ImagePile: an Alternative for Vertical Results Lists of IR-Systems

Saskia M. Akkersdijk, Merel Brandon, Hanna Jochmann-Mannak, Djoerd Hiemstra, Theo Huibers

Centre for Telematics and Information Technology
University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands

ABSTRACT
Recent work shows that children are very well capable of searching with Google, due to their familiarity with the interface. However, children do have difficulties with the vertical list representation of the results. In this paper, we present an alternative result representation for a touch interface, the ImagePile. The ImagePile displays the results as a pile of images where the user navigates through via horizontal swiping. This representation was tested on a search engine for the EMMA child hospital's library. Using a within subject experiment, both representations were tested with children to compare the usability of both systems. The vertical representation was perceived as easier to use, but the ImagePile system was considered more fun to use. Also, with the ImagePile system more relevant results we chosen by the children, and they were more aware of the number of results.

General Terms
Your general terms must be any of the following 16 designated terms: Design, Experimentation, Human Factors.

Keywords
Children, search-interface,

1. INTRODUCTION
In a time where Internet provides access to a large amount of the world's information, the accessibility of IR-systems for children is extremely important.

The Puppy-IR project, an European funded project that focuses on the accessibility of IR-systems for children, has developed a demo-system that should allow young patients to search for book and DVD's of the EMMA child hospital's library. The library contains books, DVDs, and objects (such as 3D models of limbs and organs) about the hospital, the human body, diseases, or treatments. This information is very valuable for the children since it helps them to understand and communicate about their situation.

This demo-system has a Google-like interface and the results are displayed in a vertical results list, accompanied with pictures (Fig.1). In this paper we will use the term "Vertical Results List" to refer to this demo-system.

They found that children's performance on Google is better and that Google was by far most preferred. Children found it difficult though to determine which items from the large lists of results that Google provides are relevant. This problem might be due to the children's unawareness of the accuracy of the information source [1]. The problem did not occur with the other systems because they use more contained repositories and did not present results from the WWW. Druin et al. [5] also found that children have problems determining the relevance of the items. Children also have difficulties with the vertical list representations that most IR systems (like Google) use. Some of these problems are:

1. Problems with vertical scrolling [4,5,6]
2. Problems with subpages [4]
3. Problems with distracting items [5]
4. Problems with controlling a mouse [1]

Most of the papers describing the problems emphasize the need for a different design of the result pages [1,5]. Druin et al. [5] suggest "a single page of results with less text, fewer links, and no scrolling required" as an alternative representation for children in the 7-11 age range.

Figure 1. The Vertical Results List system

In this paper, we suggest an alternative representation of the results, the ImagePile. Because of the difficulties that children might have with using a mouse and scrolling [1,5] we decided to design the system for a touch screen. The EMMA child hospital is also planning to put computers with a touch screen in all patient rooms. The ImagePile system displays the result items in a horizontal row of covers, and supports a swiping gesture for navigation through the results. The next section will provide an extensive description of the design.

Both the usability of the ImagePile and that of the Vertical Results List are explored using a user study with eight children

aged eight to eleven at a Dutch primary school. We wanted to know whether the problems described above also occur with the specific application for the IR system for the library of the Emma Children Hospital. And whether the problems are still accurate, since some of the research project that found these problems are performed a long time ago.

The main research question is: *Is the ImagePile representation easier and more accessible to use than the Vertical Results List?*

To answer this question we divided it in several sub-questions:

- Which problems occur when children use the Vertical Result List representation?
- Which problems occur when children use the ImagePile representation?
- Which representation do the children find easier to use?
- Which representation do the children find more fun to use?
- Which representation do the children prefer?

Section 3 describes the method used to evaluate the systems. Section 4 explains the results and section 5 discusses the findings and the used method. Finally section 6 will provide some suggestions for future work.

2. IMAGEPILE DESIGN

The goal of the project was to design an alternative result representation for the Vertical Results List that improves the usability for children. We focused on children aged eight to eleven. Children aged twelve or older have similar cognitive and motor skills as adults [7,8], we therefore expect them to have less problems with the Vertical Results List. Children younger than eight have in general difficulties with writing, typing and reading, therefore a system designed for this age-group would need more radical changes [7,8].

