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 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.

![Figure 1. Example of the ‘smiley-scale’](image)

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

<table>
<thead>
<tr>
<th>Name</th>
<th>Arrangement of information</th>
<th>Type of search engine / Primary search tool?</th>
<th>Menu structure / navigation</th>
<th>Clutter / density</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Schoolbieb.nl</td>
<td>Traditional</td>
<td>Google-like / search engine / primary search tool?</td>
<td>Taxonomical / menu in words</td>
<td>Little clutter / low density</td>
</tr>
<tr>
<td>2. Willemwever.nl</td>
<td>Build of blocks</td>
<td>Question-answering system / primary search tool</td>
<td>Abstract main menu in words / basic categories on 2nd level</td>
<td>A lot of clutter / high density</td>
</tr>
<tr>
<td>3. Kids.kennisnet.nl</td>
<td>Metaphorical</td>
<td>Google-like / search engine / navigation metaphor</td>
<td>Abstract menu categories</td>
<td>Little clutter / high density</td>
</tr>
<tr>
<td>4. Wikikids.nl</td>
<td>Traditional, but a lot of text</td>
<td>Google-like, option to get direct results page / primary search tool</td>
<td>Abstract menu categories</td>
<td>Medium amount of clutter and density</td>
</tr>
<tr>
<td>5. Google</td>
<td>Minimalistic</td>
<td>Google-like / primary search tool</td>
<td>No possibility to browse through categories</td>
<td>No clutter and very low density</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Condition 3</th>
<th>Condition 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1 Birth-day Queen</td>
<td>Google.nl</td>
<td>Google.nl</td>
<td>Google.nl</td>
<td>Google.nl</td>
</tr>
<tr>
<td>Task 2 Animals</td>
<td>Traditional interface (kangaroo)</td>
<td>Question-answering interface (Mozart)</td>
<td>Metaphorical navigation (shark)</td>
<td>Textual interface (dolphin)</td>
</tr>
<tr>
<td>Task 3 Arts or Music</td>
<td>Question-answering interface (Rembrandt)</td>
<td>Traditional interface (Beethoven)</td>
<td>Textual interface (Mozart)</td>
<td>Metaphorical navigation (Van Gogh)</td>
</tr>
<tr>
<td>Task 4 Sports</td>
<td>Metaphorical navigation (hockey)</td>
<td>Textual interface (soccer)</td>
<td>Traditional interface (basketball)</td>
<td>Question-answering interface (gymnastics)</td>
</tr>
<tr>
<td>Task 5 Health</td>
<td>Textual interface (hay fever)</td>
<td>Metaphorical navigation (head lice)</td>
<td>Question-answering interface (travel-sickness)</td>
<td>Traditional interface (braces)</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Success score:</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Comprehension</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Required help</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

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 layout, 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 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 (F1,31) = 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).

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.

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).

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
This research is funded by the European Community’s Seventh Framework Programme FP7/2007-2013 under grant agreement no. 231507, called ‘PuppyIR’. The research is also part of the research program called ‘The digital youth library’ that is funded by The Netherlands Public Library Association (VOB).

9. REFERENCES