American Journal of Engineering Research (AJER)
e-ISSN : 2320-0847 p-ISSN : 2320-0936
Volume-02, Issue-12, pp-98-109
www.ajer.org

Research Paper Open Access

Design, Analysis and Implementation of a Search Reporter System (SRS) Connected to a Search Engine Such as Google Using Phonetic Algorithm

Md. Palash Uddin[1], Mst. Deloara Khushi[2], Fahmida Akter[3], Md. Fazle Rabbi[4]

[1&4]*Computer Science and Information Technology, Hajee Mohammad Danesh Science and Technology University (HSTU), Dinajpur, Bangladesh.*

[2]*Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh.*

[3]*Computer Science and Engineering, East Delta University, Chittagong, Bangladesh.*

Abstract: - A web search engine is an information retrieval system on the World Wide Web showing a list of necessary and less necessary URLs against the searching keywords elapsing a more time and requiring some analysis to get the required URL. The web creates new challenges for required information retrieval because the amount of information on the web is growing rapidly. Thus the target of SRS connected to a search engine such as Google is to get the required information though the searching process more easily, effectively and efficiently. For this, admin should insert the searching keywords in the SRS and then the SRS connected to any search engine will get all the titles, URLs and descriptions and then check and count the keywords in the web pages of the URLs. Then these retrieved information and an id associated with each keyword are stored in the database. Now, if a user searches for the keyword, then the SRS loads the search results from database and then ranks the pages based on its highest number of matching keywords. More significantly, a phonetic algorithm namely Metaphone algorithm is used in the SRS to eliminate the problem of spelling errors in the keywords given by the users. Admin can update existing keyword with the rapid updating of information based on the keyword and also add new keywords that are not found by the users. The SRS with necessary analysis has been implemented using appropriate and latest demanding tools and technologies such as Metaphone algorithm, HTML, PHP, JavaScript, CSS, MYSQ, Apache server etc.

Keywords: - *Information Retrieval, Metaphone Algorithm, Phonetic Algorithm, Search Engine, Search Reporter, Spelling Errors*

I. INTRODUCTION

Search Reporter System (SRS) along with a search engine offers that when we search for any keyword, then it provides the search result exactly related to the keywords. The SRS along with a search engine is very user friendly and simple also. A user can easily search and get the informative result from this system. The SRS is designed in such a way that a user will never feel boring with the system. If a user searches for a keyword, then the search reporter loads the search results from any search engine such as Google and then ranking the pages based on its information that means the most informative page will have rank 1, then the second informative page rank 2 and so on. In this way a user get the main information which he/she is actually looking for. There are three steps in this system: search any keyword in the system, get the results as they want from Google based on ranking, and open the links and get the best information. It provides an outstanding web based search interface. The salient features of the SRS are given below:

- It saves large amount of time.
- It offers the informative result without analysis all the results get from Google.
- It helps the users who actually don't know about searching.

1.1 History of Search Engine and Development of SRS

In the summer of 1993, no search engine existed for the web, though numerous specialized catalogues were maintained by hand. Oscar Nierstrasz at the University of Geneva wrote a series of Perl scripts that periodically mirrored these pages and rewrote them into a standard format. This formed the basis for W3Catalog, the web's first primitive search engine, released on September 2, 1993 [1]. The web's second search engine Aliweb appeared in November 1993. One of the first "all text" crawler-based search engines was WebCrawler, which came out in 1994. Google adopted the idea of selling search terms in 1998, from a small search engine company named goto.com. Around 2000, Google's search engine rose to prominence [1]. The company achieved better results for many searches with an innovation called PageRank. By 2000, Yahoo! was providing search services based on Inktomi's search engine. Yahoo! acquired Inktomi in 2002, and Overture (which owned Allthe Web and AltaVista) in 2003. Yahoo! switched to Google's search engine until 2004, when it launched its own search engine based on the combined technologies of its acquisitions. Microsoft's rebranded search engine, Bing, was launched on June 1, 2009. On July 29, 2009, Yahoo! and Microsoft finalized a deal in which Yahoo! Search would be powered by Microsoft Bing technology. By the passing of time the use of search engine is increasing. As increased use of search engine for searching information, a system has been developed that helps users to search information. When a person wants to search anything he simply places his words in search engine. Then search engine returns him relevant information according to his/her words based on many more criteria. But user has to extract their necessary information after doing much analysis as search engines can't give the exact information manually. This makes searching for any information very time consuming. Then we thought that we may develop search reporter system so that users may search and get any information manually and which is not time consuming. Search engines use many criteria such as SEO (Search Engine Optimization), searching and returning information but we choose primarily only the words that are given for searching. On the stage of developing the SRS at first, admin places a keyword in the field that is defined for him. Then the system which is connected to any search engine such as Google will get all the titles, URLs and descriptions and then check and count the keyword in the web pages of the URLs. Then the titles, URLs, descriptions, number of matches of the keyword and an associated id against the keyword are stored in database. Now, if a user searches for a keyword, then the search reporter loads the search results from database and then ranking the pages based on its highest number of matching keywords. As search engines updated their information day by day, so the admin needs to update the database of SRS day by day so that user gets the updated information from the SRS.

