Impact of Faceted Search on Search Performance of Children

BY

KASHIF AMAN ULLAH

Department of Computer Sciences
Quaid-i-Azam University
Islamabad, Pakistan
March 2011

Impact of Faceted Search on Search Performance of Children

BY

KASHIF AMAN ULLAH

SUPERVISED BY

DR. MUHAMMAD AFZAL BHATTI

Department of Computer Sciences
Quaid-i-Azam University
Islamabad, Pakistan
March 2011

Impact of Faceted Search on Search Performance of Children

BY

Kashif Aman Ullah

A Dissertation Submitted in the Partial Fulfillment for the Degree of

MASTERS OF PHILOSOPHY

IN
Computer Sciences

SUPERVISED BY
DR. Muhammad Afzal Bhatti

Department of Computer Sciences
Quaid-i-Azam University
Islamabad, Pakistan
March 2011

Impact of Faceted Search on Search Performance of Children

By
Kashif Aman Ullah

CERTIFICATE

A THESIS SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF THE MASTER OF PHILOSOPHY

We accept this dissertation as conforming to the required standards

1. _____
Prof. Dr. Muhammad Afzal Bhatti
(Chairman)

2. _____
Prof. Dr. Muhammad Afzal Bhatti
(Supervisor)

3. _____
Prof. Dr. Arshad Ali Shahid
(External Examiner)

Department of Computer Sciences
Quaid-i-Azam University Islamabad, Pakistan
March 2011

Declaration

I hereby declare that this dissertation is the presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly with due reference to the literature and acknowledgement of collaborative research and discussions.

This work was done under the guidance of Professor Dr. Muhammad Afzal Bhatti, Chairman, Department of Computer Sciences, Quaid-i-Azam University. Islamabad.

Dated: March 25th, 2011

Kashif Aman Ullah

Dedicated to

My Parents

Their continuous struggles and efforts made me achieve everything. I owe my every achievement to them

Abstract

Web searching is one of the second most common activities over the Internet. Search engines have revolutionized the way information is pursued. We all know that Google is the most widely used search engine and sets the standards for other search systems. People have become so much used to Google's style of searching that it is really hard for them to adapt to another style. But this approach to searching has various issues especially for a certain group of the Internet population i.e. children. Children face many difficulties in finding their desired contents because of the style of search engines. It is because children have specialized needs and most search engines are designed for elder people. So the style of searching which is most popular across the web is not the best one for children.

Faceted search is a specialized form of searching which allows the users to search the domain on the basis of its attributes. This quality of faceted search encouraged us to apply it to design an interface for children. The objective was to propose such a search

system which solves major problems of children which they face while using keyword based search engines.

Our hypothesis is that faceted search will help and improve the search performance of children when compared to query based searching. We proposed a faceted search interface to verify our hypothesis. We tested our proposed system with children and compared the results with keyword based search engine. The results of analysis and comparisons with a query based search system showed that children performed much better on the faceted search interface and most of their problems were solved.

Acknowledgements

First of all, thanks to Almighty ALLAH for giving me the strength and courage to accomplish my goal. He has been the biggest source of strength for me.

I would first like to owe my deepest gratitude to supervisor and chairman of Computer Science Department Dr. Muhammad Afzal Bhatti for his sincere support and the confidence he showed in me throughout the research. Without his guidance, motivation, and support it would have been impossible to remain firm in tough situations. His willingness to motivate contributed tremendously to my research.

I am very thankful to my teachers Dr. Shuaib Karim, Dr. Onaiza Maqbool, Dr. Abdul Sattar, Dr. Khalid Saleem, Dr. Hassan Mahmood, and Dr. Asif Ali for teaching me interesting courses and countless productive things and give me opportunities to build my future effectively. All the teachers guided me during the research especially Dr. Khalid Saleem who helped me in my survey.

My family has been the biggest support throughout the research. The prayers of my parents and support of my brothers made the complete phase of research easy for me. They stood by me all the way.

I am obliged many of my colleagues and friends. They provided me great support in many ways throughout the research. Special thanks to Danish Saif Talpur who was with me throughout the research and helped me in achieving my goals. I would also like to pay my gratitude to Zafar Saeed, Adil Javaid, Adeel Ahmad, and Javeria Kanwal for their support and valuable suggestions.

I would like to especially mention the name of Fahad Adeel Qazi whose valuable suggestion proved really beneficial in designing my prototype. I am really grateful to my very dear friends Awais Amanat, Alia Waqar, Ali Shayan, and Syed Muhammad Jaffar whose moral support encouraged me at every step.

And finally, I would like to thanks all my juniors and university friends particularly Latif Saif Talpur, Mohsin Ansari, Nauman Ali Khan, Rashid Naseem, Imran Zeb, Sohail Pervaiz, Mudassar Ali Khan, and Hanny Tufail whose company made this journey a truly memorable one.

Contents

List of Figures

Chapter 1

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 connect to the 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 was a common feature (Schraefel 2009). 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.

1.1 Modes of Searching the Web

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

1.1.1 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.

1.1.2 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 results. But the basic philosophy or model to searching remains the same i.e. either searching or browsing. According to Bates information seeking is performed in four ways (Bates 2005)

1. First 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.
2. 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.
3. 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.
4. 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.

1.2 Motivation for 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

1.2.1 Definition of Exploratory Search

What exploratory search actually is? Marchionini 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 (Marchionini 2006)

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.

1.3 What is Faceted Search?

Faceted search is the most prominent form of implementation of exploratory search (White 2009). 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. Faceted search 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.

1.4 Faceted Search for Children

Faceted search due to its inherent advantages can prove to be a very useful 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

1.5 Problems Faced by Children While Searching

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

1.5.1 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.

1.5.2 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 (Solomon 1993).

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.

1.5.3 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 (King and Alloway 1993). 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 (Crook and Bennett 2007). 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.

1.5.4 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. 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 (Naidu 2005). They simply can't comprehend the linear list of huge result set and they do not 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

1.6 Objectives

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.

The objective of our research is to exploit faceted search for solving the problems faced by children in information seeking process. With the help of faceted search it is possible to overcome the above mentioned problems. To validate our hypothesis that faceted search is appropriate for children we have selected a particular domain of Science. We will design a prototype based on faceted search and will evaluate that interface for the validation of our hypothesis.

Hypothesis

"Faceted search will help and improve the searching performance of the children when compared to query based searching"

1.7 Summary

In this chapter we gave an overview of the problem, problem area and proposed work. We touched upon what are the problems are with current search engines and why they are not suitable for children. We discussed the issues children face when they search for the information and proposed that faceted search could solve these problems. In the end we illustrated the objective and hypothesis of our study.

1.8 Thesis Outline

Rest of the dissertation is structured as follows

Chapter 2 describes the related research work to our field of study. In Chapter 3 we discuss our proposed interface based on faceted search. Chapter 4 describes the sampling technique, research design and experimental setup of our work. Chapter 5 presents the detailed analysis and evaluation of our results. Chapter 6 discusses research contributions and future work.

Chapter 2

Related Research Work

Searching the Internet for information is one of the most (if not the most) popular activity. People use Internet for various reasons but searching is one of the most common and popular reason. Searching is a widely and extensively researched issue and still remains the one. In recent times it has become an industry and Google has emerged as the clear leader. But others like Yahoo, Bing, etc. are also giving a stiff competition and trying hard to win people by introducing different tweaks and optimizations in the interfaces and implementation. Also the information available over the Internet is increasing exponentially and it is becoming harder every day to find desired information.

Most search engines do well in finding known and factual information e.g. Capital of a country, information about a movie, data of birth, and other information like this.

Searching problems catering such information have a very good success rate and modern search engines have seemed to solve the problem of general information searching.

But things start going wrong when someone doesn't know what he is searching or what should he search? When someone tries to learn about a new domain or a subject but he has no idea about the field then what he does, he writes a small query related to his domain and tries to find the information using his preferred search engine. The problem with this approach is that a small query will return many results but since the user has no general idea about the domain, he starts exploring different results without knowing what he should search. This is where the contemporary search engines lack in their functionalities. It is not that developers haven't yet realized this problem; in fact a lot of work is being done in this area. But it still remains an issue to search and explore.

2.1 History of Exploratory Search

Exploratory search has been around for quite some time in some shape or another, but it emerged as an individual and significant field of study in 2005 when a workshop on exploratory search interfaces was held at University of Maryland (White et al. 2005). This workshop brought together researchers from different domains like Information Retrieval (IR), Information Seeking (IS), Human Computer Interaction (HCI), etc and various issues related to exploratory search were discussed including but not limited to

interfaces, evaluation, cognitive processes, etc. This workshop proved to be a cornerstone for the field of exploratory search and brought to attention many new ideas and dimensions. It motivated researchers to set out on the path of exploratory search and explore these issues to find a reasonable solution to a very challenging problem.

This Workshop inspired researchers to write a special issue of Communications of the ACM in 2006 (White et al 2006). This issue discussed the progress made thus far in the field of exploratory search. It discussed the different systems supporting exploratory search and prevailing issues and directions for supporting it at bigger scale and levels.

In 2006 an ACM SIGIR workshop was held with the objective to focus on the techniques and tools necessary to evaluate existing exploratory search systems since it is imperative to analyze the impact they make on overall performance of the users while searching.

In 2008 a workshop was held by the name of "Information Seeking Support systems". Researchers realized that the term "exploratory search" can be confusing sometimes so they tried to put it under the broader category of Information Seeking Support Systems (ISSS). This workshop urged the researchers to come up with new ideas related to three main objective; better HCI models, new tools and techniques for supporting ISSS, and new techniques and way to evaluate ISSS. Researcher delivered many new ideas and various issues were also pointed out regarding models, tools, and evaluation.

In March 2009, IEEE Computer Magazine featured ISSS as the cover feature and published recent development and new directions for the future.

2.2 Explanation of Exploratory Search

Exploratory search, as the name suggests deals with the situations when the basic purpose of searching of a user is exploration. The field of exploratory search was started with the idea to improve the techniques, tools, and technologies to help searchers in the process of exploration. Aim of exploratory search is to make process of searching more engaging for user, where the user has more command and liberty to specify the information needs and play and manipulate results. Different researchers have tried to define the exploratory search in various ways. There are different terms used in literature like exploratory search, information seeking support system, interactive search, human computer information retrieval, all describing the same concept. Every type of search involves bit of exploration but there are certain properties which can be associated with exploratory search.

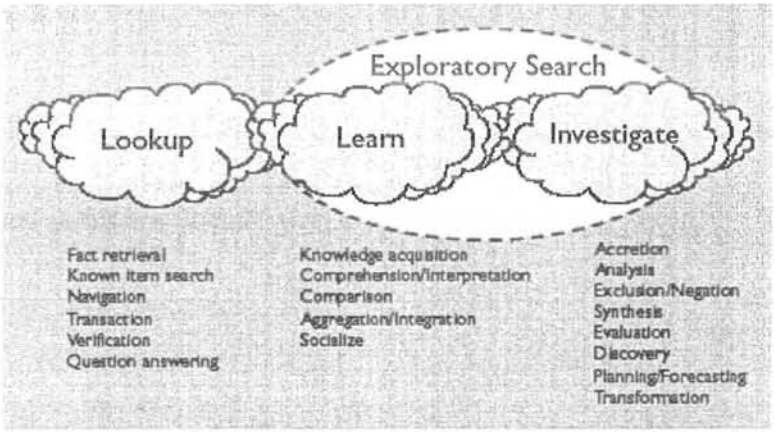

Fig 2.1 Search activities (Marchionini 2006)

White et al. (White et al 2006) stated that exploratory search is one which deals with the situations when the information need demands exploring the field while the users have no or very little knowledge about the domain they want to search and they don't know how to form the query or choose proper keywords. It also caters to the situation when the required information in not indexed properly.

Marchionini divided all types of information needs into three main categories lookup, learn, and investigate and stated that all search activities fall under these three categories (Marchionini 2006). He further described that lookup activities are the most basic type of search activities and mostly user succeed in finding their required information rather easily. Search activities in learn and investigate categories constitute exploratory search. Learning about some domain or investigating a query are the tasks which are harder to accomplish using the traditional search engine, are the main target of exploratory search.

2.3 Non-Exploratory Search

A question comes to the mind that what is not exploratory search. Daniel Tunkelang who is an active researcher in the field of exploratory search has answered this question in his one blog post (Tunkelang 2008). The definition of exploratory search tells us that ambiguous requirements, inability to form exact queries, and limited knowledge of domain are some of the characteristics of exploratory search. So, according to Daniel Tunkelang if we are sure about our requirements and know exactly what we want to complete the task at hand and have enough knowledge to build queries and ask it precisely then this type of search is considered non-exploratory search. Queries formed for exploratory search tasks don't result in either success or failure and if a task is completed by giving a query and getting an exact result in return then this task in non-exploratory. The reason is that exploratory search is a consistent process of learning and knowledge building. The tasks which can be completed shortly and exactly don't fall into the category of exploratory search.

2.4 Exploratory Search Systems

Many Exploratory search systems have been developed over the years.. Below is a brief overview of some of these systems.

2.4.1 Relation Browser

Relation Browser (Marchionini & Brunk 2003) aimsto provide an overview of a web space by exploring relationships among attributes in a small set and simple interaction mechanisms. It basically provides enhanced and more interactive overview of a web space as compared to simple sitemap and visualizations. It is important to provide an overview since then user is in a better place to understand the web space and it also helps him in exploration process since he has a general idea of what he can find here. So Relation Browser seeks to improve the situation further by providing additional aid and overviews.

2.4.2 Faceted Search

Faceted search is perhaps the most representative form of exploratory search. Whenever we talk about exploratory search the first thing which comes to mind is faceted browsing or faceted search. The reason is that the most systems which target exploratory search include some sort of faceted search. Examples are mSpace, Flamenco, Apple's iTunes, etc.

Faceted search allows users to search a domain based on its attributes (Schraefel 2009). It promotes learning and enables users to draw new ideas about the domain they are searching by just looking at interface. The reason is since the domain is divided into its

attributes called facets so a user can have a look at the interface and can extract certain information about the domain easily. It helps him in better understanding the domain and guides him in exploration of his required information especially when he has no substantial knowledge of the domain.

A brief description of the few popular faceted search systems is described below.

2.4.3 Flamenco

Flamenco (Yee et al. 2003) is a faceted browser which makes use of rich metadata and hierarchical categorization to help user to guide to his required information. FLAMENCO stands for FLexible information Access using MEtadata in Novel COmbinations. The information is organized in such a way that exploration of domain becomes rather easy and user can form and refine queries in more assured manner. User is presented with multiple choices to make his selection which enhances the chances of user finding his required information in any of the facets. Text based search is also provided along with faceted search.

