Overview On Methods Of Searching The Web

Introduction

World Wide Web (WWW) is the ultimate source of information. It has taken over the books, newspaper, and any other paper based material. It has become the first preference for many people for finding their required information. Curiosity is built into human nature. People always want to explore and find new information about different things. A person goes through the process of learning since his birth to his death. All the progress man has made so far is the result of learning and his curious nature to explore more. Before the advent of Internet, people used to quench their thirst of information through the use of books, magazines, journal, newspapers, etc. But all these things were not in the reach of every man and also they cost much more. And then Internet happened which revolutionized the way people lived their lives. Internet has made the world's knowledge only a few finger tips away. Now, if someone has any information need all he has to do is just log in to Internet, open the search engine of his choice, enter a keyword and here you go. He is presented with millions of results matching his query, so far so good. But is all this process that easy as it sounds? The answer is No. There is no doubt that Searching technology has come a long way. A lot of effort and progress has already been made in the web search paradigm and it is still continuing. No wonder the searching is the second most popular and common activity over the internet after email. Despite all that, there are still lots of issues with web searching. We cannot say by any means that, it is a solved problem. It still can be really frustrating and time consuming to find relevant information.

In the early days of Web, when there was no prominent search engine, serendipitous browsing (Schraefel 2009) was a common feature. People used to navigate the web and noticed the interesting pages and items that come along. And then search engines came which changed the way people used to find the information. They have become so popular that people don't even know what other means were used in the past to find information.

1.2 Modes of Searching the Web

Searching the web for some particular information can be performed in two ways:

1.2.1 Keyword Search

This is the most common and widely used form of the web searching. All the popular and major search engines employ this form of searching. It has become so popular and people have got used to it so much that it is very hard to think of any other way. Google, Yahoo, Bing, etc. all use this type of searching. User is given a simple text box where he can enter anything and after hitting search button he is presented with many results depending on the query.

There is a variation to this which is called fielded search where user is given options of different fields and user enters a keyword after selecting the field of his choice like title, author, etc. and search is performed for only those documents which are related to the selected field.

1.2.2 Browsing

In this type of searching user is presented with already defined fields and categories and he scans the pages for the information of his interests.

Many search engines try to bring about certain tweaks, improvements, and optimizations but they achieve this by changing interface, improving algorithms, modifying presentation of the information and result. But the basic philosophy or model to searching remains the same i.e. either searching or browsing. According to Bates (Bates 2005) information seeking is performed in four ways. Number one is by actually searching and looking for the information when a user has a clear and well defined goal and he knows where to look for this information. Second is by monitoring the information. This happens when a user has a main interest in something but he does not actually look for the information instead grabs that information when he comes across something interesting in life. The third way is browsing for information when the user has no specific goal and interest and he is just exploring the information space just to gain knowledge or to achieve some vague goals.

Fourth way of information seeking is by being aware of information. Users have no clear goals and directions and do not look for information but absorb certain knowledge and information unconsciously throughout the life.

Keyword base searching model which is followed by many popular search engines like Google, Yahoo, Bing, etc. has become so common and prominent that now people can't think of any other way to search information. People have become so used to seeing a search box and enter a term and get a ranked list of results in return that it is very difficult to convince them to search differently. It is true that Google and other search engines do succeed in finding and fulfilling basic information needs. Most of the queries given to a search engine are very common queries like name of capital of country, population of a country, date of birth of a famous person, and queries like that. Google and other search engines specialize at finding such information and they return the required results within milliseconds. So we can say that problem of finding known and common information is almost solved.

1.3 Why Exploratory Search?

When it comes to the vague and complex queries, exploratory search comes into the picture. Few years back researchers realized that the traditional model of searching is not ideal for all situations. The reason is simple. User is presented with a simple search box and he is expected to translate his information need in few words accurately. But it

is not that easy all the times to express the information need in few words and even people cannot find the words to express the query in some situations. It is similar to the case when a patient goes to the doctor and tells him that he has pain in his arm. Now the doctor has to distinguish his pain from many possibilities ranging from heart pain to a simple bug bite. So this situation motivated the idea of exploratory search.