We formulated the following requirements for the new system based on the literature described in section 1 and some usability guidelines. The most important requirements are:

1) The navigation and interaction with the system should be intuitive.
2) The system should support image oriented scanning of the results.
3) Information of a result must be sufficient to determine the relevance
4) The amount of results visible at once should be no more than fifteen items.
5) The system should use a minimum of subpages
6) The relevance ordering of the results should be clear.
7) It must be possible to focus on the information of one relevant result, without being hampered by other results

Based on these requirements we for chose a design that presents the result items in a horizontal row of covers among fifteen alternative designs. We call this the ImagePile, this system uses of a touch interface.

The ImagePile always focuses on one item, the picture of the item that has the focus is in the center of the screen, with a description beneath it. In contrast to the vertical results list system the ImagePile system only shows the short description of the book, but not the publication year, author, ISBN, or category. By making a swiping gesture on a touch interface the user can browse through the items. Moving a finger from right to left moves the focus to the right, the opposite movement moves the focus to the left. The number of items that is moved depends on the length of the swiping gesture. The swiping can be done at every position on the screen. It is also possible to click on a cover to focus on it directly. The first time a user enters the results page the first most relevant result is displayed in the middle of the screen, the other results are visualized as a horizontal row of covers on the right side. The items are ordered on relevance from left to right. A maximum of fifteen result covers are displayed at once in the screen. When there are more than seven results left or right of the focused item, the remaining item are replace by a single icon symbolizing a stack of images (see the gray icons on both sides in Fig. 2). As a result, all items are displayed in a single page.

Figure 2. The ImagePile system

The ImagePile system was build upon the existing search engine of the Vertical Results List[1]. This search engine was created using PF/Tijah[2]. We have created a different interface for the result representation, using CSS3 and Javascript to present the result items in a horizontal row of covers.

3. METHOD

To evaluate the usability of the ImagePile system we conducted a user study at a Dutch primary school. The experiment was of an explorative nature and is meant to provide a direction for future research. We tested the systems with eight children aged 8-11. Because of the small number of subjects and the explorative nature of the experiment, a within subject design was used and qualitative data was collected. During the experimental condition the children used the ImagePile system on an iPad and during the control condition they used the Vertical Results List on a laptop. Each child had to perform three tasks on the first system and three tasks on the second system. We eliminated an order effect that could arise when always providing one of the conditions first, by alternating the order of the conditions between the children. Four children worked with the ImagePile first.

[1] http://pathfinder.cs.utwente.nl/puppyir/

[2] http://dbappl.cs.utwente.nl/pftijah/

3.1 Participants

The usability of the system was tested with eight children aged eight to eleven (four male, four female). Their parents gave their permission. The group of children that first worked with the ImagePile and the group that first worked with the Vertical Results List were matched on their gender and grade. The children attend a Montesori school and were selected from grade five to eight. Table 1 shows how the children were divided over the two groups.

Table 1. Characteristics of the participants

grade	5		6		7		8	
group	A	B	A	B	A	B	A	B
gender	F	M	M	F	F	M	M	F

3.2 Procedure

The experiment took place in a room of a Dutch primary school. We used a protocol to keep the difference between the sessions and conditions as small as possible. This protocol for example described how to deal with questions of the children about the navigation.

One by one the children were retrieved from their classrooms for an experiment. We told the child that we designed both systems (to prevent social desirable answers in favor of the system we designed) and that we wanted to find out what kind of improvements the systems need to make them easier and more fun for children. We used a screen recorder for the laptop and a hand camera for the iPad to record the actions of the children, this was motivated to them. After the introduction the children's experience (e.g. frequency of use) with computers, search engines (especially with Google), and touch screens was questioned using a questionnaire.

Subsequently, we gave the children three tasks on both systems. After each task the children were asked to rank the system on difficulty and enjoyability. For this evaluation by the children we used the "smiley face" 5 point Likert scales (Fig. 3) as used in [1,4] and derived from the Wong-Baker pain rating scale [9].

When the motivation of the rating was unclear, the children were asked to motivate their rating.

