Information Search for Children Using Faceted Navigation

Kashif Aman Ullah
Department of Information Technology
University of Gujrat,
Gujrat, Pakistan
kashif.amanullah@uog.edu.pk

Muhammad Aksam Iftikhar
Department of Computing and Information Sciences,
Pakistan Institute of Engineering and Applied Sciences,
Islamabad, Pakistan
aksam.iftikhar@gmail.com

Abstract—Information search is one of the most common activities over Internet. Several search engines crawl the web round the clock and fetch useful information according to users’ interest. To date, most of these search engines exploit the keyword based searching procedure which might be inappropriate for some naive users such as children. It may be very intricate for children to search their desired content effectively using the traditional ‘keyword based’ search mechanism. They may fail to map the desired concept into appropriate keywords. To circumvent this situation, Faceted navigation is proposed as an effective alternate search method for children. Faceted navigation, also termed as Faceted search, is a specialized form of exploratory search that allows the users to search a domain on the basis of its attributes. We have performed several experiments to compare faceted search to the traditional query-based search mechanism. In particular, a faceted search based test interface has been designed. Children were asked to perform some pre-designated tasks on this system and their performance was analyzed on the said faceted search interface. A quantitative comparison was made with a query based search system using few quantitative performance metrics: keyword and spelling mistakes, number of missed pages and average accuracy. It is evident from experimental results that children performed much better on the proposed faceted search interface than query based search system.

Keywords: faceted search; exploratory search; information search for children; keyword-based search

I. INTRODUCTION

World Wide Web (WWW) is, no doubt, an imperative source of information in current era. Internet has made the world’s knowledge only a few finger tips away. Any information, available on internet, can be found with help of different search engines. Search engines, as compared to the serendipitous browsing [1] in early years of web, have enabled a very convenient method of information searching for web users. The only demanding task is to map your desired information into appropriate keywords for a search engine. However, this process is sometimes not as simple as it sounds. It can be very difficult to define proper keywords and formulate appropriate query. Therefore, finding relevant information, in context of keyword based information searching, can happen to be really frustrating and time consuming.

Children comprise a big chunk of population on the web. They usually use internet for recreational activities like games, movies, music etc. Using internet academic purpose like learning counting, alphabets, solving assignments etc. is also very common among children. They have to rely on popular search engines supported query based searching to find their desired content. Most web pages and search engines are developed by setting focus on adult users [2]. But, they pose certain challenges to children like limited vocabulary and grammar, typing and spelling errors, getting over the information overload and interpreting the search results [3, 4]. These search engines do not help children much in formulating queries or avoiding mistakes [5]. Children need to learn and adapt themselves to situations [6]. They are still immature users of web, undergoing through the process of cognitive, physical, emotional and social development [7, 8, 9]. Therefore, an alternate search mechanism is highly desirable, particularly for children, which can alleviate the problems associated with traditional keyword/query based searching mechanism. The terms keyword and query are used alternatively throughout this paper, because a query, effectively, is a set of one or more keywords.

Faceted search is an alternate mechanism to hunt for desired content in the multitude of information on web. It promotes learning and enables users to draw new ideas about a domain/field, which they are searching, by just looking at the interface. It is based on relating your query to the attributes of the domain. Each field or domain has certain properties and attributes which makes it distinguishable from other fields. Faceted search makes use of this fact and enables people to search the information using the attributes of a particular domain [1]. A domain, divided into its characteristic attributes, known as facets, is easier to explore and search than the traditional keyword/query based searching. This is because, an overview of the field is provided in the form of facets and the concept of recognition is employed rather than recall as in query based search systems [3]. This is particularly relevant when the user has no or very limited knowledge about the domain. Faceted search, in this case, provides a potential starting point.

