Visual Exploration of Health Information for Children

Frans van der Sluis[1], Sergio Duarte Torres[2], Djoerd Hiemstra[2], Betsy van Dijk[1], and Frea Kruisinga[3]

[1] Department of Human Media Interaction, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[2] Database Group, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[3] Emma Children Hospital, University Medical Center Amsterdam, The Netherlands

Abstract. Children experience several difficulties retrieving information using current Information Retrieval (IR) systems. Particularly, children struggle to find the right keywords to construct queries given their lack of domain knowledge. This problem is even more critical in the case of the specialized health domain. In this work we present a novel method to address this problem using a cross-media search interface in which the textual data is searched through visual images. This solution aims to solve the recall and recognition problem which is salient for health information, by replacing the need for a vocabulary with the easy task of recognising the different body parts.

1 Introduction

Using health information is often problematic, being complex to find and difficult to read. For children, this becomes even more problematic. Often pediatric health information is not written at an appropriate readability level [1] and is difficult to retrieve for childen because of a lack of domain knowledge [2], whilst the benefits of pediatric health information to children are extra salient: e.g., in particular for ill children, allowing to overcome uncertainty about their disease.

Van der Sluis and Van Dijk [2] identified four salient problems children have with IR systems, one of which is the vocabulary problem. Numerous studies show children have difficulties in choosing the right words. Often, they misspell their keywords or use keywords that are too broad or too narrow [3]. These findings are often attributed to a lack of vocabulary, which is a known problem in IR research; i.e., the vocabulary problem [4].

Visual search interfaces allow to alleviate all facets of the vocabulary problem for children through removing the free recall of words. Tag clouds are well-known examples of a visual search interface [5]. A tag cloud gives a representation of the word frequency for the most used words in the underlying corpus. However, in the case of health information for children, also the recognition of commonly used words is likely not to be helpful enough.

P. Clough et al. (Eds.): ECIR 2011, LNCS 6611, pp. 788–792, 2011.

To overcome both the recognition and the recall problem, this demo will present a cross-media retrieval interface. This interface reduces the need for the recognition of words by using images to represent easy recognizable body parts. Moreover, to make more complex concepts available, users can zoom-in on certain body parts to make their search more specific. Hence, children can explore health related information through visually exploring the different body parts.

2 Design

The User Interface (UI) of the demo is shown in Figure 1. Two main functionalities are involved in this UI: visual search and results presentation, both will be further elaborated in the following paragraphs. Moreover, we will first describe the data set and the performed data enrichment allowing for a better mapping on the search metaphor.

Fig. 1. Illustration of the user interface

2.1 Data Set

The data utilized in this work consists of metadata describing the items available in the library of one of the largest children's hospitals in the Netherlands. These items represent a collection of diverse media types as books, dvds, coloring pages, etc. The purpose of the hospital's library is to provide trusted information for children about health, diseases, treatments and other physical and emotional issues that commonly arise in the hospital. Currently, the library consists of 560 items and each item is described with title, age appropriateness range, a short description, and an image.

To connect the interface with the search engine, we first determined the body parts and the level of granularity needed for the interface. For this purpose a dictionary of body parts was built based on the Wikipedia categories containing articles with body parts (e.g., *Anatomy*). The identification was carried out by simply matching the entries of the dictionary with the title and description of the items. Given that the recall of this approach is low since few items explicitely mention the body parts in the meta-data, we enriched the data by adding a list of body parts that are related to the content of the items.

Table 1. Top-5 frequencies per body part obtained by applying simple matching (a) and after enriching the data (b)

(a) Simple matching.

Body part	Frequency
Brain	15
Body	16
Head	6
Eyes	4
Ears	3

(b) Enriched data.

Body Part	Frequency
Brain	21
Body	19
Nose	12
Lungs	9
Eyes	7

This process was performed automatically by first identifying the entities from the metadata using Wikipedia, in a similar fashion as in [6]. We establish a relation between an item and a body part if the Wikipedia article associated to one of its entities is connected in the Wikipedia link structure with the article associated to the body part. For example, if one of the entities mentioned in the item description is *deaf*, our system is able to relate the item with the body part *ear* since its Wikipedia article is referred from the article *Deaf*. Using this method allowed us to increase the coverage from 28% to 53%. Table 1 illustrates the most frequent body parts found in our dataset before and after applying the method described to enrich the data.