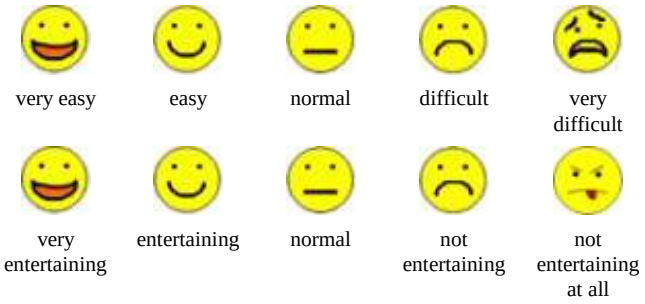

Figure 2. Example of the "smiley-scale" used for difficulty and enjoyability raking.

When the children finished their tasks on both systems we asked them to compare the systems. The children denoted which system they perceived as the most difficult and which system they found the most fun to use. We asked what they preferred, the horizontal or the vertical representation. The children were invited to motivate their answers. The experiment ended with ideas of the children for improving the system. The total experiment lasted on average 30 minutes.

3.3 Tasks

The children were given three types of tasks per system. Since we used a within subject design we needed two comparable tasks per task type. The first two tasks were goal directed. For example: "John visited the doctor. The doctor has told him that there is something wrong with his digestion. John doesn't know what digestion means. Can you find a book for John that explains him the meaning of digestion? Type 'digestion'." (In which John is a fictional child) The query was predefined and children were helped if needed to type the query. The first task resulted in two to four results and required no scrolling with the Vertical Results List system. The second task had eleven to fifteen results and had two result pages with the Vertical Results List. We wanted to see if there were different problems observed for small and large numbers of results. The third task was an open-ended task. For example: "Search for something you want to know about hospitals?". We used this open-ended task to see how children interact with the system when they do not have a specific task. We believe that this could provide a more realistic view on how the children would use the system.

3.4 Data analysis method

We have collected data from four different data sources. Firstly, an important data source is our observations of the interaction with the system, and the search strategies used by the children. Observations were notated by one of the authors during the experiment. Using the recorded data the experiments were analyzed again by both authors and notes were taken using an observation schema. The observations were quantified, for example by counting the amount of children that used a certain search strategy. Secondly we have gathered data about perceived ease of use, and enjoyability of the systems with the "smiley scale" questionnaire. We compared the mean scores for the systems. Thirdly, we gave the children some questions to compare the systems. These results where quantified by counting the amount of children than made a certain choice. Finally, we looked at how relevant the selected items, given the task, of the children were. Per task we rated all the results on a five point Likert scale (1= not relevant at all, 5 = very relevant). Because we saw that some children tended to select a item on their own preference or didn't remembered the total task while selecting an item, we also looked at the motivation given by the children to select an item. We rated the selected items for the tasks on their relevance given the motivation of the child on 5 point Likert scale (1= not relevant at all, 5 = very relevant). We conducted a T-test to see whether differences in relevance between the systems were significant.

4. RESULTS

In this section we will present the most important results of our experiment. We will start with some general observations, followed by a specific section for each system. Finally, we will state the given opinions of the children and discuss the task relevance.

Of the eight subjects only one of them was inexperienced in searching the internet, six of them know Google, and used search

engines more then ten times. Six of the subjects have used some kind of touch screen before, varying between two times to more than ten times.

With both systems the subjects did not need help to find out how to navigate through the results. However, with both system children tried to 'click' on a cover (which did not do anything). We identified three different search strategies used with both systems. Firstly, scanning multiple results and selecting one of them (scan-all). Secondly, scanning until a relevant result is found (scan-until), and finally, a combination based on the number of results (scan-combi).

4.1 Vertical Results List

Using the vertical result list, all search strategies were applied by different participants. Two of them applied the scan-all search strategy, five of the children applied the scan-until method, and one used the scan-combi strategy. However, the children who applied the scan-all method were not aware that there was another result page. Neither did the other children, or this was not required for their search strategy. Of the eight children participating in our experiment only one of them was fully aware of the number of results. An example of this was one girl who realized after asking by one of the authors: "Are these all the results?" that there were more then four results, and used the scrollbar to go to them. After seeing all the results on the first result page she changed her final answer. After this, when being asked again: "Are these all the results?", she noticed the next result page.

