* appears to be important as well. The interface can reduce any of the problems, such as cognitive overload or disorientation [23]. For example, visual information search metaphors have been investigated as an alternative to the query interface [25].

The information resources available are likely to be very influential on the eventual success of an information search. Thus, the *data corpus* should contain the information needed and in a format needed. The corpus often differs from the whole web to a library to a subset of websites.

The *media type* can be different as well; e.g., video, picture, text, or hypertext [23]. Different media types put different demands on the user. For example, Homer et al. [16] show that learners who have a visual learning preference had less cognitive load when learning from videos. Furthermore, Rouet and Levonen [32] showed hypertext can cause feelings of disorientation and creates complex cognitive demands [32].

Text characteristics can help readers in finding information and understanding the text. At least three types exists. First, *signaling devices* are such as titles and (sub)headings. Second, *typography* like bold, italic, font use, and capitalization. And, third are *structural elements*; e.g., graphics, (sub)sections, tabel of contents, and indexes. These elements have been shown to interact with, amongst others, age and number of search terms, in students success to find information [8].

The system dimensions form the methods to act for the system to alleviate the problems of Section 1. For example, the next section will show how, through changing the retrieved information sources, the system can aim at improving the actualized user variables.

3. CHILD-CENTERED RELEVANCE

The previous section has presented a framework containing numerous variables of influence when performing an information search. These variables have been related to age as well, indicating why children sometimes have problems with using an IR system (See Section 1). Using the framework, this section will look at a notion of relevance suited for children. This child-centered relevance looks at characteristics of information beyond topicality, salient to improve the actualized user dimensions (e.g., the experience). Several possible lines to influence the actualized user dimensions are, per dimension:

Table 6: Interrelations of system dimensions.

Concept	Interrelations
data corpus	search performance**
media type	learning preference** [16], affective state** [32]
text characteristics	search performance** [8]
interface	perceived problem*, affective state* [23]

*Note. * weak relation, ** strong relation*

- Perceived problem. The *complexity* of the retrieved documents, influencing the perceived problem.

- Motivation. The *interestingness* of information, next to task performance a core determinant of intrinsic motivation [30].

- Affective state. The *affective value* of the information and interactivity of a website influence the affective state of the user.

Please note that this is not supposed to be a comprehensive nor prioritized list. Each of the italic written concepts will be elaborated.

The presented parts of relevance have been confirmed by other studies. For example, Barry and Schamber [4] showed numerous relevance criteria: scope, validity, clarity, currency, tangibility, quality, accessibility, availability, verification, and affectiveness. [15] showed interestingness, accessibility, and language are among children's relevance criteria. Moreover, each of the parts are intrinsically related to the actualized user dimensions, enhancing the affective state, motivation, and reducing the perceived problem of the user.

3.1 Complexity

The perceived problem can be influenced by reducing the complexity of a search. When incorporating this into the notion of relevance, the expectation is that retrieving an information object close to a user's skills and knowledge (See Section 2.3.1) will reduce the perceived complexity of the information need. Moreover, adjusting the complexity of the search results can be especially salient for the beginning of a search, where the user often has less structure in his information problem. For example, if the user has an unstructured information problem for a relatively unknown topic, the scope of the retrieved information objects should be quite large, whereas the used vocabulary should not be too specific. Numerous studies confirm that understandability, comprehension, and complexity are among the core relevane criteria users apply to documents [4]. Moreover, navigational complexity of a website is also related directly to the user experience [26].

The following characteristics have already been identified and linked to the complexity of a document: readability, entropy, semantics/scope, and coherence. Automatic extraction of these charaterics has been shown feasible [38].

3.2 Interestingness

Interest is, together with a sense of control in the perceived problem, core to intrinsic motivation [30]. Interest is often viewed as a cognitive construct; "*liking and willful engagement in a cognitive activity*" [35, p. 23]. Moreover, interests differ between people and between groups of people; e.g., children have different interests than adults. Especially for children, who have less metacognitive skills (e.g., self-regulation) or motivation, interest is, thus, an important part of (child-centered) relevance.

Of particular relevance to IR is knowledge-based interest. The knowledge-based interest refers to the influence of previous knowledge, which has been found to enhance interest, though not always. Novelty of information is a strong interacting factor; when new information can either enhance or assimilate into an existing knowledge scheme, knowledge-based interest is at its optimum [41]. Several approaches have already been undertaken to get some estimation of a user's knowledge-based interest; e.g., through analyzing a user's query log using an ontology [39].

3.3 Affective Value

There are many emotions which (can) occur when searching, which partly originate from the processing of a search result: uncertainty, optimism, confusion, frustration, doubt, clarity, confidence, relief, and (dis)satisfaction. As such, the affective value of an information object can be considered the most direct antecedent to user experience available, having several salient indirect consequences as well (see Section 2.3.2). Moreover, it has been identified as part of relevance [e.g., 4].

For a text, detecting the affective value is mostly pursued with the aim of opinion mining. However, some studies have given attention to pure affect analyses, ranging from a classification success of a low 47.14% to a high 80.98% [1]. These are often achieved by applying a variant of calculating the affective valence of each word. The differences in classification success is attributable to the number of differentiated classes of emotion, the used corpus, and the used method of analysis. Moreover, not only the text influences the emotions. For example, Lin et al. [24] found indications for a relationship between the interactivity of a website and the enjoyment (among which positive affect) it brings about. And, Lin et al. [24] also showed a relation between the use of multimedia and enjoyment.

4. DISCUSSION

As indicated in Section 1, current IR systems are not optimal for children. To support this statement, four groups of problems were introduced, concerning the mental model, vocabulary, (chaotic) search behavior, and relevance judgements. These problems could (partly) be explained by dissecting core factors relevant in IR usage, concerning the human, system, interaction, and context. For example, the problems concerning the mental model can be explained by the domain knowledge of the user; the vocabulary problem is related to the literacy skills and domain knowledge; the chaotic search behavior is partly explained by self-regulation or motivation, and partly by learning styles; and, finally, rel-

evance judgements are related to the epistemological believes of the user. All of the mentioned attributes improve with age: e.g., knowledge and (meta)cognitive skills increase with age, making it clear why these problems are often found in children's IR usage.

Section 3 introduced one direction to address the identified groups of problems, through indicating those aspects of information objects which are in particular salient for children's IR experience. As such, trying to achieve a notion of, so-coined, child-centered relevance. Three parts of child-centered relevance were elaborated: complexity, interestingness, and affective value. Though an implementation of any of these is not (yet) straightforward, the feasability has been illustrated. Through adjusting the complexity, interestingness, and affective value to the user, the perceived problem, motivation, and affective state can (theoretically) be influenced. This immediately shows the need for further research: it is yet unclear if child-centered relevance can indeed influence these variables, and with that, the user experience of IR for children. Moreover, the balance between these variables is also an unsolved issue: how complex, novel, and positive should an information object be, compared to topicality, to be relevant to an un-experienced user?

The child-centered relevance looks at the search results and the (actualized) user variables. However, instead of focussing on the information objects and the (actualized) user variables, implications can be derived for other aspects of the IR system as well. The introduced HCI-model (interaction, context, user, system) illustrates the full length of possibilities. For example, when looking at the system dimensions, the interface can be altered, including the interaction paradigm (e.g., query-based) and results presentation (e.g., synopsis of search results). Moreover, the interaction can be extended to support every step of the search process (See Table 2). Lohmann et al. [25] give an example of an explorative interface for images, supporting a fully different approach to accessing and presenting information objects. Therefore, the overview given in Section 2 mainly serves a broader view on IR, a necessity when retrieving information for children, or broader, users with special needs.

It is key for children's development to give them access to information understandable to them, in a way which suits their view of the world. However, current IR systems do not support children well: they require complicated queries, and often retrieve inapprioriate results in a format unsuitable for children. This paper aimed to facilitate the creation of child-centered information access, based on an understanding of the behaviour and needs of children. It is part of a larger project, PuppyIR, which constructs an Open Source Framework that will provide the infrastructure to develop child-focused and child-friendly components to be deployed within child information services.

Acknowledgements

References

[1] A. Abbasi, H. Chen, S. Thoms, and T. Fu. Affect analysis of web forums and blogs using correlation ensembles. *IEEE Trans. on Knowl. and Data Eng.*, 20(9):1168–1180, 2008.

[2] M. Alexandersson and L. Limberg. Constructing meaning through information artifacts. *The New Review of Information Behaviour Research*, 4(1):17–31, 2003.

[3] I. Arapakis, J. M. Jose, and P. D. Gray. Affective feedback: an investigation into the role of emotions in the information seeking process. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 395–402, New York, USA, 2008. ACM.

[4] C. L. Barry and L. Schamber. Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing & Management*, 34(2-3):219 – 236, 1998.

[5] D. Bilal. Children's use of the yahooligans! web search engine: 1. cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51(7):646–665, 2000.

[6] D. Bilal. Children's use of the yahooligans! web search engine: Ii. cognitive and physical behaviors on research tasks. *Journal of the American Society for Information Science and Technology*, 52(2):118–136, 2001.

[7] C. L. Borgman. The user's mental model of an information retrieval system: an experiment on a prototype online catalog. *Int. J. Hum.-Comput. Stud.*, 51(2):435–452, 1999.

[8] G. Brown. Searching informational texts: Text and task characteristics that affect performance. *Reading Online*, 7(2), 2003. Retrieved March 11, 2009, from www.readingonline/articles/brown.html.

[9] C. S. Bruce. *The Seven Faces of Information Literacy*. Auslib Press, Adelaide, 1997.

[10] J. Coiro and E. Dobler. Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the internet. *Reading Research Quarterly*, 42(2):214–257, 2007.

[11] E. T. Dresang. The information-seeking behavior of youth in the digital environment. *Library Trends*, 54(2):178–196, 2005.

[12] A. Druin, G. Revelle, B. B. Bederson, J. P. Hourcade, A. Farber, J. Lee, and D. Campbell. A collaborative digital library for children. *J. Comp. Assisted Learning*, 19(2):239–248, 2003.

[13] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. The vocabulary problem in human-system communication. *Commun. ACM*, 30(11):964–971, 1987.

[14] J. Heinstrom. *Fast Surfers, Broad Scanners and Deep Divers - Personality and Information Seeking Behaviour*. PhD thesis, Abo (Turku): Abo Akademi University Press, 2002. Retrieved 10 March, 2009 from http://www.abo.fi/ jheinstr/thesis.htm.

[15] S. G. Hirsh. Children's relevance criteria and information seeking on electronic resources. *J. Am. Soc. Inf. Sci.*, 50(14):1265–1283, 1999.

[16] B. D. Homer, J. L. Plass, and L. Blake. The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior*, 24(3):786 – 797, 2008. Instructional Support for Enhancing Students' Information Problem Solving Ability.

[17] F. Hultgren and L. Limberg. A study of research on children's information behaviour in a school context. *The New Review of Information Behaviour Research*, 4(1):1–15, 2003.

[18] A. Isen and J. Reeve. The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion*, 29(4):295–323, 2005.

[19] D. W. Johnson and F. P. Johnson. *Joining together: Group theory and group skills.* Boston, MA: Allyn & Bacon., 8 edition, 2003.

[20] D. H. Jonassen. Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4):63–85, 2000.

[21] Y. Kafai and M. J. Bates. Internet web-searching instruction in the elementary classroom: Building a foundation for information literacy. *School Library Media Quarterly*, 25(2):103–114, 1997.

[22] C. C. Kuhlthau. *Seeking Meaning: A Process Approach to Library and Information Services.* Norwood, NJ: Ablex Pub. Corp., 2004.

[23] A. W. Lazonder and J.-F. Rouet. Information problem solving instruction: Some cognitive and metacognitive issues. *Computers in Human Behavior*, 24(3):753 – 765, 2008. Instructional Support for Enhancing Students' Information Problem Solving Ability.

[24] A. Lin, S. Gregor, and M. Ewing. Understanding the nature of online emotional experiences: a study of enjoyment as a web experience. In *ICEC '09: Proceedings of the 11th International Conference on Electronic Commerce*, pages 259–268, New York, NY, USA, 2009. ACM.

[25] S. Lohmann, P. Heim, L. Tetzlaff, T. Ertl, and J. Ziegler. Exploring relationships between annotated images with the chaingraph visualization. In T.-S. Chua, Y. Kompatsiaris, B. Mérialdo, W. Haas, G. Thallinger, and W. Bailer, editors, *SAMT*, volume 5887 of *Lecture Notes in Computer Science*, pages 16–27. Springer, 2009.

[26] C. Mathwick and E. Rigdon. Play, flow, and the online search experience. *Journal of Consumer Research*, 31(2):324–332, 2004.

[27] P. Moore. Information problem solving: A wider view of library skills. *Contemporary Educational Psychology*, 20(1):1 – 31, 1995.

[28] D. A. Norman. Some observations on mental models. In D. Gentner and A. Stevens, editors, *Mental Models*. Erlbaum Associates, Hillsdale, N.J., 1983.

[29] J. E. Ormrod. *Human Learning.* Pearson Education Inc., Englewood Cliffs, NJ, USA, 2004.

[30] J. Reeve. The interest-enjoyment distinction in intrinsic motivation. *Motivation and Emotion*, 13(2):83–103, 1989.

[31] J.-F. Rouet and B. Coutelet. The acquisition of document search strategies in grade school students. *Applied Cognitive Psychology*, 22(3):389–406, 2008.

[32] J.-F. Rouet and J. J. Levonen. Studying and learning with nonlinear documents: Empirical studies and their implications. In A. P. D. J.-F. Rouet, J. J. Levonen and R. J. Spiro, editors, *Hypertext and Cognition*, pages 9–24. Mahwah, NJ: Erlbaum, 1996.

[33] J. Schacter, G. K. W. K. Chung, and A. Dorr. Children's internet searching on complex problems: performance and process analyses. *J. Am. Soc. Inf. Sci.*, 49(9):840–849, 1998.

[34] G. Schraw, M. E. Dunkle, and L. D. Bendixen. Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology*, 9(6):p523 – 538, 1995.

[35] G. Schraw and S. Lehman. Situational interest: A review of the literature and directions for future research. *Educational Psychology Review*, 13:23–52(30), 2001.

[36] M. Tanni and E. Sormunen. A critical review of research on information behavior in assigned learning tasks. *Journal of Documentation*, 64(6):893–914, 2008.

[37] P. Vakkari. A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation*, 57:44–60, 1 January 2001.

[38] F. Van der Sluis and E. L. Van den Broek. Applying Ockham's razor to search results: Using complexity measures in information retrieval. In *Information Interaction in Context (IIiX) Symposium*, New York, USA, in press. ACM.

[39] F. Van der Sluis and E. L. Van den Broek. Modeling user knowledge from queries: Introducing a metric for knowledge. In *Proceedings of the 2010 International Conference on Active Media Technology*, Lecture Notes in Computer Science. Springer, in press.

[40] E. Whitmire. The relationship between undergraduates' epistemological beliefs, reflective judgment, and their information-seeking behavior. *Information Processing & Management*, 40(1):97 – 111, 2004.

[41] A. S. Yarlas and R. Gelman. Learning as a predictor of situational interest. In *Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, California, April 1998).*, 1998.

Assessing Fun: Young Children as Evaluators of Interactive Systems

Yusrita Mohd Yusoff
Department of Computer and
Information Sciences
University of Strathclyde
+44 (0)141 548 4311

yusrita.yusoff@cis.strath.
ac.uk

Monica Landoni
Department of Computer and
Information Sciences
University of Strathclyde
+44 (0)141 548 2934
Monica.Landoni
@cis.strath.ac.uk

Ian Ruthven
Department of Computer and
Information Sciences
University of Strathclyde
+44 (0)141 548 3098

Ian.Ruthven@cis.strath.
ac.uk

ABSTRACT
In this paper, we describe an exploratory study on the challenges of conducting usability tests with very young children aged 3 to 4 years old (nursery age) and the differences when working with older children aged 5 to 6 years old (primary school). A pilot study was conducted at local nursery and primary schools to understand and experience the challenges working with young children interacting with computer products. We report on the studies and compare the experiences of working with children of different age groups in evaluation studies of interactive systems.

Categories and Subject Descriptors
I.3.6 [**Methodology and Techniques**]: Interaction Techniques

General Terms
Measurement

Keywords
Usability and fun, evaluation, computer products, young children

1. INTRODUCTION
There are many definitions of usability. For instance, usability as defined by ISO9241 is "the effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments". Bevan et al. [1] also provides one definition of usability, the degree to which a computer system is easy to learn and effective to use. Naturally, this easiness depends on who is the user.

Many go further than these standard usability attributes. Jordan as quoted in Monk [17] noted that "usability as a concept does not seem to include (positive) feelings such as, e.g. pride, excitement or surprise". Feelings such as fun and enjoyment are rarely touched on in computer products, except in specific contexts such as computer gaming. Measuring fun, especially when children are interacting with computer products, has become an interesting and growing research topic.

In general, fun is doing activities that are enjoyable and amusing. According to Dorman [8] fun consists of elements of humour, chuckles, delight, ecstasy, gags, gaiety, happiness, jests, jokes, joy, laughter, merriment, mirth, play, pleasantries, quips, and witticism, etc. Read & MacFarlane [18] defined fun as something that children know about; they are experts. They experience it; therefore they can talk about it, describing it as excitement, play, laughter, and feeling good.

Carroll [4] suggests that fun should be included as a separate usability area because fun is not same as satisfaction. MacFarlane et al. [14] also agree that fun is not the same as satisfaction in the definition of usability by ISO 9241-11. Satisfaction is about progress towards goals and fun is not a goal-oriented. Shneiderman [25] states that designing for fun is associated with designing for children. Now more people notice the importance of fun as one of the critical success factors in determining the usability of children's application software. But Yatim [29] claims there are no specific guidelines to measure the effectiveness, efficiency, and satisfaction or fun in any game authoring tool or similar. According to Blythe et al. [2] it is a beginning of the science of enjoyable technology known as "funology".

Computer products for children are developed by adults. Therefore issues like usability and fun are very important to understand from a child's point of view. Measuring fun especially for young children has become crucial and interesting to develop appropriate and interesting computer products for children. As computer products are being developed for increasingly younger children, new evaluation techniques are necessary to help younger children take part in evaluations.

In this paper we report on an exploratory study to investigate the challenges of involving very young children in evaluations: what kinds of evaluations can very young children engage in and what differences are there from evaluations appropriate to older children?

Firstly we review the literature on children and technology, particularly on evaluation, and then we outline the context of our studies which took place in a local nursery and primary school, followed by a discussion of our findings and thoughts on directions for children-centered evaluations particularly focusing on the implications for evaluating interactive search systems.

2. CHILDREN AND TECHNOLOGY

According to Demner [5] in November 2000 almost 20 percent of all digital media users were children and the Internet is a part of child's natural environment with many children now having access to the Internet at school and/or at home. As Plowman and Stephen in Stephen [28] note, information and communication technology (ICT) is not only about desktop computers, laptops and peripherals but also interactive television, digital cameras, video cameras, DVDs, mobile telephones, games consoles, electronic keyboards and toys that simulate 'real technology' such as toy laptops or barcode readers. So children and technology are intertwined because the technology gives impact on the way the children live and learn with all ICT.

2.1 Children as Participants

As many products are designed for childrens' use, many researchers have argued that children should be involved in the software development process. Scott [24] argued that the best people to give information on the child's perspectives, actions and attitudes are children themselves. They can give honest responses if questioned about events that are meaningful to their lives. Guha et al. [13] stated that usually children are not involved in the design process until the end even though there are many roles that children can play in the design of new technology. They believed that children should be involved as equal stakeholders throughout the design process.

Children have their own needs and preferences on computer products which are different to adults. Therefore, it is important to understand how to involve children in the product/system development life cycle. But of course there are challenges working with children especially when working with children at different age groups. For example, a technique might work for a 13-year-old but it would not work or need to be modified for a 4 year-old child [13]. Many problems need to be solved when respondents are children, including problems of language use, literacy and different stages of cognitive development [24].

For children, playing is the most enjoyable activity and nowadays it is very associated with technology. Markopoulus and Bekker [16] mentioned that mostly children play and learn while interacting with technology. They use computer products such as entertainment websites at home, school, or everywhere to get information, education and entertainment. Monk et al. [17] highlighted that fun and enjoyment are becoming a major issue since (ICT) moves from office to home. Research on children and technology is becoming crucial in the last few years.

Druin [9] and Markopoulos and Bekker [15] have started paying attention to children as technology users. They focused their researches on understanding children roles in developing new technologies. For example, Druin [9] stated that children can be involved in many roles such as user, tester, informants or design partner in developing new technologies. Markopolous and Bekker [16] also discussed children's involvement in the design process based on a model introduced by Druin [10] but their focus was on involving children as a tester in usability testing methods.