2.4.4 mSpace

mSpace is a faceted browser that lets users perform not only keyword search but also faceted search (Wilson & Schraefel 2008). The domain is divided into its attributes which are shown in columns and search criteria can be specified by making selections in different columns. Columns are not fixed and can be modified i.e. new columns and attributes can be added or deleted. A useful feature present in mSpace is of backward highlighting. If a user selects any column from middle or right then not only this filter is applied to right columns but also all the related item to the left are also highlighted which makes it easier for users to make connections about the domain. Another feature is the ability to perform OR function in a column i.e. multiple items within a column can be selected. Sharing of information is also possible through the use of tagging, comments, and discussions. Preview cues can also be provided on demand which allows previewing the kind of information present in selected item and in turn help users in better exploring and querying the domain.

2.4.5 Apple's iTunes

iTunes is a multimedia application which lets users play different songs and tracks (Schraefel 2009). It provides the option to build a library from local and online collection and enables to search the collection on the basis of facets. Faceted search is performed on

the basis of three facets from music domain i.e. Genre, Artist, and Album. Selection made in any of the columns filters the results. It helps in drawing new conclusion about domains as well as provides unknown information. For example if a user selects the genre Jazz and then see an artist which he thought in the past that he only sings Pop music then this will be new discovery for him. Its interface is pretty similar to mSpace but backward highlighting feature is not provided in it.

2.5 Limitations of Current Search Engines for Children

Technology is advancing at a rapid pace and it has penetrated into our daily lives so much so that we feel helpless sometimes in its absence. Computers and Internet are now very common and can be found in many households. So children are exposed to Internet and Computers at very early stages of their lives. At early ages from 3-6 they use it mostly for playing games, watching cartoons, and other fun stuff. Then they start using it for searching and browsing information related to their school work, assignments, and other general information. But children face certain problems when they start growing and start using Internet more and more and these problems are caused by their limited skills and abilities due to their early age. Most pages and search engines are made by keeping adults in mind (Jochmann-Mannak 2009). Developers might have some thoughts about children at the back of their minds but most interfaces are developed for general audiences which

are mostly adults. Designers don't usually think that children have special needs and will suffer from certain problems while using this interface.

Searching for the required information seems an easy and straight forward task; just go to the search engine of your choice, type in a query, press Enter, and there you are, you have the desired results. But, as simple as it might sound, it is not that easy especially for children. People have to go through a complete process to find their desired information. When a child wants to explore a certain piece of information then first of all he needs to know what he wants to search. Since, children have limited vocabulary they find it difficult to choose a proper keyword for their information need. Their requirement might be clear in their mind but it is difficult for them to put it in two or three words. So they decide to put a longer query in natural language style when in doubt. Children are not good in typing. They have difficulties in finding the right keys which makes the process of typing very slow. Also they make a lot of spelling and typing mistakes (Druin et al. 2009) which results in wrong keyword being input to the search engine which will obviously return wrong results. When a query is input to a search engine it returns millions of result against that query. Children even adults cannot handle this unnecessary overload of information and it is also a known fact that most relevant results are usually found on the first page. So users really go to second or third result page. So this irrelevant load of information confuses the children and makes the task of comprehending the

results difficult. So basically children suffer from following problem while using a search engine to find information:

- Choosing appropriate keywords to form queries due to limited vocabulary
- Inefficiency in typing which makes the process of typing slow and also results in typing errors
- Spelling mistakes because of limited knowledge of grammar
- Comprehension of results due to the overload of information

As mentioned above, children suffer from these problems due to negligence of developers to specifically support children in the process of searching. We cannot expect children to follow the same steps for searching the information which are taken by adults. Their age factor demands specific consideration. They require special adjustments to be made to interface to enable them to search in a better way. Faceted search seems to provide the building block to cater to the problems of children in their hunt for information. So it's worth putting some efforts to investigate the impact faceted search can have on the search performance of children. It will be interesting to know the results because if the results are positive then researchers and developers will have a new line of direction to work. They can target specifically children and provide them with the improved interfaces based on faceted search since children are fast becoming a major part of Internet population.

Faceted search has the structure and properties to not only accommodate the above mentioned problems faced by children when searching but also makes the process of searching more interactive and intuitive. The ability of faceted search to split the domain in its attributes or facets automatically solves few of the problems that children have to deal with when they start exploring for information. Since the domain is already divided in various facets children don't have to think about keywords, they don't type so they don't make spelling mistakes. Everything is presented in front of them and they just have to make appropriate selections. Also, the presentation of results is also done in a more instinctive and comprehensible manner and helps children in make connections, and hence enables them to perceive results and make better selections. So faceted looks promising in the forefront but it needs to be analyzed and evaluated to see how much helpful it actually is.

2.6 Related Work

The faceted search has received a lot attention in the recent past. A lot of research is currently in progress to enhance the applications of faceted search. It has become a kind of fashion to provide faceted search along with typical keyword search. There are even talks of supporting faceted for open web (Teevan et al. 2008) but there are still many challenges in doing so as the authors have also identified some of these challenges. We

can find various examples of faceted search in our daily use of Internet. Sometimes we notice it and many other times we don't even know that it's faceted search or something different from typical search. EBAY, Amazon, ACM, IEEE, etc. are few examples of websites where we can search for our desired items using the combination of faceted and keyword search.

It is generally thought that faceted search can work well in specific domains since these domains have clearly defined metadata and lend them well for the implementation of faceted search. This is the reason why faceted search is very popular in the domains of shopping, online stores, groceries, music, images, etc. because these domains allow defining facets clearly and different items can be searched on the basis of these facets. For example in an online store we can find a mobile phone or any other item of our choice on the basis of different facets like price, manufacturer, operating system, etc. along with the combinations of keyword search. What actually happen is that user generally enters a tentative query and results are presented in front of him on the basis of that query. After that user can refine these results according to different categories or facets provided. Similarly, in digital libraries like ACM user can give a simple query related to the document he wants to find and matching documents are returned. After that, user can refine the results on the basis of different facets like people, publication, conference, etc. The advantage of all this is that user is never really stuck anywhere. He has something to work with all the time. In this way more interaction is provided to play

with results of the initial query. It's not like typical search engines where hundreds of results are available but all you can do is to check each of the items manually to see if it fits the requirements. So faceted search not only saves the time and efforts, it's also more effective and efficient if implemented properly.

Researchers have worked on various different projects related to faceted research over the past few years and have generated lot of ideas and directions to work on in the coming periods. Below is a brief overview the work that has been done related to the faceted search.

Faceted Search or faceted browsing both terms point to the same thing and both are used interchangeably throughout the text. According to Wilson faceted browsing falls into two main styles (Wilson & Schraefel 2008)

(i) Traditional faceted search

(ii) Directional column faceted search.

In traditional style of faceted search, collections are filtered automatically on the basis of the selections made in different facets and these filtered results can be refined further. FLAMENCO uses this style of faceted browsing

Fig 2.2 FLAMENCO Faceted Browser on Nobel Prize Winners

In directional column faceted browser the facets are presented column wise and any selection made to the facets in left column filters the result in right columns. Examples of systems which implement this type of faceted search are mSpace and iTunes.

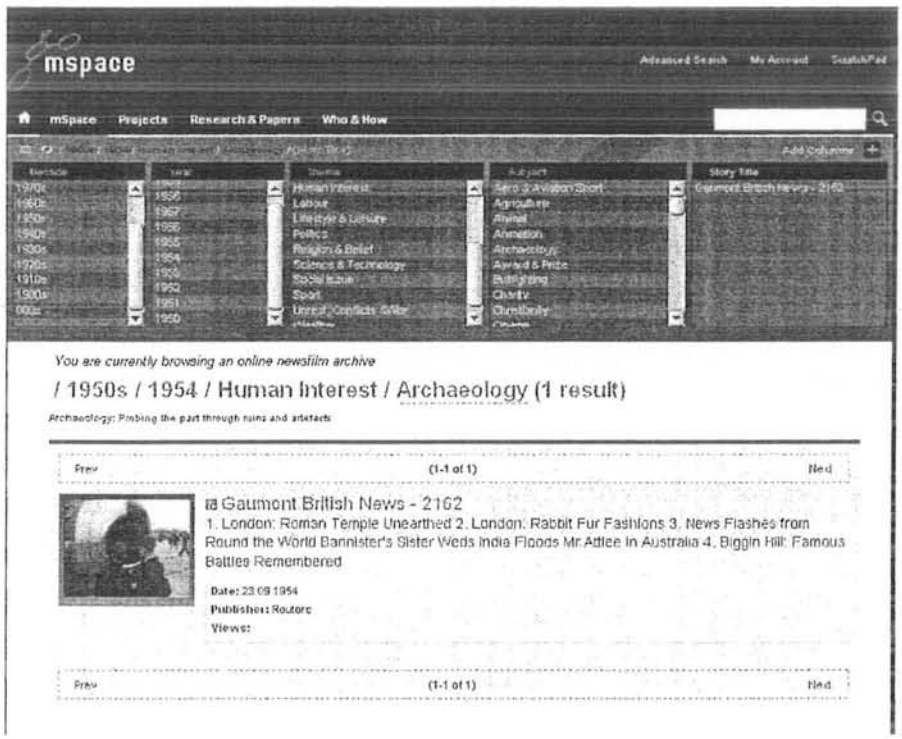

Fig 2.3 mSpace Faceted Browser on Online Newsfilm Archive

Social networking has become very popular in the past few years and it is now a fashion statement to use facebook, twitter, LinkedIn, etc. It's hard to imagine now that whoever uses Internet doesn't use any social networking website. So it is becoming increasingly important for social networking websites to provide tools to its users to find people they know e.g. their friends, class-mates, coworkers, etc. accurately and speedily.

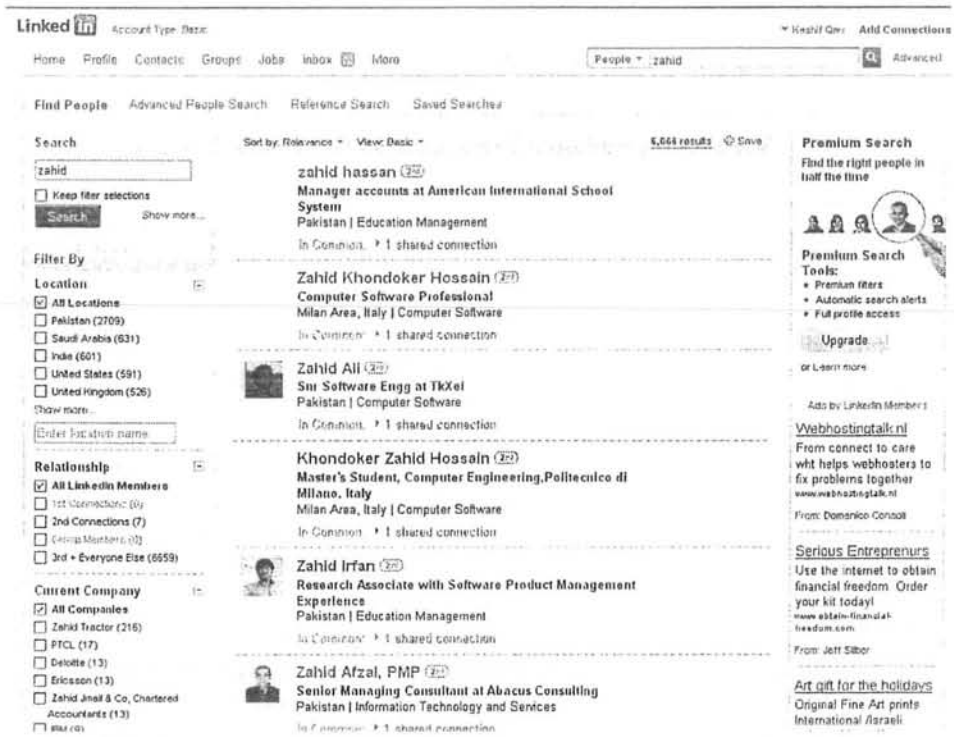

Fig 2.4 LinkedIn Faceted People Search

LinkedIn has taken up the challenge and realizing the potential of faceted search they have provided the faceted search to support people search. Different facets are provided and the users can manipulate the results on the basis of these facets until he/she finally finds the desired target. It is easy to understand and use and gives user more control over the results set and hence is much more interactive than typical static keyword based search.

Mainstream search engines like Google, Yahoo, Bing, etc. are now planning about providing exploratory search. No one is directly providing it yet but there are talks of and

even some efforts exerted on exploratory search for WWW. But there is still a long way to go to get any closer to the aim of providing faceted search for the open web as identified by the researchers at Microsoft (Teevan et al. 2008). The very nature of open web makes it a tough candidate to apply faceted search. The diverse nature of information available on the web and the lack of quality metadata for each document are few of the many hurdles in the way of faceted search for open web. Google tried its stint at faceted search by implementing Google Squared. What Google Squared does is it tries to find simple facts related to any term entered as input query by traversing web and the results are provided in an organized fashion. This organized presentation of information helps user in drawing an image of the domain and this way user gets better idea of the fields and hence is in a better position to perform further actions. Quite recently, Google has launched Google Instant to support users in the process of exploration.

One interested application of faceted navigation can be found on Nuggetize (www.nuggetize.com) webpage. Nuggetize traverses the web to find the most relevant information related to the query provided, analyzes each piece of information comprehensively, and organizes and presents the information in a categorized manner called nuggets. The idea is to help user get started in their quest of information by providing them exact and to the point information. It gives users more options to interact with the results and reformulate their queries so users are not stuck at a point and don't feel helpless, as they don't have to think about what to do now. Users can also provide

feedback about whether a result matches their requirements or not and the results are refreshed accordingly.