In its history, faceted search has observed frequent application to web searching. Relation Browser [10], FLexible information Access using MEtadata in Novel COBinations (FLAMENCO) [11] and mSpace [12] are just to name a few. These systems not only allow keyword based search but also
provide rich metadata based hierarchical categorization to help user guide to his required information. Some of the said systems use traditional faceted search style (e.g. FLAMENCO) while others use directional column faceted search (e.g. mSpace) [10]. Some faceted search interfaces are also specifically designed for children. KidsClick, for example, is a search engine designed by librarians especially for children [13]. It presents some material relevant to children in certain categories and sub categories so that children can find the required information by exploring the appropriate content. An interface design[14] based on PuppyIR[15] framework makes use of query expansion and search moderation techniques to help children find what they are searching. Knowledge Journey [16] is another web based search interface for children which provides audio based search as well as text and menu based search. However, the most similar work to that of ours is the International Children Digital Library (ICDL) [17] interface designed for finding children books in a faceted search manner. It is based on different facets or attributes of the books so children can search a book on the basis of its properties. It provides a visually appealing and interactive interface for children. Above mentioned Interfaces provide an alternate search mechanism more suitable to children requirements. But, such interfaces need empirical validation so that an insight is developed into their inherent advantages and decisive comprehension can be established regarding such interfaces. We have strived to contribute in this direction by quantitatively validating a faceted search based interface for children and comparing its performance with traditional query based searching mechanism.

In this research, we propose that faceted search should be employed as an alternate searching mechanism for children. We have designed a faceted search base prototype interface for children and compared its performance with traditional query based searching in order to validate the performance of the proposed searching method. The quantitative measures computed for both the system verify that faceted search is a preferred searching mechanism choice for children.

Rest of the paper is organized as follows. Section II describes the proposed scheme of searching and the prototype interface. Section III describes the experimental setup established to validate the proposed searching scheme. Section IV presents the experimental results and discussion. Finally, Section V concludes this dissertation.

II. PROPOSED SCHEME

We are motivated by several characteristics of the faceted search mechanism. The problems faced by children using traditional keyword/query based search systems include spelling and typing mistakes, limited vocabulary and inability to formulate appropriate query and interpret search results. A naive child may enter the cumbersome world of internet to find data relevant to his requirements. But, he may go through several hurdles until he reaches the desired content. First of all, deciding on what keywords to search for is an arduous task for him. Next, even if he chooses correct keywords luckily, the common typo hindrance is still there. Moreover, immature typing skills also add to complexity of the search process. Finally, after getting search results against his assumingly correct keywords, the child is posed to, probably, the most difficult task of interpreting the search results. The search results by traditional query based search engines are usually in order of thousands. It is no less than a nightmare for a child to scan these results in order to interpret and comprehend the results.

The limitations, associated with traditional query based searching, demand for a more convenient searching mechanism. We have proposed faceted search as an alternate searching mechanism for children in order to overcome the said limitations. Faceted search is a sub-category of exploratory search which enables users to search a domain, based on its attributes [18]. In order to validate the effectiveness of faceted search method for children, we have developed a specialized faceted search based prototype interface for children, named as Kidzz Search. The Kidzz search interface is a prototype version of the proposed faceted search system which will help children to find their desired content quickly and effectively. In the following text, we describe different elements of the proposed Kidzz search interface.

As shown in Figure 1, the Kidzz search interface comprises of mainly two portions, i.e. choose options and results sections. In the choose options section, the user is presented with various categorizations of the domain in which he is searching for some information. Whereas, the results section shows findings corresponding to the selection made by user in the choose options pane.

The essence of faceted search is to divide the domain in its attributes and perform search on the basis of these attributes. For the prototype interface, the domain of science is divided into four main categories: space, planets, human body and earth. These categories are by no means exhaustive and are just selected for demonstration purpose. An exhaustive list of domains and their corresponding categories requires considerable effort and is beyond the scope of this work.
To apply faceted search to the enormous amount of heterogeneous web pages, we need extensive metadata. But, web pages don’t contain enough metadata for the successful application of faceted search at large scale [19]. There are meta tags available in the web pages but they contain little information about the page and are usually neglected by most of the web developers. Therefore, in order to alleviate the problem of metadata availability, we have used different headings present in a web page as metadata. Headings dictate important information about the content of a page and can be used as facets in our faceted search interface to filter out the results. We have considered headings present in our data set (web pages) as attributes of the domain and presented them as facets in our proposed interface.

Figure 2 shows two facets under the space category in the science domain i.e. size and composition. The results section in the figure shows search results against selection of size and composition facets. Similarly, other search results are presented in the results pane for each selection in the options pane and facets narrow down these search results. The search results are presented in the results pane along with the facets which resulted in hitting the result; as in Figure [2] size and composition are displayed to the right of search results. The user can navigate directly to the location where the facet appears in the page.

