A Study of Children's Search Query Formulation Habits

Nicholas Vanderschantz
Department of Computer Science
The University of Waikato
New Zealand
vtwoz@waikato.ac.nz

Annika Hinze
Department of Computer Science
The University of Waikato
New Zealand
hinze@waikato.ac.nz

The strategies children use for digital information search in educational settings are rarely explored. Open questions remain on such fundamental issues as to which information-seeking strategies children employ, how they construct queries, and if the strategies that are taught are effective when using modern search engines. We conducted an observation study with school children to gain insights into these questions. As a result of this study, we identified query-creation and query-reformulation strategies that children use.

web search, children, digital information behaviour, internet search engines, information search

1. INTRODUCTION

New Zealand (NZ) classrooms are often well equipped with interactive communication technologies including Internet resources, eBooks, and information search software on both desktop and mobile computing devices (Vanderschantz et al. 2014a). Inquiry-based learning is a core value of the NZ curriculum and children regularly conduct information search during their educational pursuits. The systems for information search used by children in NZ classrooms and homes appear to have been developed with adults in mind (Timpany and Vanderschantz 2011). van der Sluis and van Dijk (2010) report that these adult-oriented systems do not suit children's information seeking needs because they require complicated search procedures and return results that might not be appropriate for children.

In this paper we seek to provide further insight into the issues children encounter during Internet search. We conducted an observational study with school children aged 9 to 13 years old to gain insights into their practice of information search in an educational context. This study was a task-based usability observation with children in their schools.

2. RELATED WORK

Most of the related work is somewhat dated and focuses on children searching for printed artifacts (e.g., Moore 1995) or using Internet Search Engines (ISEs) that utilise less robust search algorithms than available today (e.g., Bilal 2000). More recent investigations have included larger samples of anonymous data from information search logs (e.g., Duarte Torres et al. 2010) from which information needs and purposes are difficult to ascertain.

van der Sluis and van Dijk (2010) provide a survey of research pertaining to children's Information Retrieval (IR) system use. They identify four categories of issues that children face when using IR systems: 1) children have an insufficient mental model of IR systems; 2) children encounter vocabulary problems; 3) children display chaotic search behaviour; 4) children struggle with making sound relevance judgments. The first two issues are most pertinent to the study reported here.

van der Sluis and van Dijk (2010) report that the ISEs children use do not suit their information-seeking needs because they require complex knowledge about query formulation. Additionally, ISEs often provide results that do not answer children's information needs. The standard response to this problem is the development of specialised child-centred IR systems (e.g., Druin et al. 2003; Lingnau et al. 2010). For a number of reasons, these child-centred systems are not used in NZ classrooms.

3. STUDY METHOD

A series of structured observation studies were conducted by a single researcher to explore how children search for digital information using an IS

Participants. Our study was conducted in the Waikato School District, in the central North Island of NZ. Participants included children from two primary schools and a single intermediate school within Years 5 to 8. In total, 24 boys and 26 girls between 9 and 13 years old participated in this study.

Procedure. The observation study consisted of a set of five search tasks conducted by each student in the order listed (see Table 1). These tasks were developed to be educationally appropriate for children in these school year levels. Tasks 1, 2 and 3 were read aloud to the child with all instructions given by the researcher. Tasks 4 and 5 were given to the children as a printed hand-out.

Table 1: Five search tasks

1	Do you have a favourite book or sport or musician? ... Can you please search for information about ...
2	Where do possums live and are they a pest in New Zealand?
3	Mount Cook is a mountain in the South Island of New Zealand; let's find facts about Mount Cook.
4	How many rings does Uranus have, and what are they made of?
5	The Cricket World Cup will be held in New Zealand in 2015, find facts about the Cricket World Cup.

Apparatus. The children used school computers and the Google Chrome web browser that they were all familiar with. Observations were carried out in library or classroom environments while teachers, librarians, and classes shared the spaces. Each school provided different workstations. Sessions typically took between 30 and 45 minutes.

Data Collection. Sessions were video recorded (over the participant's shoulder), and additional handwritten notes were taken by the researcher. The researcher used ELAN software to perform qualitative annotation of video and audio moments and intervals with codes for head movements, mouse and hand movements, text entered onto the screen, mouse clicks, and mouse location information. For example, we coded head movements as follows: each time a child would move their eyes from the screen to the keyboard or mouse during search query entry the researcher placed the HM code on the ELAN timeline. These HM codes were tallied to give a per child count for each of the five search tasks and data was able to be later aggregated across cohorts.

4. RESULTS

We discuss here the results of our observation studies of children's Internet search.