Two out of eight children used the scrollbar to navigate through the results, the other six used the mouse wheel.

There were no problems mentioned by the children about the navigation through the results.

4.2 ImagePile

Using the ImagePile, all search strategies were applied by different subjects too. Five of them applied the scan-all search strategy, one of the children applied the scan-until¬ method, and two used the scan-combi strategy. All children were aware that there were more results, even if they did not see them, independent of their search strategy.

For navigating through the results six of the subjects used a swiping motion to navigate through the results, two of the children used a combination of swiping and 'clicking'.

There were a few problems mentioned by the children, which could be categorized in three items. First, a problem which was experienced by seven out of the eight children, is that the system did go one or more covers further than they wanted the system to go. Secondly, one of them used his fingernail, which did not work, and finally, there was a problem with 'clicking' on the correct point (precision).

4.3 Opinions about the systems

We asked the children to rate the difficulty and enjoyability of both systems after each task. The vertical results list received an average score of 2.1 on a five points scale for both the difficulty and enjoyability. The ImagePile received an average score of 2.3 out of 5 for both the difficulty and enjoyability. However, this is not a significant difference. The comments given by the children with their rating were mostly about the tasks, not about the system. For the vertical results list the relevant remarks were as followed: "*easy because there was more information about the books (category, release year)*", "*easy because of typing on the keyboard*" and "*I liked typing on the keyboard*". For the ImagePile the relevant remarks were as followed: "*difficult because I wanted to go to the next picture but the system goes to a picture further than that*", "*difficult because sometimes you skip a results, but that also makes it easy to navigate back to were you found a relevant result*", "*there were many results*", "*easy because of the good descriptions*", "*fun to explore*", "*funny*" and "*I liked it better, because you can do everything on the screen*".

Seven out of eight children found the vertical results list easier to use than the ImagePile, but everyone enjoyed the ImagePile more. Examples for the reasons given by their choice of the difficulty are: "*I am more used to the computer*", "*Typing on the keyboard is easier*", "*I did want to go to the next picture but the system goes to a picture further than that*" and "*Because I am used to typing on the I-pad. Because sometimes I send messages to my father using my mothers iPhone. I don't use the computer that often.*". Examples of comments give by their choice for the enjoyability are: "*More special, more new, more fun, exploring how it works*", "*To use the screen to search was easy*", "*Fun to control it with your fingers*".

The last question was which system they preferred. Five out of eight preferred the ImagePile, and the other three preferred the vertical results list. The explanations given with their preference will be further discussed in the discussion.

4.4 Task relevance

Afterwards we evaluated the chosen books for each task. We gave marks for relevance between 1 (non relevant) and 5 (relevant), and their explanations given with the choice also between 1 (explanation did not agree with their choice) and 5 (explanation did agree with their choice). The ratings were evaluated with a paired t-test and the results can be found in table 2 and 3. For the reasoning and relevance the difference between the systems is not significant. The exception for this was the relevance of the third task. However when we did take all tasks into account, we get a significance of 0.057 at the relevance, which is almost significant.

Table 2. Statistics of the relevance

Paired Samples Statistics & Test for the relevance				
Tasks	ImagePile (S1) Vertical Results List (S2)	Mean	Std. Deviation	Sig. (2-tailed)
1	S1	5,0	0,0	
	S2	4,5	0,9	0,170
2	S1	2,5	1,3	
	S2	2,4	1,1	0,930
3	S1	3,9	1,1	
	S2	2,3	1,4	0,042
all	S1	3,8	1,4	
	S2	3,1	1,5	0,057

Table 3. Statistics of the reasoning

Paired Samples Statistics & Test for the reasoning				
Tasks	ImagePile (S1) Vertical Results List (S2)	Mean	Std. Deviation	Sig. (2-tailed)
1	S1	5,0	0,0	
	S2	4,3	1,4	0,170
2	S1	2,6	1,3	
	S2	2,5	0,9	0,850
3	S1	3,1	1,8	
	S2	2,4	1,3	0,310
all	S1	3,6	1,6	
	S2	3,1	1,4	0,130

5. DISCUSSION AND CONCLUSIONS

We will continue with discussing the opinions given by the children as comments during the rating of both systems. This will give us more inside in their opinions.