The model is shown in Figure 1.
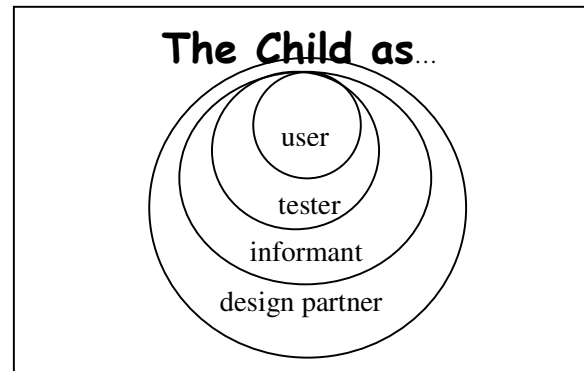
Figure 1. The four roles that children may have in the design of new technologies. Figure adopted from Druin (2002)

2.2 Children as Evaluators

Many researchers have conducted evaluations involving children as evaluators and using different type of evaluation techniques. Previous works have involved children as evaluators to examine, in particular, the usability of computer products [6, 12, 27, 29 and 30]. Children are involved in many ways in evaluation sessions: for example they are required to perform predefined tasks and answer pre and post questionnaires in a lab. Increasingly, evaluations with children are conducted outside labs as interactive technology become more mobile [21].

Read and Markopoulus [21] suggest that different methods should be used in different locations. Some methods such as diary studies, think-aloud methods, surveys, and Wizard of Oz techniques have been used with older children. Read and Markopoulus also describe the Fun Toolkit - a survey method to obtain children's opinion on technology. The Fun Toolkit is a survey instrument or a tool that was developed by Read. It started from a concept (v1) and being developed, used, and reviewed until becoming a Fun Toolkit (V3). The Fun Toolkit comprises of three instruments, the Smileyometer, the Fun Sorter, and the Again Again Table and was carefully designed to be Fun, Fast, and Fair [22]. Some researchers have used other methods like talk aloud (adapted from think aloud) and observation when involving young children as evaluators [7].

The Smileyometer is the first instrument in the Fun Toolkit and is the one most used. It is based on Visual Analogue Scales (VAS) and uses a 1-5 Likert scale and pictorial representations that can help children to identity their feelings or opinions. Faces with supporting text under it are represented horizontally to the children and they are asked to tick only one face. The faces in the Fun Toolkit were co-designed with children aged eight and nine and can be used before and after the children experience the computer technology. The Smileyometer features are easy and quick to complete and requires limited reading and no writing ability [22]. But [20] revealed that the Smileyometer was a useful tool for older children compared to young children. It is because too many young children tended to choose the high values and so the data had little variability.

The Fun Sorter is a tool used to compare a set of related technologies or products. It is based on a repertory grid and made up of n+ 1 columns (where n is the number of items being compared), and m+1 rows (where m is the number of constructs

being used). There are different ways of completing the Fun Sorter. First children interpret the construct then write a description of the technology in blank spaces. But for children with poor reading and writing abilities, they place picture cards (pre-prepared) on an empty grid after interpreting the construct. Few suggestions are given in order to use the Fun Sorter. The use of constructs needs special attention since children are unpredictable in understanding words. It is also recommended that each construct presented individually for younger children aged less than 8 years old. One important thing is make sure children know what the cards represent if picture cards are used. This tool is the most challenging because children require to position and rank items to the construct. The good point is it can be made that no writing is required. Besides, it is fast and fun to complete especially when stick cards are used [22]. But the intention of the Fun Sorter is to record a children's opinions of the technology or activity, to gain a measure of the child's engagement [20].

The Again Again Table is a simple table consists of four columns and n + 1 rows (where n is the number of activities under comparison). Child needs to tick either 'yes', 'maybe', or 'no' for each activity or product. The table should be presented in a single sheet after the children have experienced all the technologies. This tool is most useful if three or more products or activities are being compared. In order to improve validity, the first column can be presented in different orders for different children. It is advisable to minimize the rows (items to compare) as to avoid children from being bored. This table is easy and quick to complete, no writing activity involved, and only has one question to be answered, "Do you want to do it again"? Thus this tool is very suitable to younger children [22].

For interactive search systems evaluation is a particular concern, and interactive systems must be evaluated with end-users – the people for whom the system is constructed [23]. Without such end-users we cannot understand how well the interface supports the user, how usable the system is or evaluate how well the system supports the user in completing a whole search [3]. With children this is difficult. Literature describing children's involvement with usability studies, e.g. [9], points at the difficulty of gathering valid feedback since verbal communication, both in understanding and formulating sentences, is not as effective as with adults. Consequently, evaluators have been forced to seek methods, ranging from interpreting free drawings [11] to using collections grids with "smiles" instead of grades [19]. Children also have problems in expressing their feelings in terms of satisfaction [9, 19]. A third element emerging from previous studies [26] is the discrepancy between reported and observed usability when children are asked to provide subjective feedback versus direct observation.

Thus it is clear that standard tools for user evaluations are not directly applicable to children's evaluation; questionnaires require higher degrees of literacy than is common in young children, interviews require high degrees of reflection and techniques such as think-aloud require high degrees of cognitive dexterity [20]. Neither can we expect children to engage in standard experimental procedures such as searching on artificial search tasks, searching for controlled amounts of time, or engaging in procedures such as training or debriefing.

Rather, for conducting user evaluations with children we need to (a) develop evaluation methodologies that allow children to interact naturally with the system being evaluated whilst retaining some experimental control, (b) understand how children express notions of satisfaction with a system and (c) understand what metrics are appropriate for children's search systems.

There is particularly a dearth of work on engaging very young children in usability evaluations. This might be because of physical and mental abilities; limitations of these mean that some researchers think young children are not capable of being involved in usability evaluation. In this paper we explore the challenges of involving very young children in usability evaluations with particular reference to evaluating the fun of an information system.

3. THE STUDY

An exploratory study on evaluating young children interacting with an edutainment website was conducted at a local nursery school and a local primary school. The purpose of the study was to understand and experience the challenges of working with young children, aged 3 to 4 years old at the nursery and 5 to 6 years old at the school and any possible differences when working with children of different ages. Eight nursery children and five school children voluntarily participated in the study.

3.1 The Nursery Background

The Nursery is located in the UK and offers two sessions, morning and afternoon. The morning session starts at 8.45am and finishes at 11.45am. The afternoon session starts at 1.00pm and finishes at 4.00pm with 10 permanent staff.

The capacity of the nursery is 80/80. It means for each session, the maximum number of children is 80. In session 2008/2009, there are 79 children attending the morning and afternoon nursery session, genders were equally represented in each session.

The nursery is a diverse school with children from many nationalities. Besides English, there are various languages spoken by the children in the nursery such as Urdu, Punjabi, Malay, Mirpuri, Pushto, Arabic and Farsi.

In the nursery, there are four rooms fully-equipped with toys, books, and other children's' material but only Room 1 and Room 2 have computers. Room 1 was the place where the study was conducted. Even though there are three computers available in the room but only one computer (in the middle) with a speaker was used in the study.

3.2 The School Background

The Primary School is also in the UK. The school has 15 teaching and five support staff at the moment. It also has pupils come from different minority ethnic communities such as Pakistan, Malay, Czech, and Arabic. The working capacity of the school is 260. But in the current session 2008/2009, the present roll is 219 pupils, which are 122 boys and 97 girls.

The school starts at 9.00 am and finishes at 3.00 pm. There are two slots of breaks, one in the morning and another in the afternoon. In the school, there are seven classrooms for Primary 1 to Primary 7. The Primary 1 classroom, which is located on the first floor, was the place where the study was conducted. There were 12 boys and 15 girls in the class and all of them can speak English. The classroom was provided with 2 personal computers.

3.3 The CBeebies Website

During the study the children were asked to interact with the CBeebies website. The CBeebies website is based on a very popular children television channel in UK. Figure 2 shows the main site as used in the study. There are 18 main links on the left handside of the screen such as, Home, All CBeebies Characters, Fun and Games, Stories and Rhymes, Print and Colour, Make and Do, Music and Songs and many more. At Home screen, contents on the right handside changes regularly. This interactive website that contains multimedia elements like graphic, audio, video, animation, and text can be accessed through URL http://www.bbc.co.uk/cbeebies. For the study purposes, the children were asked to play/explore the Fun and Games section only, Figure 3.

Figure 2. CBeebies Websites Screenshot

Figure 3. CBeebies Websites: Fun and Games Screenshot

3.4 Procedure

Prior to the study we obtained ethical permission from the Local Education Authority, Departmental Ethics Committee and parental consent forms. The latter was required to allow children to take part in the study. We discussed the issue of reward with the Headteachers who felt this would not be appropriate so no reward for participation was given in the nursery study.

3.5 Methodology

Both studies comprised of five main activities:

1. recruiting the children

2. introducing the researcher

3. asking volunteer children to play/explore the Fun and Games section in CBeebies website for 5 minutes,

4. interviewing each participant for about 5 minutes,

5. asking the child to draw a character that represented what they enjoyed about the game

In the sections that follow we describe how these stages were accomplished in the two locations and why they were important. In both locations we followed methodologies that were acceptable to the nursery and school. Although this results in differences in recruitment and methodologies, it is important for real-life studies to fit with the constraints imposed by the participating organizations.

3.5.1 Greeting

Nursery: The researcher made several visits to the nursery prior to the study to familiarize herself with the nursery environment and to familiarize the nursery pupils with her presence in the nursery.

In the nursery, rather than employ direct recruitment the nursery staff suggested that the researcher wait at the computer desk until an interested child came to play computer games. This suggestion was agreed by the researcher. The nursery staff were also a good source of knowledge as to which children were good at using computers from their daily observations of the children.

On the day the study was conducted, the weather was warm and sunny. Most of the children enjoyed playing bicycles and scooters outside the nursery building and showed less interest in playing inside. Due to an outbreak of Swine Flu in a nearby primary school, some of the children who were most able to use computers were absent from the nursery.

School: A Pupil Support Assistant (PSA) was assigned by school's headteacher to help the researcher at the primary school. Based on the returned parental consent forms, the researcher was asked to select 5 children to take part in the study. Before the study was started, all the Primary 1 (first year) children were taken to the gymnasium for a physical exercise class. Then the PSA took children one-by-one from the gymnasium to participate in the study. The participant selection at school was done systematically and took a shorter time to accomplish.

3.5.2 Introduction of researcher

Nursery: The researcher was a familiar person in the nursery but not personally known to all children. The researcher introduced herself informally to each of the children, who participated in the study by asking questions,

14

"Do you know me..?" and then answering it by herself,

"I am Mrs X..."

School: In the school, the researcher was introduced by the class teacher formally in front of the class before the study was conducted. This was a standard method of introducing new people to the children in the school.

3.5.3 Ask volunteer child to play/explore
In both locations, the researcher showed them a laminated-screenshot of the CBeebies websites and the children were asked to choose a game to play with. This meant the children choosing a game with which they were familiar. As we were interested in evaluation methodology, rather than evaluating a specific product, we felt this was a fair limitation.

We set a target of 5 minutes to play with the game because it was presumed that young children might lose focus in a longer period. It also to make sure the study at nursery can be finished before snack time, around 11.00am. At school, the study was stopped for 15 minutes for playtime or morning interval at 10.45am.

Each child spent another 10 minutes for interviewing and drawing.

3.5.4 Interview
In both locations, if the child remained long enough to be interviewed we asked a range of open and closed questions. These questions were to explore what kinds of questions children of different ages were comfortable answering and what kinds of responses they were willing to give. The questions were deliberately conversational in nature, starting with closed questions which are easier to answer. The questions were as follows:

1. Have you seen this program before?
2. Have you used this program before?

These two closed questions were to gain insight into a child's previous experience which may be useful for contextualising the responses to later questions and for exploring what the child found fun about a game.

3. Do you like to play game from this website?
4. Which game do you like to play?.
5. Why do you like to play this game?

These questions are on general experience of using this popular site and were asked if the children was familiar with the site. The question block starts with a closed question, leading to simple choice question and finally an open question.

6. Do you like the colours used?
7. Do you like to hear songs from this game?
8. How do you feel after playing this game?

These questions explore what aspects of a program or game might be enjoyable to a child. We are particularly interested in the evaluation of fun from a child's perspective and wanted to explore what judgements a child may give through the use of open questions.

9. Do you want to recommend this game to your friend? Why?

This question tested a child's ability to identify, express, and share their emotions of having fun by telling other people. In this question, friends are highlighted because of the importance as the closest person for them to play with.

10. Can you draw the character that you like most from the site?

The final question, really a task, asked the child to draw the character they liked most from their exploration. This exploratory activity might be useful to identify whether children having fun interacting with the game. Their enjoyment of playing self-chosen game can be transformed into a cartoon character drawing explicitly on a piece of paper.

4. FINDINGS
In this section, we summarise the outcomes from each study in sections 4.1 and 4.2, with particular attention to the final task in section 4.3 and draw some comparisons in section 4.4.

4.1 Findings from Nursery Study
10 questions were planned but which questions were asked was determined by the child's mood and ability to answer. We took care not to place any pressure on any child or to continue if it became clear that a child was becoming bored or did not understand questions. Some children did struggle with physical limitations such as hand and eye coordination in using computers. A particular issue, which we will return to later is the child's mood.

Questions 1 - 4 were easily answered by the children. They were very familiar with the websites. In fact, they can directly go to the page without any help. The children said that they watch CBeebies TV channel at their home almost every day. Only one participant did not want to play any games from the CBeebies websites and chose another game.

The open question 5 was more difficult for children to answer and was not asked to all participants. It was clear that open reflective questions were difficult for very young children to answer. Similarly other open questions such as question 8 and 9, which were only asked to children that showed ability to communicate and reason, were difficult to answer.

It was particularly difficult for children to reason about emotions. Even though many computer programs and games are designed to be fun and enjoyable, very young children could express enjoyment but not reason about it.

4.2 Findings from School Study
10 questions were prepared for the study and all of them were asked during the study to all participants. But only two participants could understand and answer all the questions.

Questions 1 - 4 were easily and confidently answered by the children. They were very familiar with the websites. One participant managed to go to the CBeebies websites by clicking *Favorites Center*. Questions 6 – 7 also can be answered by all of them.

The open question 5 was answered by two children only. The other children had difficulties to give reasons as were other open questions such as question 8 and part of question 9. But the children were more confident in their responses by saying they didn't know or did not have an answer. Overall the children showed a greater ability to understand and communicate.

4.3 Drawing

The final task we asked the children to engage in was to draw their favourite character from the game they choose to play with. This was an attempt to see if we could learn something about what children enjoyed about a game from an associated activity. The quality of the drawing here was not important – and most very young children naturally could not produce recognizable drawings – rather we wanted to create a stimulus for discussing their experience of the game.

In the nursery most of the children could not answer question 10. All of them were unable to draw except one girl. The others were only able to colour the paper that had been given to them and engaged in little discussion related to the game. One child, when asked about the drawing, said the character he liked most is *Batman*, which is not in the CBeebies websites and out of context.

In the school, however, all of the participants could draw a character related to the experience of the game. They were able to draw the cartoon character even though it was not exactly same as seen on the computer screen. In particular a drawing was so good to be easily recognized by the researcher.

4.4 Comparisons between the Two Groups of Children

The aim of this exploratory study is to understand what are the major differences between working with young children (school) and very young children (nursery) when evaluating information systems. As more information and particularly search systems are being created for very young children it is important to understand the challenges of evaluation by such children and how best to engage them in the process of evaluation. In this section we will summarise some of the major trends from our study.

4.4.1 Recruitment

Recruitment is a challenge when working with very young children. In the nursery context, where the main activities are play, children's participation had to be voluntary. Hence only children who were interested took part and their involvement ceased when they were bored. One child participant got bored playing after two minutes and walked away to play with other things in the room. As noted before, other environmental distractions such as good weather or interesting toys made computers less attractive. Another participant refused to play any CBeebies game but would play other games. Other children were more shy and took longer to approach the researcher, although were interested to join the study.

We deliberately chose a real-life setting to conduct the evaluations as children's use of computer products naturally takes place within environments where there are choices of activities. If a very young child becomes bored or has more interesting activities – particularly those that involve other children – then they can quickly lose interest in the evaluation. Although this means that evaluations with very young children may often be snap-shots of interactions with computer products being picked up and quickly dropped, this does lend realism to the evaluation compared to the actual use of a computer product.

Recruitment also relates to child's confidence in the researcher. In our case, the researcher took care to become a familiar part of the nursery environment. However, we did notice that some children took longer to trust the researcher than others and the researcher did devote time to engaging with children in other activities, such as singing songs, to help engender a trust relationship with children.

In the school context all children were comfortable with computers and the school was happy to assist in direct recruitment within the class. This will not be the case in all schools but the context of a school – where children are expected to learn as well as play - does mean that children are becoming used to engaging in activities that they have not chosen themselves.

4.4.2 Verbalisation

All children in the school environment were better at verbalising and general communications. All could choose a game and explain why they chose it. They could not answer all the open-ended questions but some could answer the most difficult questions and give reasons for their answers, e.g. why they would recommend a game to friends, which could be used to gain additional information on the attractive features of a game.

In the nursery environment, the children were more reluctant to answer questions and at least one child would use nodding rather than verbalizing responses. In same environment, two children gained confidence from participating together. Very young children also had more difficulty in understanding questions.

When working with very young children, therefore, it may be necessary to have different means of asking questions and to carefully consider what kinds of questions children may be able to answer.

4.4.3 Evaluation as a process

A particular issue that arose was the degree to which the process of evaluation can be separated from the process of interaction with a computer product. Often evaluation techniques are separated to the act of interaction, i.e. the evaluation takes place after the interaction. Alternatives that can be used at the same time as interaction, such as think-aloud are not suitable for very young children due to the need to verbalise and reason.

When we asked the school children to draw a favourite character from the game, most could carry out this task and could discuss the character with reference to the game. However with the children in the nursery, this task largely failed and the act of drawing was seen as a different activity to the game. This raises questions about how to connect evaluation to the experience of interacting with a product.

5. LIMITATIONS AND FUTURE DIRECTIONS

This study was a small exploratory study conducted within one nursery and one school environment. We only used one website, although many games were available from this site, and carried out only one round of evaluations. Nevertheless, we believe that the tentative findings are of interest in pointing to some difficulties in working with a distinct group of computer users. This is a challenging, but rewarding, group to work with and it was clear that they have specific needs in terms of evaluation. We are continuing to work with the nursery school pupils to explore what kinds of evaluation are attractive and useful to them in evaluating products designed for their use. Specifically, we are investigating methods that enable them to express emotional reactions to computer products.

6. CONCLUSIONS

Our study focused on computer games. This was to provide children with a familiar computer product so that we could concentrate on the process of evaluation. However, we believe our findings on evaluation are relevant to any interactive product such as search systems. Increasingly researchers are examining search systems for use specifically by children. Usually these systems assume a certain level of literacy and so concentrate on older children. However, younger children are often able to use computers and may want to search for information. The information they wish and the methods that are appropriate to enable them to search may be very different from those of other users, which means that we need to have methods of evaluating systems for this group of children.

Our study was aimed at understanding the challenges of working with young children. Obviously it was difficult to get data from the young children. They can easily feel bored, do not understand some questions, cannot necessarily reason about experience, may experience language barriers due to low vocabularies and may have physical limitations such as hand and eye coordination in using computers. This has implications for the design of search systems for children but also for evaluation: evaluations of search systems with very young children cannot rely on the relatively open-ended data gathering methods (such as interviews and think-aloud) common in search evaluations of older people. Neither can search evaluations rely so strictly on the comparative experimental method commonly seen in IR evaluations where the same participants operate two or more versions of a system for fixed times and on given search tasks. Our experience suggests that, given very young children are emotionally driven, evaluation techniques will require to be flexible in coping with children's emotional states (including boredom and shyness), and focus on concepts accessible, understandable and interesting to children. We are exploring such approaches now.

While the study conducted at the primary school with five volunteer participants had indicated that there were possible differences when working with children of different ages. Children at primary school are more confident, easily can understand questions and instructions, and also have better communications skills.

Several studies should be conducted in the nursery to obtain more and richer data from the young children. Our first study had the additional merit to break the ice with the children and let them familiarise with the researcher. We expect that if several studies are conducted involving young children, perhaps the process of getting data from them becomes easier and their ability to contribute to computer product development becomes stronger.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Bevan, N., Kirakowski and J., Maissel, J. 1991. What is usability? In *Proceeding of the Fourth International Conference on HCI*. Stuttgart, Germany. 1991.

[2] Blythe, M., Hassenzahl, M., and Wright, P. 2004. Introduction. *interactions* 11, 5 (Sep. 2004), 36-37. DOI= http://doi.acm.org/10.1145/1015530.1015545

[3] Borgman, C. L., Hirsh, S. G. and Hiller, J. "Rethinking online monitoring methods for information retrieval systems: from search product to search process". *Journal of the American Society for Information Science*, 47,7, 568-583. 1996.

[4] Carroll, J. M. 2004. Beyond fun. *interactions* 11, 5 (Sep. 2004), 38-40. DOI= http://doi.acm.org/10.1145/1015530.1015547

[5] Demner, D. (2001). Children on the Internet. Retrieved May 27, 2009, from University of Maryland, Universal Usability in Practice Web site from http://otal.umd.edu/uupractice/children/

[6] Donker, A., and Markopoulos, P. Assessing the effectiveness of usability evaluation methods for children. Paper presented at the PC-HCI2001, Patras, Greece. 2001.