1.2 Present Search Engine

Generally, a search engine is software code that is designed to search for information on the World Wide Web [1]. The search results are generally presented in a line of results often referred to as search engine results pages (SERP's). The structure diagram of present system is shown below:

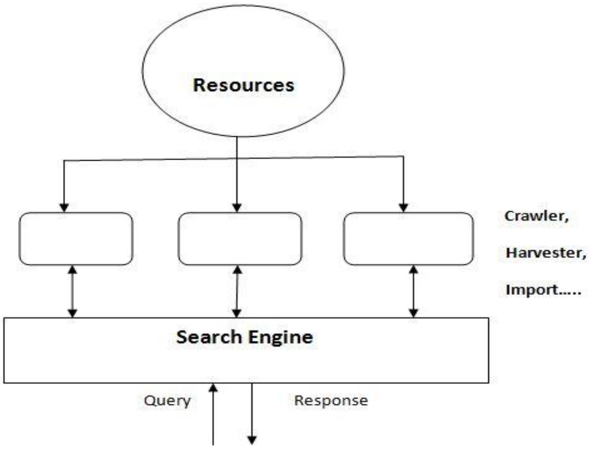


Figure 1: Structure Diagram of Search Engine

The structure diagram of present system shows that it contains resources such as server. Then information from resources come to crawler, harvester and import. A Web Crawler is a computer program that browses the World Wide Web in a methodical, automated manner or in an orderly fashion. Web crawlers are mainly used to create a copy of all the visited pages for later processing by a search engine that will index the downloaded pages to provide fast searches [2]. A Web Harvester extracts every single word every time it accesses a webpage. Additionally, a web harvester stores every single page harvested as a separate version in our database. It has two main advantages. These are Analytic capabilities and Versioning of Web Pages [3]. A

Web Import imports information from server and provide it to search engine. Users request information from search engine by query and search engine them information by response.

An information flow diagram (IFD) shows the relationship between external and internal information flows between organizations. It also shows the relationship between the internal departments and sub-systems [4]. The IFD for present system is shown below:

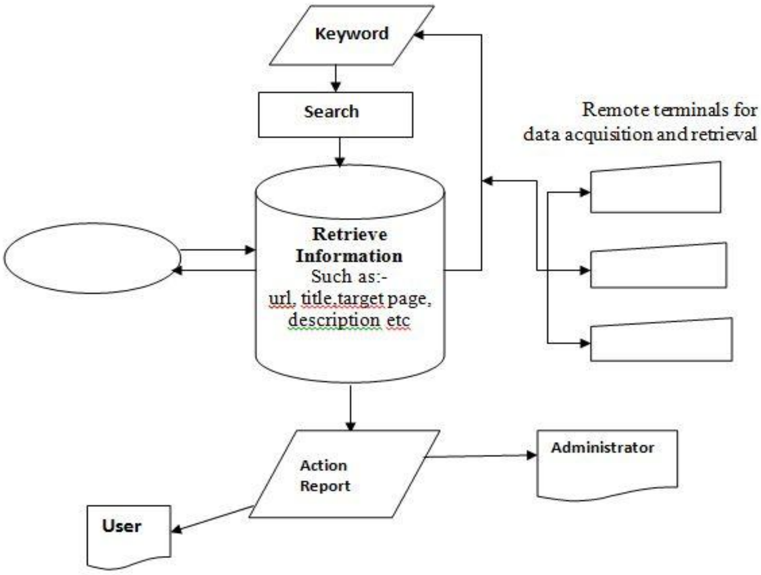

Keyword
Search
Remote terminals for data acquisition and retrieval
Retrieve Information Such as:- url, title,target page, description etc
Action Report
Administrator
User

Figure 2: IFD of Search Engine

1.3 Limitation of the Present System

- **Time consuming**: In present system, information retrieval is very time consuming.
- **Complex queries**: The queries are very complex in present system.
- **Less user interactivity**: In present system, user interactivity is very poor.
- **Scam**: As more visitors , there is a chance to more scam and sometimes it might become cost efficient as more and more people to advertise a website online.
- **Learning curve**: Using engines does involve a learning curve. Many beginning Internet users, because of these disadvantages, become discouraged and frustrated.
- **Sophistication**: Regardless of the growing sophistication, many well thought-out search phrases produce list after list of irrelevant web pages. The typical search still requires sifting through dirt to find the gems.
- **Overload**: Search engine creates information overload.