Comments given during the tasks and given together with their ratings imply that the ratings were sometimes based more on the tasks than on the systems. An example of this was a child who said "easy because the query was not too difficult to spell". Therefore, ratings about the difficulty are not always accurate and all comments given become more valuable than the ratings. The same goes for the enjoyability. An example for this is "*In the description there was information I did not know before, I like that.*".

There were several difficulties with the navigation of the ImagePile, the most important being that the system went one or more covers further than they wanted the system to go, which was experienced by seven of the eight children. The ImagePile system was designed to go to the next cover with a small gesture and several covers further with a bigger gesture. The problem experienced can be lead back to this feature which can be improved by fine-tuning the feature for children or by removing the feature so you always go to the next cover. What also contributed to this problem was the speed with which the system reacted to the gestures. We did the experiment on an iPad, which was slower in reacting to the gestures than a laptop with touch (which was not available at that time). The subjects sometimes repeated the gesture or enlarged it, because they thought it had not worked, while the system was processing it. This lead to going further than wanted. The second problem was with 'clicking' on the correct point (precision). This problem was related to the area in which the children had to "click" to type or to go to the next cover. To be accurate at "clicking" at the correct point, motor skills are involved, which are less developed in children than in adults. To improve this, a larger screen can be used so the area will increase in size, or the number of covers can be decreased which will also lead to a bigger area. The last problem was the usage of fingernails for the gestures which is related to the touch screen used. This can be solved by using another type of touch screen with which this is possible.

The search strategy used by the children changed depending on the system used. Six of eight children did not use the same method on both systems. With the vertical results list the scan-until method was used most and with the ImagePile the scan-all method was used most. This can be an indication that the ImagePile encouraged the children to look at more results before choosing one. However, we have to be cautious, because while they did go to the end of the list this also can be a problem. Due to the limited database this was currently not a problem. But it could be a problem when there are many results. Another difference between the systems was the awareness of the number of results. With the vertical results list children were not aware of the number of results. None of the users looked further than page one, and one was not even aware that there were more results besides the first four and you could scroll down. With the ImagePile they were aware that there were more results, even if they did not see them. To which extend looking at less results forms a problem depends on the type of search system the child is using. During Internet searching it might be a good strategy for children to look only at the top five results found by the search engine because most search engines use a good relevance ordering. But for the system discussed in this paper we think that this problem is more important, because when choosing a book, it not only has to match the information need, it also has to have a certain attraction to it, which matches the child's preference. Therefore it is more likely that a relevant item has a higher index. Moreover children have some problems with forming effective queries, which leads to less relevant results for their information need.

This led to the question if the selected results found with the ImagePile would be more relevant than of the vertical results list. For this the ratings for relevance and reasoning were created, and a paired t-test was performed. Only one of the tests gave us a significant result, but when we combined all tasks the test came back almost significant. This indicates that children while searching with the ImagePile, choose a more relevant result, then when searching with the vertical results list. However, keep in mind that the number of subjects is very low, which makes it impossible to make any definitive statements (this goes for all results).

As stated before, the comments given with the rating for difficulty and enjoyment were of more use than the ratings themselves. The comments given during the rating for the difficulty were mostly related to their previous experiences. For example: "*More used to the computer*", "*Typing on the keyboard is easier*", '*I think I like the mouse*" and "*Because I am used to typing on the I-pad. Because I sometimes send messages to my father using my mothers I-phone. I don't use the computer that often.*". Other than their previous experiences the comment: "*I did want to go to the next picture but the system goes to a picture further than that*", was the comment most made. This indicates that previous experiences influence their perception of how difficult it is. Something similar counts for the enjoyability. With the comments given during the rating of the enjoyability the most heard one was: "*More special, more new, more fun, exploring how it works*". Followed by: "*To use the screen to search was easy*", and "*Fun/more pleasant to control it with your fingers*". This indicates that the unfamiliarity, newness of the system and the touching of the screen is what they enjoy most.