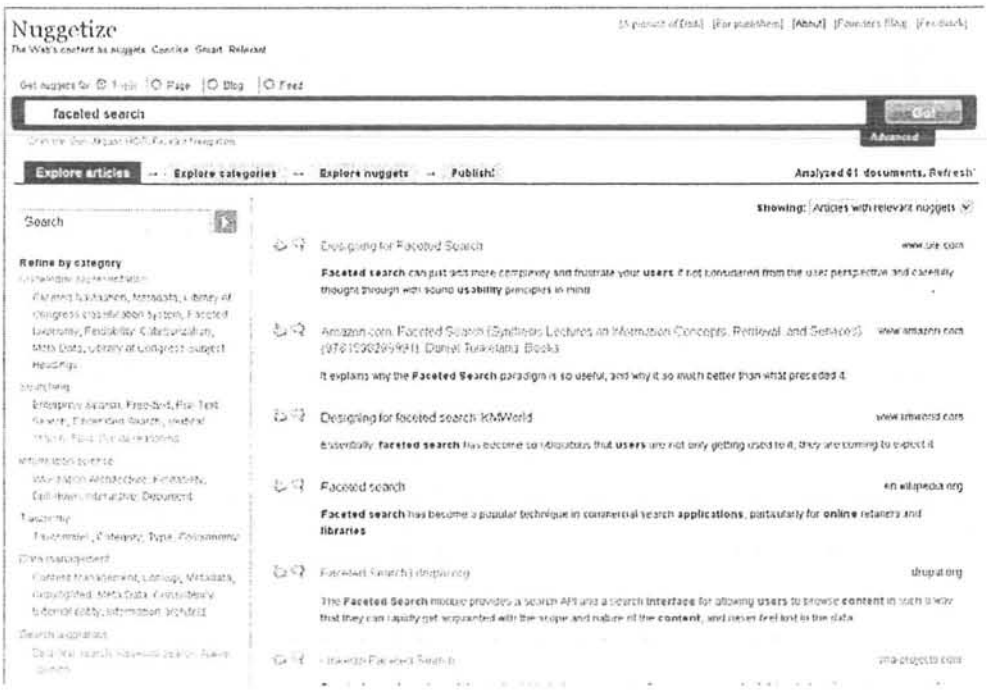

Fig 2.5 Nuggets for the query "faceted search"

Elastic lists are used to improve already existing faceted datasets by providing a dynamic and more visually appealing view (Stefaner and Muller 2007). It assigns weights to the items and various comparisons can be made at a glance. Animations are used when different filters are applied so that users get the idea of what is going on as compared to an abrupt change in view which happens in faceted interfaces without elastic lists.

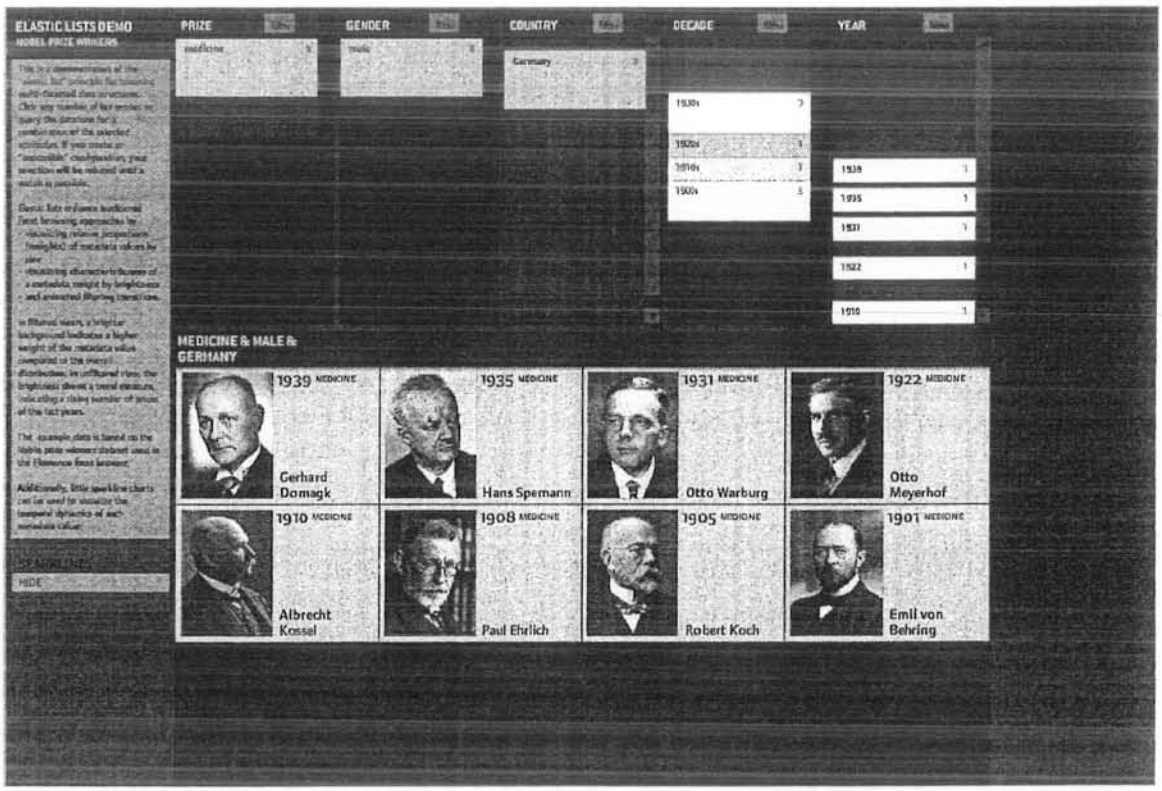

Fig 2.6 Elastic Lists on Nobel Prize Dataset

2.7 Interfaces for Children

When it comes to the web searching, children are never the central figure. Search engines are rarely developed keeping the special needs of children in mind. Grownups almost always play the role of target users by default. It is not that children have been ignored altogether. There are few notable researchers who are very active in the field of designing

interfaces specifically for children considering all their special needs and demands. Even in the past people have made efforts in this direction but this work is very limited when compared with the work done for adults. Despite, search engines which have been designed specifically for children are not comprehensive and do not help children much more than any traditional search engine will do. All they do is filtering of the results according to the needs and age of children, and objectionable pages are omitted. These search engines do not help children much in formulating queries or avoiding mistakes (van Kalsbeek and de Wit 2007).

The problem with traditional search engines is that they are designed implicitly for adults keeping their needs and abilities in mind. But there is a vast difference between the cognitive and analytical skills of children and adults. Both neither think nor act alike. Kids have limited set of mental and other expertise as they are still in the process of learning and growth and their skills are not yet developed to the full potential.

Researchers believe that children feel more comfortable in using browsing based search interfaces rather than keyword based search engines. The reason they give is that browsing tools are based on the concept of recognition and kids feel at ease to use such interfaces because it does not require a lot of effort on the part of children to look for information (Borgman et al. 1995).

Why children face problems in using keyword based or traditional search engines? People have explored and observed the searching habits of children and various difficulties they face and have found out the key problem areas. According to Borgman et al. children face problems in retrieving information from information retrieval systems due to their lack of skills in the following disciplines:

- Typing
- Spelling
- Vocabulary
- Alphabetizing
- Boolean Logic (Borgman et al. 1995)

Bilal and Kirby suggested that children need to learn more and adapt themselves to situations. They need to understand each task after careful analysis and evaluation and should plan accordingly to put them in a position to make better use of information retrieval system (Bilal and Kirby 2002).

As mentioned above, one of the main reasons of why children face many problems while using the search engines to find information is that search engines are designed by elder people for grown ups mainly. On the other hand, children are still in a phase where their developmental growth consisting of many aspects like cognitive, physical, emotional, and social is still in progress (Cooper 2005). It is not yet reached at a point where it is

comparable to the abilities possessed by adults at a much older age. So it is unfair to assume that children will be able to utilize the facilities provided by the search engines in the same way as adults. Even many adults face various problems in using the search engines even though they are much more experienced and skilled. So, how can we expect children to be good at it?

Druin et al. studied the keyword searching behavior and patterns of children aging 7-11 and found vital areas where the research is lacking (Druin et al. 2009). They closely worked with children and observed the problem areas for children where they face difficulties and commit errors. They made valuable and important revelations. They established that despite all the advancement in technology children still are hampered by the very basic problems i.e. typing mistakes, spelling errors, and keyword selection due to lack of vocabulary. Even the aids provided by search engines to minimize such problem such as auto complete and spelling suggestions features are ignored by children since children are busy looking at keyboard to avoid mistakes hence cannot focus on the screen. Children are also not good at complex queries when they have to search more than one item at a time. Children are not comfortable with so many results returned against a single query which confuses them and they don't usually go beyond first few results or the first page. All these things create a sense of frustration in children and they want more help from search engines. Druin et al. reported that children and their parents want the search interfaces to be more innovative and interactive (Druin et al. 2009). They

want interfaces which have prefixed and predefined categories so that they can make choices and search for information by clicking and don't have to type. Children also want that their queries should return limited set of results which cater to their needs and irrelevant results are minimized.

There are examples of few interfaces which are designed specifically for children such as KidsClick, International Children Digital Library (ICDL), Ask for Kids, Yahoo! Kids, etc. Some of these are good while some are not quite to the mark. But none of these systems or any other system fully captures the needs of children and provides the best solution. Below is a brief description of some of these systems.

KidsClick! (www.kidsclick.org) is a search engine designed by librarians especially for children. Its specialty is that some children related material and pages are assigned to relevant categories which are further divided into sub categories. So that children can find the required information by exploring the appropriate categories and subcategories.

Fig 2.7 KidClick! Web Search for Kids

When a child clicks a specific category the links of relevant pages assigned to that category are shown. It's a neat, simple and easy to understand interface and a keyword search is also provided along with categorical search. Its disadvantage is that these choices are not exclusive and do not cover a lot of information. Also it is a static interface with no updates with no changes to predefined categories or pages overtime.

Fig 2.8 Search results for category "Internet" at KidsClick

Safe Search for Kids powered by Google (www.safesearchkids.com) is keyword based search engines and returns only those pages which are relevant to the needs of children by using the safe filter.

Fig 2.9 Google Safe Search for Kids

It has a very simple interface just like Google. It provides a ranked list of results depending on the keyword. Google directory for kids and teens is also provided along with search engine. It also has the disadvantage of being static and not much of interaction is provided for children.

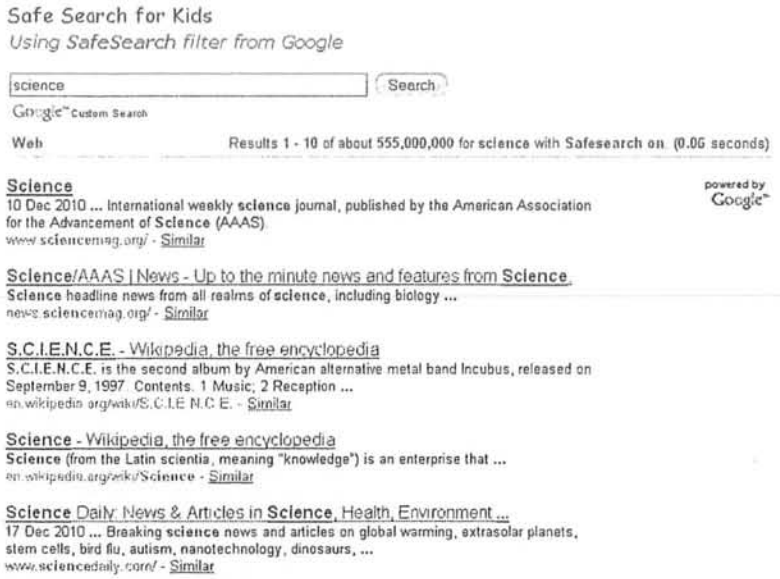

Fig 2.10 Result of query "science" at Google Safe Search for Kids

Search engines mentioned above and many other designed for children are mainly static and lack the quality of interaction and refinement of search results. Children themselves have to find the appropriate information from the list of ranked result which is not an easy task for them. International Children Digital Library is designed specially for children and consists of books for kids in many languages and related to many subjects. Its search engine is devised keeping the special needs of children in mind and the thought process they follow. It is based on different facets or attributes of the books so children can search a book on the basis of its properties. Children can search a book on the basis of its color, its length, its subject, and age group. It is more visually appealing to children.

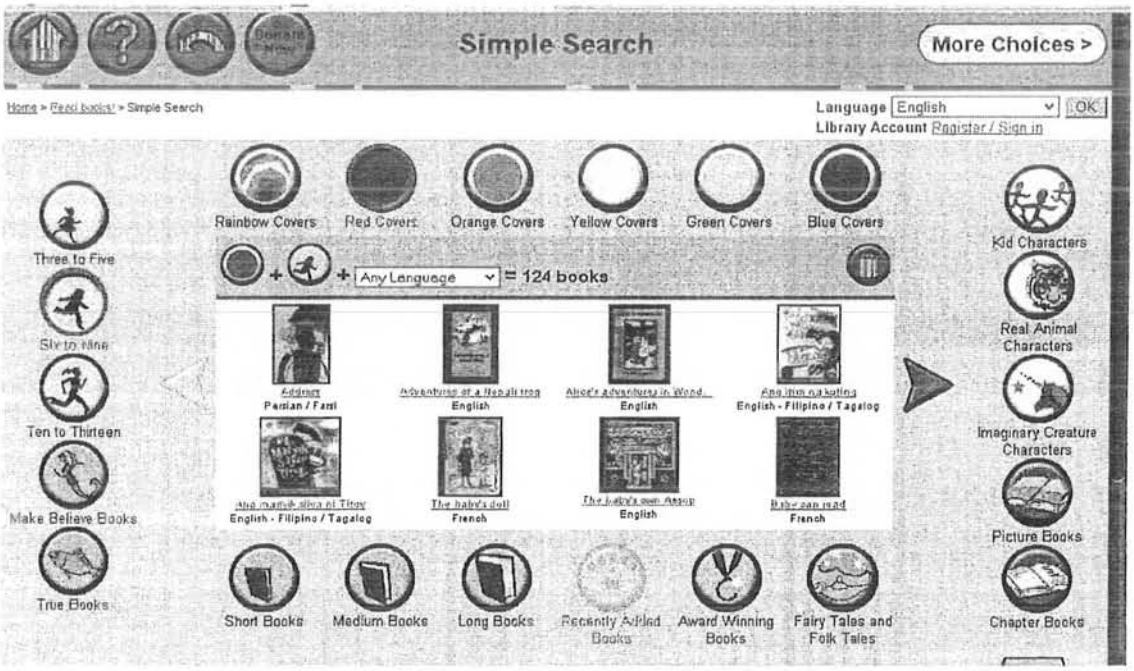

Fig 2.11 Simple search for Red Cover and Six to Nine age group books at ICDL

An advanced search interface is also provided where many other facets or options are provided along with the keyword search. It provides the chance to children to be more interactive with the results and they can refine their queries on the basis of many other choices available. But there is a drawback that it takes some time for children to getting used to such type of interface because they are accustomed to using traditional search engines. But there is no doubt that this interface does a good job of facilitating children in their information search.