[7] Donker, A. and Reitsma, P. 2004. Usability testing with young children. In *Proceedings of the 2004 Conference on interaction Design and Children: Building A Community* (Maryland, June 01 - 03, 2004). IDC '04. ACM, New York, NY, 43-48. DOI= http://doi.acm.org/10.1145/1017833.1017839

[8] Dorman, C. (2000). Engaging consumers, using humour in electronic commerce. Computers and Fun 3, BCS HCI Group, York, England.

[9] Druin, A., Bederson, B., Boltman, A., Miura, A., Knotts-Callahan, D., and Platt, M. 1998. Children as our technology design partners. In *the Design of Children's Technology*, A. Druin, Ed. Morgan Kaufmann Series In Interactive Technologies. Morgan Kaufmann Publishers, San Francisco, CA, 51-72

[10] Druin, A. "The Role of Children in the Design of New Technology". *Behaviour and Information Technology*, 21, 1, 1-25, 2002.

[11] Druin, A., Weeks, A., Massey, S., and Bederson, B. B. 2007. Children's interests and concerns when using the international children's digital library: a four-country case study. In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries* (Vancouver, BC, Canada, June 18 - 23, 2007). JCDL '07. ACM, New York, NY, 167-176. DOI= http://doi.acm.org/10.1145/1255175.1255207

[12] Egloff, T. H. 2004. Edutainment: a case study of interactive cd-rom playsets. *Comput. Entertain.* 2, 1 (Jan. 2004), 13-13. DOI= http://doi.acm.org/10.1145/973801.973822

[13] Guha, M. L., Druin, A., Chipman, G., Fails, J. A., Simms, S., and Farber, A. 2004. Mixing ideas: a new technique for working with young children as design partners. In *Proceedings of the 2004 Conference on interaction Design and Children: Building A Community* (Maryland, June 01 - 03, 2004). IDC '04. ACM, New York, NY, 35-42. DOI= http://doi.acm.org/10.1145/1017833.1017838

[14] MacFarlane, S., Sim, G., and Horton, M. 2005. Assessing usability and fun in educational software. In *Proceedings of the 2005 Conference on interaction Design and Children* (Boulder, Colorado, June 08 - 10, 2005). IDC '05. ACM, New York, NY, 103-109. DOI= http://doi.acm.org/10.1145/1109540.1109554

[15] Markopoulos, P. and Bekker, M. 2002. How to compare usability testing methods with child participants. *In Proceedings of the International Workshop Interaction Design and Children (IDC) 2002*. Eindhoven, The Netherlands.

[16] Markopoulos, P. and Bekker, M. "Interaction Design and Children". *Interacting with Computers,* 15, 2, 141 – 149, 2003.

[17] Monk, A., Hassenzahl, M., Blythe, M., and Reed, D. 2002. Funology: designing enjoyment. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA, April 20 - 25, 2002). CHI '02. ACM, New York, NY, 924-925. DOI= http://doi.acm.org/10.1145/506443.506661

[18] Read, J.C. and MacFarlane, S. (2000). Measuring Fun – usability testing for children. Computers and Fun 3, BCS HCI Group, York, England.

[19] Read, J. C., MacFarlane, S. J., and Casey, C. Measuring the Usability of Text Input Methods for Children. Paper presented at the HCI2001. Lille, France. 2001.

[20] Read, J. C. and MacFarlane, S. 2006. Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceedings of the 2006 Conference on interaction Design and Children* (Tampere, Finland, June 07 - 09, 2006). IDC '06. ACM, New York, NY, 81-88. DOI= http://doi.acm.org/10.1145/1139073.1139096

[21] Read, J. C. and Markopoulos, P. 2008. LIFELONG INTERACTIONS Understanding children's interactions: evaluating children's interactive products. *interactions* 15, 6 (Nov. 2008), 26-29. DOI= http://doi.acm.org/10.1145/1409040.1409047

[22] Read, J. C. "Validating the Fun Toolkit: an instrument for measuring children's opinions of technology". *Cognition, Technology & Work*, 10, 2, 119 – 128, 2008.

[23] Saracevic, T. 1995. Evaluation of evaluation in information retrieval. In *Proceedings of the 18th Annual international ACM SIGIR Conference on Research and Development in information Retrieval* (Seattle, Washington, United States, July 09 - 13, 1995). E. A. Fox, P. Ingwersen, and R. Fidel, Eds. SIGIR '95. ACM, New York, NY, 138-146. DOI= http://doi.acm.org/10.1145/215206.215351

[24] Scott, J. 2000. Children as Respondents: The Challenge for Quantitative Methods. In P. Christensen and A. James. Research with Children: Perspectives and Practices. RoutledgeFalmer. London, UK.

[25] Shneiderman, B. 2004. Designing for fun: how can we design user interfaces to be more fun?. *interactions* 11, 5 (Sep. 2004), 48-50. DOI= http://doi.acm.org/10.1145/1015530.1015552

[26] Sim, G., MacFarlane, S. and Horton, M. 2005. Evaluating usability, fun and learning in educational software for children. *In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2005*. Chesapeake, VA: AACE.

[27] Sim, G., MacFarlane, S., and Read, J. 2006. All work and no play: Measuring fun, usability, and learning in software for children. *Comput. Educ.* 46, 3 (Apr. 2006), 235-248. DOI= http://dx.doi.org/10.1016/j.compedu.2005.11.021

[28] Stephen, C., McPake, J., Plowman, L. and Berch-Heyman, S. "Learning from the children: Exploring preschool children's encounters with ICT at home". *Journal of Early Childhood Research*, 6, 2, 99 – 117, 2008.

[29] Yatim, M. 2008. Usability and fun evaluation of a game authoring tool. In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications.

[30] Zaman, B. Evaluating games with children. Paper presented at the Proceedings of Interact 2005 Workshop on Child computer Interaction: Methodological Research, Rome, Italy. 2005.

Text Simplification for Children

Jan De Belder
Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A
B-3001 Heverlee, Belgium
jan.debelder@cs.kuleuven.be

Marie-Francine Moens
Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A
B-3001 Heverlee, Belgium
sien.moens@cs.kuleuven.be

ABSTRACT

The goal in this paper is to automatically transform text into a simpler text, so that it is easier to understand by children. We perform syntactic simplification, i.e. the splitting of sentences, and lexical simplification, i.e. replacing difficult words with easier synonyms. We test the performance of this approach for each component separately on a per sentence basis, and globally with the automatic construction of simplified news articles and encyclopedia articles. By including information from a language model in the lexical simplification step, we obtain better results over a baseline method. The syntactic simplification shows that some phenomena are difficult to recognize by a parser, and that errors are often introduced. Although the reading difficulty goes down, it still doesn't reach the required level for young children.

Keywords

text simplification, readability

1. INTRODUCTION

The Internet contains a wealth of information, but only a small fraction of that information is suited for the reading level of children. Especially in the last decade, a lot of research has been put into automatically assigning a measure of readability to text, and retrieving documents that are suited for a predetermined reading level. This paper addresses a related issue, that arises when a document with the right reading level can't be found: rewrite the text so that it does become suited, according to an external readability measure. We introduce a method that takes complicated text as input, and generates a text that is simpler and easier to understand for children.

Text simplification may serve many purposes, and has been researched with very different objectives in mind. Originally, the purpose was to break down long sentences in order to improve the accuracy of parsers [4, 26]. Text simplification was also used to automatically make text more understandable by aphasic readers [3], or readers with low literacy skills [2]. Yet another application is the simplification of text as a preprocessing step for other NLP tasks, such as Relation Extraction [16], Semantic Role Labeling [27] and Machine Translation [21].

The goal of most research on text simplification is to make the text as simple as possible. Only [23] and [2] first train a classifier that decides whether or not a sentence is too difficult, and if it is the case then a rule based system is applied to attempt to simplify the sentence. The problem with training a classifier is that annotated training data is needed, and even then the decisions are made on the level of individual sentences, not on the level of the entire document. The problem with simplifying as much as possible is that the text might become too easy: we want the text to fit the reading level of a child as good as possible, rather than making it overly simple.

By casting the problem as an Integer Linear Programming (ILP) problem, we can find a global solution (i.e. choice of simplifications) so that the entire text satisfies certain conditions regarding the reading difficulty. These conditions can be modeled through the objective function and constraints.

In the next section we will discuss relevant work. Section 3 introduces the different parts of the method. In section 4 we evaluate the two main components of the system (lexical and syntactic simplification), and also evaluate how well it is able to reduce the reading difficulty. We end with conclusions in section 5 and indications for future work in section 6.

2. RELATED WORK

2.1 Automatic text simplification

Relevant work started with [4], where sentences were split into shorter sentences by using suppertagging (a weak form of parsing), in order to speed up parsers and improve accuracy. Research in this direction continued with [26], making use of shallow preprocessing and taking hints from punctuation. In this work was also attention for the regeneration stage, so the sentences that were split form a coherent piece of text.

Simplification in order to make text more accessible for aphasic readers was done in [3], in the PSET project. Long sentences and passive constructions are hard to understand for people with aphasia, and these phenomena were simplified making use of the output of a parser and a set of rules thereon. Anaphoric expressions were replaced by there antecedents. The PSET project also had attention for lexical simplification, by replacing difficult words with ones that are easier to understand [8]. The method was evaluated on news

articles from a local newspaper. Similarly, the PorSimples project tries to automatically simplify Brazilian Portuguese text for people with low literacy skills [2]. A first step is classifying each sentence as easy enough or too difficult, using many of the features common in predicting the readability of text. If a sentence is classified as too difficult, a rule based system tries to simplify the sentence. In contrast, the approach we present in this paper decides which sentences to simplify on the level of the entire document, instead of on a per sentence basis.

In a more general setting, the method in [9] can rewrite text by using a Synchronous Tree Adjoining Grammar and large set of paraphrases on this grammar, defined as tree transformations. A global constrained solution is found with Integer Linear Programming (ILP), although only one transformation can be applied to every sentence.

[17] introduced a method to simplify sentences for information seeking applications by extracting Easy Access Sentences. These can loosely be defined as grammatical sentences with one finite verb. Although it has attention for semantically problematic environments, such as conditional constructions, this method generates rather dull texts that are suited for information seeking applications, but not for children. The method in [27] is based on the reduction of parse trees, by applying a set of 242 rules on them, in order to obtain a less 'noisy' dataset for Semantic Role Labeling. However, much information is lost in this process (e.g. modal verbs), and the meaning of the sentences is likely to be altered. The method in [16] serves as a preprocessing step for Relation Extraction, and is based on the Link Grammar parser. Longer chunks of the sentence are fed to the parser until an 'S-link' is found, meaning that that part of the sentence forms a sentence by itself.

2.2 Lexical simplification

Lexical simplification has been performed in [8]. The simplification there consists of lexicon substitution. All words are looked up in WordNet [10] and their synonyms (synsets) are retrieved. For all the synonyms, including the original word, the method looks up the Kucera-Francis frequency [18] in a psycholinguistic dictionary [25]. If one of the synonyms has a higher Kucera-Francis frequency, it is an indication that this synonym is easier, as more frequent words are better known than less frequent ones. If one of the synonyms has a higher frequency than the original, the latter is replaced by the most frequent synonym.

However, in [19] the authors reported that this method often generated "weird sounding" sentences. A possible explanation is that every word can have different unrelated synonyms, because a word can have different meanings. In [8], the argument against using Word Sense Disambiguation (WSD) was that a difficult word might have only one specialized meaning.

2.3 Reading level assessment

Several researchers have demonstrated the effectiveness of machine learning approaches over the traditional measures such as the Flesch-Kincaid and Dale-Chall readability tests. These traditional measures usually are a linear combination of the average sentence length, average number of syllables, and the number of 'difficult' words, e.g. words with 3+ syllables or words that are not in a basic word list. The more advanced features used nowadays range from lexical

[5] to syntactic [23, 13], and even coherence [1]. Also cognitively motivated features [11] and discourse structure [24] have been used.

An interesting difference between all these methods is the target audience. [3] focuses on patients with aphasia, although no explicit attempt at identifying the difficult parts of text was made. [22] aims at foreign and second language learners. [11] focuses on people with a cognitive disability. These last two have made use of data obtained from Weekly Reader[1], a magazine with an edition aimed at children of different grades, and thus ideal to train and test reading level assessment approaches.

3. METHOD

Our method consists of three components. The first two are the lexical and syntactic simplification of text. The third component concerns choosing the right set of simplifications that were generated by the previous components.

3.1 Lexical simplification

In the lexical simplification step the aim is to replace difficult words and expressions with simpler ones. This task is closely related to paraphrasing and machine translation, with as source language English, and as target language 'simple' English. Unfortunately, whereas there are parallel corpora available for paraphrasing and machine translation, a similar parallel corpus to learn simplifying expressions from is not available. For this reason we focus our attention on an easier task, the lexical substitution of individual words.

As mentioned in section 2.2, using the most frequent synonyms does not always generate the correct substitutions. Our approach uses a limited form of Word Sense Disambiguation to alleviate this problem. The main idea is that we not only generate alternative words from WordNet, but combine this with a language model [7]. The Latent Words Language model models both language in terms of consecutive words and the contextual meaning of the words as latent variables in a Bayesian network. In a training phase the model learns for every word a probabilistic set of synonyms and related words (i.e. the latent words) from a large, unlabeled training corpus. So rather than taking simply the synonyms from WordNet, we take the intersection with the words generated by the language model (see figure 1 for a graphical representation). Because of the one sense per context phenomenon [28], this gives reasonable grounds to assume the substitutions are correct.

Alternatively, another approach could be to use a standard trigram language model, and ignore the synonyms that have a language model probability below a certain threshold.

What remains is the problem of ranking the different candidates in the intersection of WordNet and the language model, in order to select the easiest. An indication of how easy a word is, could be obtained by looking at the *Age of Acquisition* rating, available from the Oxford psycholinguistic database [25]. Unfortunately, many words lack this rating, so like in previous work we use the Kucera-Francis frequency. The word with the highest frequency is chosen to replace the original word, if it has a higher frequency than the original word.
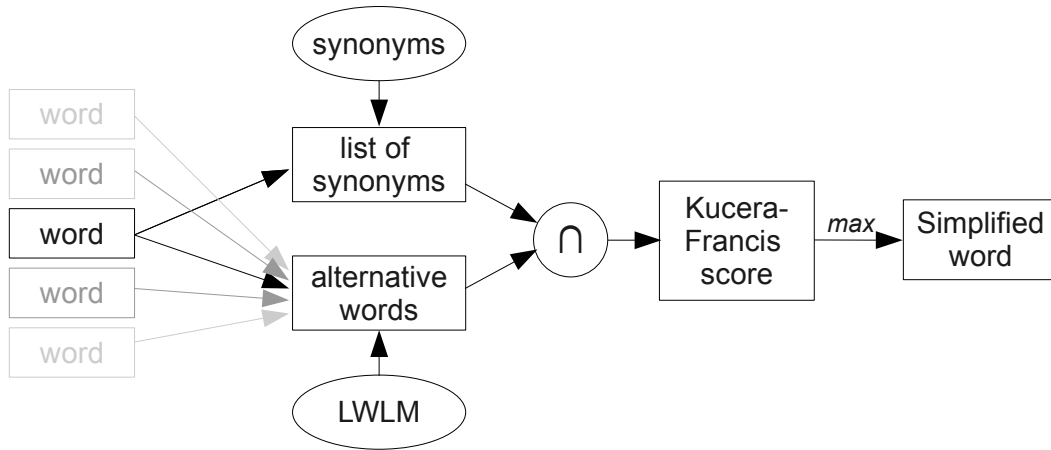
---

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

Figure 1: Schematic representation of the lexical simplification

3.2 Syntactic simplification

Previous work has relied on rule based systems to simplify a certain number of syntactic constructions. This is also the approach we follow in this paper. Constructions that are typically simplified are relative clauses, appositions, passive voice, and conjunctions [3], but also constructions such as subordinate clauses and if-then structures [26], which are inspired by Rhetorical Structure Theory (RST).

We use the Stanford parser [6] to perform a syntactic analysis of the input sentences. It has a rich annotation scheme that marks several structures that we aim to simplify. We selected the following set of operations to simplify the sentences:[2]

- Appositions: when an apposition is encountered, it is converted into a new sentence, by introducing an auxiliary verb. The clause it is attached to is copied and made the subject of the new sentence.
  Example: John Smith, a New York taxi driver, won the lottery.
  Becomes: John Smith is a New York taxi driver. John Smith won the lottery.

- Relative clauses: the wh-word is replaced with the word it refers to, and the clause is turned into a new sentence.
  Example: The mayor, who recently got a divorce, is getting married again.
  Becomes: The mayor recently got a divorce. The mayor is getting married again.

- Prefix subordination: this simplification also involves the introduction of new words, slightly based on RST.
  Example: Although it is raining, the sun is shining.
  Becomes: It is raining. But the sun is shining.

- Infix coordination and subordination: trivially, two parts of a sentence connected by 'and' are split into two sentences. If the subject of the first sentence is

also the subject of the second, the Stanford parser detects this, and the subject is duplicated. Next to *and*, two sentences conjoined by words such as *although*, *but*, *because*, ... are also split.

If a sentence can be simplified, and is split into two sentences, then we try to apply the rules again to both of the new sentences. We maintain a list of all possible combinations of rules that can be applied. Thus in this phase, we simply generate all possible simplifications of every input sentence. The actual decision of which rules to apply to which sentences is made by the method described in the next section.

3.3 Optimizing the choice of simplifications

Before starting the section on the Integer Linear Programming formulation, we will first motivate our choice of variable to optimize in order to make the text fit for a reader of a certain age. Afterwards we will extend this to a more general scenario, to incorporate more features.

Numerous features have been used in assessing the difficulty of text. One that recurs many times is the average sentence length. This feature has often been used in the traditional readability measures; the easiness with which it could be calculated probably played an important role. Still, in current research average sentence length is an important feature when training a classifier for readability assessment. It must be noted that in [24], the average sentence length feature is not significantly correlated with the readability, whereas in [11] this feature was found to be significantly different between original and simplified texts.

3.3.1 Integer Linear Programming

A Linear Programming problem consists of decision variables and an objective function, that is a linear combination of the decision variables. Solving the problem means finding an assignment for these variables, so that the objective function is maximized (or minimized). The decision variables can be bounded by linear constraints. In the case of Integer Linear Programming, the decision variables are also constrained to take only integer values. ILP has often been used to find a global solution, for example for dependency parsing [20] and multi-document summarization [12]. One

---

[2]Those that we did not choose to simplify did not occur in the data (if-then constructions), or did not have a significant effect on the readability measure used in the evaluation (activation of passive voice.)

of the first applications of ILP in Natural Language Processing was in the work of [9], whose goal is somehow similar to ours. His goal was to apply paraphrases to sentences in a text, so that the text as a whole conforms to a set of guidelines (e.g., a conference paper that can be no longer than 8 pages). The paraphrases are defined over a Synchronous Tree Adjoining Grammar (STAG). Each paraphrase has a *cost* to apply, and the goal is to make the text conform to the guidelines with a minimal cost. In contrast, in our research the objective function serves to make the text fit a certain age as good as possible. We also take it a step further, by seeing how this is related to research in readability assessment.

3.3.2 Finding a global solution

At the end of the previous step (see section 3.2), we have for every sentence a list of alternative formulations, that can replace the original sentence. For each of these alternatives, we can calculate the influence this will have on the text as a whole. Focussing on the average sentence length, the relevant features that will be influenced by each alternative are the number of sentences and the number of words.

Suppose the original text has S sentences and W words, and sentence $i, \forall i = 1 \ldots S$ has n_i possible alternatives, indicated by $a_{i1} \ldots a_{in_i}$, and a_{i0} the original sentence[3]. The a_{ij} variables can only be zero or one (a value of one meaning the corresponding alternative should be used), and for a fixed i exactly one of the a_{ij} variables must be one (there can only be one alternative chosen). We can calculate for each a_{ij} the influence this will have on the average sentence length, by calculating the difference in number of sentences, Δs_{ij}, and the difference in number of words, Δw_{ij}, compared to the original sentence. To illustrate with the example of the first rule in section 3.2: the application of this rule ($a_{10} = 0, a_{11} = 1$) would result in an increase of 1 in the number of sentences ($\Delta s_{11} = +1$), and an increase in number of words by 3 ($\Delta s_{11} = +3$).

Stating that the average sentence length should be at most m words per sentence can then be written with the formula:

$$\frac{W + \sum_{ij} a_{ij}\Delta w_{ij}}{S + \sum_{ij} a_{ij}\Delta s_{ij}} \leq m \qquad (1)$$

By rearranging, this equation can be rewritten to the following form:

$$\sum_{ij}(\Delta w_{ij} - m\Delta s_{ij})a_{ij} \leq Sm - W \qquad (2)$$

With the following constraints:

$$a_{ij} \in \{0,1\} \qquad (3)$$

$$\sum_{j=0}^{n_i} a_{ij} = 1, \forall i \qquad (4)$$

The left hand side of equation 2 can be minimized by using it as the objective function in the ILP formulation, with the constraints from equations 3 and 4. Defining a lower bound on the average sentence length can be done trivially by using equation 2 with a $\geq$ sign instead of the $\leq$ sign, in the form of another constraint. This way the average sentence length isn't made too small, and the text overly simple.