1.3.1 Definition of Exploratory Search

What exploratory search actually is? Marchionini (Marchionini 2006) described it in terms of information seeking problem context having the attributes of being:

Open ended

Persistent

Multifaceted

And information seeking processes described by being:

Opportunistic

Iterative

Multi tactical.

What it means is that when a user is searching for some information which is loosely defined i.e. does not have specific targets, requires time and effort to be solved i.e. user cannot just enter a query and find the required information, and constitutes of multiple components which need to be explored to fulfill the information need. In other words, exploratory search corresponds to information seeking process which requires multiple iterations to satisfy a user need and user has to constantly come up with different ideas to go deep into the domain and find the information.

Exploratory search is important due to various reasons. The obvious inclination towards keyword search has made people think that this is the best way to search for information. This is not true in many situations and people realized this when they failed to find required information in certain cases using keyword search.

The amount of information over the Internet is increasing exponentially every year. A single query given to a search engine results in hundreds of documents and hence it's really difficult to find the relevant documents. This problem becomes worse when user is not sure what he is actually looking for i.e. he is just exploring for the sake of knowledge, fun, or learning. These issues made researchers to come up with the idea of exploratory search to aid searchers in such situations. The aim of exploratory search

is to provide a solution to the problems found in the existing interfaces when the task of user is the exploration of information.

There are various reasons why people search for information. The reason could be the school assignment, homework, medical information, infotainment, or user could just be exploring the information space for no specific reason. The goal could be as simple as finding the name of the capital of a country, name of the president, or it could require thorough understanding of multiple fields to come up with the right answer. If the search task has the properties of being vague, complex, and users have no reasonable idea of how and what to find then the task falls under the category of exploratory search. Exploratory search requires enhanced user-system communication. Currently, when user submits the query to the search engine his work is almost finished as he just has to find the appropriate results presented in front of him. There is little or no interaction provided by the search engines. Search engines developers have realized this mistake and they are constantly coming up with different and new ideas. For example, suggestions provided by search engines when a user is typing a query, search pad provided by Yahoo, which keeps track of search activities, etc. Similarly there are many other small improvements developers working on to support exploratory search. For exploratory search to be successful there is a great need for enhanced mechanisms to improve the communication between users and the system through the use of better presentation of results, enhanced query reformulation, improved visualizations and visual queues, and other techniques. What these techniques do is they try to engage the user in the process of searching. Currently users don't have to do a lot while searching and most of the work is done by the search engines. By making this process more interactive the user will be focused more towards the task and in turn speed and performance will be improved.

Exploratory search has become quite popular in the recent times and is an active field or research these days. There have been quite a few implementation of systems following exploratory search and even popular search engines are giving it a thought that how to merge exploratory search in their existing setup. All this goes on to show the importance of exploratory search.

1.4 What is Faceted Search?

Faceted search (White 2009) is the most prominent form of implementation of exploratory search. Many applications of exploratory search use faceted search as the main tool. Faceted search enables people to perform tasks which are not easy to perform using keyword based search. Each field or domain has certain properties and attributes which makes it distinguishable from other fields. For example if a search engine is to be developed which searches the books, then there are certain properties which can be associated with each book like, title, authors, publish date, cover, etc. So what faceted search does is make use of this fact and enables people to search the information using the attributes of that particular domain. There are many advantages of using this approach. First when a domain is divided into its attributes then it is easier to

explore and perform search because an overview of the field is provided in the form of facets which helps in understanding and learning new things about the domain. This type of interface is particularly useful in the case when the user has no or very limited knowledge about the domain and he does not know where to get started. Faceted search provides a good starting point.