1.4 Proposed SRS

To minimize the problems as specified above we are proposing the following structure and the IFD of the SRS.

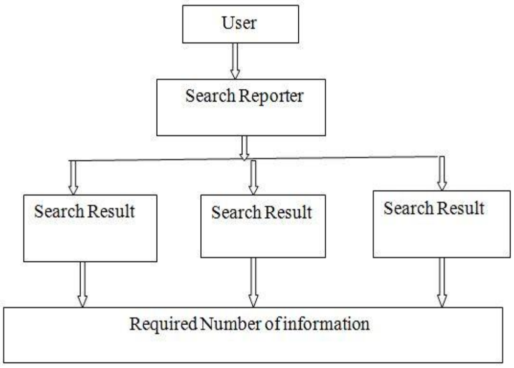


Figure 3: Structure Diagram of SRS

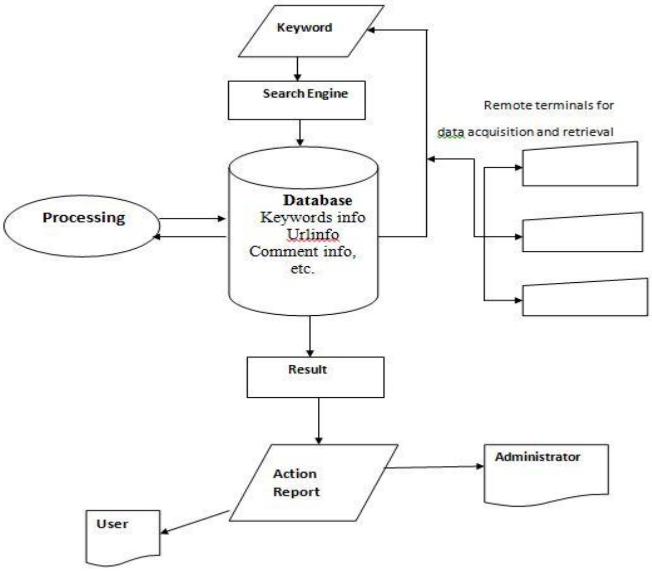


Figure 4: IFD of SRS

1.5 Objectives of SRS

The SRS is an open system. Here the user who wants to search anything, he/she just needs to visit the website and searches the keyword which he/she actually wants. The main objectives of this system are:

- to save time
- to help those users who don't know about searching
- to give the users meaningful information which they need
- to place the less important information in the last position

Also the SRS possesses the following advantages:

- It offers instant services, no need of time consuming
- The new users can easily understand the system and maintain it easily
- The system enables the users to precisely describe the information that they seek
- The SRS reduces the problem of sophistication and scam of present search engines.

II. DESIGN METHODOLOGY

The purpose of system design is to create a technical solution that serves both the user and the admin. The system should be designed in such a way that is very flexible to use for both the administrator and the user. The preparation of the environment needed to build the system, the testing of the system and the migration and the preparation of the data that will ultimately be used by the system are equally important. In addition to designing the technical solution, system design is the time to initiate focused planning efforts for both the testing and data preparation activates. The SRS is a real life problem solving application. Both the admin section and the user section are designed in such a way that both parties enjoy the facilities of the application.

2.1 Modular Design

The whole system is divided into two parts i.e. the user and the admin section. That is why, the modular design of the system is also divided into two modular diagrams.

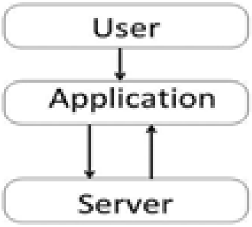


Figure 5(a): Modular diagram for user

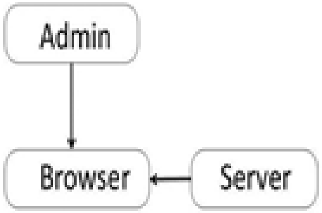


Figure 5(b): Modular diagram for admin

2.2 Use-case Diagram

It covers the whole SRS and how it works. It makes easier to communicate between the user and the system developers. The two main components of a use case diagram are actors and use cases as specified in the following diagram.