---

[3]Note that a_{ij} can consist of more than one sentence for $j > 0$.

3.3.3 Extension to more general features

A limitation of this method is that it is not possible to minimize a linear combination of averages, what would be needed for optimization towards e.g. the Flesh-Kincaid score. Because of the two averages in this formula (average sentence length and average syllables per word), the optimization problem becomes a Quadratic Programming problem, which is harder to solve.[4]

It is possible to optimize towards features that are not averages. For example, suppose that we can measure the difficulty of a text by a linear combination of the total number of sentences and the total number of words:

$$\text{difficulty} = \alpha W + \beta S$$

We can then use a similar ILP formulation as in equation 1, so that the difficulty can be minimized by choosing optimal assignments for the variables a_{ij}:

$$\alpha(W + \sum_{ij} \Delta w_{ij}a_{ij}) + \beta(S + \sum_{ij} \Delta s_{ij}a_{ij}) \leq \text{difficulty}$$

Which can be rewritten to:

$$\sum_{ij}(\alpha\Delta w_{ij} + \beta\Delta s_{ij})a_{ij} \leq \text{difficulty} - \alpha W - \beta S$$

with α and β the model parameters, originating from, for example, a linear regression model. Linear regression has been used often in predicting the reading difficulty (e.g. [11, 14]). As long as the features are defined as a total, rather than an average, it is possible to write this in the ILP formulation, and optimize for a certain difficulty. Also the statistical language modeling approach from [5] can be formulated in this way.

In the case that averages are still needed, an alternative solution would be to define upper and lower bounds on each of these features separately, e.g. by taking the average $\mu \pm$ the standard deviation σ, estimated from training data. If the resulting ILP is infeasible, i.e. it is impossible to solve, then the constraints can iteratively be relaxed to fall between $\mu \pm \gamma\sigma$, with $\gamma \geq 1$, until the ILP problem becomes feasible.

4. EVALUATION

4.1 Data

A problem with simplifying text and assessing the reading difficulty of text, is that there is no standard dataset. Because the intended audience is often different (children, students learning a foreign language, people with intellectual disabilities, . . .), or the data is protected by copyrights, finding a suitable dataset is not easy. Furthermore, for future research on the simplification of text, it would be convenient if there is a dataset that consists of an original version and a simplified version, so that the latter can be used as a gold standard.

With these objectives in mind, we used data from two publicly available sources, from two different domains. The first part comes from Wikipedia articles. We use the abstracts of the articles on the list of "100 articles every Wikipedia should have". 50 were randomly chosen for the evaluation, the remainder was used for development. Simpler versions of the articles can be found on Simple Wikipedia[5], although

---

[4]See [9] for details.
[5]http://simple.wikipedia.org

corpus	baseline (synonyms)	our method (+lang. model)
Wikipedia	53.2%	65.0%
Literacyworks	45.9%	57.6%

Table 1: Results of the lexical simplification in terms of precision

Operation	Wikipedia		Literacyworks	
Appositions	23/39	58.9%	35/72	48.6%
Relative clauses	12/20	60.0%	15/35	42.8%
Prefix subordination	2/3	0%	0/0	/
Infix coordination and subordination	30/43	69.7%	78/112	69.6%
Total	67/105	63.8%	128/219	58.5%

Table 2: Accuracy of the syntactic simplification (number correct / number that matched the rule)

Property	Wikipedia	Literacyworks
Nb. of articles	50	50
Nb. of sentences	552	1219
Nb. simplifiable	105	219
Percentage simplifiable	19.0%	18.0%

Table 3: Statistics of the used text data

the similarity between both versions of the same article is rather low.

The second part of the data comes from the Literacyworks website[6]. It contains news articles from CNN, and every article is accompanied by an abridged version. The abridged version is a simplified form of the original, which is easier to read for students and people that learn English. We randomly selected 50 articles from this set for evaluation.

So in total we have 100 articles, from two different domains.

4.2 Lexical simplification

For the evaluation of the lexical simplification, we randomly selected 180 simplifications from each domain. As a baseline, we compare with the simpler method from [8], discussed in section 2.2. In short, by using a language model we add a weak form of Word Sense Disambiguation to the baseline method, which consists of only selecting the most frequent synonym given by WordNet. The language model was trained on the Reuters corpus.

The evaluation was done using Amazon's Mechanical Turk. Each lexical substitution was graded by three persons, who were asked to indicate whether the substitution was correct or not. The majority vote was taken as the correct answer.

4.2.1 Discussion

It is clear from the results in table 1 that our method, in the third column, outperforms the baseline, shown in the second column. The latter is often too eager to replace words, where our method also looks at the context and makes less errors. This can be illustrated with the following example:

1. Authorities employ (use) various mechanisms to regulate certain behaviors in general.

2. In 2007, about one third of the world 's workers were employed (used) in agriculture.

In sentence 1, both methods replace the word *employ* by the word *use*, which is correct. But in sentence 2, the word *employ* is used in a different context, and the baseline method still replaces it, whereas our method does not.

Table 1 only shows the precision. Empirically, we noticed that the recall is rather low: the most difficult words in the texts are often not replaced. An explanation for this could be that the most difficult words don't have synonyms that are easier to understand. To give a clearer view on this matter, we decided to check how many words are replaceable by a 'simple' word. We started with a list of 3836 unique simple words: the union of the 3000 basic words from the Dale-Chall readability measure and the list of Basic English words that Simple Wikipedia recommends using[7]. For each

word, we used WordNet to retrieve the synonyms, thereby ignoring the retrieved synonyms that were already on the initial list of simple words. The total number of unique synonyms was a surprisingly low 10864. Thus, simplifying a text so that it consists entirely out of words from the list of 3836 simple words, is only possible when the input is already limited to the list of 10864 words. Words not in this latter list will not have a synonym, and can not be simplified to a word in the list of simple words. A solution would be to insert elaborations in the text, that explain the meaning of these words, or to leave out the difficult parts by using summarization techniques.

Finally, this experiment only gives an indication on the validity the substitutions, and not of the simplification. Evaluating the latter would require a more extensive evaluation, with children as test subjects (see [15] for example).

4.3 Syntactic simplification

We used the same 100 articles from the lexical simplification experiment. We also evaluated the system with Amazon's Mechanical Turk, asking the judges to indicate if the two resulting sentences[8] were still correct English. Again, we used the majority vote out of 3 opinions. To keep the answers simple, we only worked with a binary choice: correct or not correct. The results of the syntactic simplification are in table 2, and details about the data sets are in table 3. The average pairwise inter-annotator agreement was measured with the kappa statistic, and amounts to 0.7, which is reasonable to draw conclusions from.

4.3.1 Discussion

From the results in table 2 it is clear that many errors are made. A lot of the syntactic constructions that we want to simplify are also difficult to recognize for parsers. The task for the parser is made extra hard, because usually long sentences need to be simplified. A lot of the problems come from detecting the boundaries, e.g. finding the clauses that are connected by *and* or finding the end of appositions. The Stanford parser also has problems with lists, separated by commas, as in "I went to Spain, Italy, and Switzerland", in which Italy would be marked as an apposition of Spain.

---

[6]http://literacynet.org/cnnsf/

[7]http://simple.wikipedia.org/wiki/
Wikipedia:Basic_English_combined_wordlist

[8]In most cases, if a sentence could be simplified, it was by only one rule. See section 4.4 for more details.

Property	Wikipedia	Literacyworks
original avg. sentence length	21.6	17.3
minimal avg. sentence length	18.0	14.6
original Flesch-Kincaid grade level	16.2	10.8
minimal Flesch-Kincaid grade level	14.1	9.3

Table 4: The lower bound on average sentence length (words per sentence), and the Flesch-Kincaid grade level. Averaged over 50 articles per type of text

These results might be an indication that the original idea behind text simplification, as a preprocessing step *before* parsing [4, 26], could be worth revisiting. But, since a sentence does not have to be simplified, an easier solution is to analyze it with different parsers, and leave it intact if the difference between the output of the parsers is too large.

4.4 ILP evaluation

To investigate to what end we can simplify the text for a given age, we first let the ILP model make a text that is as simple as possible, by minimizing the average sentence length. These results can be found in table 4, showing the original average sentence length, and the average sentence length after the simplifications. It is clear that the average sentence length is still very high, especially for the Wikipedia articles. The results on the Literacyworks data are better, but still not good enough for the younger children.

When we also include the lexical simplification, we can calculate the Flesch-Kincaid grade level. This is defined as:

$$0.39 \frac{\#words}{\#sentences} + 11.8 \frac{\#syllables}{\#words} - 15.59$$

The result is a grade level, based on the U.S. education system. Grade 8 corresponds to age 13-14. As can be seen in table 4, the new Wikipedia articles are still far away from a level that is suited for children. The simplified news articles from Literacyworks come closer to the 8th grade, but are still not quite simple enough. That is why we will not perform further evaluation of the global result at this moment, but first put more research into the simplification operations.

For completeness, in figure 2 is a histogram representation of the number of choices that the ILP solver has for each sentence. It is clear that in most cases only one simplification operation can be applied, giving a choice between using the original sentence, or the simplified version. Sometimes an absurdly high number of alternative sentences are generated, the reason for which lies in the interpreting of comma separated lists as appositions, as discussed in section 4.3.1.

5. CONCLUSION

In this paper we have presented a set of methods to simplify text, and simplify the text so that it should better fit the age of the child reading the text. We thereby make an attempt to close the gap between predicting the difficulty of text, and the actual simplification. We improved the accuracy of the lexical simplification with a 11.7% absolute increase, by using a language model to perform a weak form of Word Sense Disambiguation. We implemented a system to split sentences based on the syntax. We relied on the out-
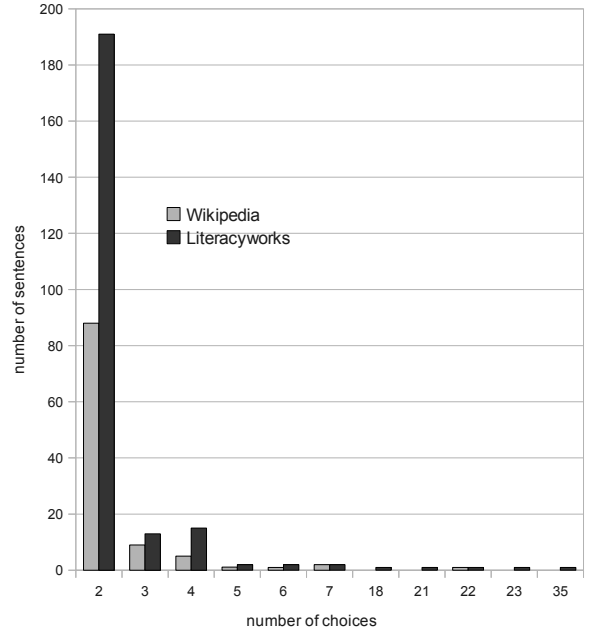
Figure 2: Histogram representation of the number of choices for each input sentence

put of a dependency parser, but it made several mistakes. It appears the constructions we want to simplify to make text more readable for children, are also difficult to understand by parsers.

On a document level, we used an Integer Linear Programming approach to find an optimal choice of simplification operations to perform on the text. We can constrain the ILP formulation to let the features of the text fall between certain boundaries, specific for the age or reading skills of the reader. Unfortunately, with the set of simplification operations we used, it was not possible to reduce the reading difficulty enough for children, at least not without removing information from the text. Simplification of the abstracts of Wikipedia articles resulted in an average decrease of around 2 grade levels according to the Flesch-Kincaid grade level formula, simplification of news articles resulted in a 1.5 grade level decrease. The lexical simplification was unable to simplify the most difficult words, mostly because there is no simple synonym for them. These result show that there are still a lot of possibilities in the field of text simplification.

6. FUTURE WORK

Ignoring the errors that were introduced in the simplification process, existing techniques to simplify text are inadequate to reach a level that is suitable for children. There are however other techniques that can be used as well.

The easiest solution is to incorporate summarization. The disadvantage is that information will be lost from the original text. But in return, it is possible to remove (parts of) sentences that contain difficult words, and to make sentences shorter to decrease e.g. the average sentence length. Care must be taken not to remove parts of the sentence that are being referred to elsewhere in the document, but with the necessary preprocessing steps this can easily be incorporated

in the ILP formulation, in the form of additional constraints.

There are also ways to make the text easier, without discarding information. It is possible to introduce elaborations for difficult words. Another option is to do the inverse of a technique in multi-document summarization, i.e. sentence fusion. In sentence fusion, two partly overlapping sentences from different documents are merged together, to create a new sentence that contains the information from both of the sentences. The inverse can be done to make text easier: split a sentence to create partly overlapping sentences, with each a part of the information (e.g. each new sentence has the same subject and verb, but different prepositional phrases).

Ideally, the method should also be able to let the children learn. Rather than making text more understandable, and not let the children learn new words, it would be better from an educational point of view if we could automatically transform the text around a difficult word so that the meaning becomes clear from the context, and the children learn something new.

Finally, we aim to further evaluate these methods, also on Web texts, and see how they can be used in an interactive setting, in order to make the Internet more accessible for children.

Acknowledgments

7. REFERENCES

[1] R. Barzilay and M. Lapata. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34, 2008.

[2] A. Candido Jr, E. Maziero, C. Gasperin, T. Pardo, L. Specia, and S. Aluisio. Supporting the adaptation of texts for poor literacy readers: a text simplification editor for Brazilian Portuguese. In *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 34–42. Association for Computational Linguistics, 2009.

[3] J. Carroll, G. Minnen, Y. Canning, S. Devlin, and J. Tait. Practical simplification of English newspaper text to assist aphasic readers. In *Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pages 7–10. Citeseer, 1998.

[4] R. Chandrasekar and B. Srinivas. Automatic induction of rules for text simplification. *Knowledge Based Systems*, 10(3):183–190, 1997.

[5] K. Collins-Thompson and J. Callan. Predicting reading difficulty with statistical language models. *Journal of the American Society for Information Science and Technology*, 56(13):1448–1462, 2005.

[6] M.-C. de Marneffe, B. MacCartney, and C. D. Manning. Generating typed dependency parses from phrase structure parses. In *In Proceedings of LREC-06*, pages 449–454, 2006.

[7] K. Deschacht and M.-F. Moens. The Latent Words Language Model. In *Proceedings of the 18th Annual Belgian-Dutch Conference on Machine Learning*, 2009.

[8] S. Devlin and J. Tait. The use of a psycholinguistic database in the simplification of text for aphasic readers. *Linguistic Databases*, pages 161–173, 1998.

[9] M. Dras. *Tree adjoining grammar and the reluctant paraphrasing of text*. PhD thesis, Macquarie University NSW 2109 Australia, 1999.

[10] C. Fellbaum. *WordNet: An electronic lexical database*. MIT press Cambridge, MA, 1998.

[11] L. Feng, N. Elhadad, and M. Huenerfauth. Cognitively motivated features for readability assessment. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 229–237. Association for Computational Linguistics, 2009.

[12] D. Gillick and B. Favre. A scalable global model for summarization. In *Proceedings of the Workshop on Integer Linear Programming for Natural Langauge Processing*, pages 10–18. Association for Computational Linguistics, 2009.

[13] M. Heilman, K. Collins-Thompson, J. Callan, and M. Eskenazi. Combining lexical and grammatical features to improve readability measures for first and second language texts. In *Proceedings of NAACL HLT*, pages 460–467, 2007.

[14] M. Heilman, K. Collins-Thompson, and M. Eskenazi. An analysis of statistical models and features for reading difficulty prediction. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, pages 71–79. Association for Computational Linguistics, 2008.

[15] M. Huenerfauth, L. Feng, and N. Elhadad. Comparing evaluation techniques for text readability software for adults with intellectual disabilities. In *Proceeding of the Eleventh International ACM SIGACCESS Conference on Computers and Accessibility*, pages 3–10. ACM, 2009.

[16] S. Jonnalagadda, L. Tari, J. Hakenberg, C. Baral, and G. Gonzalez. Towards effective sentence simplification for automatic processing of biomedical text. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, pages 177–180. Association for Computational Linguistics, 2009.

[17] B. Klebanov, K. Knight, and D. Marcu. Text simplification for information-seeking applications. *On the Move to Meaningful Internet Systems, Lecture Notes in Computer Science*, pages 735–747, 2004.

[18] H. Kučera and W. Francis. *Computational analysis of present-day American English*. University Press of New England, 1967.

[19] P. Lal and S. Ruger. Extract-based summarization with simplification. In *DUC 2002: Workshop on Text Summarization, July 11–12, 2002, Philadelphia, PA, USA*, 2002.

[20] A. Martins, N. Smith, and E. Xing. Concise integer linear programming formulations for dependency parsing. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACLIJCNLP 09), Singapore*, 2009.

[21] F. Oliveira, F. Wong, and I. Hong. Systematic

processing of long sentences in rule based Portuguese-Chinese machine translation. *Computational Linguistics and Intelligent Text Processing*, pages 417–426, 2010.

[22] S. Petersen. *Natural language processing tools for reading level assessment and text simplification for bilingual education.* PhD thesis, University of Washington, 2007.

[23] S. Petersen and M. Ostendorf. A machine learning approach to reading level assessment. *Computer Speech & Language*, 23(1):89–106, 2009.

[24] E. Pitler and A. Nenkova. Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 186–195. Association for Computational Linguistics, 2008.

[25] P. Quinlan. *The Oxford psycholinguistic database.* Oxford University Press Oxford, 1992.

[26] A. Siddharthan. Syntactic simplification and text cohesion. *Research on Language & Computation*, 4(1):77–109, 2006.

[27] D. Vickrey and D. Koller. Sentence simplification for semantic role labeling. In *Proceedings of ACL-08: HLT*, pages 344–352, Columbus, Ohio, June 2008. Association for Computational Linguistics.

[28] D. Yarowsky. One sense per collocation. In *Proceedings of the Workshop on Human Language Technology*, page 271. Association for Computational Linguistics, 1993.

Children searching information on the Internet: Performance on children's interfaces compared to Google

Hanna Jochmann-Mannak, Theo Huibers
Human Media Interaction,
University of Twente
PO Box 217, 7500 AE Enschede

{h.e.mannak, t.w.c.huibers}@ewi.utwente.nl

Leo Lentz, Ted Sanders
Utrecht Institute of Linguistics
UiL-OTS, Utrecht University
Trans 10, 3512 JK Utrecht

{l.r.lentz, t.j.m.sanders}@uu.nl

ABSTRACT

Children frequently make use of the Internet to search for information. However, research shows that children experience many problems with searching and browsing the web. The last decade numerous search environments have been developed, especially for children. Do these search interfaces support children in effective information-seeking? And do these interfaces add value to today's popular search engines, such as Google? In this explorative study, we compared children's search performance on four interfaces designed for children, with their performance on Google. We found that the children did not perform better on these interfaces than on Google. This study also uncovered several problems that children experienced with these search interfaces, which can be of use for designers of future search interfaces for children.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information search and retrieval – *Query formulation, Search process and, Selection process;* H.5.2 [**Information interfaces and presentation**]: User Interfaces - *Graphical user interfaces (GUI), Natural language, Screen design and User-centered design*

General Terms

Performance, Design, Human Factors.

Keywords

Information-seeking behaviour, searching, browsing, navigation children.

1. INTRODUCTION

Everyday, digital media play a more important role in our society and in children's lives. Besides playing games, children use digital media for social gatherings, to create their own digital space and to find information as a support for learning. While recent studies in the U.S. report that 74% of children ages 8-18 years have access to the Internet [9], in the Netherlands, practically all children are online nowadays [5].

Researchers report all kind of problems children encounter during information-seeking, because they are confronted with information systems that are designed by and for adults. Therefore, the last decade, many digital environments have been developed with a child-friendly interface, especially for children. Do these search environments really support children in effective information-seeking? And are these systems consistent with children's cognitive needs and skills?

Although research uncovered several critical problems that children experience when using an ubiquitous keyword interface, such as Google [9], almost 80% of children ages 8-12 in the Netherlands in 2008 used the Google-search engine to find information on the Internet [15]. Two years later, this percentage will most likely even be higher. Apparently, Dutch children prefer using Google as their primary source for information-seeking. What does that mean for there information-seeking behaviour on child-friendly interfaces that are developed especially for children?

In this paper, we report an explorative study on how children search information for a school assignment, by offering them closed, fact-based search tasks on four different Dutch informational websites, especially designed for children. In this study, we used children's search performance on Google as a benchmark for their performance on the search interfaces for children.

We will give an answer to the following two research questions:

1. With what type of interface do children perform the search tasks best?
2. What problems do children experience while conducting the search tasks on the interfaces and what characteristics of the interfaces do these problems relate to?

With the knowledge gained from this study, we hope to contribute to the research and development of digital search interfaces that support children in effective information-seeking.

2. RELATED RESEARCH

A general assumption is made by researchers that browsing-oriented search tools are better suited to the abilities and skills of children than keyword search tools [6]. They say that browsing imposes less cognitive load than searching, because more knowledge is needed to retrieve terms from memory than simply to recognize offered terms.