Numbers of Queries. We coded the queries that children constructed once the child pressed the *Enter* key or clicked the search button. When a child started a query and changed their mind before pressing the *Enter* key, we coded and counted only the query that was submitted. Overall, the 50 children worked through 250 tasks, posting 550 queries. Y5&6 children made a total of 360 queries while Y7&8 children made only a total of 190 queries. The average number of queries constructed by Y5&6 children was 10.59 and 11.63 by Y7&8 children (approximately 1 query more per Y7&8 child than Y5&6 child across their observations).

Query Types & Reformulations. Children used both Natural Language (NL) and Keyword (KW) query construction strategies during our study. We identified two types of NL queries used during our study, both *Natural language sentences* (NLS), and *Natural language questions* (NLQ). The observed children did not construct Boolean search queries (e.g., using operators AND and OR) and only used the terms "and" and "or" as part of NL queries. We explored how children proceeded when their first search query did not resolve their investigation. Overall Y5&6 children and Y7&8 children were equally likely to reformulate their query. Only five students conducted all five of the search tasks without reformulating a query. Reformulations were used most often when the initial query was a KW. KW queries were most used initially, and most needed reformulating. NL queries were used more in reformulated queries than in initial queries.

Query Expansion & Related Searches. All but one child claimed that they saw a query expansion suggestion in the search box drop down and that they use query suggestions when conducting Internet searches. 46 of 50 children used query expansions at some point during the observation studies; only four children (all Y5&6) did not use a query expansion during any of their observed searches. No students at Y5&6 and only three students at Y7&8 were observed to use Related Searches in the SERP. 15 of 16 Y7&8 students claimed to have seen the related searches during this session while only 14 of 34 Y5&6 students claimed to have seen these. No Y5&6 student claimed to use Related Searchers during their normal search habits.

Link Selection. We found that the older children visited more websites to complete the tasks for all five of the tasks. The average number of minutes to complete a task was similar for Y5&6 and Y7&8 children, i.e, the older children were able to visit more websites in a shorter amount of time.

First Link Visited. We examined how many children visited the first link in the results list as the first   link they clicked. Of the 250 initial searches children visited the first link 84 times. We also examined how many children visited Wikipedia as the first link they clicked. Of the 250 initial searches children visited Wikipedia as the first link 50 times. In total, a Wikipedia link was visited 99 times by children during their search tasks. Only 11 students did not visit a Wikipedia link during any of their tasks.

Advertisements or Sponsored Links. Across all five tasks, nine children visited an advertisement (also known as a sponsored link) as the first link they clicked. In total 20 children visited advertisement during their studies. With 3 of 8 Y8 children visiting advertisements, this shows that advertisements can be confusing for children at all levels of our study.

Repeat Link Visits. We counted the number of times a student visited the same link before adjusting their search query. A total of 14 children visited the same link twice during their first query creation. We noted that eight Y5&6, six Y7&8 children visited   the same link twice. Children also visited links from a pull-box or sidebar and then revisited the same website via the search results list. Some children noticed that the information found in the pull-box or sidebar was the answer or a part of the answer (and announced this to the researcher), yet they still visited the same website via the search results list. Children appeared not to realise that the information has come from the same source.

Time on Task. Time is a common measure for usability studies, however, we argue that the speed it takes a child to complete a task or series of tasks for educational purposes may not be of educational concern and therefore, is a low priority measure for our study. We coded the length of time a search took from the time that the child began to type to the point at which the child indicated or declared that they were satisfied that they had completed the task. The Y7&8 children averaged less time than Y5&6 children to conduct most of the tasks.

Keyboard Use & Mouse Use. Many of the children did not touch type, nor did they keep their hands on the keyboard or mouse during searching or reading. This required the child to regularly move their eyes from the computer screen to the input devices and back. We counted the number of times a child looked up from their fingers to review what they were entering into the search box for each task.   All children looked up from the keyboard to the computer screen at least twice before firing a query.

5. DISCUSSION

We scrutinized the query-construction and search-term-adjustment phase of children's internet search.

Query Formulation. Literature on query formulation by children discusses the difficulty children experience when constructing KW queries and the incorrect use of NL query formulations (e.g., Bilal 2000). Our own related research (Vanderschantz et al. 2014b) found that children still believe that KWs are more appropriate than NL queries. The findings of our present study show many instances where children successfully found information using both NL & KW queries. Given the number of children who successfully used NL queries, clearly these search patterns are useful in modern ISEs. ISEs could better clarify this improved ability to handle NL.

Query Reformulation is an area that received significant discussion in the literature with many issues remaining unresolved. Older children than those in our study are reported to reformulate search queries a number of times in studies that observed them using dated internet search engines (Bilal 2000). This is evidenced in the present study whereby the modern ISE we tested did not explicitly teach users to construct or reconstruct queries. Interface developments for ISEs that illustrate successful query formulation practices will alleviate some of the difficulties experienced by children.