III. EXPERIMENTAL SETUP

The proposed faceted search interface has been validated by performing several experiments and obtaining quantitative measures of performance. The performance of the faceted search interface was compared to that of a query based searching; the Sphider searching system. Query based searching system will be described shortly in more detail. The data set in this work consists of a subset of pages on open web which are relevant to the domain of science and its categories, already presented in Section II. Target subjects of this research were 58 Children with diverse background and possessing certain predetermined characteristics. They were asked to perform certain tasks on each of the searching system. In particular, children studying in some school at the age 8-13 were considered. A primitive knowledge of internet and search engine was also considered. After the children performed the assigned tasks for some time, the effectiveness of each system was measured using few performance metrics which will be described shortly.

A. Sphider

Sphider is a lightweight open source keyword-based web search engine that can be embedded in a web site to enable search functionality. It provides features similar to the most contemporary search engine, so we employed it to compare with the proposed prototype faceted search interface. Figure 3 shows a preview of the results returned by Sphider when the keyword sun was entered. We have configured Sphider to search for the input query from our restricted data set, instead of the whole world wide web.

B. Performance Measures

Quantitative performance measures provide an opportunity to compare the effectiveness of different elements of a study in a concrete manner. Decisive information can be elicited from such numerical comparison. We have employed three...
quantitative measures for comparing the effectiveness and efficiency of the Kidzz search and Sphider systems. They are described in the following text.

1) **Accuracy**

Accuracy is mapped to the ability of children to find the exact information they are looking for. More often they are able to find their desired information, higher will be the accuracy and vice versa. During information search, it may happen that children are not able to find their desired information due to two reasons. 1. The information is really not present on page because the search query was unable to find relevant content. 2. The information could not be found due to information overload. Both types of inaccuracies are considered when measuring overall accuracy of a particular system. Discerning the two causes of inaccuracy of a system helps comparing the relative contribution of each and extracting useful results. Accuracy can be computed by following simple expression.

\[
\text{Accuracy} = \frac{N_c}{N_{\text{total}}} \times 100
\]

Where, \(N_c\) represents total number of correct results retrieved and \(N_{\text{total}}\) represents total number of queries.

2) **Time Efficiency**

It relates to the amount of time it takes for children to reach the place which contains the accurate information, being sought by children. Time efficiency of searching is an important factor because a delayed information availability increases the frustration and desperation of children which may result into relinquishing the search process altogether.

3) **Error Count**

Error count refers to the number of errors committed by children in typing and spelling the keywords during search process. Children may incorrectly spell keywords and make typing mistakes. These factors contribute to the deterioration of search results relevance. N.b. error count is only feasible in case of query based interface. For faceted search interface, it is implicit that error count will be 0, because no typing is required by children. Therefore, the proposed faceted search provides an inherent advantage in terms of error count.

C. **Tasks**

The tasks assigned to children, while using the systems, relate to the domain under consideration. All the tasks demand children for finding certain information about different categories in the domain. When using Sphider, children tried to map the information required by each task to corresponding keywords. On the other hand, children used facets for different information required, in case of the faceted search based interface. Following tasks were selected regarding the preselected *science* domain.

- Describe the composition of solar system and its major components. How much mass is distributed to Sun?
- Which is the largest asteroid and what is its size?
- What are the different types of meteorites?
- What is the Size of Sun? Find the mean surface temperature of Sun.
- How gravity was discovered, who discovered it and under what circumstances?
- Compare the volumes of planets Mercury, Mars and Neptune. Which is the biggest planet out of these three planets?
- Describe the various nervous systems present in human brain.
- How human heart functions? Describe its structure and major parts.
- Describe the different component of the blood and few of their characteristics. What are the different blood groups and which one is the most common?
- What are the major functions of human skin? How skin is divided into different layers and what are the various colors of the skin?
- What are various parts of human eye?
- What are the different types of volcanoes? Present a comparison.
- Which are the ten longest rivers in the world? Explain the topography of rivers.
- What are fault-block Mountains?