2.1 Searching versus browsing

In 1998, Schacter et al. [16] found that with both highly specific and vague search tasks, children sought information by using browsing strategies. In their research on children's internet searching on complex problems, they reported the following: "Children are reactive searchers who do not systematically plan or employ elaborated analytic search strategies." They found that the structure of the task (ill- or well-defined) played a role in children's search behaviour. The number of key word searching used in the well-defined tasks was significantly greater than the number of key word searching in the ill-defined tasks. Well-defined tasks provided a more concrete structure upon which to generate queries. However, they also found that the children performed poorly on the well-defined tasks. When the tasks were vague and abstract, children performed more successfully than when the tasks were specific and concrete. They conclude that children lack highly skilled analytic searching strategies, which are not needed when browsing for information in ill-defined tasks [16].

In the beginning of 1998, however, Google did no yet exist and browsing was a more important strategy to find information in those days. The development of popular search engines, such as Google, might have a positive effect on the development of children's analytic searching skills.

In her research on the use of the Yahooligans! Web Search Engine in 2000, Bilal [1,2] found that most of the children preferred using keywords to search for information, but that they were better at finding information by browsing. However, this result might be due to the design of the search engine on the interface that was used in this study, which did not work properly for children.

By tracking the web logs of The International Children's Digital Library (ICDL) in 2003, Druin [8] found that approximately 75% of the searches used category search (browsing), 15% used place search (by selecting a location on an interface of a globe) and just over 10% of the searches used keyword search.

Finally, in 2006, Hutchinson et al. [11] found that children are capable of using both keyword search and category browsing, but generally, they prefer and are more successful with category browsing. They explain this finding in relation to children's 'natural tendency to explore': "Young children tend not to plan out their searches, but simply react to the results they receive from the IR-system. Generally, their search strategies are not analytical and do not aim precisely at one goal. Instead, they make associations while browsing. This is a trial-and-error strategy."

2.2 Children's information-seeking problems

Researchers often find that children experience difficulties while using both searching and browsing tools. These tools do not take into account children's cognitive and motor skills.

Motor skills

Concerning motor skills, children can have difficulties using a mouse. The smaller the object to be clicked on, the longer it takes for a child to click on it [10]. Second, many children have difficulty with typing. They are not yet capable of typing without looking at the keyboard, termed touch-typing. Instead, they 'hunt and peck' on the keyboard for the correct keys [6]. This is why typing for children often takes a long time and can lead to frustration.

Difficulties with searching and browsing

Usually, formulating a search query is difficult for children, because they have little knowledge to 'recall' concepts or terms from their long-term memory [6,10]. Besides, for searching relevant documents using keyword search, correct spelling, spacing and punctuation are needed. Children often make spelling errors [6]. Deciding on a single keyword is also difficult for a child, because children tend to use a full natural language query.

With browsing, children first of all have trouble finding the right category, because they have little domain-knowledge to decide which category is optimum. In addition, problems with browsing tools are mostly the result of a lack of vocabulary knowledge. Children often have difficulties understanding abstract, top-level headings, because their vocabulary knowledge is not yet sufficient to understand such terms [11]. Children are able to use hierarchies to locate information. However, they may experience difficulty in conceptualizing abstract concepts and traversing deep multilevel hierarchical structures. The deeper the hierarchies, the more likely children are to become lost [4]. In their research on de design of web directories for children, Bilal and Wang found that children's conceptual structures (the way knowledge is organized in their minds) are more similar to each other for concrete than for abstract categories. Principles used to map the relationship among concepts are based on a concrete approach (perceptual, situational and experiential, whereas often the approach used in directories is abstract (e.g. discipline oriented) [4].

2.3 Model for web navigation

To examine children's digital information-seeking, a model is needed that simulates web navigation such as the Comprehension-based Linked Model of Deliberate Search (CoLiDeS) [12]. This model assumes that comprehension of texts and images is the core process underlying Web navigation. It is inspired by the concept of 'information scent' (semantic relevance of screen objects to users' goals) and emphasizes the semantic dimension of Web navigation; that is, it is assumed that the process of relevance assessment is central to web navigation. Information scent is measured based on three factors: semantic similarity, frequency and literal matching. Semantic similarity is calculated based on co-occurrences between words and documents with the aid of a machine learning technique called latent semantic analysis (LSA).

Juvina and van Oostendorp [12] show that not only semantic but also structural (spatial) knowledge is involved in navigating the Web. That is why they developed the model called CoLiDeS+ that uses 'information scent' to account for user's judgments of relevance (semantic dimension) and 'path adequacy' (the semantic similarity between a navigation path and a user's goal) to account for the user's efficiency in traversing a Web structure (structural dimension).

3. METHOD

In the spring of 2010, we conducted a study to explore how children search for fact-based information on several Dutch informational websites, especially designed for children. The purpose of this study is to gather both qualitative and quantitative data that can help us formulate hypotheses about children's interactions with digital search interfaces. All used methods, procedures and instruments were pilot-tested in the lab of our department before conducting the actual explorative study.

3.1 Participants

For our study, we approached a primary school in the Dutch region Utrecht, from which we knew that the Internet is an important and frequently used instrument to find information for school assignments in the classroom. We wrote a letter to the parents of 35 children from two classrooms and asked them for their consent for participation of their child. Only three of the parents did not give their consent. In total, 32 children participated in our study; 27 children from a classroom with children from fourth to sixth grade and five children from a classroom with children from first to third grade. From the children that participated, three were 8 years old, eight 9 years old, ten 10 years old, eight 11 years old and four were 12 years old. Eleven of them were boys and twenty-one of them were girls.

3.2 Data collection methods

The study was conducted by the first author of this paper (the test instructor), in a quiet room in the school during school hours. Each child participated individually and the duration of the sessions per child ranged from 30 to 45 minutes. Our data collection methods were both quantitative and qualitative. The quantitative data was collected through a questionnaire that had to be filled in by every child at the beginning of the session. Further, the task performance was measured by recording children's navigation paths during the search sessions.

Quantitative data collection

The questionnaire was a profile survey in which children were asked about demographic data, such as age, grade and gender. They were asked about their computer experience: frequency of PC use, frequency of using the Internet, activities on the internet (such as playing games, watching movies, etcetera), and frequency of online information-seeking. Further, they were asked about their prior experience with the interfaces for children that were selected for this study. Finally, they were given a free-recall task to measure their prior knowledge of the subjects that would be used in the search tasks. In this task, children had to tell what they knew about the four main subjects from the search tasks that would be given to them during task performance. For example, the child was asked: "Can you tell me what you know about sharks?" When the child stopped talking, the test instructor asked once more: "Is there more that you can tell me that you know about sharks?" Prior knowledge was not measured for the subject of the Google-task.

After performing each search task, the child was asked to evaluate the difficulty of performing the search task on that particular interface on a 'smiley-scale' with evaluations from 'very easy' to 'very difficult' (see Figure 1). At the end of every session, the test instructor asked the child to rank the websites from 1 to 5; the website that the child definitely would use the next time for information-seeking had to be ranked as 1, the one that the child would use after that had to be ranked as 2, etcetera.
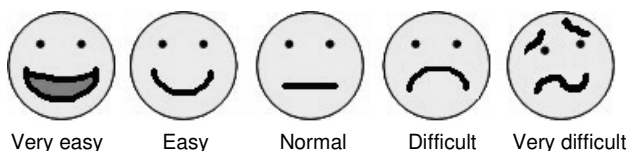

| Very easy | Easy | Normal | Difficult | Very difficult |

Figure 1. Example of the 'smiley-scale'

Qualitative data collection

We collected the qualitative data by using a structured observation method to observe children's performance on the research tasks, during which notes were made of remarkable observations. However, we did realize that children are often afraid to fail or to do something wrong. When they do not understand something, they might 'hide' this problem from the test instructor during the task performance. Therefore, we wanted to stimulate the children to express their feelings out loud during the performance. However, because we do not think that most children are very well capable of thinking aloud during their sessions [13], interventions were made during the sessions, by asking neutral questions after chunks of the task performance, termed post-task interviews.

The questions in the post-task interviews were written down in a strict protocol. In this way, we hoped to prevent that the test instructor led the children towards particular items on the websites or opinions about these items. Also, with the help of these strict protocols, we wanted to standardize the dialogues within the different sessions between the test instructor and the children. For example, the test instructor asked the children in the post-task interviews to explain their evaluations on the smiley-scales per search task and at the end of the sessions, she asked the children to provide a reason for their ranking (see Section 3.2) of the websites.

During the sessions, all browser activities were recorded and, more importantly, the children's eye movements on the screen during the task performance were recorded using the Tobii Eyetracker and the software named Studio. This eye-tracker is a free standing, non-invasive device which can be set up in front of any interface. Also, a video display of the child in front of the computer and an audio recording of the spoken comments of both the children and the test instructor, were recorded during the sessions.

3.3 Procedure

At the start of every session, the child was asked to sit behind the computer screen (Tobii eyetracker screen). Every step of the procedure was written down in a strict protocol, so that the procedure would be the same for every child. First, the test instructor explained the goal of the research session to the child and the tasks that the child would be asked to conduct on the different websites. After that, she asked the child to fill in the questionnaire about prior experience with computers and the Internet. She then took the free-recall task as described in Section 3.2.

The next step in the session was the calibration of the eye-tracker. After the calibration, the child started with the actual search tasks. Every search task within the sessions started on a very simple 'start page' with links to the five websites. Between each task, the child returned to this 'start page' with the browser's home button.

The test instructor offered the tasks to the child verbally, to prevent the children from 'typing over' keywords in stead of thinking about the formulation of the queries and the spelling of the words. For example, the test instructor asked the child: "Rembrandt was a famous Dutch painter and one of his most famous paintings is called 'De Nachtwacht'. Can you find the reason why he made this painting on the website 'willemwever.nl'?" During task performance, the test instructor

sat next to the child to reassure the child if necessary and to ask questions during the post-task interviews.

3.4 Interface selection

We selected four interfaces for our explorative study on the basis of several criteria. First of all, we wanted to conduct research with children in the Netherlands. Therefore, the websites had to be in Dutch. Second, we wanted the content of the websites to be comparable. That is, they had to represent the real information world in basic main categories, such as animals, sports, music, arts, nature, history, etcetera.

However, the selected interfaces also had to differ on several important interface dimensions:

1. Arrangement of information on the webpage
2. Type of search engine and the way in which search results are displayed
3. Menu structure: the way information is structured throughout the website
4. The amount of clutter and density

We selected the following interfaces for our study, which varied most from each other on the dimensions mentioned above, as described in Table 1.

Google was selected to function as a baseline in our study. Every child was offered the same search task to be conducted on Google, so that search performance on Google could be compared between the children and a baseline could be set for 'search skills' in our study.

Table 1. Selected interfaces and their interface characteristics

Name	Arrange-ment of infor-mation	Type of search engine / Primary search tool?	Menu structure / navigation	Clutter / density
1. School bieb.nl	Traditional	Google-like / search engine *not* primary search tool	Taxonomical / menu in words	Little clutter / low density
2. Willem wever .nl	Build of blocks	Question-answering system / primary search tool	Abstract main menu in words / basic categories on 2nd level	A lot of clutter / high density
3. Kids. kennis net.nl	Metapho-rical	Google-like / search engine *not* primary search tool	Abstract main categories / *navigation metaphor*	Little clutter / high density
4. Wiki kids.nl	Traditional, but a lot of text	Google-like, option to get direct results page / primary search tool	Abstract menu categories	Medium amount of clutter and density
5. Google	Minimalistic	Google-like / primary search tool	No possibility to browse through categories	No clutter and very low density

For the rest of this paper, we will use these labels for the selected interfaces:

1. Traditional interface *(schoolbieb.nl)*
2. Question-answering interface *(willemwever.nl)*
3. Metaphorical navigation interface *(kids.kennisnet.nl)*
4. Textual interface *(wikikids.nl)*
5. Google

3.5 Tasks

Four different search tasks were formulated for each website. The tasks were fact-based and not classroom related. Each child conducted one of the four search tasks per website, so every task per website was conducted by eight children. In every condition the websites were visited in a different order.

Every child was offered the task on Google first, to set the baseline. After that, the children were offered a task about an animal, a task about arts or music, a task about sports and at the end a task about health (see Table 2). Every task belongs to the same domain and is formulated on the same level of abstraction. Also, the amount of effort needed to conduct the tasks was the same for the different tasks per website, to make them comparable to each other within the websites. The different tasks per website should have the same effect on children's search performance on these websites, so that the nature of the tasks will not be a confounding variable in this study.

The Google-task was inspired by the complex, multi-step task of Druin et al [9] in which children had to find an answer to the question: "Which day of the week will the Vice-President's birthday be on next year?" However, because none of the children in their study was able to find an answer to that question, we decided to make the question a bit less complex. We first asked the children to the date of our Queen's birthday. Only when children could easily find an answer to that question, the child was asked to find the day of the week on which her birthday would be next year.

Table 2. Task distribution over websites

	Condition 1	Condition 2	Condition 3	Condition 4
Task 1 Birth-day Queen	Google.nl	Google.nl	Google.nl	Google.nl
Task 2 Animals	Traditional interface (kangaroo)	Question-answering interface (giraffe)	Metaphori-cal navigation (shark)	Textual interface (dolphin)
Task 3 Arts or Music	Question-answering interface (Rem-brandt)	Traditional interface (Beethoven)	Textual interface (Mozart)	Metaphori-cal navigation (van Gogh)
Task 4 Sports	Metaphori-cal navigation (hockey)	Textual interface (soccer)	Traditional interface (basketball)	Question-answering interface (gymnast-tics)
Task 5 Health	Textual interface (hay fever)	Metaphori-cal navigation (head lice)	Question-answering interface (travel-sickness)	Traditional interface (braces)

3.6 Collected data

The data collected consists of 16 hours of video and audio footage of the children's browser activities, eye movements over the screen and a video and audio display of the children in front of the screen during their research session.

For each participant, we also collected data from the profile survey, the free-recall task and the difficulty evaluations on the smiley-scales per search task. In total, we collected 96 pages of notes and comments made by the test instructor during the sessions.

4. DATA ANALYSIS METHOD

For the analysis of our data, we decided to use a top-down approach. First, we analyzed the quantitative data. After that, we analyzed the qualitative data to understand the process and outcomes of the children's search performances and to explain the outcomes of the quantitative results.

4.1 Quantitative data analysis

As mentioned in Section 3.2, we measured quantitative data before the search tasks were conducted through a questionnaire and a free-recall task. After the search tasks were conducted, children's difficulty-scores were measured for each search task and they were asked to rank the interfaces for future use.

During the search performance, we measured the following variables per search task:

1. Amount of events (clicks and submitted queries)
2. Deviation of the optimum navigation path
3. Amount of time needed to conduct the search tasks
4. Success in finding the relevant information
5. Used search strategy (searching or browsing)

We determined the optimum navigation path by counting the amount of clicks needed that brought us to the right information on the websites in the most efficient way. 'Success' was measured by judging three variables of success: the success of navigating to the information, the success of comprehending the content that the children passed along the way, and the amount of help children required from the test instructor. The calculation of the success scores is presented in the following table.

Table 3. Calculation of success scores

Success score:	8	7	6	5	4	3	2	1
Navigation	+	+	+	+	-	-	-	-
Comprehension	+	+	-	-	+	+	-	-
Required help	-	+	-	+	-	+	-	+

With these data, we could determine whether there were differences on the performance and evaluation scores between particular groups of children (e.g. age-groups), between the search tasks or between the visited interfaces. We could also check whether there were confounding variables, such as computer experience or prior subject knowledge.

4.2 Qualitative data analysis

For the qualitative data analysis, we studied the video and audio footage from the 32 sessions. We registered all positive and negative observations concerning the search performance. With 'positive observations', we mean observations of actions that led to successes in navigation or comprehension and with 'negative observations', we mean observations of actions that led to navigation or comprehension failures. These observations were related to the system characteristics (such as lay-out, navigation tools or search engine characteristics) of the interfaces.

We did not impose categories for analysis before we started the qualitative data collection. Instead of that, we developed categories inductively after all the data was collected by the test instructor, by categorizing and sorting all the positive and negative qualitative observations. We discovered that most developed categories could be assigned to one of the constituent processes in the Comprehension-based Linked Model of Deliberate Search (CoLiDeS) [12] as described in Section 2.3. This stage model to simulate navigation on the web, divides the search process in several constituent processes: attending a webpage, parsing a webpage, focusing on an area, selecting a relevant entry, etcetera. Although this model only covers a browsing strategy, we think it can also be useful for a keyword searching strategy.

5. RESULTS

In this section, we will first discuss the results from the quantitative data. After that, we will try to explain some of these outcomes by describing the most important observations within the qualitative data.

5.1 Search performance on the interfaces

Before addressing the search performance on the different interfaces in our study, we will first address whether there were differences between different groups of children. Therefore, we conducted a one-way ANOVA to compare the means on several variables for different age groups, grades and gender.

There are no significant differences between different age groups of children, concerning computer experience and Internet experience and prior knowledge of the subjects in the search tasks. There is only a difference in the frequency of Internet use between school grades ($F_{(3,29)} = 3.25$, p = .036); the higher the grade, the more the children make use of the Internet. Concerning experience, there is only one significant difference between genders ($F_{(1,31)} = 10.33$, p = .003); relatively more of the boys watch movies on the Internet.

We did not find significant differences for the performance variables between different age groups, different grades or genders. This is against our expectations, because we would expect a correlation between age and performance. We only found significant differences between the time children from different grades needed to conduct the tasks on two of the interfaces, the traditional interface ($F_{(3,28)} = 5.74$, p = .003) and the question-answering interface ($F_{(3,28)} = 8.36$, p = .000); the higher the grade, the less time the children needed to conduct the tasks on these interfaces.

On what type of interface do children perform the search tasks best?

In this section, we will give an answer to the first main question: *With what type of interface do children perform the search tasks best?*

On each website, four different tasks were conducted within the four conditions. These tasks are comparable in complexity and require the same amount of clicks within the optimum navigation path. Therefore, new performance variables could be composed by combining the performance scores from the four tasks per interface to one score for each of the interfaces. We looked at the differences in performance between the websites by calculating a 'repeated-measures ANOVA' for each of these compound variables.

The average amount of time needed to conduct the tasks is largest with the metaphorical navigation interface and smallest with Google, followed by the question-answering interface and the traditional interface (see Figure 2). The time needed to conduct the tasks on Google is significantly smaller than the time needed to conduct the tasks on the other interfaces ($F(3.23, 96.92 = 16.28, p = .042$). And the time needed to conduct the tasks on the metaphorical navigation interface is significantly larger than on the other interfaces ($F(3.23, 96.92 = 16.28, p = .000$).
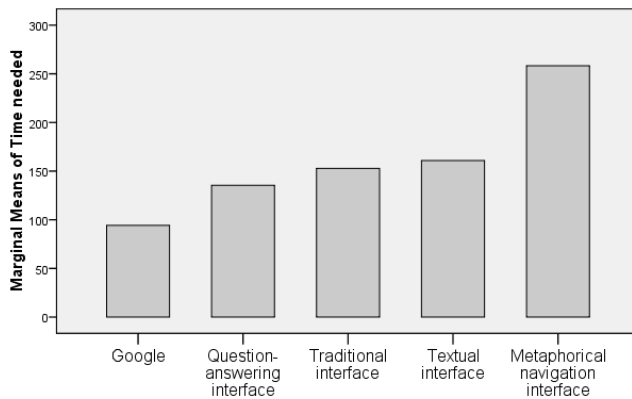

Figure 2. Results of the mean amount of time (in seconds) needed on the different interfaces

Also, the average deviation from the optimum navigation path is largest with the metaphorical navigation interface and smallest with Google and the question-answering interface. The deviation of the optimum path is significantly larger for the tasks on the metaphorical navigation interface than for the tasks on the other interfaces ($F(2.54, 76.25 = 19.34, p = .001$).

Finally, the success scores are most high on Google, followed by the question-answering interface and the traditional interface and the success scores are lowest on the metaphorical navigation interface (see Figure 3). As described in Section 4.1, success scores were rated on a scale from 1 to 8 in which 1 is the lowest success score and 8 is the highest success score. The success scores achieved for the tasks on the metaphorical navigation interface and the textual interface are significantly lower than the success scores achieved on the other interfaces ($F(2.76, 82.92) = 24.19, p = .000$).