1.5 Faceted Search for Children

Faceted search due to its inherent advantages can prove to be a very helpful tool for children to help them in finding information over the Internet. Children make a big chunk of population on Web. Due to advancement in technology and dramatic drop in prices over the past few years, many people can afford computer and Internet now. Many people have computer and Internet at home which plays a part in exposing children very early in their lives to wired world which has the potential of being helpful or harmful at the same time. At early ages when children just start using and exploring the Internet they don't have enough knowledge and experience. They are mostly attracted by the colors and the very mysterious nature and curiosity level surrounding it. Many children start from playing games and involve in other fun activities like learning alphabets, counting, etc. and they are usually guided by their parents or other family members. They don't usually engage in difficult tasks like browsing, searching, etc. When they start growing and get a better understanding of things they start expanding their activities. They start using the search engines for finding information related to their school work, assignments, sports, music, games, etc. It is at this point when they come across different problems due to their lack of skills and very nature of search engines in particular and Internet in general.

1.6 Problems Faced by Children While Searching

Let's take a close look at what these problems are.

1.6.1 Limited Vocabulary and Grammar

At early ages, children have limited knowledge of vocabulary and grammar. They lack the ability to think deep and choose most appropriate words describing a situation. They use the words they find most easy to use and which first come to their minds. Now keyword selection has been and remains both critical and harder part of searching process. Even adults find it difficult to come up with a proper keyword describing the essence of the information need then how can we expect children to be good at it. The reason is simple. People are not always clear what they want, and even if they are it is difficult sometime to translate that need to a few words in a way that is understandable for the search engine. A search engine will definitely accept the input and present the results according to the keyword, but what if the keyword was wrong at the first place. This shows the importance of finding a proper keyword, because choosing a keyword is first step in finding the information and if a user starts from making mistake then the whole process will be a futile effort. The situation is even more dangerous when a user

thinks that the whole process has gone right and he is satisfied with the knowledge he has gained but the reality is totally different. This situation is really difficult to handle since there are no benchmarks or standards to evaluate that whether the knowledge being gained is intended for the current context or not. Same thing applies to children as well.

At small ages, children's mind is not developed and matured like elder people. Finding the right words to use as keywords is always a problem for them. Children often complain that they cannot find their desired information. The reason is that they do not provide a proper keyword. Faceted search can play a part to counter this situation. If children can be provided with clearly defined facets with properly defined hierarchy then the task of searching can become easier. If children can find important information about a particular domain at a glance it will obviously enhance their understanding and help them to make better choices and decisions about that domain. Children will be able to quickly scan the options provided to them and make a selection in a more efficient manner.

1.6.2 Spelling Mistakes

Another problem faced by children is that they are not good at spelling the words i.e. even if they are able to make up a proper keyword in the mind somehow, they cannot often spell it correctly. So they are faced by the challenge of not only thinking a right word but also to spell it correctly.

With the increase in use of computers, typing has taken over the writing. There are already talks of paper less environments i.e. to transfer all the written materials to digital forms so that there is no need for papers anymore. So now there are very few written files or papers as compared to the past. Most of these tasks are now performed over the computer by using word processing software like MS Word, OpenOffice Writer, etc. All these applications provide aid in spelling corrections by highlighting spelling mistakes. Despite all the advantages of this shift to digital world, it has a downside to it. People are not using their analytical skills to the optimum level and trusting too much the aids provided by the application they are working on as they know that even if they make any spelling mistake it will be pointed out. The obvious disadvantage of this approach is when a user makes a logical mistake rather than a syntactical and the application can't catch that. The point is this dependency on digital application work against users sometimes. This is also one of the reasons why not only children but also adults face difficulties in spelling the words. Nowadays people just type without thinking much about the spellings since they know that their word processing software will highlight any mistake. So if an occasion arise and that person has to write something himself without using any aid then he does face a lot of problems and makes quite a few spelling and other mistakes. So this problem also shows when people input a keyword to a search engine and they make mistakes in spelling. Many search engines have provided the option of automatic query suggestions to handle such situations. But it is not always sure that it will work because if a user is not sure about the spelling he may

chose a similar word which has the totally different meaning. So the end result may turn out to be very worthless. The very fact that information is provided in a very structured manner in a faceted interface makes it a candidate solution in such situations. The presence of all the domain information in a categorized and hierarchical manner eliminates the need to write a query and in turn there is no chance of spelling mistakes. All the children have to do is to choose a right category and its sub category to find their required information. All the fuss of first thinking an appropriate keyword and then spelling it correctly is avoided by using faceted interface.