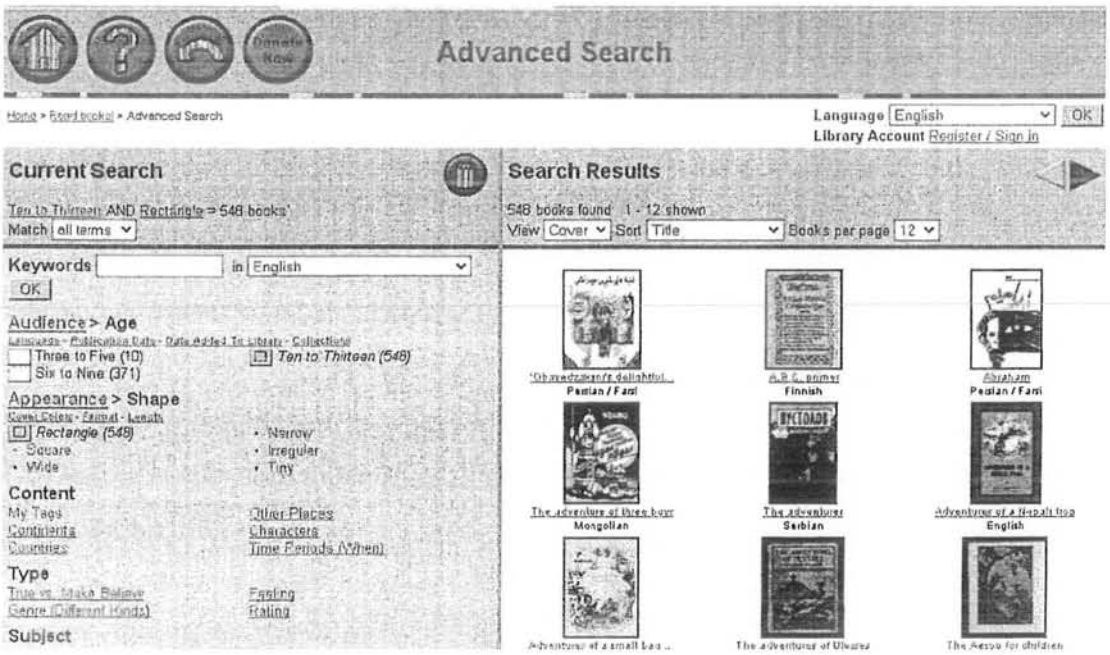

Fig 2.12 Advance search for books of ten to thirteen age group and rectangle shape at ICDL

2.8 Summary

Many researchers have evaluated the usefulness and effectiveness of the different exploratory and faceted search systems. Evaluation typically involves around the different tasks defined to test the efficiency of the system or improvements made to the interface. Some studies show many systems have the potential to be very successful exploratory search systems and worth the effort and attention exploratory search is

getting. Some studies have failed to show the desired results due to various reasons e.g. the lack of growth of technology, improper implementation, etc. Even though faceted search might be more effective and useful in certain situations, it is not preferred over typical search by users (Kules et al. 2009). The reasons could be many depending upon the context and situation e.g. the lack of familiarity of users, partiality towards typical style of searching, or the absence of the elements of an interface for searchers.

There is no doubt that faceted navigation has potential. But interfaces developed so far do not make use of this potential to a full extent. It has been hit and miss kind of journey so far. Also, faceted search is still limited to particulars domains like item selection, stores, shopping, etc. and works best for them. So if we want to take the faceted search to the next level then there is still a lot of work to be done as there is a great room for improvement in faceted search interfaces.

Chapter 3

Proposed Work

In the previous chapters the focus has been on the following points

- Faceted navigation and faceted search interfaces
- Children's use of Internet and search engines
- Problems faced by children while using search engines to find information

Based on the comprehensive literature survey, we have reached the point where we can safely say that children are still struggling from the very basic problems of spelling and typing mistakes, limited vocabulary, and making sense of results. Solutions exist here and there but each of them is far from best and leaves a lot of room for improvement. In the

previous chapters we discussed the features and advantages of faceted navigation and it is evident that faceted search has certain features and qualities that it can prove helpful in solving the long lasting problems of children. It is just the matter of how to use it in a proper way. Now, we will present our proposed interface and its various elements and we will explain how these elements cater to the various problems of the children they encounter throughout their searching process.

3.1 Motivating Scenario

One of the major activities for children over the Internet is searching the web for the information related to their school work e.g. quizzes, exams, books, general knowledge, etc. He/she use different resources mainly search engines to find related information. Now let's think of a scenario considering steps a child goes through to search for the required information. This scenario is described below.

Asjad studies in the 5th grade. In his science class the topic of today was space and teacher told the class about asteroids. At the end of the class teacher gave the class an assignment to go and search on the internet for more information about asteroids and come with a short essay in the next class. Asjad connects to the Internet after reaching home and opens his favorite search engine which is Google for most children. He is not good in typing so he type with one hand while looking at the keyboard. He starts typing

and enters the spelling he remembers but he starts the word "asteroids" with the alphabet "e" instead of "a". He ends up entering the term "estroid" instead of asteroid in the search box. When he completes the spellings he looks up the screen and see the suggestions provided by Google and spots that the term he entered is also present so he becomes happy that his spellings are correct. Now he presses enter and he is shown a long list of result. After examining the results for a while he realizes that these results are strange and not what he was expecting. He thinks he probably made mistake in spelling but he is also confused since he had selected the term from the suggestions provided by Google and after entering the query no further suggestions appeared in the "Did you mean" tag which made him believe that he had entered the correct term. This scenario is shown in the following figure.

Fig 3.1 Asjad entering the term "estroid" in search box and suggestions provided by Google

Above figure shows the case when Asjad is actually searching for "asteroids" but he is entering the spellings incorrectly and even suggestions provided are confirming his spellings so he mistakenly thinks that his spelling are correct but search results show that he is wrong. This situation can occur with anyone and especially with children when they think they are correct about the spellings and evidences also confirm their belief. So, when the results don't turn out to be as expected, they think that there is no information available related to that particular topic and they become frustrated. They don't realize that information is available and the fault is on their side. At this point if they don't alter their keyword thinking that there is nothing wrong with the spellings or may be this thought doesn't even come to their minds, then this situation can seriously hamper the beliefs of children about search engines and how they work.

Coming back to the example of Asjad, after he realizes that he is making mistakes in spellings he confirms the spellings from his textbook and notebook and catches the mistake he is making. So after a considerable effort he is finally able to enter the correct term.

Fig 3.2 Search results for the term "estroid"

Asjad thinks that after entering the correct spelling and getting the relevant results, the difficult part is over. But he is wrong. His actual and harder task is just starting. Now that he has entered the correct keyword and a long list of results is in front of him, a lot of questions are popping up in his mind e.g. why so many results, which of these are relevant and important for me, what points to focus and what to leave, and many other questions like these. Asjad realizes that getting a lot of results is very easy but comprehending and analyzing these results is very difficult task. He starts checking the results one by one and after a brief look he decides which page to keep and which to discard. He finds that not every result is relevant to his requirements. After scanning first

two pages of results he thinks that he has gathered enough material and also further he explores results they are turning more irrelevant. He thinks why there are so many unrelated results and wishes that all the results very exactly according to his demands, which is off course a bit too ambitious to think. Now, Asjad moves on to the results he has short-listed for the essay assignment. He notices that there is no structured or standard format of the web pages. Few pages are written as simple text without any formatting of headings, while some pages are well organized and present the information in an ordered manner. So obviously he finds it easy to comprehend the pages written in a structured method, because it is simpler to single out the important concepts in the presence of heading since, headings obviously signify important things about the topic. On the other hand, in case of simple text he has to read the whole text, and it is totally up to him to think and decide which points are important and which are not. Somehow Asjad manages to prepare an essay after handling all the hurdles that came into his way.

Based on the above scenario we can clearly see the problems Asjad had to face during the task of searching and these problems are the same which have been highlighted earlier:

- Spelling mistakes due to lack of vocabulary which in turn makes the task of keyword selection harder
- Typing errors made due to less developed skills

- Understanding and analyzing result set due to so many results and unstructured format of web pages

3.2 A Faceted Search Interface for children

We are now in a good position to introduce our interface based on the above scenario and the material covered earlier. We will try to justify how this interface caters to the various problems highlighted previously. I have named my interface as Kidzz Search. In Fig 3.3, we can clearly see the different elements of the interface. We will now give the description of the functionality of these elements i.e. what do they do, their importance, etc.

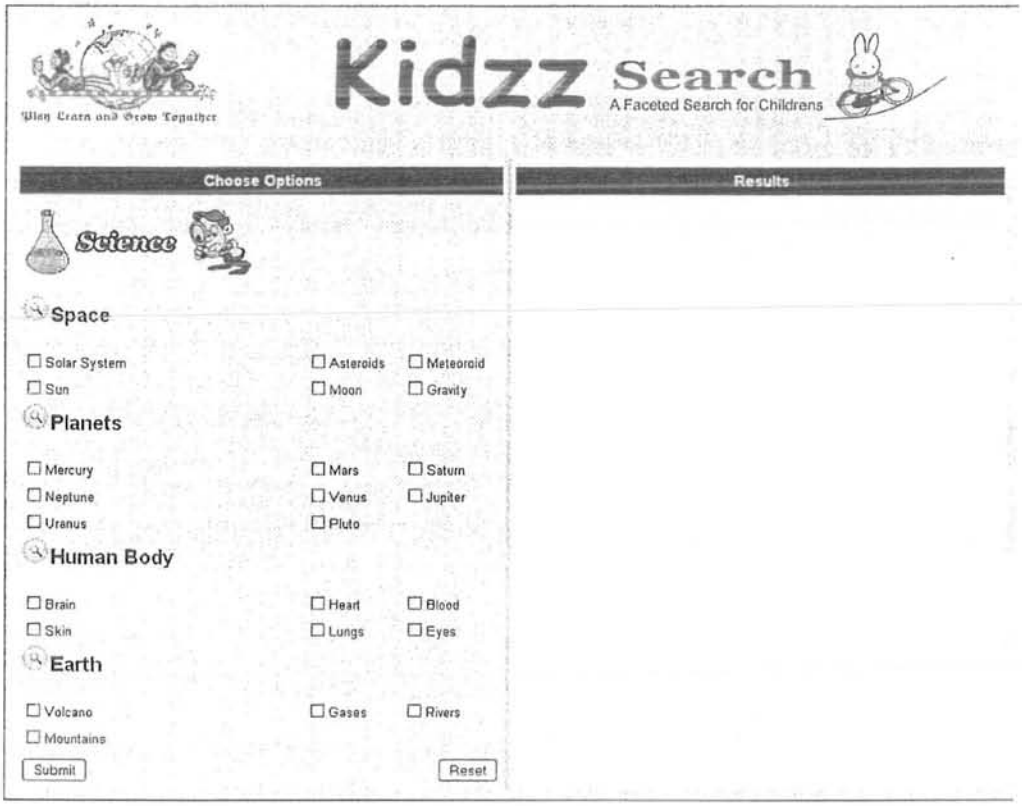

Fig 3.3 Kidzz Search Interface

3.3 Structure and Data Collection

A glance at the interface tells a few things about it. The page has two sections

1. Choose Options
2. Results

Choose Options section has few categories and check boxes under each category. Result section is empty when no selection has been made. We can see that Choose Options section has the main heading "Science" at the top. This is because all the pages which are stored and categorized are related to the subject science. Science is the subject which children start learning from early classed in their schools. It offers them the opportunity to explore and build upon their knowledge related to various domains of life since science involves everything from small parts of our body to the biggest secrets of this universe. So whenever a child learns a new and interesting concept from his class or hears about it from somewhere then he obviously wants to find more information about it. He goes to the Internet and starts searching about this particular topic using the search engine of his choice. So we have focused on the subject of science for the purpose of prototype creation and evaluation.

The next task is to apply the faceted search successfully. The essence of faceted search is to divide the domain in its attributes and perform search on the basis of those attributes. Here the domain of science is divided into four main categories Space, Planets, Human Body, Earth. These are by no means a complete and exhaustive list of categories. There can be a lot more other categories. These categories are only a sample from many categories. Each of these major categories is further subdivided into sub categories. Again there can be many other sub categories but only a small sample is taken for the prototyping level. These categories are made after analyzing different websites

(kidsclick.org, nsdl.org) and few text books of children. The pages related to each category are found from different resources (mainly from kidsclick.org, nsdl.org) and simple web searching using different search engines Google, Yahoo, Bing.

The sub categories under each main category are shown as check boxes. Whenever a checkbox is checked the results related to that category are shown in the result section as shown in the image below.

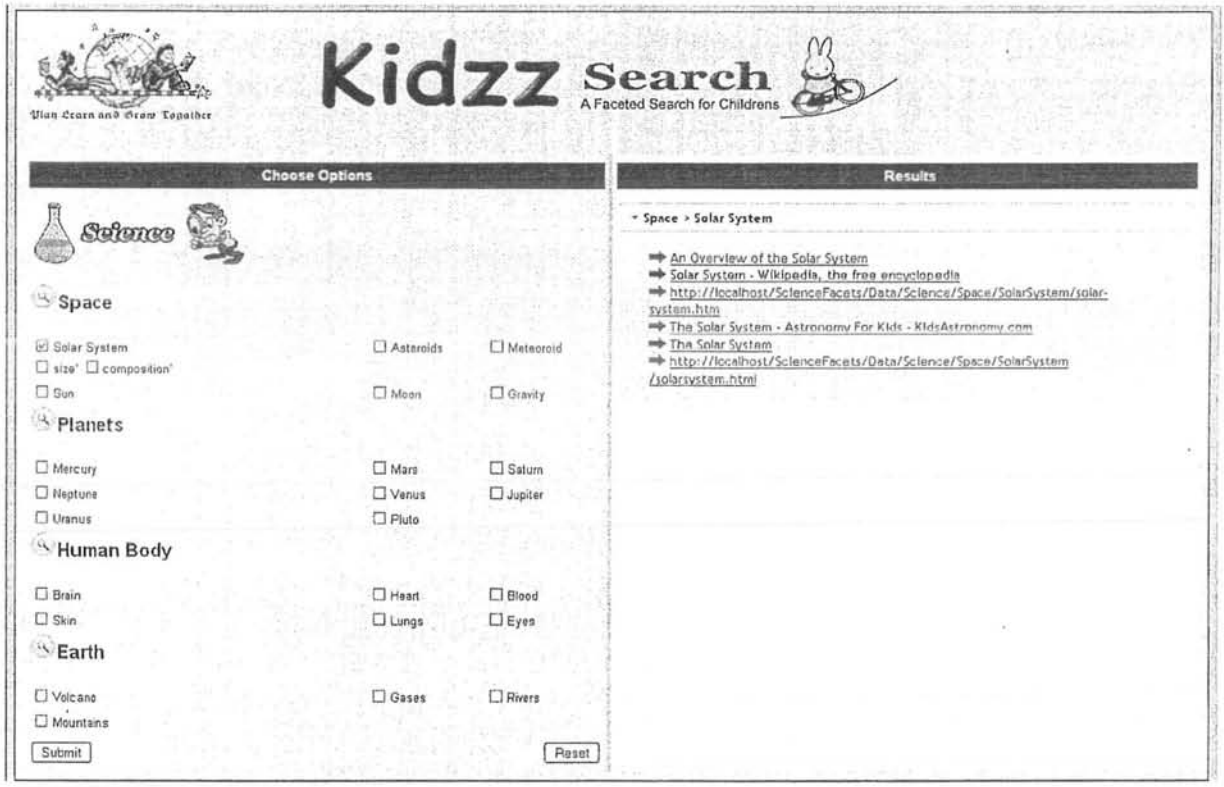

Fig 3.4 Solar System is checked and the results are shown

We can see the results in the Result section which are presented in a structure manner as heading at the top of the results goes like "Space > Solar System" which immediately helps in understanding that these results are related to Solar System in the Space category. We can also see that two further checkboxes have appeared. These are the facets which are used to filter the results. We will discuss these in details a bit later.