The search is performed by constructing a query based on the body part that is clicked. This query is send to the search engine. We employed the Pf/Tijah engine [7] to index the title, description and the augmented body parts found. The communication between the UI and search engine is performed via the open search protocol [8] which also allows other parties to safely search our data.

2.2 Visual Search and Results Presentation

The search-part of the UI uses an illustration of a body as a search metaphor. As indicated in Section 1, such a metaphor is expected to reduce the vocabulary problem, which is particularly salient with pediatric health information. As Table 1 illustrates, selections can range from the whole body to the brain. This considerable difference in the level of detail creates the need for a zoom-in. We solved this need by using one image which contains a great level of detail, allowing for an image zoom-in to very specific parts (e.g., the ears). Moreover, the image changes when zooming in to highlight the relevant aspects at that level of details (e.g., the brains).

The interaction has been kept deliberately simple: a point and click paradigm is used for both selecting and zooming in, using feedback to make the functionality intuitive. Certain parts of the image are highlighted at any time, indicating a click on the highlighted part will give search results on that part. Moreover, when clicking on a highlighted area, the image will automatically zoom-in, leading to new highlighted areas.

3 Discussion and Future Work

This demo presents a novel way to solve a challenging retrieval problem: making often highly specialized health information searchable by less experienced users. Although the presented visual metaphor cannot cover all the information in the data set it is highly useful for exploring a particular domain of information.

The presented system can work for data written in any language, not relying on the use of words to search, and can work for different visual metaphors and thus different domains as well: any data set can be enriched with its association to a visualized concept using the Wikipedia-based method presented in Section 2.1. However, the reliance on a fixed visual metaphor makes the system fairly query-specific. The idea of basic level categories, indicating that people have common basic concepts for which clear representative images can be found [9], and the availability of resources such as open clip-art can alleviate this problem in the long run.

The presented design incorporated basic findings on interaction of children with IR systems, making it a truly user-centered design, aimed at some of the most salient problems in a particularly difficult domain. The presented demo is ongoing work and will be employed in a children's hospital where each patient has a touch screen to be used for entertainment and informative purposes.

Acknowledgements

This work was 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. D'Alessandro, D.M., Kingsley, P., Johnson-West, J.: The Readability of Pediatric Patient Education Materials on the World Wide Web. Arch. Pediatr. Adolesc. Med. 155, 807–812 (2001)
2. Van der Sluis, F., Van Dijk, E.M.A.G.: A closer look at children's information retrieval usage: Towards child-centered relevance. In: Proceedings of the Workshop on Accessible Search Systems held at the 33st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, New York (2010)
3. Bilal, D.: Children's use of the yahooligans! web search engine: 1. cognitive, physical, and affective behaviors on fact-based search tasks. Journal of the American Society for Information Science 51, 646–665 (2000)
4. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. Commun. ACM 30, 964–971 (1987)
5. Hassan-Montero, Y., Víctor Herrero-solana, A.: Improving tag-clouds as visual information retrieval interfaces. In: Merída, InSciT 2006 Conference (2006)
6. He, J., de Rijke, M.: An exploration of learning to link with wikipedia: Features, methods and training collection. In: Geva, S., Kamps, J., Trotman, A. (eds.) INEX 2009. LNCS, vol. 6203, pp. 324–330. Springer, Heidelberg (2010)

7. Hiemstra, D., Rode, H., van Os, R., Flokstra, J.: Pftijah: text search in an xml database system. In: Proceedings of the 2nd International Workshop on Open Source Information Retrieval (OSIR), Seattle, WA, USA, Ecole Nationale Supérieure des Mines de Saint-Etienne, pp. 12–17 (2006)
8. Clinton, D.: Specification opensearch 1.1 draft 4 (2007)
9. Hoenkamp, E.: Why information retrieval needs cognitive science: A call to arms. In: Proceedings of the 27th Annual Conference of the Cognitive Science Society, vol. 2005, pp. 965–970 (2005)