The role of Google as a baseline for the search performance in general is very clear in the results. The children needed the least time and clicks and were most successful in conducting the task with Google, compared to the other four interfaces, as can also be seen in Figures 2 and 3.
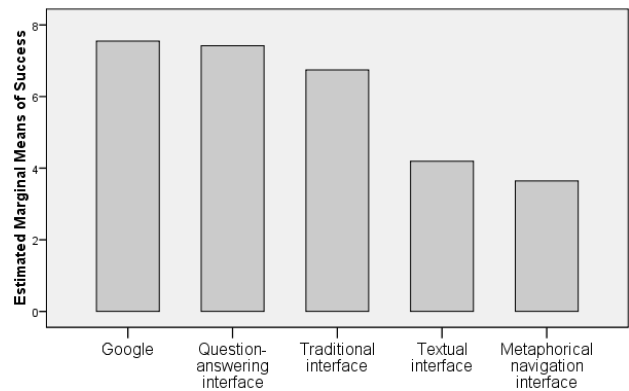
Figure 3. Results of the mean success scores on the different interfaces on a scale from 1 to 8 (1=lowest success score, 8=highest success score)

What type of interface do children prefer?
As with the performance variables, we also composed difficulty variables for each interface and calculated a repeated-measures ANOVA for these compound variables of difficulty scores. In general, the children evaluated the tasks on Google, the question-answering interface and the traditional interface as more easy than the tasks on the highly textual interface and the metaphorical navigation interface. The difficulty scores for the textual and the metaphorical navigation interface were significantly higher than the difficulty scores on the other interfaces ($F(3.22,102.98) = 23.45, p = .002$).
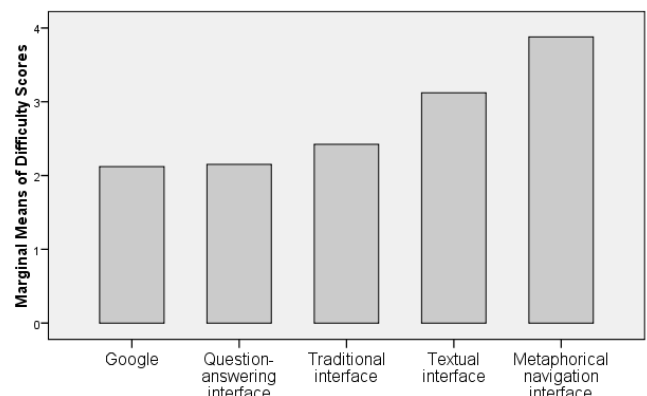
Figure 4. Results of the mean difficulty scores on the different interfaces on a scale from 1 to 5 (1=very easy, 5=very difficult)

With the ranking of the interfaces for future use, the children almost unanimously ranked the metaphorical navigation interface as least preferred for future use and the textual interface as second least preferred for future use. Google was almost unanimously chosen as the most preferred search interface for future use, followed by the traditional and the question-answering interfaces. In Section 5.2, the type of problems that the children experienced with the metaphorical navigation interface and the highly textual interface will be described, which can explain the fact that children found the tasks on these interfaces more difficult and preferred these interfaces less than the other interfaces.

What search strategies do children use?
Almost all children used the search engines as their main search strategy on the different interfaces. The only exception was the metaphorical navigation interface. Most children did not find the

search engine within the navigational metaphor on this interface, because it was 'hidden' on an unusual location on the screen.

5.2 Problems children experience with the interfaces

In this section, we will give an answer to our second research question:

What problems do children experience while conducting the search tasks on the interfaces and to what characteristics of the interfaces do these problems relate?

Parsing problems
As mentioned in the CoLiDeS model [12], after a web page is attended to, a web page is parsed in several areas and the relevant area is focused on. We recognized different forms of 'parsing' in the children's search sessions. Some children indeed looked at all areas of a web page, before focusing on a relevant area (e.g. on the main menu). Others only scanned a few items, before focusing on a particular area.

Also, many children went straight to the relevant area with their eyes, without parsing other areas of the web page. They 'shortened' the parse process, because they had clear expectations about the page arrangement and looked at the area that is conventional for that item (e.g. the conventional location for the search engine is at the top right corner of the webpage).

However, very often children experienced problems with parsing web pages. They did not see relevant items, because they 'parsed' the page too quickly or because items were placed on unexpected locations.

Mine-sweeping navigation metaphors
In his study on the usability of children's websites, Nielsen [14] found that children were willing to indulge in mine-sweeping behaviour. However, we found exactly the opposite in our study. The homepage of the metaphorical navigation interface presented a navigation metaphor in which the child sailed in a boat and could visit different islands by clicking on them. However, many children did not understand this form of navigation. With the island 'Know', for example, there were two kinds of problems. The first problem was that this label 'Know' was too abstract. Children expected to find information here about almost everything. The second problem was that the subcategory images on this 'island' did not attract attention because they had no visible words explaining their meaning. The children had to scrub the screen with the mouse to find the labels that belonged to these subcategory images. In our study, this type of navigation not only proved to be ineffective for children, but also caused a lot of frustration.

Looping navigation style
The children in our study often went back to pages they already had visited before, although they had not found the relevant information there. We also saw this 'looping' behavior while children processed search results from the search engine. Bilal and Kirby [3] reported the same results in their study on children's search behavior. They found that most children had a "loopy" navigation style. They explain that this "loopy" style can be caused by children's lower cognitive recall, because the web imposes memory overload that reduces recall during navigation.

Home as 'comfort zone'
Most children went all the way back to the homepage, when they started a new task within the same interface. Navigating to a new page from a deeper page, was often too complicated for them. Problems arose when there was no clear home button, as was the case on the textual interface. However, also when there were clear home buttons, children found it easier to use the browser's back button to go back to the homepage, which is an inefficient way to go back. This search strategy to go back a couple of times or back to the home page was already mentioned by Chen [7], who termed this strategy "going back to the comfort zone".

Failing search engines because of natural language queries
Many children used natural language when formulating their queries in a search engine, especially the younger children. With Google and the question-answering interface, using natural language did not cause any problems. However, the search engines on the other interfaces did not work well with natural language queries. The children often did not understand that the problems were caused by using a whole sentence, and tried to adjust the spelling of the words in their sentence. They did not think about bringing the query back to one keyword.

Spelling and typing
Spelling turned out to be a major obstacle for the children. Particularly with the interfaces that did not offer spelling corrections. The children frequently asked the test instructor if their spelling was correct, as Druin et al [9] also found in their study. In our study most children gratefully made use of the spelling correction tool 'Did you mean' in Google. They seemed to be very experienced using this tool. Many children immediately clicked on the spelling suggestion after the search results were displayed. Our results concerning the use of the 'Did you mean' tool are more positive than the results that Druin et al [9] found. They found that these tools were not always discovered by the children.

Typing also caused a lot of problems. Most of the children had to 'hunt & peck' for the right keys and did not notice when they made typing mistakes. Only two girls (ages 10 and 11) were able to use the touch-typing method. These results confirm the findings from Druin et al [9] that familiarity with technology still has not allowed children to become proficient at typing. However, we do think that familiarity with Google allows children to overcome problems with spelling more and more.

Query suggestions
Two of the interfaces in our study (the textual interface and Google) offered query suggestions in a drop-down box while typing a word in the search box. Although children had their eyes on the keyboard while typing, children did look at the screen while typing quite often and many children took notice and made use of the query suggestions when offered. Some children used the query suggestions when they were not sure of the right spelling by checking whether the right keyword would come up. Others even used the query suggestions as 'type help' so that they had to type only a few letters. For example, one boy only typed in the letters 'moz' on the textual interface and then clicked on the query suggestion 'mozart' that appeared in the drop-down box below.

These results are opposite to the findings of Druin et al [9]. Almost all children in their study did not notice and did not take advantage of the offered query suggestions, because of the critical

disconnect from keyboard to screen while typing queries. Our findings suggest that mere familiarity with the Google technology has allowed children to become more proficient and take more advantage of the offered query suggestions.

The output of search engines
The most important problem that the children experienced on Google, was deciding what results were relevant. Particularly, many children found it difficult to determine the appropriateness of the source of a search result. One child, for example, interpreted the outcome of a poll as a fact and used it as an answer to the search question. This problem almost did not occur on the other interfaces, which were more contained repositories and did not present results from all of the World Wide Web.

A more remarkable problem was experienced by the children on the interfaces with search results similar to Google. While these children had no trouble at all working with Google, they did have problems with the results pages of these other interfaces, because they did not recognize the results as such. They thought that the summary or snippet was all the information there was to get and did not understand they could click on the results to read more about the subject. Apparently, they did not relate the functionality of these search engines to the functionality of Google, whom they were familiar with.

This problem did not happen with the question-answering interface, because this interface presented the search results with one sentence marked clickable and with a picture in front of each result. With this format, the children knew exactly that they could click on the result for more information.

We saw the same positive effect of the use of images with categories or subcategories. When images were placed in front of subcategories (as was the case on the traditional interface), the children more easily recognized the categories as clickable and scanned the list of categories by looking at the pictures.

6. CONCLUSION
In this study, we found great differences in performance on the different interfaces. The children performed most poorly on the metaphorical navigation interface and after that their performance was poorest on the textual interface. Their performance was much better on the traditional interface and they performed best on the question-answering interface and on Google.

The most important reason for their poor performance on the metaphorical navigation interface, was that the children did not understand this type of navigation. It took a lot of time for them to understand how to navigate on this interface and especially on the navigation pages, where they had to 'mine-sweep' the screen to discover subcategories. Another reason for the low performance on this metaphorical interface, were the abstract main categories. The children had a lot of trouble selecting the right category for their search tasks from these categories.

Most children could not find the search engine on the metaphorical navigation interface. And when children did find it, most of the time it did not lead them to a relevant results page, because the search engine did not accept the natural language queries of the children.

Performance on the textual interface also turned out to be quite low. The main problem was that the children found it hard to parse the high textual homepage to locate relevant items. Children

also experienced a lot of problems on this interface, because they could not go back 'home' easily. They did not know that the logo was also the home-button. Furthermore, children experienced a lot of problems with processing the search results. Often, they did not recognize the search results as clickable. Although this interface contained a very smart feature by directly presenting a relevant search results page after submitting a query, this feature did not work for most children, because they entered natural language queries or made spelling errors.

The traditional interface resulted in much better performance. Although most children used the search engine on this interface, some children could also browse quite easily through the menu structure on this interface. The pictures used in front of the sub menu worked quite well for the children. However, some of them experienced problems with the search results, because they did not recognize the results as clickable.

Of all four children's interfaces, the question-answering interface resulted in the best performance. Children immediately saw that they could submit a question at the top of the screen and the search engine could handle natural language queries quite well. The children also recognized the search results immediately as clickable, because there was only one sentence presented for each result and there were pictures in front of each result.

6.1 Search strategies
In contrast to previous research on children's search strategies, we found that children used more searching than browsing strategies while performing informational search tasks. They also preferred using a search engine rather than browsing the main categories.

Druin et al [9] described the same development of children using search engines. They uncovered several critical problems that children experience using search engines, such as problems with spelling and typing. Tools designed to make searching easier for children went unnoticed by the children in this study.

However, our findings are much more positive concerning children's search performance on Google. In our study, the children did take advantage of the tools, such as the spelling correction tool 'Did you mean' and the query suggestion tool that appeared in a drop-down box while typing a query.

The only problems children experienced with Google, concerned judging the relevance of search results for their search task. This problem did not occur on the other interfaces, because of their smaller, more contained content.

7. DISCUSSION AND FUTURE WORK
The results of this study raise all sorts of hypotheses about children's search behaviour on digital search interfaces, on which we can base future studies. Considering future search interfaces for children, we can suggest some design directions based on the results of this study.

First of all, the effect that Google has on today's children should not be underestimated. We should keep in mind that experience with search systems and search conventions that arise from these experiences, cause changes in children's search behaviour and strategies over time.

Designers should be careful with well-meant, child-friendly designs, because they might not work for children. An important example of such an interface is the navigation metaphor in which

children had to mine-sweep to find subcategories. In this study, we found that it is not easy to design a search interface for children that adds value to searching with Google.

We did find some directions in this study to add value to Google for children, such as adding pictures to search results or categories. Also, making search results as simple as possible (e.g. with one sentence) and making clear that the results are clickable, supports children in effective information-seeking.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] Bilal, D. Children's Use of the Yahooligans! Web Search Engine. III. Cognitive and Physical Behaviors on Fully Self-Generated Search Tasks. Journal of the American Society for information science and technology, 53, 13 (2002), 1170-1183; 1170.

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

[3] Bilal, D. and Kirby, J. Differences and similarities in information seeking: children and adults as Web users. Information processing and management, 38, 5 (25 October 2002), 649-670; 649.

[4] Bilal, D. and Wang, P. Children's conceptual structures of science categories and the design of Web directories. Journal of the American Society for Information Science and Technology, 56, 12 (2005), 1303-1313; 1303. DOI=10.1002/asi.20216.

[5] Boot, L., Borgdorff, M., Pardoen, J. and Pijpers, R. Mijn kind op internet 9-12 jaar. Stichting Mijn Kind Online, Den Haag, 2010.

[6] Borgman, C. L., Hirsh, S. G., Walter, V. A. and Gallagher, A. L. Children's Searching behavior on browsing and keyword online catalogs: The Science Library Catalog Project. Journal of the American Society for information Science, 46, 9 (January 31, 1995 1995), 663-684; 663.

[7] Chen, S. L. Searching the Online Catalog and the World Wide Web. Journal of Educational Media & Library Sciences, 41, 1 (2003), 29-43; 29.

[8] Druin, A. What children can teach us: developing digital libraries for children with children. The library quarterly, 75, 1 (2003), 20-41; 20.

[9] Druin, A., Hutchinson, H., Foss, E., Hatley, L., Golub, E., Leigh Guha, M. and Fails, J. How children search the internet with keyword interfaces. In Anonymous 8th International Conference on Interaction Design and Children. (Como, Italy,). ACM, , 2009, 89-96.

[10] Hutchinson, H., Druin, A., Bederson, B. B., Reuter, K., Rose, A. and Weeks, A. C. How do I find blue books about dogs? The errors and frustrations of young digital library users. In Anonymous International Conference on Human-Computer Interaction (HCI). (Las Vegas, Nevada, USA). Lawrence Erlbaum Associates, Inc., 2005.

[11] Hutchinson, H. B., Bederson, B. B. and Druin, A. The evolution of the international children's digital library searching and browsing interface. In Anonymous The 2006 conference on Interaction design for children. (Tampere, Finland). ACM Press, , 2006, 105-112.

[12] Juvina, I. and Oostendorp, H. v. Modeling Semantic and Structural Knowledge in Web Navigation. Discourse Processes, 45(2008), 346-364.

[13] Markopoulos, P., Read, J., MacFarlane, S. and Hoysniemi, J. Evaluating Children's Interactive Products. Elsevier, Burlington, 2008.

[14] Nielsen, J. and Gilutz, S. Usability of Websites for Children: 70 design guidelines based on usability studies with kids. Nielsen Norman Group, 2002.

[15] Pijpers, R., Marteijn, T., Bosman, M., Berg, v. d.,W. and Dijkerman, E. Klik en klaar. Een onderzoek naar surfgedrag en usability bij kinderen. Stichting Mijn Kind Online, Den Haag, 2008.

[16] Schacter, J., Chung, G. K. W. K. and Dorr, A. Children's Internet Searching on Complex Problems: Performance and Process Analyses. Journal of the American Society for information science, 49, 9 (September 11, 1997 1998), 840-849; 840.

WebMark: A Rapid Internet Browsing Technique for Visually Impaired Web Users

Hesham M. Kamel

Faculty of Information Technology (FIT)
UAE University, UAE
+001(510) 333-1103
hesham{@cal.berkeley.edu, @uaeu.ac.ae}

ABSTRACT

The continuous growth in the size and use of the World Wide Web (WWW) imposes new methods of design and customization of internet browsers. Personalization of Internet Access (PIA) as we define it is the process of customizing a Web navigational method to the needs of each specific user or set of users. In this paper we present WebMark, a PIA technique that makes use of the internet data content, user's personal preference and Internet Explorer (IE). This user-control method provides visually impaired users with a faster navigational procedure to easily search and access internet content via voice synthesis feedback.

WebMark allows blind users, while browsing the internet, to randomly mark desired Web pages temporarily without using the browser's feature of bookmarking them. This method gives the users the flexibility to quickly revisit and access these marked pages during the same browsing session. Consequently, the user can spend less time browsing more information by skipping a large number of undesired previously visited pages. In addition, WebMark gives the user the flexibility to mark links for faster navigation to information on Web pages that contain a large number of links.

KEYWORDS

WWW; visually impaired; search; WebMark; blind; hypermedia design; accessibility; Internet.

INTRODUCTION

Navigation of complex hypermedia environments, of which the web is the most apparent example, has long been considered a major issue in hypermedia design and usability literature [1]. The potentially complex navigation task is further complicated if the user happens to be visually impaired, since the richness of visual navigational cues presented to a sighted user are not appropriate nor accessible to a visually impaired user [2]. Individuals with physical disabilities, at times, are simply overlooked by fully capable people or their needs are addressed as an afterthought especially in the information Technology field [11, 12]. Efforts have focused on supporting the 'sensory translation' of visual textual content to either audio or touch (through Braille), rather than enhancing web navigational methods [3].

The lack of proper organization in web browsers forces Web users to make random choices in storing bookmarks while browsing [22]. In addition, the history list of pages visited that is stored in the browser cache could become unreasonably long in a short period of time. This feature is not too practical for page relocation access, since it forces the visually impaired user to rely on recalling the title of the desired page, a process that violates the Heuristic concept of 'recognition rather than recall [16].

The problem arises when a visually impaired person browses a large number of web pages, a course that is inevitable while the user surfs the Internet. A finite number of them may turn out to be the focus of interest, which they might not be consecutive. The task of singling out a preferred page becomes time consuming, once the user is concerned with identifying these pages of interest amongst a large number of undesired previously browsed pages. Giving this scenario, it is impractical to expect blind users recognize visual cues on the pages (e.g., colors, images or page layout). These cues, on the other hand, allow sighted users to quickly and easily scan pages to locate the information of interest. Studies show that the use of visual aesthetics can transform a wall of dry text into a presentation that users will enthusiastically approach [17, 18].

This problem of having visually impaired users revisit desired and non-desired pages becomes inefficient, while trying to relocate a previously visited page. Unfortunately, special browsing features that are built in the most popular screen-readers (e.g., Windoweyes [19] and Jaws [7]) do not provide the blind users the means to separate unwanted information from a desired one.

In this paper, we introduce the concept of WebMark, a Personalization of Internet Access (PIA) technique, which overcomes some of the problems described above. Through WebMark, we present a method to assist visually impaired web users with navigating the internet in a faster manner. WebMark is one of the main

features of a talking browser called "WebSight" that conveys layout of Web pages. The WebMark feature gives blind users the ability to dynamically select web pages by marking Internet content (e.g., pages and links) while browsing the net to easily relocate them later in the same browsing session.

In the following sections, we first cover some related work of Internet accessibility techniques and blindness. Next, we briefly introduce The WebSight project entailing the WebMark mechanism in more details. We follow this section by a description of the WebMark complete usability method via examples. Next, we cover future work then draw our conclusion.

RELATED WORK

Visually impaired people have difficulties accessing the Web, either because of the inappropriately designed Web pages or the lack of total accessibility provided by the current available technologies. This lack of accessibility leads to poor web navigational support for visually impaired users. Blind individuals usually access the Web, by using screen readers [7] or specialized browsers [9]. For these access technologies to work properly, Web pages must be appropriately designed as well as encoded in valid HTML that conforms to various accessibility guidelines. Some progress has been made in accessible computing since 1992 (such as increased research and publication or user empowerment in the form of increased user participation in Information Technology (IT) design and development) [13]. Moreover, there remains an over-arching bias that is evident in the lag of accessible technology behind this barrier, which is designed for fully capable users [14]. The W3C Web Accessibility Initiative and others recognize these difficulties and provide guidelines to promote accessibility on the Web [9, 10].

Harper [6] introduced the notion of travel and mobility to improve the accessibility of Web pages for visually impaired. He drew the analogy between virtual travel and travel in the physical world. Where, the notion of travel extends navigation and orientation to include environment, mobility and the purpose of the journey. Navigation suggests an opportunity of movement within the local environment. Orientation is the knowledge of the basic spatial relationships between objects within the environment, and the objects and the traveler. Mobility is the ease and condense at which travel can be accomplished. Environment is the context in which the traveler journeys through and includes the way the landscape rendered and perceived [9]. Travel objects are environmental elements that are used during a journey. On the Web, these objects deliver by the page design and the browser.

Yesilada [20] has presented a travel analysis framework and demonstrated that the travel analysis process in the frame-work can be applied automatically as well as manually. This framework based upon a model of real world travel. It is used for identifying travel objects on Web pages and classifying them by concerning the roles of travel objects in the model of real world travel.

Eirinaki [21] introduces SEWeP, a web personalization system. Eirinaki associates Web usage and content knowledge, by enhancing the information in the Web usage logs with semantics derived from the content of the Web site's pages. The enhanced Web logs, called C-Logs are then used as input to the Web mining process, resulting in the creation of a broader set of recommendation.

WEBMARK NAVIGATIONAL ACCESS AND WEBSIGHT QUICK VIEW

In this section, we briefly introduce the WebSight system approach, which overcomes the layout barrier of Web pages. WebSight integrates WebMark as one of its main features. The WebSight project is a Microsoft Internet Explorer (IE) plugin. Using the Microsoft Speech API, the WebSight prototype provides visually-impaired individuals a method of conceptualizing the layout of a given webpage. Using this method, WebSight builds its own data structure of the HTML Code, preserving the unique spatial orientation of on-screen data. Then, it renders the content in the same fashion as sighted individuals would perceive it via voice synthesis.

The interface is composed of a 3 by 3 grid [4, 5]. The cells of this grid are labeled in the same fashion as a telephone keypad, where the top-left cell is 1 and the bottom-right cell is 9. As the user navigates from one cell to another, WebSight renders each content of that cell (e.g., a link, an image etc.) along with its spatial orientation in relation to the surrounding content (see Figure 1).