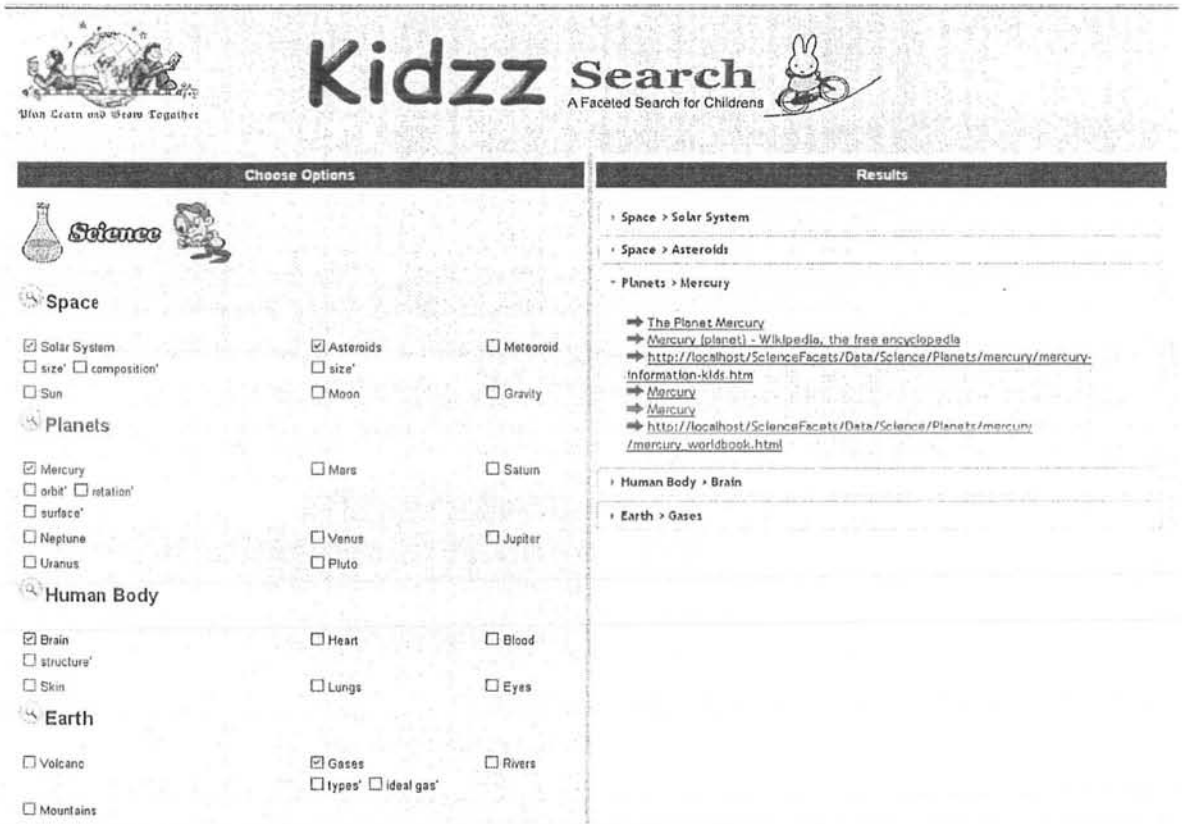

Fig 3.5 Kidzz Search when multiple options have been selected

What happens when user selects multiple options? This situation is depicted in the above figure. With each selected option its relevant results and corresponding headings appear in the Results section. User can check the results of any category from Results section. We can also see the relevant facets of each subcategory under it.

3.4 Facets and their Effect

We have seen that whenever a sub category is selected few options appear under it. We have called them facets but the question remains what is the purpose of these facets? We discussed briefly that facets are used to filter the results of the sub category they belong. Now, we will discuss the working and effects of these facets in details.

As we know that Web consists of enormous amount of web pages which are heterogeneous in nature. To apply faceted search to webpages we need metadata, but web pages don't contain enough metadata for the successful application of faceted search at a large level (Teevan et al. 2008). There are Meta Tags available in the webpages but they don't contain enough information about a webpage. The only metadata we have about a webpage is its title or sometimes some keywords or description about the page provided by the author of the page. Other than that we know nothing about contents of a webpage in terms of metadata. It is not considered mandatory to provide metadata after designing the website. Developers are focused mostly on the design aspects of the process and they

think their job is done as soon as the website is working. Also the task of providing accurate and quality metadata is not easy. It requires time and effort. Automatically generating quality metadata for webpages is also hard with the current state of the technology due to the diverse nature of the webpages and the inability to extract metadata which consistently applies to a large group of webpages (Teevan et al. 2008).

Keeping all these things in mind, the question arises how to apply faceted search to the webpages? If we analyze webpages it will become evident that there is no particular structure to the layout of the information. There are no standards to design webpages and everyone creates pages according to their preferences and requirements. Some pages only consist of written text without any formatting, headings, images. Some pages are written in a structured manner with proper headings and formatting. It is obviously easy to read and comprehend text which written in a structured manner as compared to the text which is written in a flow without any breaks or headings etc. It is even harder for children to read, analyze, and pick important points from plain text. Even with the pages written and presented in a formatted manner, it is difficult for children to digest the information quickly. They take time to read and reach the desired point in the text. Analytical and cognitive skills of children are not yet reached the point where they can cope with the information load easily hence the process of exploring information is slow in their case.

Here we have presented a brief analysis of the situation over the open web and problems which occur in understanding information due the unstructured nature of the webpages.

The reason of this analysis was to reach a point where the hurdles in the path of successful application of faceted search become evident.

Coming back to the structure of webpages, if a webpage consists of heading then, these headings say something important about that particular topic. Headings are an indication of some important information coming up. Presence of headings in a page signifies that there is some important information under each heading otherwise there was no need to write that information in the separate space. Since we know the basic idea of faceted search i.e. to search the domain on the basis of its attributes, and we also know that open web consists of unstructured data and not enough metadata is available to apply faceted search on the open web, hence we can make use of the headings within a page and utilize it as metadata. So idea is to use headings found within webpages as facets to apply filters to the results within each section. It makes sense since headings are vital piece of information about text and hence can be counted as attributes of the domain. In Kidzz Search interface we have made use of this fact and applied faceted search on the basis of the headings found within the pages in our dataset. Whenever a subcategory is selected its results are shown in Results section but also the relevant facets associated to that particular subcategory also appear below the selected sub category.

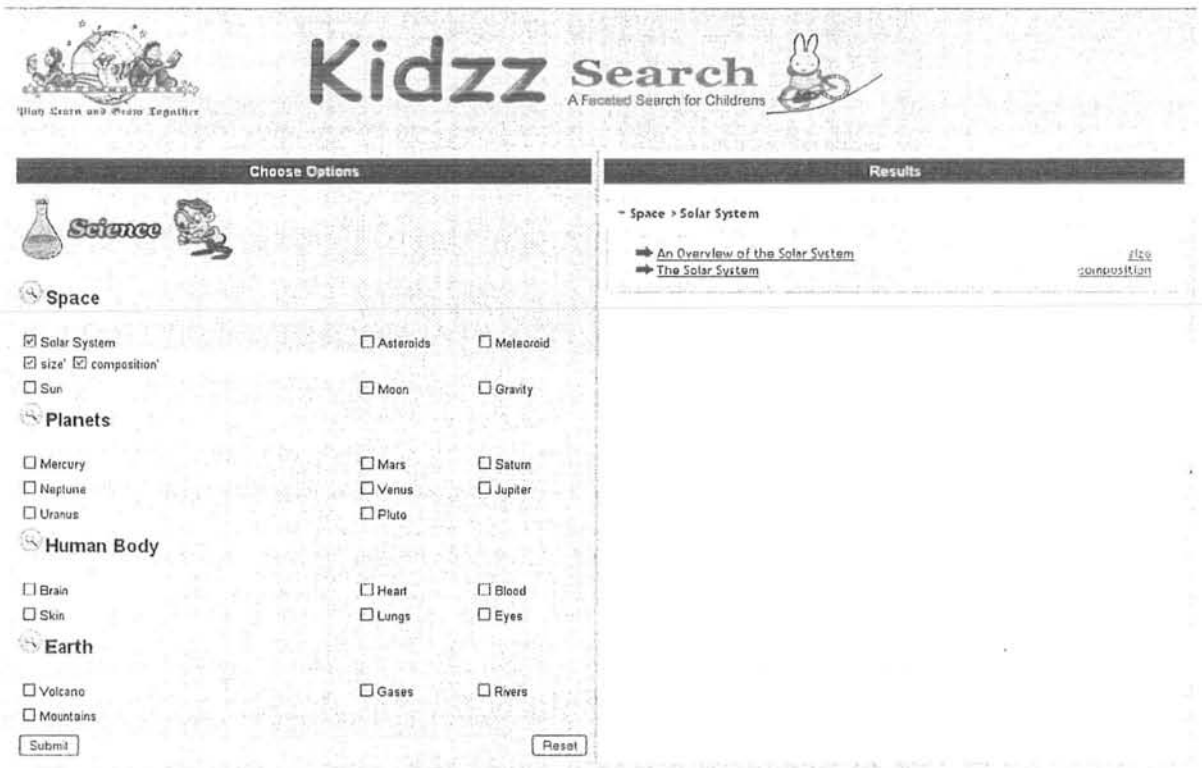

Fig 3.6 Results after facets selection

In Fig 3.4 we can see the state of the system when the Solar System is selected and its corresponding results are shown in Results section. We can see Solar System has two facets; size and composition. The effect of selecting the option "size" is that it will filter the results in the Solar System section and only those results will be shown which have the heading "size" in their contents. Also a label will appear of the name "size" in front of

all those results which have this heading. This label is clickable and it takes directly to place where the heading of "size" and its related material is written within the page.

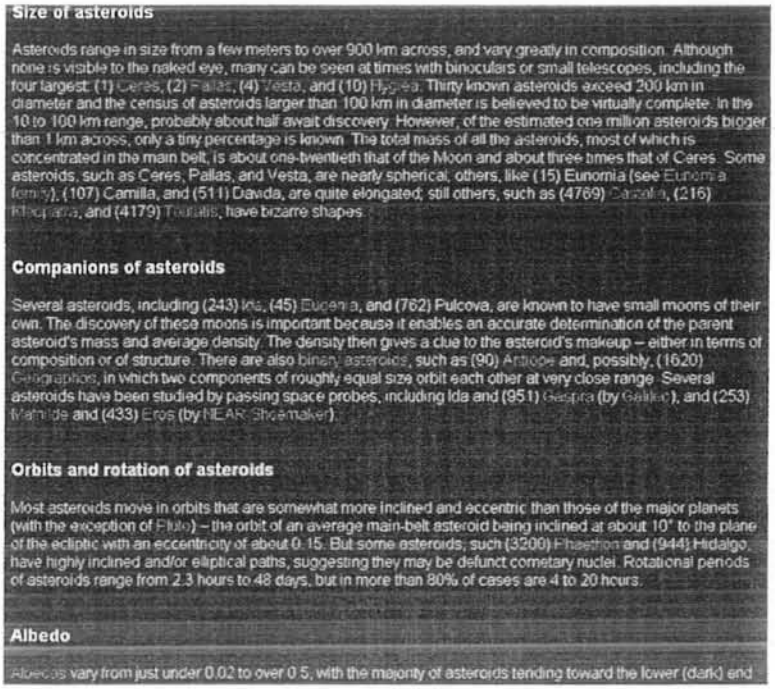

Fig 3.7 Clicking "Size" takes directly to the related heading

Fig 3.6 shows that both the facets of Solar System have been selected and the results have been filtered accordingly in the Results section. Associated labels of each result have been shown in front of them. Fig 3.7 shows that when the label of "size" is clicked the page related to that label is opened and focus is shifted directly to the place where the heading of "size" is located. This is how Kidzz Search interface works.

3.5 Issues Handled by Kidzz Searches

We will now discuss the advantages of using our designed faceted search interface, type of problems it solves, and how it helps children in finding information easily and quickly.

3.5.1 No Need for Query Input or Keyword Selection

One of the difficult tasks for children is finding proper keywords at the time of searching. This problem is solved by eliminating the need to enter query altogether. Since all the information can be searched on the basis of selecting appropriate categories so there is no need for the search box. So the time which is consumed in this activity is minimized hence using this searching interface is time efficient. All the children have to do is to find the appropriate categories and the results related to those categories are displayed right away. Some would argue that how omitting keyword search counts as advantage. Selecting the relevant category is easier for children than thinking and reformulating queries again and again since children are more comfortable with recognizing objects than recalling the information (Borgman et al. 1995). Further help is provided to children by showing relevant facets which apply more filters and narrow down the results. Therefore, if sufficient information is covered in all the categories and options and it gives a good overview of the domain and directs towards the desired information then keyword search can be regarded as beneficial especially in case of children.