1.6.3 Typing Mistakes

Problems faced by children while searching are not just limited to the problems mentioned above. There are also few other issues which should be considered if researchers want to help children in a more assured manner in finding the information.

In early stages of their lives children are not good at typing using keyboard. They have to constantly look down to the keyboard to type any words. Their typing speed is also very slow and they usually type using one or two fingers. So, all these problems make the process of searching quite slow and since children constantly look down to the keyboard and their focus is always shifted so they often miss the query suggestions and other hints provided by the searching engine and hence they end up in making mistakes. And this ultimately results in the frustration of children as they are not able to find what they want.

Faceted search can help here too. Since children just have to make selection from predefined categories hence there is no need of typing at all. All the pain of choosing and typing a suitable keyword is avoided using this approach.

1.6.4 Making Sense of Results

Even if children somehow succeed in thinking a proper keyword and writing it properly without making any spelling mistakes, the task ahead is even more bigger and challenging one. Inputting just a single query in the search box brings up about million of results out of which perhaps only few are relevant and meaningful which are usually found on the first result page. Children usually get confused with so many choices and results. They simply cant comprehend the linear list of huge result set and they don't usually go beyond the first few pages which is valid for adults too because it is an established fact that most relevant results are found on first few pages and most probably on the first page itself. This overwhelming amount of information related to a simple query causes children confusion and chaos. So there is a need to facilitate children in a better way. The results should be presented to children in such a way that it is easy for them to understand and find the relevant information. If the results are categorized according to their attributes with the help of faceted search then it may help children in easing their information hunt. Since the children will know the choices they

have made while selecting the appropriate facets so naturally they will look for the resultant information in the relevant categories and will find it there.

1.7 Proposed Idea

We tried to present few of the reasons why faceted search can prove beneficial to children. Above mentioned points provide enough encouragement to work on a faceted search interface which will guide children in the process of their searching. The idea we will be working on is described in the following lines.

Children usually start using the Internet as an educational resource when they start going to school. In early stages they use Internet more for fun purposes like playing games, movies, music, etc. But as the time passes and they make progress in life and school they start using Internet more and more for their school work, assignments, quizzes, etc. They start experiencing different search engines to find information related to academics and non academics topic for the sake of general knowledge and entertainment. Since many children spend a reasonable time on Internet searching for the academic information so we will be focusing on this domain. Since children have under developed skills at this early age so they usually search for clues to find their required information instead of working their way out themselves e.g. if a child is assigned a task to find information related to the topic space then most probably he will go to the Google type space in the query box and will get many results in return. But now he doesn't know what to do next? Space is very broad topic so the child will get confused how to move ahead. So if a child is provided already defined categories it will make much easier for him to navigate and find information. So we have selected a particular domain of Science and have divided it into its various facets like space, planets, human body, and earth similar to style they find in their text book in order to support familiarity. Each facet is further divided into its attributes to make the task of information exploration easier for children. Children can select any attribute. When children make their selection to find their required information further help is provided in the presentation of results by dividing the results in their respective facets. Now our task is to evaluate that interface to find whether it helps and improves the performance of children in supporting them and finding their required information quickly and easily when compared to the traditional way of searching. To carry out this task we will evaluate the interface by selecting few children who will use the interface to complete different tasks. These children will also perform the same tasks using Google or any other search engine. Then these results will be compared to see the difference that whether the hypothesis is justified or not.