Our own related interview studies indicated that children were not able to easily identify promising query-reformulation strategies (Vanderschantz et al. 2014b). In our present study reformulations were required more often when the first query formulation used a keyword search. Children were observed to make fewer reformulations with NL queries. It may be necessary to further assist children in recognizing when natural language is an effective query process and to show them what to do when natural language searching fails. Children's difficulties in constructing good search queries has been linked to their lack of vocabulary or lack of cognitive structure (e.g., Druin et al. 2009; Bilal 2000). Our finding that keyword queries require more reformulations than natural language queries further supports this view.

Query support. Druin et al. (2009) reported that live spelling and query suggestions present in ISE search box interfaces did not assist children in their study because the children were looking at   the keyboard when typing. These findings differ from our observation that children shifted their attention from the keyboard to the screen often during query creation. Therefore, these children were more likely to review these live suggestions and take advantage of query expansions or spelling

corrections offered by the ISE. Related searches were indeed used seldom in our study (only three overall). A significant design floor in the implementation of related searches by many current ISEs is the location of related searches a significant distance from the search box (Vanderschantz and Hinze 2017).

Internet Search Procedures. In Vanderschantz et al. (2014b) children's query formulations were reported as likely to be very broad and involve a combination of KW and NL. Our present study confirmed this with observations of children creating broad search queries, and queries that were very similar to the task instructions. It was stated in (Vanderschantz et al. 2014b) that children would often click the top link or visit a known website (i.e. Wikipedia). Children in our present study often clicked the top link, and numerous children visited Wikipedia. If the child was satisfied with the answer, a second decision was required regarding triangulation and confirmation of the answer using multiple sources (Vanderschantz et al. 2014b). Here we observed very little evidence of triangulation.

6. CONCLUSION

This paper presented an analysis of children's query creation and use of search engine results pages in classrooms. We conducted a user-observation study with Year 5&6 and Year 7&8 children (aged 9 to 13 years old) and identified query-creation and reformulation strategies that children used. The children encountered a range of issues, in particular while creating searches as well as when triaging SERP lists. Our results emphasise the need for educating both children and teachers in how to construct better search queries as well as the need for ISEs to better assist users in query formulation. We observed that the traditional strong focus on KW searches is potentially holding children back, and that natural language queries were very successful for these children. Reformulation of queries was needed most for keyword queries. Because natural language questions required least reformulations, we suggest educating children to use natural language questions or natural language sentences. Finally, we identified the need for search engine interfaces to assist with formulating appropriate natural language queries as well as supporting triaging search engine results lists.

7. REFERENCES

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

Druin, A., Bederson, B. B., Weeks, A., Farber, A., Grosjean, J., Guha, M. L., Hourcade, J. P., Lee, J., Liao, S., Reuter, K., et al. (2003). The international children's digital library: Description and analysis of first use. Technical Report CS-TR- 4433 UMIACS, University of Maryland.

Druin, A., Foss, E., Hatley, L., Golub, E., Guha, M. L., Fails, J., and Hutchinson, H. (2009). How children search the internet with keyword interfaces. In *International conference on interaction design and children*, pages 89–96. ACM.

Duarte Torres, S., Hiemstra, D., and Serdyukov, P. (2010). An analysis of queries intended to search information for children. In *Symposium information interaction in context*, pages 235–244. ACM.

Lingnau, A., Ruthven, I., Landoni, M., and Van Der Sluis, F. (2010). Interactive search interfaces for young children-the puppyir approach. In *10th International Conference on Advanced Learning Technologies (ICALT)*, pages 389–390. IEEE.

Moore, P. (1995). Information problem solving: A wider view of library skills. *Contemporary educational psychology*, 20(1):1–31.

Timpany, C. and Vanderschantz, N. (2011). Learning outcome dependency on contemporary ict in the new zealand middle school classroom. In *12th Annual Conference of the New Zealand Chapter of the ACM Special Interest Group on Computer-Human Interaction*, pages 65–72. ACM.

van der Sluis, F. and van Dijk, E. M. A. G. (2010). A closer look at children's information retrieval usage: Towards child-centered relevance. In *Workshop on Accessible Search Systems, at ACM SIGIR 2010)*, pages 3–10.

Vanderschantz, N. and Hinze, A. (2017). How kids see search: A visual analysis of internet search engines. In *2015 British HCI Conference*. ACM.

Vanderschantz, N., Hinze, A., and Cunningham, S. J. (2014a). Current educational technology use for digital information acquisition by young new zealand children. In *Australasian Computer Science Conference*, pages 125–134. ACS.

Vanderschantz, N., Hinze, A., and Cunningham, S. J. (2014b). "Sometimes the internet reads the question wrong": Children's search strategies & difficulties. *American Society for Information Science and Technology*, 51(1):1–10.