3.5.2 Typing and Spellings Mistakes are Avoided

Browsing based search engines have an added advantage, since children don't have to type so they don't make spellings and typing mistakes at all. Spellings and typing mistakes are other big issues for children other than keyword selection. They spend a lot of time in finding the required information because of the mistakes they make in spellings and typing, as once they enter the incorrect word and realize their mistakes after checking the results, they have to go back and do it all over again. So a lot of time is consumed in all these activities. Kidzz Search saves this time by providing the interface which is free from typing and spellings mistakes since there is no need for keyword searc

3.5.3 Hierarchical Representation Helps in Making Association and Sense of Results

It is very common to enter a query in search box and get thousands or millions of results. After that it is totally up to the searcher to make connections between the results and the query he entered. There is no explicit help provided by the search engines. Contemporary search engines now do try to provide some filters by dividing the results into different broad categories like Videos, Images, News, and etc. but when it comes to the actual contents of the results no filtration or categorization is provided. This situation becomes

more serious when children perform the searching. Children are at such a level where they cannot process a lot of information as their cognitive skills are not matured enough. Search engines show the results in a set of 10 results per page but research has shown that even this information is more than they can absorb. As the amount of information grows, children have issues in establishing relevancy of results (Large and Beheshti 2000). In many cases they failed to do so and in turn believe that information does not exist. Kidzz Search suggests a solution to this problem as well. We have implemented the system such as that each category in Options section has a relevant category in the Results section and the each selection made to the right side brings the result in the left side and these results are only for the associated category. For the purpose of providing visual cues the title of each heading in the Results section has a title which shows the related category to which these results correspond. By arranging and presenting the information and results in this hierarchical way children are helped greatly since they can make strong association between the selections they made and the results they got in return. This representation also does not overload them with the information and they can easily understand what is happening

3.5.4 Ease of Navigation Improves Satisfaction and Outcomes

Excess of everything is bad and it also applies to the amount of information related to a particular topic which is available to a person when he/she is looking for it. Excess of information is even worse in case of children. Children find a result after completing

many difficult steps i.e. after overcoming the challenges of keyword selection, spelling and typing mistakes, and analyzing results for desired information. They finally reach a result after fighting hard against many deficiencies they have due to their undercook abilities. At this point what they want is that they are presented with their required information, but what they actually get in many cases is a lot of text. We discussed earlier that sometimes this text is formatted and many times unstructured and written in a sequence. This unformatted and unstructured text poses a great challenge to children. They actually have to read this whole text to find the exact piece of information they have been looking for. Now they don't know whether this information is present in the current page or not but they give it a try. But children are just not good enough at such a small stage to grasp the ideas presented in the text. They read through the lines reading few words and skipping many and cannot focus on all the words. As a result, they usually end up assuming that this information is not available on the page even if it is present and they have missed it (Livingstone et al. 2005). Structured representation of text doesn't also guarantee that children will locate the target position. If there are a lot of headings on the page, they might miss the intended heading, or they might not know the meaning of words which actually represent their desired information. Kidzz Search proposes a solution to this problem by defining facets, which are actually created by extracting the headings from the result pages associated to each category. Each facet selection creates a label in front of each associated result, and clicking this label takes directly to the place of heading. These facets are defined in simple language which is easily understandable

for children. This way, children are saved from the efforts this whole process takes and which often results in the frustration which builds due to the failure to find information. As a result children become more satisfied as their efforts are now producing the desired outcomes and that too with the minimal efforts.

3.5.5 Facets Provide an Overview of the Domain

One important aspect of the faceted search is that it provides an overview of the domain just by looking at the facets and helps in developing an image of the domain in the mind. This in turn puts users in a better position to make decisions as he/she knows what to expect, and what sort of information is present. This is especially helpful in situations when users have no general idea about a particular domain so they face issues in defining queries. Facets provide the initial knowledge of domain to let users decide themselves what they are looking for. In case of Kidzz Search facets are the headings of the pages in the data set, and headings obviously say something about the contents of the page it belong. So children are helped here by providing a synopsis of the pages in the shape of facets and in this way they get a good idea of the information available on the pages.

3.5.6 Opportunities to Learn and Knowledge Building

Faceted Search facilitates learning and knowledge building. It indirectly builds the knowledge about the domain being explored and provides the opportunities to learn at each step. Facets are the attributes of the relevant domain which are chosen carefully to represent the domain in a best possible manner. So when people use a faceted search interface for information finding they have to closely observe the facets provided to accurately direct the system towards their required information. This process of correctly selecting the relevant facets ultimately helps users by building their knowledge about that domain, as they now know certain facts about that field which may have been unknown before using the faceted search system. Since facets in Kidzz Search come from the headings within pages from dataset, they contribute to the knowledge of children and add up to that knowledge.

3.5.7 Fewer Results Pose less Information Load on Children

Keyword based search engines work on the basis such algorithms which make them return many results. Now some would consider it a good thing but it is not so good particularly in case of children. Children examine only first page of the results in most cases and that too only first few results of the first page. If they don't find the information they are looking for from these results they leave the task assuming that this information

is not available. This behavior of children is justified since most relevant results are usually available on the first page and other pages rarely contain any valid result in most cases. Children will be comfortable with fewer but more accurate results which could satisfy their information needs. This is exactly the case implemented in Kidzz Search, as children get only few results as the outcome of each selection they make. These results are further pruned when facet selection is made. So each query results in fewer matches and children are not bombarded with the long list of results.

3.6 Summary

In this chapter we presented the detailed overview of our proposed interface Kidzz Search. We described in details the salient features of our proposed systems and how facets are generated and what effect they have. In the end we illustrated what are the issues which are handled by our proposed system.

Chapter 4

Experimental Design

4.1 Hypothesis

Our hypothesis for this work is

"Faceted Search will help and improve the searching performance of the children when compared to query based searching"

There are few things which are clear from the above hypothesis statement

- There are two interfaces, one is query based and the other is our proposed interface i.e. Kidzz Search
- Children will get the chance to use both the interfaces and their performance will be measured and compared against some predefined tasks

- After the analysis of the results the conclusion will be drawn that whether the hypothesis is justified or not

In this chapter we will discuss all above mentioned steps in detail. First of all the experimental setup and research design is described. After that the tasks are defined which form the basis for the analysis phase. Then the details of the evaluation process are given which include the details of the survey and interview conducted from children. In the end the analysis of the outcomes is performed and results and conclusions are drawn.

4.2 Measures for Evaluation

Three important measures around which the evaluation of our hypothesis revolves are

- Accuracy
- Time Efficiency
- Error Count

4.2.1 Accuracy

Accuracy is related to the ability of children to find the exact information they are looking for. It often happens that children do succeed in finding a relevant result from the list of the results but when they examine the page by opening it they cannot find their

required information due to either information is not present on the page or they cannot find it because of load of information. Accuracy will be measured by assigning different tasks to children and they will be asked to complete these tasks on both the interfaces. Comparing the results of tasks on both the interfaces will show that which interface has enable the children to find more accurate information.

4.2.2 Error Count

Error count illustrates the number of errors committed by children in typing, spellings, and keyword finding while searching. Children commonly commit many mistakes in these departments which naturally obstruct the flow of the whole process. Children's performance will be measured against the assigned task to check which interface outweighs the other in the area of error count.

4.2.3 Time Efficiency

Time efficiency relates to the amount of time it takes for children to reach the place which contains the accurate information being sought by children. Time is an important factor as it directly affects the performance of the children since the more time it takes them to find the information, the more they become frustrated and loose hope and stop

going after the information after a point. Time efficiency will also be measured similarly by allocating various tasks to children, which they will perform on both interfaces. Their progress will be recorded and later it will be analyzed and compared to determine which interface is more time efficient.

4.3 Sampling Technique

Sampling techniques are used to select a subset of the actual population in order to draw certain conclusions about the sample population which then can be generalized to the actual population (Ross 1978).

4.3.1 Sample Population

It consists of a subset of the population which is taken from the actual population in order to save time, effort, and resources since it's extremely difficult or even impossible to conduct the study on the actual population. All the treatments which are important for the study are performed on the sample population and results are drawn on the basis of these treatments.

4.3.2 Actual Population

Actual population consists of all possible subsets in the population. The results compiled for sample population can be generalized to actual population.

4.3.3 Sampling for Evaluating Kidzz Search

Since the objective of our study is to explore the impact faceted search will have on the search performance of children, so children are the main subjects of our study. But not all the children qualify; they possess certain traits to be considered for the population. These traits are

- Children belong to age group 8-13
- They are studying in some school
- They have some knowledge of Internet and search engine or can use them when assigned some tasks

So children qualifying above mentioned features consist of the actual population of our research.

There are various sampling techniques used for drawing samples from the actual population. We have used the convenience sampling technique in our study for selecting the sample. This technique, as the name suggests, works on the principle of convenience.

The researcher selects that subset of the population whose subjects are easy to reach and can be contacted easily. We have used this technique since its simple, time saving, and cost efficient. Actual population is all school going children in the age group 8-13 who are competent enough to use Internet. Our accessible population is all school going children in the age group 8-13 living in the residential area of Quaid-i-Azam University. So our sample size is 15.

4.4 Experimental Design

Research design selection is an important step for obtaining the convincing conclusions to the research question. There are various research designs and the selection of the best suited design depends upon the situation and requirements. We have selected One Group Posttest Design Only Design for our study. In this design we measure the same group for both the treatments

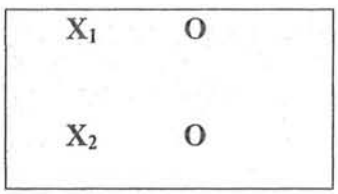

Fig 4.1 One Group Posttest Only Design

The reason for selecting this method is that the composition of our study is such that this design is the most suitable one. Since the purpose of our study is to find the impact of faceted search as compared to the general search, and whether it bring any improvement in the search performance of children. So it is imperative that we perform posttest on the group for both treatments

4.5 Task Creation

For the successful evaluation of any search interface, one essential condition is designing search tasks which are well suited to the situation at hand. Since our search interface is a faceted one and designing search task for such type of interfaces is more challenging than usual, as these tasks must involve exploratory search (Kules and Capra 2008). Exploratory search tasks have various desirable qualities (Kules and Capra 2008)

- Represent uncertain or ambiguous information needs
- Knowledge acquisition, comparison, or discovery are involved
- Uncertainty about what, where, and how of the required information

Since the subjects in our target population are all children, and we are designing and evaluating this interface especially for them, so we have to consider their special needs. We cannot expect them to perform and complete exploratory tasks like elder people who

have much more matured skills. We must be aware of the mental level and abilities of children while designing search tasks. Our dataset includes pages of Science subjects. At small ages children are mostly interested in finding information related to major concepts of any domain which they read in their books or hear from somewhere. They don't usually go into much detail.

4.5.1 List of Tasks

We have designed following search task after taking all above things into account.

We have created following search tasks

1. Describe the composition of solar system and its major components. How much mass is distributed to Sun?
2. What is the size of an asteroid? Which is the largest asteroid and what is its size?
3. What are the different types of meteorites? Describe the description of Stony Iron meteorites.
4. What is the Size of Sun? Find the mean surface temperature of Sun.
5. How gravity was discovered, who discovered it and under what circumstances?
6. Find and compare the volumes of planets Mercury, Mars and Neptune. Which is the biggest planet out of these three planets
7. Describe the various nervous systems present in human brain.

8. How human heart functions? Describe its structure and major parts.
9. 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?
10. What are the major functions of human skin? How skin is divided into different layers and what are the various colors of the skin?
11. What are various parts of human eye?
12. What are the different types of volcanoes? Present a comparison.
13. Which are the ten longest rivers in the world? Explain the topography of rivers.
14. What are fault-block Mountains? Provide a brief description

These tasks are defined to assess the performance of children against the measures we have defined. Each task contributes towards each measure by providing input.

4.6 Experimental Setup

Experimental setup for our study is as follows

4.6.1 Participants

Sample size of our study was 15. All children belonged to the age group 8-13. Their division with respect to their age is as follows

Age	No. of Participants
8-10	2
11-12	10
13	3
Total	**15**

Out of these 8 children, 5 were boys and three were girls. All these children were interviewed individually. Each child was assigned three tasks to complete using our proposed prototype. One open source search engine Sphider (www.sphider.eu) was also used for comparison purposes.

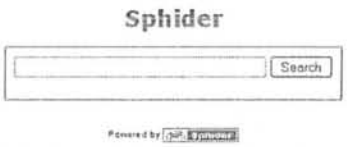

Fig 4.2 Sphider Open Source Search Engine

It's a simple and easy to use keyword based search engine and it functionality is pretty similar to the most contemporary search engine and provides most of the common search features which are present in any keyword based search engine.

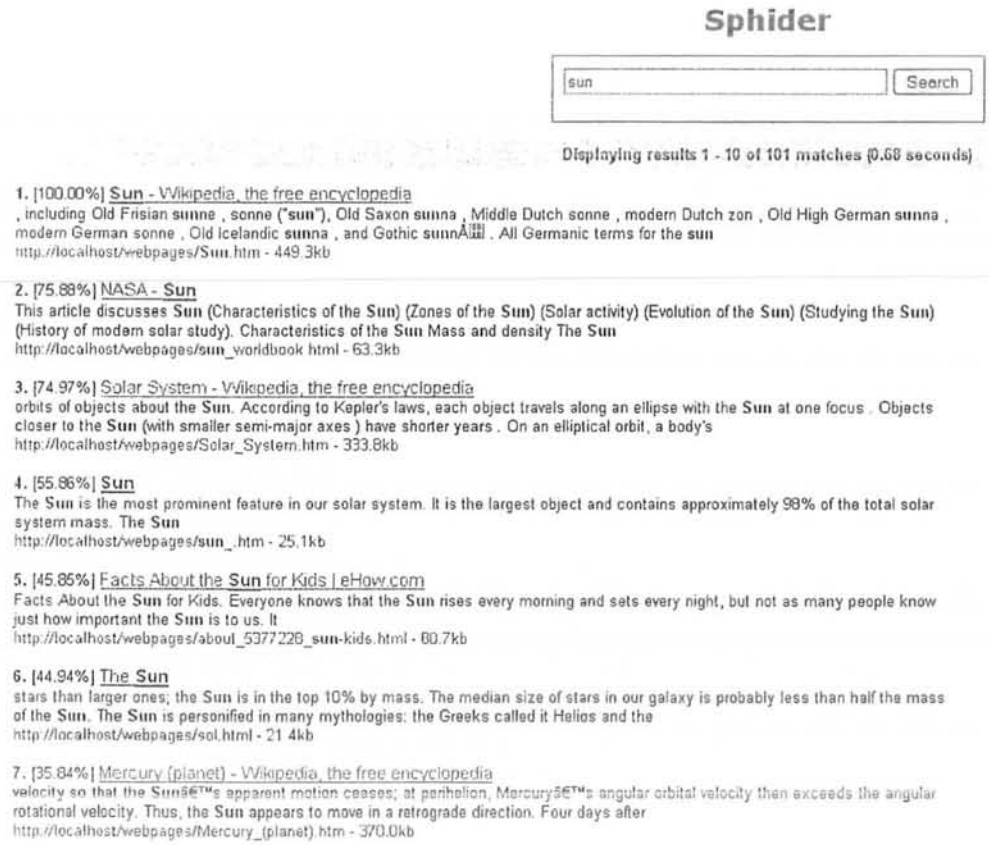

Fig 4.3 Search Results in Sphider

Each child was allotted a PC at the time of interview and they had to complete the assigned tasks on both the interfaces. Almost all the children were most familiar with Google search engine, and used it primarily for all their search needs. So, Google was also used as the reference for children to explain the working sometimes. Screen recording software was used to record all the activities performed by children during the completion of their tasks. Freez Screen Video Capture

(http://www.smallvideosoft.com/screen-video-capture/) was used for this purpose. It is free and easy to use software. It records all the activities performed on computer by capturing the screen. The main reason to use this software was that it does not slow down the system while recording like other software.