As the user moves from one link to another, WebSight announces which cell the new link is currently in and the label of that link and its relative position to the previous link. For example, using predefined navigational keys, if the user moves the mouse cursor from "Computer Science" (in cell 5) to "Electrical Engineering and Computer Science" (in cell 7), WebSight will announce "Cell 7, Down Left, Link Electrical Engineering and Computer Science" (see Figure 1). The WebSight Project has other vital features and functionalities that is under consideration for publication with complete usability testing and results.
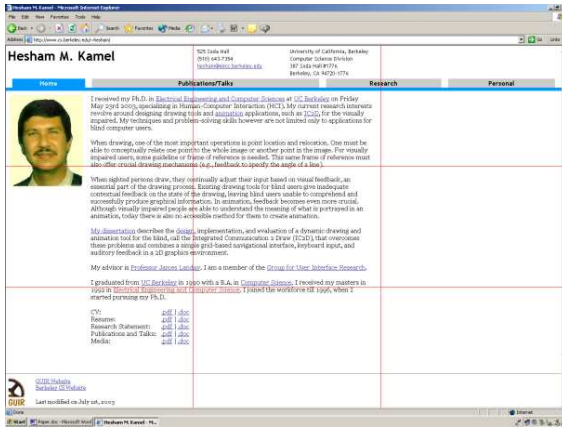
Figure 1. Spatial rendering of web information by WebSight

The mechanism of WebMark

Utilizing this technique of WebMark, the blind user while browsing the internet, has the ability to mark random Web pages and different links on any given page and then access these information at a later time during the same session. The user also has the choice to store unlimited marked items in the WebMark temporary storage data structure.

Key Function "*p*" provides information regarding the marked pages and links, "*alt-p*" marks the current page, "*ctrl-p*" goes to the next marked page, "*ctrl-shift-p*" goes to the previously marked page, "*alt-l*" marks current link, "*ctrl-l*" goes to the next marked link, and "*ctrl-shift-l*" goes to the previously marked link (see Table 1).

Two cursor navigational techniques

WebMark (as in the WebSight project) has two different cursors: a virtual and a mouse cursor. The mouse cursor assists the blind user with navigating and hearing the rendering of the on-screen content including identification of the marked items. Whereas the virtual cursor provides the user with a method to easily access the desired content (e.g., marked Web pages and links) the user can navigate the mouse cursor by The WebSight function keys. The user can navigate the WebMark cursor by the function keys listed below. Table 1 reflects the main function keys that are employed in WebMark where users are able to mark/navigate pages and links.

| Key	Function
P	Provides information of marked pages and links
alt-p	Marks current page
ctrl-p	Go to next marked page
ctrl-shift-p	Go to previous marked page
alt-l	Marks current link
ctrl-l	Go to next marked link
ctrl-shift-l	Go to previous marked link

Table 1: WebMark function keys

Marking web pages

The user presses "*alt-p*" to mark a page. WebMark announces "page marked" followed by the HTML title of the page. When a page is marked, WebSight captures the title of the page as set using the <TITLE> HTML tag. The user can navigate to a previously marked page, or to a next marked page (e.g., pressing *ctrl-shift-*p or *ctrl-p)*, WebMark announces "previously marked page" and "next marked page" respectively. WebSight then loads in the given page in the active window.

WebMark logic for marking pages

A user can mark 10 pages and then navigate the virtual cursor back an X number of pages (where X <= 10) by pressing 'ctrl-shift-p'. If the user adds a new marked page by pressing "*alt-p*" at that location, the marked page is then inserted at the end of the list and not between the position X and X+1. Therefore, in the data structure, the mouse cursor is at the 11th marked page (see figure 2).
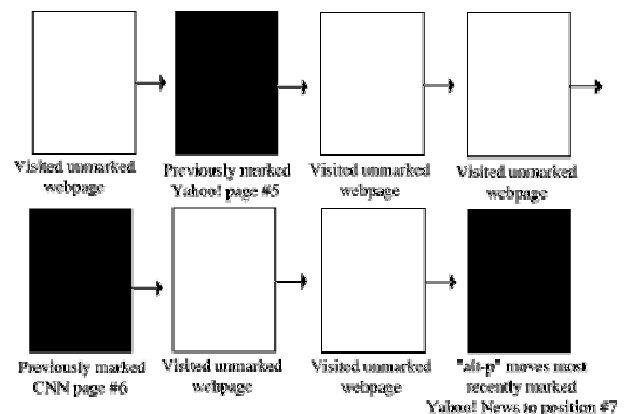

Figure 2: Marked links are added sequentially at the end of the list.

This logic support for adding a new marked page at the end of the list is considered, because if Yahoo happens to be our 5th marked page, and the user navigates to a marked paged before Yahoo (e.g., the 3rd marked page) then we try to add a new marked page. If the user recalls that Yahoo was the 5th marked page, then the insertion after the 3rd marked page will increase the position of all pages following the 3rd page by 1, including Yahoo. Thereby, confusing the user's conceptual model of the marked page list, since moving the cursor forward twice from position 3 will not reach the Yahoo WebMarked page. For example, the user navigates back to the 5th marked page (i.e., moving the cursor back to the Yahoo! WebMarked page), where the 6th WebMark is CNN. If the cursor is on the Yahoo! WebMark and the user clicks on Yahoo! News. The system adds the new WebMarked news page to the

marked page list. However, it will be added to the end of the WebMark list instead of being the 6th marked page for the reason previously mentioned (see Figure 2).

The user has 10 marked pages where the 5th marked page is Yahoo! Next, the user navigates back to Yahoo! via the WebMark page list. If on this page, the user navigates to Yahoo! News via a link on Yahoo! And the user marks this page, the position of the Yahoo! News page moves to position 11, where the mouse cursor is located. Therefore, if the user presses the previous marked page function key "ctrl-shift-p", the virtual cursor will navigate to the 10th marked page. If the user goes back manually to a marked page (using the location bar or a bookmark), the position of the mouse cursor in the marked page list still remains where it was previously.

When the mouse cursor is at CNN, the 10th marked page, and the user navigates back to Yahoo, at position 5, via the browser back button, the cursor will move to position 5 where the virtual cursor still remains. If the user tries to go back to the previously marked page using the "ctrl-shift-p", they will be going back from CNN at position 10 to the previously marked page at position 9 (see Figure 3).
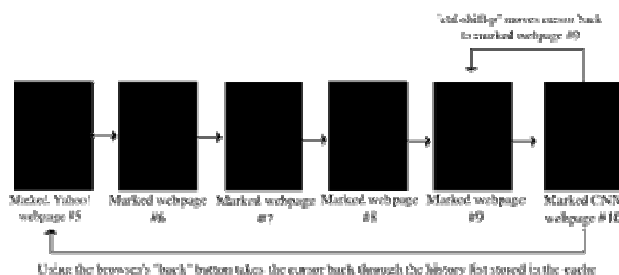
Figure 3: Use of the browser "back" button vs. the WebMark technique

Marking links
To mark a link, the user presses 'alt-l' then WebMark announces "Link Marked". To navigate to a previously marked link or to a next marked link, the user presses "ctrl-shift-l" or "ctrl-l" then WebMark announces "Previous marked link" and "Next marked link" respectively.

WebMark logic for marking links using two cursors
Links 1, 3, and 5 have been marked and the **mouse cursor** is currently at the 7th link on the page. The user then navigates backwards to the beginning of the marked linked list by pressing "ctrl-shift-l", moving the **virtual cursor** to link 1, with the goal of adding link 2. If the user then presses the key for the next link, and tries to mark that link, the link marked is 7, not 2, because the mouse cursor is at the 7th link. We made the mouse cursor not to follow the virtual cursor. This is done because of the following scenario. Suppose the

mouse cursor was at the 100th link and links 1, 3, and 5 were marked and the user navigates back to link 3 using the 'ctrl-shift-3' key and we had the mouse cursor move with the virtual cursor, then the user would lose the position of the 100th link. As a result, the only way for the user to return to the 100th link, is to navigate forward manually through all the links.

When a user marks a link, the link is placed at the end of the marked link list, regardless of where the link is on the page.

Adding links at the end of the stack
We provide this functionality of adding links at the end of the marked list of links, since there are many popular pages with hundreds of links (e.g., the CNN website). The user has the ability to browse through the page and mark the links he/she is interested in. It is not vital for the user to sequentially go through the page more than once to single out and WebMark the links they are interested in. When sighted users visit a link, the color of the link usually changes from blue to red. However, here the user does not have to visit a link to have it marked (or selected).

FUTURE WORK
We would like to conduct a full user study of the WebMark mechanism. Some significant dependent variables to measure are: a specific marked page retrieval time, recognition of marked links on a giving page and the contribution of the marked list (WebMark) to the overall conceptual model of the system. We also like to discover the effect on the user's mental model by giving him/her the ability to insert the new marked page at the virtual cursor location in the marked page list, instead of adding it at the end of the list. Also, announcing the current WebMark number in addition to the total page number in the marked list is worth implementing for ease of navigation. Furthermore, it is worth investigating the implementation of feedback conveying the number of marked links and its affect on the user's mental model.
An issue to consider in the logical design would be to check if the user browses unmarked pages, leaving the marked page list, will it be better to:
- Return the user to the same WebMark location in the marked page list
- Take the user's cursor to the end of the marked page list
- Put the user's cursor at the beginning of the marked page list.

CONCLUSION
In this research we have presented WebMark, a marking technique of Web pages and links that separates desired from non-desired data. This allows the user to concentrate on the content rather than how to arrive at it. Moreover, WebMark entails a technique that provides personalized feedback, by transforming a collection of hideously linked pages or a large set of

links into a manageable finite set of data. The feedback is provided in an appropriate manner, giving the user explicit orientation information, such that navigational information can be detected in a timely manner. In other words, a user can make a choice as to whether they want to be at the current location, and if not, how to best attempt to get to their perceived destination [2]. When there are too many bookmarked pages, retrieval becomes time consuming for a visually impaired individual. In addition, the history list of pages visited that is stored in the browser cache can become too long in a short period of time.

WebMark gives blind users the ability to dynamically select and mark web content instead of using the page bookmarking feature of the browser. This browser feature is not too practical for most blind users, since the user would have to remember the name of the desired page he/she wants to revisit. It is a non-trivial task when a blind user tries to remember each title of the desired pages while he/she navigates the Internet. To visualize such an operation with its detailed steps could become an impossible mission. Given the technique of WebMark, our design intention is to provide a method for blind users to overcome the classical problem of "recall over recognition". This procedure can assist the visually impaired user to avoid navigational delays and data-access obstacles while browsing the Internet.

The method of marking pages while browsing the Internet, and traversing them at a later time, leaves out much of the undesired pages to step through. Moreover, it speeds the process of browsing by not relying on feedback of visual content. In WebMark, we presented a simple navigational method and a set of effortless functionalities to allow users mark their desired Internet data. In general, the more the user visits Internet pages, the more the mental model of the user can become a Web!

ACKNOWLEDGMENTS

REFERENCES

1. McKnight C, Dillon A and Richardson J. Hypertext in Context. Cambridge University Press (1991).

2. Petrie H., Morley S., Majoe D. Initial Design and Evaluation of an Interface to Hypermedia Systems for Blind Users. In Proc. of the Eighth International ACM Conference on Hypertext and Hypermedia 1997, 48-56.

3. Jones S. and Cockburn A. A Study of Navigational Support by Two World Wide Web Browsing Applications. In Proc. ACM Conference on Hypertext and Hypermedia 1996, 161-169.

4. Kamel, H.M. and Landay, J.A. The Integrated Communication 2 Draw (IC2D): A Drawing Program for the Visually Impaired. In Extended Abstracts of the ACM Conference on Human Factors in Computing Systems 1999, 222-223.

5. Kamel, H.M. and Landay, J.A. Sketching Images Eyes-free: A Grid-based Dynamic Drawing Tool for the Blind. In Proc. ACM Conference on Assistive Technology 2002, 33-40.

6. S. Harper. Web Mobility for Visually Impaired Surfers. PhD thesis, The University of Manchester, 2001.

7. Freedom Scientific, JAWS for Windows. http://www.freedomscientific.com/fs_products/software_jaws.asp, last accessed January 2010.

8. C. Asakawa and T. Itoh. User interface of a home page reader. In The Third Annual ACM Conference on Assistive Technologies ASSETS'98, pages 149–156, 1998.

9. K. P. Coyne and J. Nielsen. Beyond alt text: Making the web easy to use for users with disabilities. Nielson Norman Group, October 2001.

10. J. Gunderson and I. Jacobs. User agent accessibility guidelines 1.0. World Wide Web Consortium (W3C), 1999.

11. MULLER, M.J., WHARTON, C., MCIVER, J., WILLIAM J. AND LAUX, L. 1997. Toward an HCI research and practice agenda based on human needs and social responsibility. In CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems ACM, New York, NY, USA, 155-161.

12. GLINERT, E.P. AND YORK, B.W. 1992. Computers and people with disabilities. Commun. ACM 35, 32--35.

13. LADNER, R.E. 2008. Access and Empowerment: Commentary on ``Computers and People with Disabilities". ACM Trans. Access. Comput. 1, 1-5.

14. EDWARDS, A.D.N. 2008. Keeping Up with Technology: Commentary on ``Computers and People with Disabilities". ACM Trans. Access. Comput. 1, 1-5.

15. A. FRANZ. Searching the Web by Voice

16. www.useit.com/papers/heuristic/heuristic_list.html, last accessed on June 1 2010.

17. Faraday, P. and Sutcliffe, A. Making contact points between text and images. In Proceedings of the 6th ACM International Conference on Multimedia (MULTIMEDIA'98). ACM, New York, 29–37 1

18. Hoffos, S. and Sharpless, G. The CD-I Designers Guide. McGraw-Hill, New York 1991.

19. GW Micro, Windoweyes, http://www.gwmicro.com/Window-Eyes/, last accessed on June 2010.

20. Y. Yesilada, R. Stevens, and C. Goble, "A Foundation for Tool Based Mobility Support for Visually Impaired Web Users," presented at Proc. of the Twelfth International World Wide Web Conference (WWW 2003) C Budapest, Hungary, 2003.

21. Eirinaki, M., Vazirgiannis, M., and Varlamis, I. 2003. SEWeP: using site semantics and a taxonomy to enhance the Web personalization process. In *Proceedings of the Ninth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining* (Washington, D.C., August 24 - 27, 2003). KDD '03. ACM, New York, NY, 99-108.

22. Mathur, P. and Karahalios, K. 2009. Using bookmark visualizations for self-reflection and navigation. In *Proceedings of the 27th international Conference Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 4657-4662.

The Mediated Information Needs of Children on the Autism Spectrum Disorder (ASD)

Dania Bilal

School of Information Sciences, The University of Tennessee, 1345 Circle Park COM 451, Knoxville, TN 37996. Phone: (865) 974-3689. Email: dania@utk.edu

ABSTRACT

This paper presents the preliminary results of a pilot study that explored the digital information landscape of children on the autism spectrum disorder (ASD) from the mediators' perspectives (i.e., parents, teachers, therapists, and family volunteers). A web-based survey questionnaire generated responses from seventeen mediators that were mostly parents of children with ASD. Quantitative and qualitative data analyses uncovered many characteristics of this landscape. Parents were information seeking proxies for their children with ASD. They used a variety of information retrieval systems (IRs) including online databases; web search engines; social media (e.g., blogs, YouTube, Facebook); and toys and games in support of their children's everyday life and academic needs. Parent assessed the children's skills in using IRs and other technologies, and rated their skills on the six steps of the research process identified by the American Association of School Librarians (AASL), among other set of skills, as essential for becoming information literate to meet the challenges of the 21st century information environment (see Appendix). Most parents elucidated interface design features that could support the children's information seeking goals. The results have implications for designing interfaces that meet the needs of young users with ASD, as well as for providing user-centered training to teach them effective information access, retrieval, and use.

Topics
Non-interactive evaluation

Keywords
Children, Autism Spectrum Disorder (ASD), Proxy information seeking, Information needs

INTRODUCTION
People with special needs, especially children on the autistic spectrum disorders (ASD) may possess "unique" information

behavior and express different information needs from children without ASD. In December 2009, the Center for Disease Control (CDC) in the United States reported that 1 in every 110 births and almost 1 in 70 boys have been diagnosed with autism, a neurological disability that affects a person's ability to communicate and interact with others (http://www.cdc.gov). The significant increase in cases of young people with ASD is also prevalent in other countries including Europe [1]. Autism appears in early childhood, usually before the age of three. It prevents children and adolescents from interacting normally with other people and affects almost every aspect of their social and psychological development.

Children are "unique" young users and not simply 'short' adults. They possess information needs, experiences, sense-making, cognitive knowledge structure, skills and knowledge, as well as a theory of mind that are very different from those of adults [2][3][4][5]. The theory of mind refers to a full range of mental states (beliefs, emotions, constructs, desires, imagination, etc.) that is normally developed in children as young as age three. Children with a theory of mind have understanding of the relationship between individuals' mental states and their overt actions [5]. While in the last two decades, we have developed adequate understanding of how "normal" children interact with and use a variety of IRs and tools in the digital environment, we lack much understanding of the information behavior of children with ASD. This paper is a first step towards filling this research gap by exploring the digital information landscape (access, needs, and use) of children with ASD from the perspectives and experiences of the mediators that communicate/interact with them in everyday life and/or on a regular basis (parents, teachers, therapists, etc.). Results gained from this study have implications for universal system interface design that is supportive of these young users. The results could also yield to developing user-centered programs to teach children with ASD effective use of IRs and other information technology tools and applications.

1.1 Autism Spectrum Disorder
Autism is one of the developmental disabilities disorders called Autistic Spectrum Disorder (ASD) or Pervasive Developmental Disorder (PDD). Autism originated from the Latin root "auto", meaning "self," and was coined by the Swiss psychiatrist, Eugene Bleuler, to describe the self-centered thinking of adult patients he observed and who believed that everything in the world focused on them. Kanner (1943) borrowed the term to describe the children he worked with and who shared similar characteristics with Bleuler's patients (i,e. repetitive behavior, emotional isolation, detachment from others including their parents, self-centricity, and attachment to objects such as toys). In 1944, Hans Asperger, an Austrian psychiatrist who worked with young adults, identified similar problems in his patients; however, he also found that the patients had normal intelligence and

began talking on time, unlike what Kanner concluded ([6]. Although the severity of the diagnosis varies, people with ASD share three common impairments: Social interaction, verbal and non verbal communication, and limited interest in activities and play (http://www.webmd.com/brain/autism/autism). In the areas of learning, people with ASD may be gifted/savant in specific subjects or severely challenged. Gifted people are called "high functioning" or Asperger's as they are verbal and possess a high intelligence (IQ). The traits children and young people with ASD possess pose many challenges for researchers, especially in relation to understanding their deficits and gifts, communicating with them to uncover their information behavior and needs, and to identify design criteria and architectures for interfaces that are most appropriate for them and meet their everyday life information needs.

1.2 Theoretical Framework

One of the main goals of human-computer interaction professionals is to develop understanding of differences among users in order to meet their needs, accommodate diversity, and design interfaces that are universally usable [7][8]. Addressing the needs of all users should be an ultimate goal for interface designers [9]. In general, designing interfaces for young people requires that the designer pay attention to their limitations in terms of cognitive developmental abilities (e.g., limited vocabulary, low memory recall, minimal level of abstraction, and experiential learning that vary with age)[10][11]. Designing for children with ASD pose additional challenges due to their "unique" cognitive challenges and difficulties that some researchers in the field of psychology have attributed to "deficits" in the theory of mind [5][1]. In the field of information sciences very little research has paid attention to the information landscape of this population with special needs and their information seeking in everyday life. The theoretical framework of Savolainen's everyday life information seeking (ELIS)[12] is also relevant to this research because it provides a natural dimension about the realities of these children's information world that is restricted by the "cognitive disorder" of ASD, which necessitates intervention from various mediators. In addition, the notion of "information seeking proxy" defined Fisher & Abrahamson [13] as "people who seek information on behalf of others without necessarily being asked or engaging in follow-up with the recipient(s)" is partially relevant here. While parents of children with ASD serve as proxy information seekers for their children with ASD, they, in many instances, engage in follow-ups with their children to solve their everyday information problems. This study explored the nature of the digital information landscape of a group of children with ASD from the mediators' perspectives with a focus on parents as information seeking proxies and as advocates for designing interfaces that are universally usable and that highly supportive of their children's information needs.

2. RESEARCH OBJECTIVES

This study addressed this overarching research question:

1. What are the characteristics of the digital information landscape of children with ASD?
   a. What IRs and other technology tools/applications do parents use to find
      information that supports the everyday life of their children with ASD?
   b. What skills do children with ASD possess in using IRs and other technology tools/applications?
   c. What skills do children with ASD possess on the research process?
2. Could parents articulate features of interface design that meet the needs of their children with ASD?

3. RELATED LITERATURE

Studies of the information behavior of children with ASD and other intellectual disabilities are almost nonexistent in literature of information sciences. This section covers selected relevant research on the software needs of people with autism, assistive technologies designed for children with ASD, and virtual reality.