Going back to the preferences of the children for one of the systems, five of eight children preferred the ImagePile. Given reasons for this were for example: "*easier to go through all results / more pleasant then scrolling*" (3x), "*You could see at once how many results there were*", "*Because you can see the next cover better*", "*Because it is more like a library*" and "*You can use your finger to just move them aside*". Reasons for preferring the vertical results list were: "*You could just scroll and with the other one you had to use your finger, which I did not find easy, because sometimes it goes to far*", "*Because you can go through them more easily*" and "*Because you can click on them faster*". This indicates that the problems with the navigation with the ImagePile have to be improved, to make a better comparison between the two systems.

Finally, we will answer the research questions. The most evident problems with the Vertical Results List are not being aware of the number of results and not being aware of the number of sub-pages. For the ImagePile, problems are that the system went one or more covers further than they wanted the system to go and precision problems. Most children did find the Vertical Results List easier to use, mostly due to their previous experiences and gesture problems with the ImagePile. All children enjoyed the ImagePile system more and five out of eight children preferred the ImagePile representation over the Vertical Results List representation. Based on these answers we can answer our main research question: Is the ImagePile representation easier and more accessible to use than the Vertical Results List? In the current state the ImagePile system does not improve the Vertical Results List, mostly because of the gesture problems. However, the items chosen with the ImagePile are more relevant. Therefore, we think that with proper modifications it can improve the Vertical Results List.

6. FUTURE WORK

Our recommendations for future work are, first of all to continue the development IR systems with a touch interface. This is something all children enjoyed, and when developed further has much potential. Parallel with this, the interaction with the touch interface had to be fine tuned for children in order to conquer the navigational problems. The most heard reason why the children did prefer the laptop or had difficulties with the ImagePile was "*I did want to go to the next picture but the system goes to a picture further than that*". The other navigation problems relating to precision can be solved using a bigger screen. The interaction would also benefit from more responsiveness.

There has to be a maximum number of results when using the ImagePile, while looking at more results can be good, it also can be a problem when you have more than 25 results and you still go to the end of the list. Therefore we recommend a maximum of 25 results.

Children were tended to wanting to click on the covers to enlarge them. This happened with both the systems tested and was not possible in either one of them. Because this happened several times we would recommend to include this in further designs.

Last of all and the most important of all is testing with more subjects. Experiments with eight children are not sufficient to make any valid statements and therefore testing needs to be done with more children.

Acknowledgements

This work is part of the PuppyIR project, which is supported by a grant of the 7th Framework ICT Programme (FP7-ICT-2007-3) of the European Union.

REFERENCES

[1] H. Jochmann-mannak, T. Huibers, L. Lentz, and T. Sanders, "Children searching information on the Internet : Performance on children ' s interfaces compared to Google," *Interfaces*, 2010.

[2] D. Bilal, "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*, vol. 51, 2000, pp. 646-665.

[3] A Large, "Information seeking in a multimedia environment by primary school students," *Library & Information Science Research*, vol. 20, 1998, pp. 343-376.

[4] B.S. Chaparro and S. Naidu, "Usability News 72 - Naidu Evaluating the Usability of Educational Websites for Children Usability News 72 - Naidu Page 2 of 7," *Children*, vol. 7, 2008, pp. 1-7.

[5] A. Druin, E. Foss, L. Hatley, E. Golub, M.L. Guha, J. Fails, and H. Hutchinson, "How children search the internet with keyword interfaces," *Proceedings of the 8th International Conference on Interaction Design and Children - IDC '09*, 2009, p. 89.

[6] D. Bilal, "Differences and similarities in information seeking: children and adults as Web users," *Information Processing & Management*, vol. 38, Sep. 2002, pp. 649-670.

[7] J.P. Hourcade, "CHAPTER 55," *Human Factors*, 2006, pp. 1446-1458.

[8] A. Bruckman and A. Bandlow, "Technologies , and Emerging Applications , edited by Julie Jacko and Andrew Sears . Lawrence Erlbaum and Associates , 2002 .," *Human-Computer Interaction*, 2002, pp. 1-39.

[9] D.L. Wong and C.M. Baker, "Pain in children: comparison of assessment scales.," *The Oklahoma Nurse*, vol. 33, 1988, p. 8.