IV. **Experimental Results**

All the children involved in the experiment were asked to perform designated tasks by hunting for their desired content. They used both the proposed Kidzz search system using faceted search mechanism and Sphider using keyword based searching to find their desired information. In the meantime, their performance was recorded in terms of performance parameters mentioned in Section III. In this section, we will discuss and analyze these performance results in order to compare the performance of both the systems.

First, we present the accuracy results obtained for both the systems after children extensively used them to perform the designated tasks. Results reveal far better performance of Kidzz search. Children were able to search their desired content in order to perform a particular task with more than 90 percent average accuracy. This result was way superior to that of Sphider, which emerged to be about 52 percent only. A graphical comparison really provide useful insights and is given in Figure 4. The graph shows prominent accuracy advantage of the Kidzz faceted search interface as compared to the keyword/query based Sphider interface. Not only overall accuracy is better for Kidzz search, but it is also more stable in accuracy as evident from the range of accuracy values of both the interfaces.

The accuracy of both the systems is affected by two factors mentioned in Section III. Information overload can be related to cognitive inability of children to digest huge amount of information. Whereas, absence of information is pertinent to the undesired feature of search engine to fetch irrelevant information against a search query. Note that absence of information can also be due to cognitive inability of children because they were not able to formulate appropriate queries to search desired content.
Figure 5 shows a comparison of number of pages missed due to information overload and absence of information while using the query based Sphider searching system. Note that the results are not presented for the faceted search based Kidzz search system, because a faceted search mechanism inherently solves these problems. From visual inspection of the graph, it is obvious that absence of information is more challenging as compared to information overload. Hence, it can be inferred that the query based search engine itself is more defying in the searching process than limited cognitive ability of children.

Next, we describe the typing and spelling mistakes committed by children while using the keyword based Sphider search system. The faceted search solves these problems intrinsically because typing is not required to search desired content. Figure 6 shows number of keyword and spelling mistakes committed by each child on Sphider. All children committed at least 2 mistakes and at most 9 mistakes. The faceted search interface eliminates the effect of these mistakes by providing an altogether alternate searching paradigm.

Finally, we describe the efficiency of both system in terms of average time taken by children to complete all tasks on Kidzz search and Sphider system. A lower average value dictates better efficiency and vice versa. Figure 7 compares the time taken by both systems graphically. It can be observed with bare eye that Kidzz search was efficient by a large scale as compared to the Sphider search system. The average time taken by all children to perform the designated tasks on Kidzz search was about 0.53 minutes, which was about 4 times better than that of Sphider i.e. 2.06 minutes. The results confirm that the proposed faceted search based Kidzz searching system better correlates with the mental requirements of children. They can elicit their desired content from the results retrieved by Kidzz search more effectively and efficiently as compared to the Sphider searching system.

V. CONCLUSION

Information searching has always been one of the most common activities over internet. Online community uses different search engines for finding their desired content. Most of these search engines employ the keyword based searching mechanism. The user is required to translate his requirement into a set of keywords, called a query, which can be input to the search engine. However, such a mapping from real word information into appropriate keywords, sometimes, is not straightforward. This may not be a trivial task for some users on internet, like children. In this study, we have proposed an
alternate searching mechanism for children to overcome such hindrances in the search process.

Children do not only have problems in mapping their desired information into appropriate keywords, but they also commit spelling and typing mistakes and often get overwhelmed by the information mess caused by traditional query based search systems. The proposed alternate search mechanism, known as faceted search, allows children to search a domain on the basis of its characteristic attributes and hence helps in alleviating the aforementioned problems. We have designed a faceted search based prototype interface to validate this fact. The performance of faceted search based prototype interface is compared with a query based open source search system, known as Sphider. Children performed several pre-designated tasks on both systems. Afterwards, the performance of children on both systems was compared with the help of several quantitative metrics like accuracy, typing and spelling errors and time taken to complete the designated tasks. Experimental results in all cases reveal the superiority of the proposed faceted searching mechanism over the traditional query based searching for children.

We can extend this research effort by validating the proposed faceted search mechanism by defining more accurate metadata. In this research, we only used headings and subheadings in web pages as metadata. More useful and relevant metadata can be obtained by some more sophisticated techniques. Intelligent techniques for crawling the web pages, for example, can be a feasible option to extract useful metadata from web pages.

REFERENCES