Putnam & Chang [14] assessed the software and technology needs of people with autism using online surveys to elucidate information about product design that meets user goals. They found that very few respondents (25%) had experience with the design of these tools, that the needs focused on social skills, academic skills, and organization skills, and that use of technology was a major strength and for people with autism. In a recent special report, Slomsky [15] elaborated on how assistive technology (particularly robots), are aiding autistic children with social development. Assistive technologies also have potential in mediating behaviors commonly associated with autism (e.g., sensory and motor impairments). The mediation includes gaining access to social situations and creating opportunities for work and play [16]. Robins & Dutenhahn [17] examined the use of robots in aiding children with social and communication skill development, as part of working on the AuRORA Project, a research endeavor on robots designed as educational and therapeutic toys for children with ASD. One of the early software applications developed for children with ASD was MusBus [18], which was designed to train autistic children to use a computer mouse. Leroy, et al. [19] designed an image digital library to assist in the development of communication skills of children with ASD. The library's interfaces were designed for use on a Pocket PC. The idea behind the image digital library emerged from the communication reality children with ASD, that is, deficit in verbal communication and reliance on pictures (e.g., Picture Exchange Communication System (PECS)) to communicate with others.

Many researchers have employed virtual reality to create assistive technologies for children with ASD. For example, Chamberlain [19] developed a 3D generic virtual environment platform for people with ASD to use as a therapy tool. Moore, et al.[20] created a collaborative virtual environment (CVE) to assist children with autism in understanding basic emotions using a human-like avatar, while Merryman, et al. [21] designed Virtual Peers (VPs) to enable the social interaction of children with ASD. Similarly, Schmidt & Schmidt [22] developed a 3D virtual learning environment (3D VLE) to support the social interaction and development of social competence of these young users. In their review of the literature on virtual environments for children with ASD, Lanyi et al. [23] concluded that VEs can enhance skill development while providing children way to play and motivation for success. VEs can be fundamental in these children's learning, speech interpretation, and communication, in general.

4. METHOD

This research was planned in three phases: Phase 1, collect information from mediators about the digital information landscape of children with ASD; phase 2, identify mediators willing to be interviewed to collect additional data that provide deeper understanding of the realities of the children's information landscape; and phase 3, gain consent from selected parents to interview their children to gather information from the children's perspectives and to learn about the challenges they face in everyday life information seeking. The ultimate goal of using this methodology is to generate a grounded theory based on the collected data. As of the writing of this paper, data collection was still in Phase 1, which is described in this section.

Due to the nature of the research project, a quantitative method was employed in Phase 1 to collect data from mediators of children with ASD. The researcher developed and pilot-tested a survey instrument for collecting these data. It is commonly known that although web-based surveys may reach a large number of people quickly and easily, they yield low return rate, and tend to generate data that may not be totally reflective of the participants' state of mind or realities, thus, affecting the reliability of the responses. However, using a survey approach was necessary to collect the baseline data in Phase 1 of this research project based on which additional, in-depth data could be gathered.

4.1 Instrument

The survey instrument consisted of four sections (A-D) with each section designed to be answered by specific mediators. Section A (14 questions) was targeted for parents of children with ASD, section B (11 questions) for general education teachers, section C (10 questions) for special education teachers, and section D (1 general, open-ended question) for other professionals (e.g., social workers, therapists, psychiatrists). The web-based instrument was created using SPSS mrInterview, an application that enables the creation of different types of surveys and that aggregates responses for ease of data coding and analysis. In general, sections A-C, questions elicited information on the participants' background/demographics, Internet access, purposes of using the Internet, types of IRs and other technologies used, assessment of children's skills in finding information, as well as difficulties children experience in finding information. In addition, each section had two open-ended questions for gathering feedback information and suggestions from the participants. The survey was posted on the web in mid fall 2009 and data collection is still in progress.

4.2 Participants

As was mentioned earlier, the first phase of data collection targeted mediators, people who could share their experiences in working with children with ASD in different settings. A minimum of 200 people were expected to participate in the survey, accounting for 20% of the 1000 volunteers associated with the East Tennessee Chapter of the Autistic Society of America (ASA-ETC). The volunteers receive the Society's weekly newsletter via email where events, announcements, and activities are described, including calls for participation in research projects. These members will be able to see the call to participate in this survey only if they open the newsletter and scroll to the last page. Only seventeen people participated in the survey: Eleven parents, two general education teachers, one special education teacher, two family volunteers, and one speech pathologist. The fact that a small sample participated in the survey may be due to two factors: 1. Not all members of the society are active and, therefore, they may not open the Newsletter they receive via email on a regular basis; 2. Hesitance on the part of some parents to give information about their children with ASD, and 3. The nature of surveys in general, and web surveys, in particular, that typically generate low return rate.

4.3 Sampling and Procedures

Participants were recruited using two approaches: 1. Contacting the director of ASA-ETC for a list of volunteers, and snowballing. To protect the confidentiality of members associated with the Society, the researcher was unable to secure the list of volunteers. This resulted in using a convenient sample. The director of the Society included the researchers' call for participation in the survey in the weekly newsletter two weeks. Due to a low return rate, the "call" was reposted in the newsletter for two additional weeks, which resulted in twelve returns. After one parent with a child on ASD contacted the researcher and expressed interest in helping with data collection, the researcher asked the parent to snowball the survey to people with interest in completing it. This approach yielded five additional returns, up to the writing of this paper, totaling seventeen participants.

5. DATA ANALYSIS

Responses to the survey from parents were selected for reporting in this paper due to the specific relationship they have with the children. All survey questions generated quantitative data with the exception of the open-ended questions that resulted in qualitative data. For the quantitative data, descriptive statistics were employed using percentages to describe the characteristics of the data. Responses to the open-ended questions were coded using open, axial, and selective coding that resulted in identifying one core conceptual category that best represented the data. This category is based on the parents' responses to the survey.

6. RESULTS

Data analysis uncovered many characteristics of the information seeking and needs of children with ASD. The first seven questions asked of parents on the survey instrument elicited demographic information. The rest of the questions (7) covered Internet access and use, the parents' assessment of their children's skills in using IRs and other technology tools/applications, and their ratings of their skills on six steps of research process. Parents were also asked to provide input about the kind of technologies and/or programs that should be developed or improved to help their children become independent learners and effective users of information.

6.1 Profile of Parents and their Children with ASD

Eleven children were listed by parents as autistic. Of these, six (54%) were high functioning or with Asperger's Disorder. The children's age ranged from 2 to 15 years old. Most parents also had children that did not have ASD

or other disabilities. Nine (81%) of the respondents were female and two were male. The parents' educational degree earned ranged from a high school diploma to a doctorate. Only three parents did not work outside the home, and four had help with caring for their children.

6.1.1 Parents' Information Seeking and Purposes

All parents except for one had an Internet access at home. When asked about the kind of information they sought regularly on the Internet to help their children with ASD, all parents (100%) used the Internet to find medical information about autism, 9 (81%) to visit websites,7 (77%) to locate appropriate extra-curricular activities, 7 (77%) to use online databases, another 77% to find physical toys (77%), 6 (54%) to support their child's school work, 3 (27%) to pursue other types of information, 2 (18%) to use social networks/media, and another 18% to identify appropriate games (Table 1). The parents information seeking extended beyond finding information in support of their children's school work to include social networking/media, toys, games, and specific websites on autism. Table 1 illustrates the purposes for using each of these technology tools or applications. Clearly, the parents served as information seeking proxies for their children. This is not surprising since children with ASD, including those that are high functioning will need assistance with understanding the information problem and how to solve it. For example, in many public school systems that support inclusive education and have programs for young people with autism are provided (e.g., many systems in the state of Tennessee), children with ASD are mainstreamed in the classroom and library programs and services. Information specialists (or school librarians) teach context-driven information literacy skills on a regular basis that children with ASD attend along with other classmates. Based on an interview with a special education teacher in one of the school systems (Bilal, 2008), the researcher has learned that the special education teacher review each assignment and modifies it a way that children with ASD can understand. Specifically, the teacher 'decomposes" each assignment, rewrites it using clear and simple step-by-step procedure with minimal textual information that is augmented with drawings of the process the child should follow to complete the assignment. The ultimate goal is to decrease anxiety and minimize the cognitive load that complex assignments may impose on the children.

6.2 Parents' Assessment of their Children's Skills in Using IRs and other Technology Tools

Parents were asked to rate their children's skills in using IRs (e.g., online databases; web search engines) and other technology tools and applications (YouTube, Facebook, Email, IM/chat software, web browsers, and mobile devices) on a scale ranging from 1-5 (1=low; 5=high) and to indicate N/A (not applicable) for the items that do not apply. As Table 2 shows, one parent (9%) did not provide an answer to any of the items listed; and a high percentage of parents (ranging from 45-72%) indicated N/A on each of the items. Children's skills in using YouTube, search engines, and web browsers received a high rating (rating=5) by 45%, 27%, and 27% of the parents, respectively. On skills in using email and mobile phones, 27% of the parents gave an average rating in each

case (rating=3). What is surprising is the high percentages of parents that marked N/A on each of these technologies, indicating the children were at a very young age to use these resources and/or a lack of adequate knowledge about their children's skills in these areas. Considering that only YouTube was rated high by nearly 50% of the parents reflects the need for additional research to test this finding and to identify reasons for use or lack of use of social media. Similar to the parents of "normal children", parents of children with disabilities including ASD may not allow their children to use social media, email, or IM to protect them from bullying. Conversely, children learn how to use online databases and search engines in upper elementary and middle schools. One way to confirm these findings would be to interview the parents and/or observe the children's interaction with selected IRs to determine whether the skills are age-related, autism-related, and/or information literacy-related (i.e., due to inadequate training in schools).

6.3 Parents' Assessment of their Children's Skills on the Research Process

Six steps of the research process were listed in the survey instrument (Table 3) and parents were asked to evaluate their children's skills on each step using a scale ranging from 1-5 (1=low; 5=high) that also included an "I don't know" option. These six steps have been recognized by professional organizations (e.g., American Association of School Librarians) as essential for becoming information literate and for meeting the challenges of the 21st century information environment. As Table 3 shows, most parents (45%) rated their children's skills on each of the six steps as moderately low (rating=2). Surprisingly, 27% of the parents marked "don't' know" on each of the steps. Only a low percentage of parents (18%) rated their children's skill on step four as good (rating=4) (reading, hearing, or viewing information and extracting what is relevant). Due to the fact that this is a pilot, exploratory study with a small sample of participants, one should not conclude that the children's skills are inadequate. It is suggested that these findings be confirmed through interviews with the parents and school teachers, as well as by examining the children's outcomes on research assignments.

6.4 Interface Design for Children with ASD

In two open-ended questions, parents were asked to express their thoughts and share their experience about the type of technology tools and/or programs that should be developed improved to help their children become independent learners and effective users of information, and to comment on the survey and other areas related to their children's lives. Most parents (8 or 72%) provided feedback and their statements were coded using open, axial, and selective techniques that resulted in one core category, interface design, which best represented the statements. Examples of these statements are listed below with certain keywords that the researcher highlighted to emphasize criteria for interface design.

[There should be] "a box that can be dragged on top of text and highlight the word(s) as an aid for reading"… "For my son, something that is visual…My daughter learns better with hands-on approaches and uses PECS [picture system] sometimes especially when she is unable to

verbally express what she wants/needs…Both children need simple and direct steps"..."Applications such as speak to write programs to support speech disabilities such as an auditory trainer…My son has a short term working memory and nothing seemed to help him"…"More autism friendly games that focus on basic skills"…"Audio-based homework sheets that are self-read and respond like an online homework system."

Although only eight parents (72%) provided insights about interface design needs, the statements provide a basis for additional data collection to gather more in-depth data about various aspects of design that system designers should address in developing universal

7. CONCLUSIONS

This paper reported the preliminary results of a pilot study that explored the digital information landscape of children with ASD from the mediators' perspectives, especially parents. Data are being collected from teachers and other professionals who interact with these children on a regular basis. The results of this study revealed many characteristics of the information landscape of these children. Parents were actively involved in their children's lives to the degree that they acted as information seeking proxies in everyday life. The parents information seeking was beyond school work to include every aspect of the children's lives and needs. What is gleaned from the results is that children with ASD seemed to be silent! information seekers for reasons that could not identified from the collected data. Future research should address the children's roles in the information seeking process to gain further understanding of the realities of their information world or landscape. Surprisingly, a high percentage of parents lacked adequate knowledge of the level of skills their children possessed in using IRs and various technology tools. It is speculated that lack of adequate knowledge of the level of these skills combined with the parents' ratings of the skills on all steps of the research process as "low" may explain the parents' role of proxy information seekers. Additional research should investigate in depth the silent role of the children and the proxy role of the parents.

The results of this study showed that parents possessed valuable thoughts and experiences about elements of design that should be reflected in interfaces designed to meet their children's needs. However, additional research is needed with a larger sample of parents and other mediators to gain deeper understanding of various elements of designing universal interfaces that are cognitively supportive of children with various level of ASD and possibly other disabilities.

This study has scraped the surface of the digital information landscape of children with ASD. Much research is still needed in this area of study. Individual, face-to-face interviews of a large number of parents of children with ASD could reveal data that surveys fail to provide, and also provide rationale for the information seeking proxy role, among other things. Similarly, interviewing children with ASD, especially ones that possess verbal skills may uncover their information needs and patterns of seeking strategies. Prior research found that 'typical" children were capable of articulating design features and draw interfaces for search engines of interest to them [23][24]. Children with Asperger's Syndrome may prove to be as articulate about design features as children without ASD. Studies are needed to explore varied skills and capabilities of children with ASD.

Based on the preliminary results of this study, information specialists should become aware of the deficiencies in information and research skills of children with ASD. While mainstreaming children with ASD in library programs and activities is commended, these children may need customized teaching and training in information seeking and using IRs and other technology tools.

8. ACKNOWLEDGMENT

9. REFERENCES

[1] Baron-Cohen, S. 2001. Theory of mind and autism: A review. http://www.autismresearchcentre.com/docs/papers/2001_bc_review.pdf. Accessed on June 3, 2010.

[2] Bilal, D. 2000. Children's use of the Yahooligans! Web search engine. I. Cognitive, physical, and affective behaviors on fact-based tasks. Journal of the American Society for Information Science, 51 (7), 646-665.

[3] Bilal, D., & Wang, P. 2005. Children's conceptual structures of science categories and the design of Web directories. Journal of the American Society for Information and Technology, 56(12), 1303-1313.

[4] Druin , A. 2002. The Role of children in the design of new technology. Behaviour and Information Technology, 21(1) 1-25.

[5] Wellman, H.M. 1990. The Child's Theory of Mind. Cambridge, MA: The MIT Press.

[6] Ritvo, E.R. 2006. Understanding the Nnature of Autism and Asperger's Disorder: Forty Years of Clinical Practice and Pioneering Research. London, UK: Jessica Kingsley Publishers.

[7] Shneiderman, B., & Plaisant C. 2009. Designing the user interface: Strategies for Effective Human-Computer Interaction. 5th ed. Boston: Addison-Wesley.

[8] Lazar, J. 2007. Universal Usability. Hoboken, NJ: Wiley.Leroy, G., et al. 2005. Digital libraries on handhelds for autistic children. JCDL'05, June 7-11, Denver, CO, p. 387.

[9] Shneiderman, B.2000. Universal usability: Pushing human-computer interaction research to empower every citizen. Communications of the ACM, 43(5), 84-91.

[10] Bilal, D., & Bachir, I. 2007. Children's interaction with international and multilingual digital libraries. I. Understanding interface system design representations. Information Processing & Management, 43(1): 47-64.

[11] Bilal, D., & Kirby, J. 2002. Differences and similarities in information seeking: Children and adults as Web users. Information Processing & Management,38 (5): 649-670.

[12] Savolainen, R. 1995. Everyday life information seeking: Approaching information seeking in the context of "Way of Life." Library and Information Science Research, 17, 259-294.

[13] Fisher, K. E., & Abrahamson, J.A. 2007. What past is prologue: Towards a general model of lay information mediary behaviour. Information Research, 12(4). http://informationr.net/ir/12-4/colis/colis15.html.

[14] Putnam, C., & Chong, L. Software and technologies designed for people with autism: What do users want? ACM SIGACCESS Conference on Computers and Accessibility, October 13-15, Halifax, Nova Scotia, pp. 3-10.

[15] Slomsky, A. 2009. "Im here to make you feel better." The Washington Post, p. HE01, March 10.

[16] Gillette, D., et al. 2007. Interactive technologies for autism. CHI 2007, April 28-may 3, San Jose, CA, pp. 2109-2112.

[17] Robins, B., & Dautenhahn, K. 2004. Interacting with robots: Can we encourage social interaction skills in children with autism? Accessibility and Computing, 80, 6-10.

[18] Leroy, et. al. 2005. Digital libraries on handhelds for autistic children. JCDL'05, June 7-11, Denver, CO, p. 387.

[19] Chamberlain, M. 2009. A 3D virtual environment development platform for ASD therapy tools. Masters Thesis, Stellenbosch University, South Africa.

[20] Moore, D., et al. 2005. Avatars and autism assistive technologies – from virtual reality to reality, A. Purski and H. Knops (Eds.), Assistive Technology: From Virtuality to Reality. Amsterdam: The Netherlands: IOS Press, pp. 442-448.

[21] Merryman, J., et al. 2008. Designing virtual peers for assessment and intervention for children with autism. IDC '08, June 11-13, Chicago, IL, pp. 81-84.

[22] Schmidt, C.., & Schmidt, M. 2008. Three-dimensional virtual learning environments for mediating social skills acquisition among individuals with autism spectrum disorders. IDC'08, June 11-13, Chicago, IL

[23] Lanyi, C.S., et al. 2006. Virtual reality in special needs early education. The International Journal of Virtual Reality, 5(4), 55-68.

[24] Large, A., Beheshti, J., & Rahman, T. 2002. Design criteria for children's Web portals: The users speak out. Journal of the American Society for Information Science and Technology, 53, 79–94.

[25] Bilal, D. 2003. Draw and tell: Children as designers of Web interfaces. Proceedings of the 66th ASIST Annual Meeting (pp. 135–141). October 19–22, Long Beach, CA.

Table 1. Type of Information, Technologies, and Purposes of Information Seeking by Parents of Children with ASD

Information sought/Sources	No. of respondents (n=11)	Percent	Purposes
School work	6	54%	Find reading and literacy materials, learning games, information on behavior techniques, study skills, reading delays, speech/vocabulary building, adaptive devices, and assistive devices. Identify ways to modify schoolwork using websites and articles to share with teachers.
Medical information	11	100%	Find biomedical and anything related to the child's medical needs. Locate current research in all related areas to autism. Find information on genetic, neurological disorders, speech pathology, and innovations. Find everything I can possibly read to understand the child's needs, medications, etc.
Extra-curricular activities	7	77%	Locate information on local opportunities to go out in the community (festivals, classes, etc), art exhibitions, and art classes. Find information about places that are "autism friendly." Locate camps, social groups, amusement parks, etc. geared to special needs. Find activities that are "accepting" of my children... that have "typical" peer interactions, etc.
Games	2	18%	Identify social skills and eye contact training games. Find games for building social skills,
Social networking	2	18%	Use Myspace and Facebook; to visit the websites devoted to autism.
Search engines	8	72%	Use Google, Yahoo, AOL, Dogpile, and Goodsearch.
Physical toys	7	77%	Find puppets/ toys that promote imaginative play…that require two people.
Databases	7	77%	Use Infotrack, PubMed, MED Scape, Tennessee Electronic Library, and ProQuest.
Websites	9	81%	Connect with Yahoo groups, Facebook, local group on ASK network, or Myspace. Use blogs on autism…YouTube, OpenDoors, Disability Pathfinder, local and regional autism societies and networks, and AOL chat.
Other information sought	3	27%	Find videos for my child to watch for fun. Locate research updates.

Table 2. Parents' ratings of their children's skills in using IRs and other tools

IRs/Technologies	Rating of skills and percentage of respondents. 1=low; 5=high						
	1	2	3	4	5	N/A	No answer
Online databases	18%		18%	9%		45%	9%
Search engines	9%		9%	9%	27%	45%	9%
YouTube	9%				45%	45%	9%
Facebook	9%			9%		72%	9%
Email	9%	9%	27%		9%	54%	9%
IM/chat	9%				9%	72%	9%
Web browsers	9%		9%	9%	27%	45%	9%
Mobile phones	27%		27%			45%	9%

Table 3. Parents' rating of their children's skills on the research process

Research Process Steps	Rating of skills and percentage of respondents (1=low; 5=high)						
	1	2	3	4	5	Don't Know	No answer
1.identify what information is needed	18%	45%	18%			27%	9%
2.Identify appropriate sources to use	18%	45%	18%			27%	9%
3.Locate sources and find information within sources	18%	45%		9%		27%	9%
4.Read, hear, or view the information and extract what is relevant	9%	45%		18%		27%	9%
5.Organize and present the information	18%	45%	18%			27%	9%
6.Judge the finished product and the process that was used to produce the product	18%	45%	18%			27%	9%