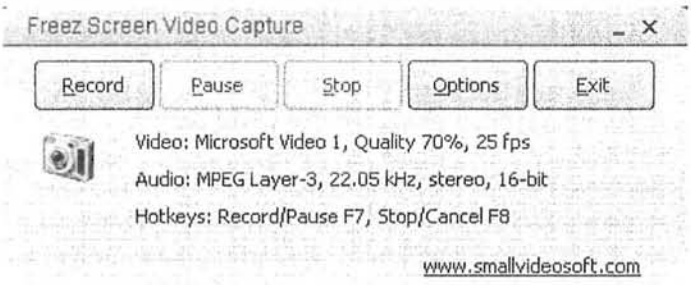

Fig 4.4 Freez Screen Video Capture Software

4.7 Test Design

Now we divert our attention to the process of interview which was conducted from children. We cannot expect children to complete the assigned tasks themselves without supervision. So the interviews were conducted individually from children under the close observation. Each child was assigned 3 tasks and he performed each task first on Sphider and then on Kidzz Search. All the activities of each child were being recorded for the purpose of analysis. Each task completed by children added some value to all the three quality measures defined earlier i.e. accuracy, time efficiency, and error count. All these

readings were noted and recorded carefully for later analysis. Each child was assigned 3 tasks randomly out of the list of 14 tasks. Following points were noted during the interview for each child.

- Time taken to complete each task
- Number of keyword and spelling mistakes made
- Number of pages each child examined to complete the task and he/she could not find the required information i.e. whether the information was not present or he/she could not locate it due to the large quantity of text

Evaluation Measure	Related Tasks
1. Accuracy 2. Time Efficiency 3. Error Count	1-14. Any 3 tasks are assigned to each child randomly

Table 4.1 Evaluation Measures and related tasks

After the completion of tasks, children were asked about the general impression related to both the search systems. They were asked which system would they prefer and why. Their responses were recorded for later analysis.

The above setup helped in measuring the parameters of our research which will be analyzed and evaluated to make a conclusion about the hypothesis.

4.8 Summary

In this chapter we gave an overview of our experimental design. Evaluation measures and sample used for evaluation of our hypothesis is described. We also discussed about the sampling technique used and children who participated in the evaluation of our interface. In the end we defined our tasks and described how the test was conducted and results were calculated.

Chapter 5

Results and Discussion

In this chapter we will present an analysis and discussion of the results which we obtained after the user experiments. We had defined three measures for evaluation i.e. Accuracy, Time Efficiency, and Error Count in the previous chapter. We also discussed the test design that we have two searching systems for the purpose of assessing our hypothesis. One is our proposed prototype Kidzz Search and the other system is Sphider which is an open source search system and it is used for the comparison purposes, Now, we will discuss the implications of the results to each evaluation measure individually. In the end we will analyze on the basis of evaluation results whether our hypothesis is justified or not.

5.1 Accuracy

We have defined accuracy as the ability of children to find the exact information they are looking for. When children hunt for the desired information using the search engine of their choice they fail to find the accurate information due to two main reasons

1. The information is not present on the page
2. The information cannot be located because too much information is written on the page

So, for measuring accuracy we have defined two more attributes

1. Non-Presence of Information
2. Information Load

Hence, we have measured accuracy by observing each child and taking note of how many pages he/she skipped. The reason for skipping the page was also noted i.e. whether the page was skipped because of non-presence or information load. In the end accuracy was calculated for each child by measuring the total number of pages which were needed to complete all the tasks and this number was divided by the number of pages actually examined for finding the information. Similarly the accuracy was computed for all children for the both the interfaces, and these results are shown in the following fig

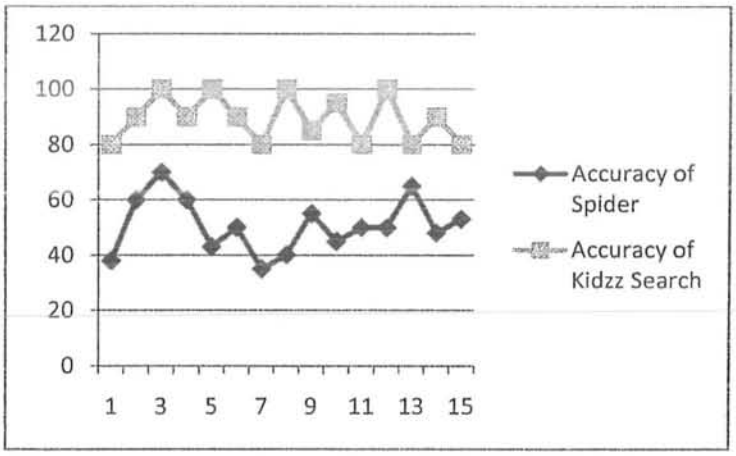

Fig 5.1 Comparison of Accuracy for both Interfaces

In Fig 5.1 we have shown a comparison of both interfaces for accuracy. We clearly see that the accuracy of Kidzz search is way above than the accuracy of Sphider. The Fig 5.1 also shows that each child has been more accurate on Kidzz search than Sphider. The property of Kidzz search of being a guided search engine makes it more accurate. Clues are provided at each step so, children are never lost. It is clear from the Fig 5.1 that in case of Kidzz search accuracy varies in the range of 80% - 100%, which is extremely good in any condition. But it must be noted that these numbers are calculated against the completion of just three tasks. It is important to understand that accuracy will reduce from 100% over the period of time. Here 100% accuracy means that, the child found all the required information for all his three tasks, without skipping or missing any pages. The obvious reason is that the information required by the child for all of his/her three tasks must have been available in the facets, which took the child directly to the relevant

place. Hence there was no need to check extra pages. It will be more relevant to compare the average accuracy for all children for both interfaces and this is shown below

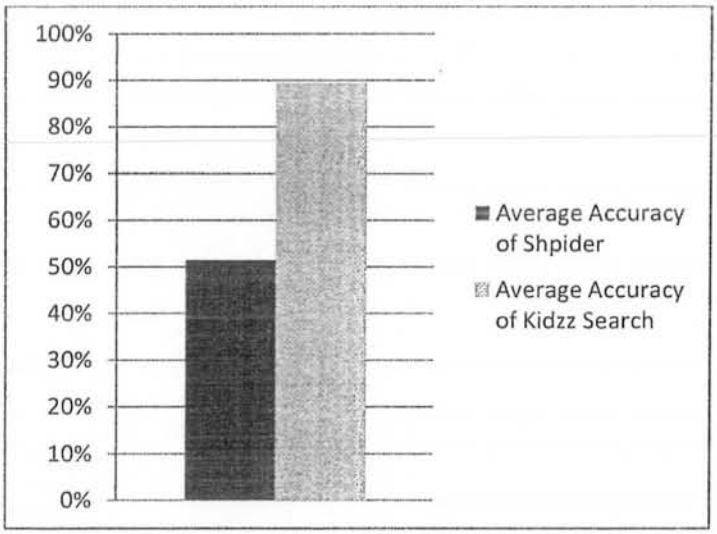

Fig 5.2 Average Accuracy for Both Search Systems

Fig 5.2 show the average accuracy for both search systems, and it can be clearly seen that Kidzz Search is ahead by big margin. Average accuracy for Kidzz Search is 89%, which is excellent. On the other hand average accuracy drops to 51%, just for three tasks in case of Sphider. So Kidzz search here is a clear winner. Average will surely decrease for more complex tasks, but there is enough evidence now that it will always be more than Sphider.

Since we defined two factors, non-presence and information load, which directly affect the accuracy, so it is pertinent to see the data related to both this factors. Figure below shows this data

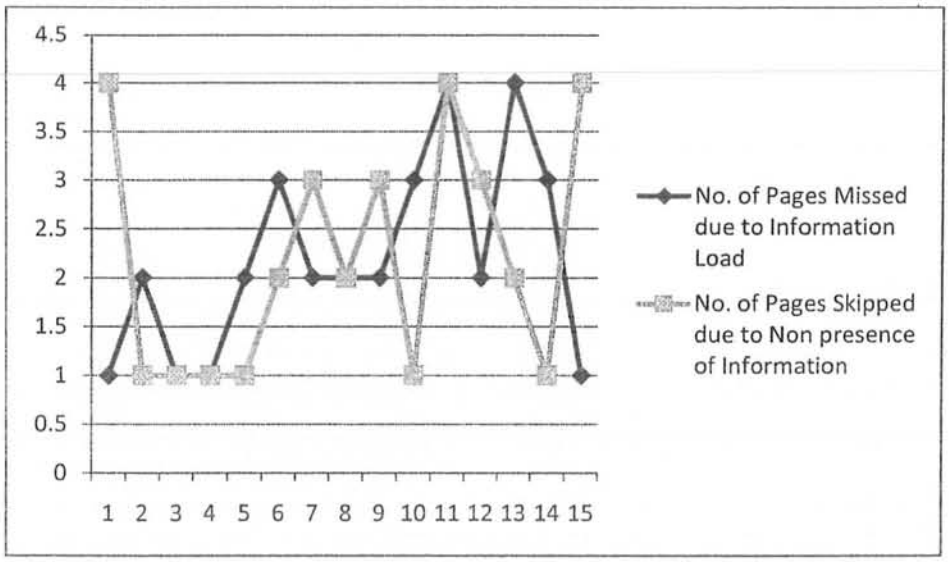

Fig 5.3 Number of pages missed and skipped using Sphider

Fig 5.3 shows the data related to number of pages each child misses due to information load, and also the data related to the number of pages skipped due to the non-presence of the information on the page, during the course of 3 tasks assigned to him. We can see that children do have to go through more than one page to find their required information even for the very simple tasks. It is a common practice for every searcher that after checking one result if the required information is not found then the searcher moves on to the next result. But the case with the children is totally different. Children don't like to

examine more than few pages of the results and get frustrated soon, and even assume that information is not available (Bilal 2000; Large and Beheshti 2005). So the condition depicted in Fig 5.3 becomes more serious automatically and accuracy becomes more crucial in case of children. Skipping pages because of non presence of information is understandable, as we cannot expect our queries to result the most relevant result at the top of the list. But if the information is present on the page and even then child misses that information due to the overload of text, then this situation is of more concers. Kidzz search has handled this case of missing pages due to information load by providing facets which take children directly to the relevant place. There is a possibility that children miss the appropriate facet from the list of facets, but this can happen rarely because the facets are kept to the minimal level. But if a child wants to search something which is not covered in the facets, then the situation becomes pretty similar to Sphider, since not all the information can be covered in facets. Facets are basically used as pointers to guide the children in their process of searching. But even if facets don't cover the required information, accuracy is not affected greatly since we have assigned the most relevant pages to their respective categories. Children don't have to go through skipping and missing many pages before they can find what they are looking for. So accuracy is built in to the Kidzz Search interface.

5.2 Error Count

Error Count has been defined as the number of errors made by children in typing, spellings, and keyword finding. These are three major concerns of children while using keyword based search engines. These errors interrupt the flow of information and cause annoyance and confusion. To calculate this measure, children were closely observed to find the number of typing and spelling errors they made and how many times they needed to change the keyword. These numbers have been shown below

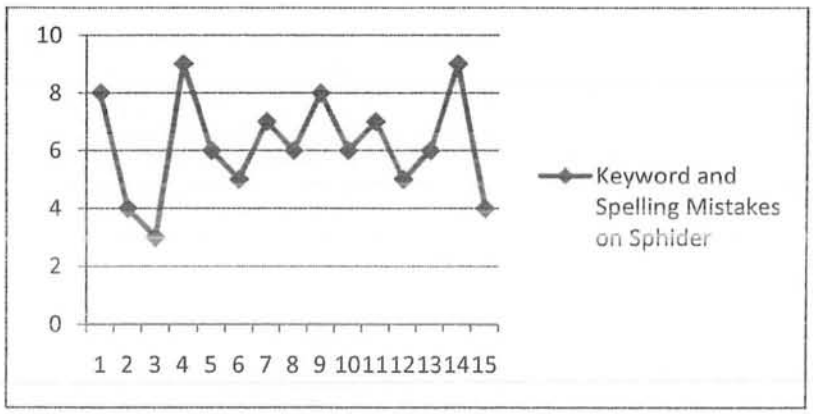

Fig 5.4 Keyword and Spelling Mistakes on Sphider

Fig 5.4 shows the number of mistakes each child made during the course of 3 tasks. On average each child committed 6 errors which is a quite a big number considering only three tasks had to be completed. These results confirm the point of view of researchers, which they have given consistently related to the issues with contemporary search

engines. These results show that keyword based search engines create various problems for children as they make many mistakes during the search, which obviously make the task of searching even harder. It was also one of the main motivating factors towards Kidzz Search as it eradicates the problems caused due to typing in the search box and it is replaced by clicking. So the source of many errors is removed, and further guidance is provided to children to make the task of searching easier and engaging.

5.3 Time Efficiency

Time efficiency was defined as the time taken by children to find the accurate information. This means that time efficiency is proportional to accuracy. This is logical because, as the number of pages examined by children grows, the time taken to complete the task also increases. But the time efficiency also depends upon the error count, since the number of errors committed by children automatically adds to the time taken to complete a task. For calculating this measure we recorded the time children took to complete each task on both the interfaces. The results of this activity are shown below

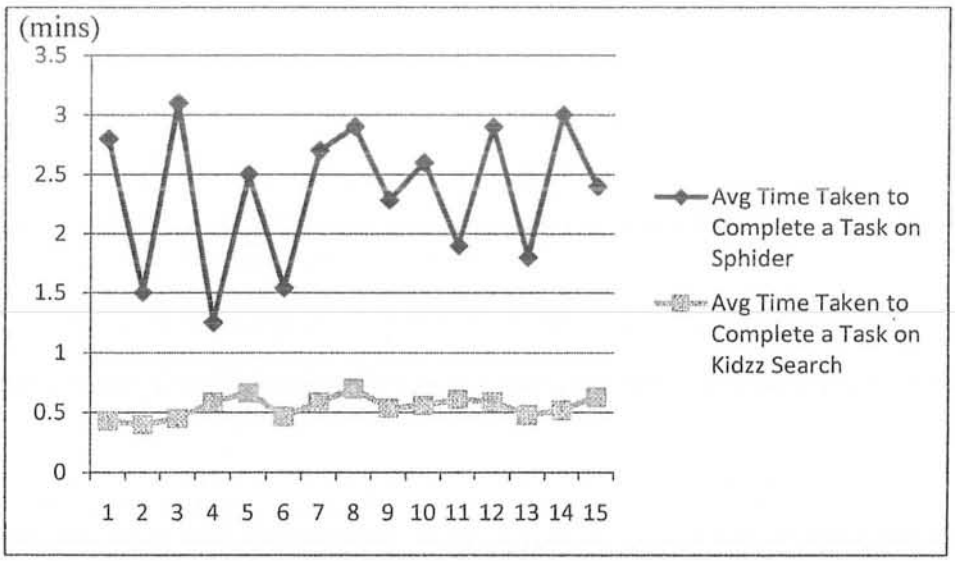

Fig 5.5 Comparison of two interfaces for average time of a single task

Fig 5.5 shows the comparison of both the interfaces for the average time it takes to complete a single task. It is evident that the average time taken on Kidzz search to complete a single task is far less than Sphider. We now examine the overall average

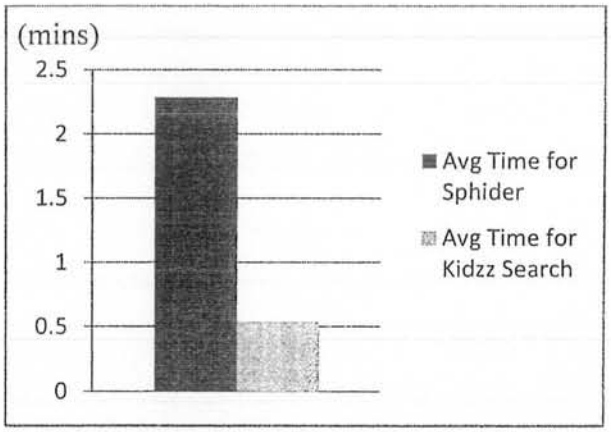

Fig 5.6 Overall average times for both interfaces for a single task

Fig 5.6 shows the overall average time for all children across all tasks. For Kidzz search the average time to complete single task is 32 seconds approximately, whereas for Sphider it shoots up to 2 minutes approximately. This is a huge difference of almost 1.5 minutes. We have already shown that Kidzz Search is far more accurate and error count is also very small as compared to Sphider. We also discussed that time efficiency depends on accuracy and error count, so it is unsurprising to see that Sphider is far too slower as compared to Kidzz Search. Hence, Kidzz search has also scored pretty highly for the measure of time efficiency

5.4 Preferred Interface of Children

We have presented an analysis and discussion of the results of user experiments. After the completion of tasks children were asked about their general opinion about both the interfaces. All the children preferred the Kidzz Search over the Sphider. Then they were asked what the reasons of their choice are, and all the children gave the identical reasons which are summarized below

- It is very easy to find information using Kidzz Search because of the way choices are organizes and results are presented
- It is very quick to locate the desired results due to its hierarchical organization

- It is quick and easy to find the required information on the result pages as the facets selection takes directly to the place where the information is written
- It is very easy to use to Kidzz Search, as there is no need to type. It is difficult to type, spell the words properly, and find appropriate keywords

Above responses of children and results of our experiments confirm our point of view which we have taken from the start i.e. keyword base searching is difficult for children. Children have also pointed out that they still suffer from all the problems which have been highlighted in the literature. All the children said that they are more comfortable with browsing base search interfaces rather than keyword based search system

5.5 Summary

We have analyzed the results of experiments. We have also discussed the responses of children related to both the interfaces. These results and responses of children combine favor our proposed interface i.e. Kidzz Search. We have already seen the comparison of both the interfaces, and how the keywords based search system, Sphider in our case, slows down the progress of children. We have also discussed the problems they faced using Sphider and how Kidzz Search handled those problems. After all these analysis we are now in the position to say that our hypothesis has been justified. We have seen the

performance of children on both interfaces, and Kidzz Search indeed improves the performance of children. Children succeed in finding more accurate information in less time without committing a lot of errors using Kidzz Search. Hence impact of faceted search on the search performance of children is positive and Kidzz search improves the search performance of children by a large margin when compared to keyword based search engines.

Chapter 6

Conclusion and Future Work

6.1 Summary and Conclusion

This dissertation started with a discussion on the contemporary search engines. We talked about the two basic modes of search i.e. keyword based search, and browsing based search and various issues with these systems. We discussed that how these issues with current search engines motivated researchers to think out of the box and the idea of exploratory search emerged as a result. A brief discussion on exploratory search then followed and it became clear that faceted search is the most commonly used form of exploratory search. We then diverted our attention towards the main goal of our research.

We discussed that existing search engines are designed according to the needs and skills of elder people. But due to the growth and development of technology and Internet, children have also become a major group over the Internet. Contemporary search engines just don't fit the needs and skills of children. Very few search engines are designed especially for children and even those search systems make a half hearted attempt.

The main goal of our research was to propose a system which was designed according to the skill level of children. We argued why such a system is needed by highlighting the issues and problems with the most popular search systems, which makes it tough for children to use these systems. When discussed the major problems children face while using these search systems. These problems included spelling and typing mistakes, finding proper keywords, making sense of results due to messy representation, selecting the most suitable results from the long list, and finding the required information on the page due to plenty of text. All these problems with the keyword based search engines laid the foundation for our proposed system. The major challenge was to take all these problems into consideration, and come up with a system which not only handles those problems but also facilitate and guide children in finding the required information.

We adapted faceted search to design such a system after careful analysis, since faceted search boasted of features which provided a solution to the above mentioned problems. The objectives of our design included to proposes such a system which not only handles the issues with the query based search engine but also helps and guides in finding

information easily and quickly. These goals were achieved with the help of faceted search. We handled the problems of spelling, typing, and keyword mistakes by providing the information in a categorized view, where children could make selections instead of typing keywords to find information. These selections provided results which were also arranged in an organized manner. This organized presentation of information helped children in finding information easily and briskly. We also provided the facility of facets. These facets help the children to go directly to pages with the required information which in turn helped in reducing the information load. We named our proposed system as Kidzz Search.

After the design of our system, the next step was to evaluate the system to confirm whether it fulfills all intended requirements. For this purpose we conducted a prototype test, in which 8 children in the age group 8-13 participated. All the children were made to complete some tasks on two interfaces. One was our proposed system Kidzz Search, and the other was an open source keyword based search engine Sphider. The results of this test were very encouraging. All the children performed much better on Kidzz search as compared to Sphider. They felt more comfortable and at ease using Kidzz Search. So our hypothesis was justified.

Summary of Results

The summary of results is given below

- Kidzz Search improved the search performance of children
- Accuracy increased (51% for Sphider and 89% for Kidzz Search). The improvement in accuracy occurred due to well organized choices and hierarchical representation of results. Also, facets played the great part in improving accuracy as children could go directly to the place which had their required information
- Children made many mistakes in Sphider (Typing and Spelling) whereas no such errors in Kidzz Search. It is because Sphider used the keyword based searching interface which is the cause of many problems for children. In Kidzz Search we replaced this interface with faceted navigation so children could find the information easily without making mistakes
- Average time to complete a task in Sphider is four times more than Kidzz Search (0.5 sec Vs. 2.3 sec). Due to the increase in accuracy and reduction in error count in Kidzz Search as compared to Sphider, the average time needed to complete a task in Kidzz Search also improved greatly. Children completed tasks in Kidzz Search with comfort and in very quick time as compared to Sphider. This fact is clearly visible from the average times for both interfaces

All children preferred Kidzz Search due to

- Organization of choices
- Presentation of results
- Desired pages are found quickly due to hierarchical organization
- Easy and quick to find the required information within the result set
- Errors are minimized

6.2 Contributions

Our proposed system Kidzz Search contributes in two main areas.

- Faceted search
- The search interfaces for children

Currently faceted search is only applied in highly specialized fields. There are talks about applying it to open web but there are still many challenges to answer. One major challenge is the lack of quality metadata on the open web. Our proposed system presents an idea to overcome this challenge. We have used the headings within webpages to be used as facets. This idea can be adapted by researchers who are aiming to provide faceted search for open web

We know that very few search systems exist which are specially designed for children. Kidzz search aims to fill this gap. We have proposed and showed that structured and guided presentation of information helps in improving search performance of children. This framework can be followed to propose future search systems for children.

6.3 Future Work

Our Work falls in two domain faceted search, and interfaces for children. Both these domains have a wide scope and there are still a lot of issues to be solved. We will now discuss some of the issues which need to be solved

- We have based our facets on the heading extracted from web pages since web pages don't contain enough metadata. This issue still needs further exploration i.e. how to extract metadata from web pages since not all web pages contain headings. Hence generating quality metadata for web pages is a challenging task.
- If a webpage contains lots of headings then showing so many headings can become messy. Further exploration is needed on how to present greater amount of headings efficiently and without causing confusion
- Facet selection is a difficult task. In our case the facets were built within the heading of pages. But when facets are defined manually then task requires a lot of consideration and thinking. This area of facet creation can be explored further.

Also when we define facets for a particular domain, it is not necessary that these are the only choices available. There may be other choices available and they may produce better results. Hence this field of defining optimum facets can be explored further. Also generating facets automatically is a big task and requires further exploration

The focus was our research was to address the problems faced by children in searching. We proposed a system to show how theses can problems can be solved. We made a small contribution at a small scale but there are still many problems to be solved to achieve the desired goal at bigger levels.

Bibliography

Bates, M. J. (2005). Are There Optimal Ways to Do Exploratory Searching? *HCIL SOH Workshop*

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 American Society for Information Science*

Bilal, D., Kirby, J. (2002). Differences and similarities in information seeking: Children and adults as Web users. *Information Processing & Management*

Borgman, C.L., Hirsh, S.G. and Walter, V.A. (1995). Children's Searching Behavior on Browsing and Keyword. *Journal of the American Society for Information Science*

Cooper, L.Z. (2005). Developmentally Appropriate Digital Environments for Young Children. *Library Trends.*

Crook, C. and Bennett (2007), L. Does using a computer disturb the organization of children's writing? *British Journal of Developmental Psychology*

Druin, A., Foss, E., Hatley, L., Golub, E., Guha, M. L., Fails, J., Hutchinson, H. (2009). How Children Search the Internet with Keyword Interfaces. *Interfaces International Conference on Interaction Design and Children*

Fraenkel, J. R. and Wallen, N. E. (1993). How to design and evaluate research in education. *McGraw-Hill.*

Jochmann-Mannak, H., Huibers, T., & Sanders, T. (2009). Children's Information Retrieval: how to support children in effective information-seeking? *The 9th Dutch-Belgian Information Retrieval Workshop*

King, J. and Alloway, N. Young children's use of microcomputer input devices. *Computers in the Schools*

Kules, B., Capra, R. (2008). Creating exploratory tasks for a faceted search interface. *Second Workshop on Human-Computer Interaction (HCIR 2008).*

Kules, B., Capra, R. (2009). Designing Exploratory Search Tasks for User Studies of Information Seeking Support Systems. *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*

Kule, B., Capra, R., Banta, M., Sierra, T. (2009). What Do Exploratory Searchers Look

at in a Faceted Search Interface? *International Conference on Digital Libraries*

Large, A. and Beheshti, J. (2000). The web as a classroom resource: Reactions from the users. Journal of the American Society for Information Science.

Large, A. and Beheshti, A(2005). Interface design, web portals, and children. *Library Trends*

Livingstone, S., Bober, M., and Helsper, E (2005). Internet literacy among children and young people: Findings from the UK Children Go Online project. London: LSE Research Online

Marchionini, G. (2006). Exploratory Search: From Finding to Understanding. *Communication of the ACM*

Marchionini, G., & Brunk, B (2003). Toward a general relation browser: A GUI for information architects. *Journal of Digital Information*

Naidu, S. Evaluating the usability of educational websites of children. *Usability News*

Schraefel, M.C. (2009). Building Knowledge: What's beyond Keyword Search? *IEEE Computer*

Solomon, P (1993). Children's information retrievalbehavior: A case analysis of an OPAC. *Journal of the American Society for Information Science*

Stefaner, M., Muller, B. (2007). Elastic Lists for Facet Browsers. *DEXA '07, 18th International Workshop*

Teevan, J., Dumais, S. T., & Gutt, Z. (2008). Challenges for Supporting Faceted Search in Large, Heterogeneous Corpora like the *Web. Proceeding of HCIR 2008*

Tunkelang, D. (2008). What is (not) Exploratory Search? *The Noisy Channel Blog Post June 2008*

Trochim, W. M. K. (2007). Research Methods. *Biztantra, An Imprint of Dreamtech Press*

Ross, K. N. (1978). Sample design for educational survey research. Quantitative Research Methods in Educational Planning, *UNESCO International Insti-tute for Educational Planning*

van Kalsbeek, M. G., de Wit, J. J. (2007). Automatic Reformulation of Children's Search Queries

White, R. W., Kules, B. & Bederson, B. (2005). Exploratory Search Interfaces: Categorization, Clustering and Beyond. *Report on XSI Workshop*

White, R. W., Kules, B., Drucker, S. M., & Schraefel, M. C. (2006). Supporting Exploratory Search. *Communication of the ACM*

Wilson, M. L., & Schraefel, M. C. (2008). A Longitudinal Study of Exploratory and Keyword Search. *8th ACM/IEEE-CS joint conference on Digital libraries*

White, R. W. (2009). Designing Information-Seeking Support Systems. *ISSS Workshop*

Yee, K. P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and browsing. *Proceedings of the SIGCHI conference on Human factors in computing systems*