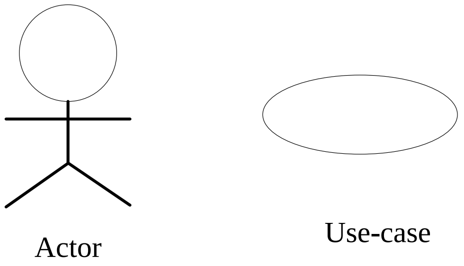


Figure 6: Symbol of Actor and Use-case

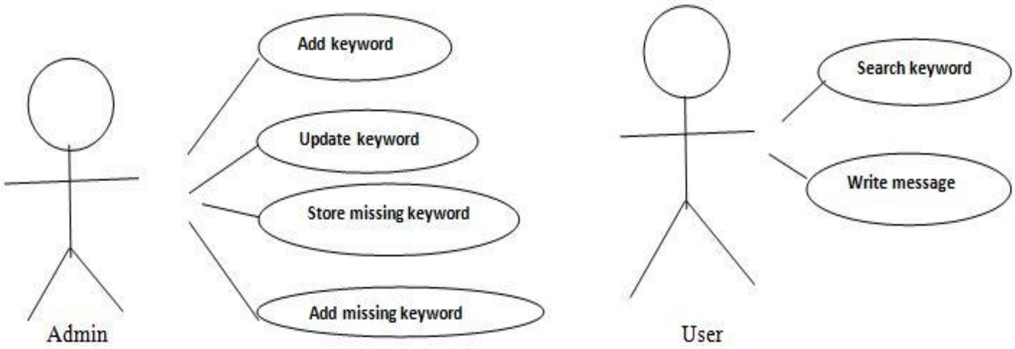


Figure 7: Use-case Diagram for Admin and User

2.3 Working Stucture

The SRS works in the following way:

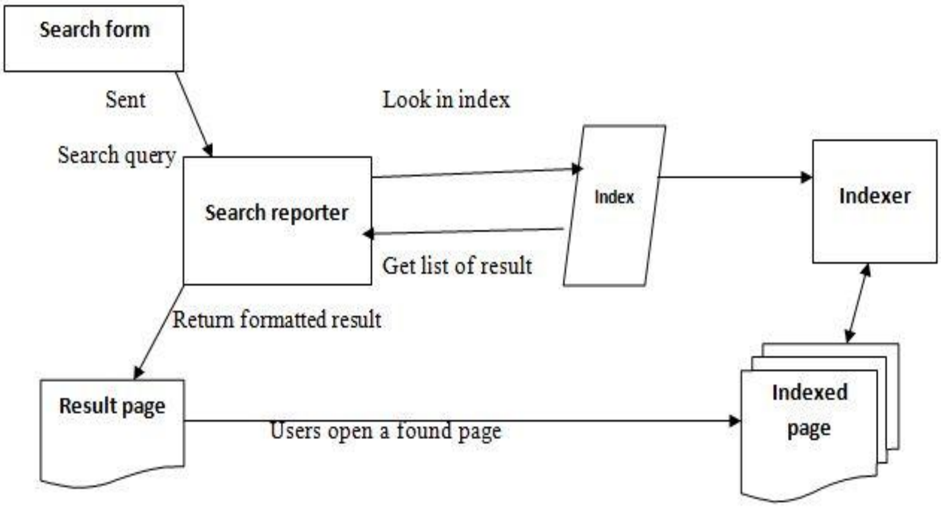


Figure 8: Working Structure of SRS

The working steps are in details:

- Create an index.
- Receive a query-a set of search terms
- Look in the index file for matches
- Gather the matching page entries and rank them by number of keyword matches
- Format the result
- Return the result page in HTML to the searcher’s web browser.

2.4 Data Flow Diagram (DFD)

The level-0, level-1, and level-2 DFD of the SRS are shown below:

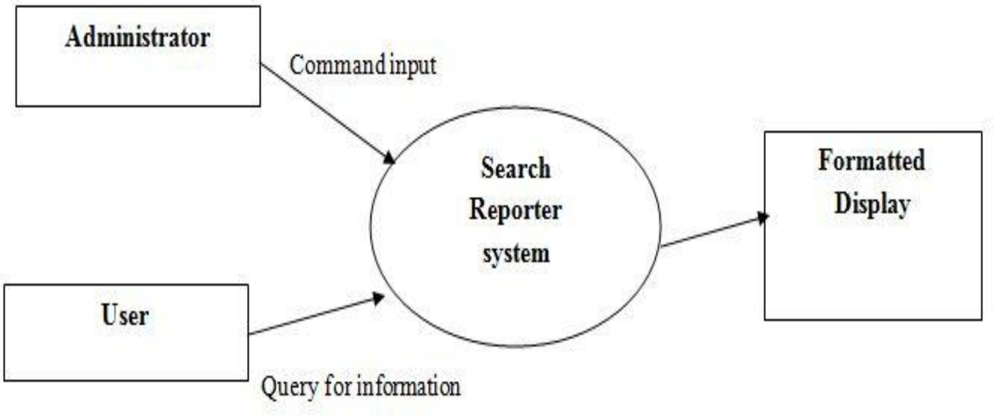


Figure 9: Level-0 Data Flow Diagram

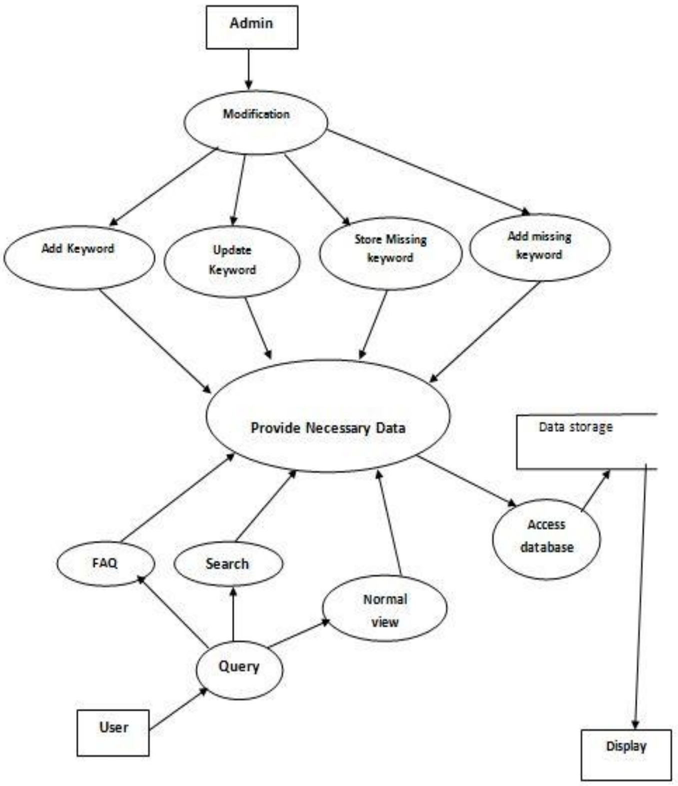


Figure 10: Level-1 Data Flow Diagram

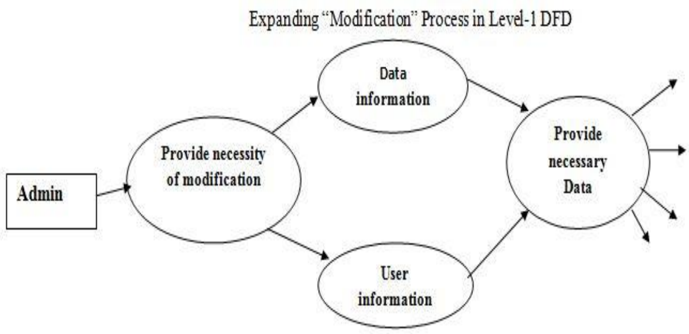

Expanding "Modification" Process in Level-1 DFD
Data information
Admin
Provide necessity of modification
Provide necessary Data
User information

Figure 11: Level-2 Data Flow Diagram

2.5 Relational Diagram

The relational or schema diagram among the table used in the SRS is shown below:

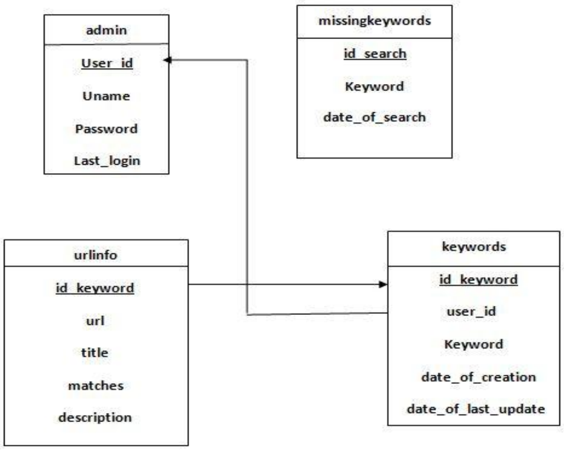


Figure 12: Relational Diagram of **SRS**

2.6 Activity Diagram

The activity at different level of the system can be shown in the following diagram:

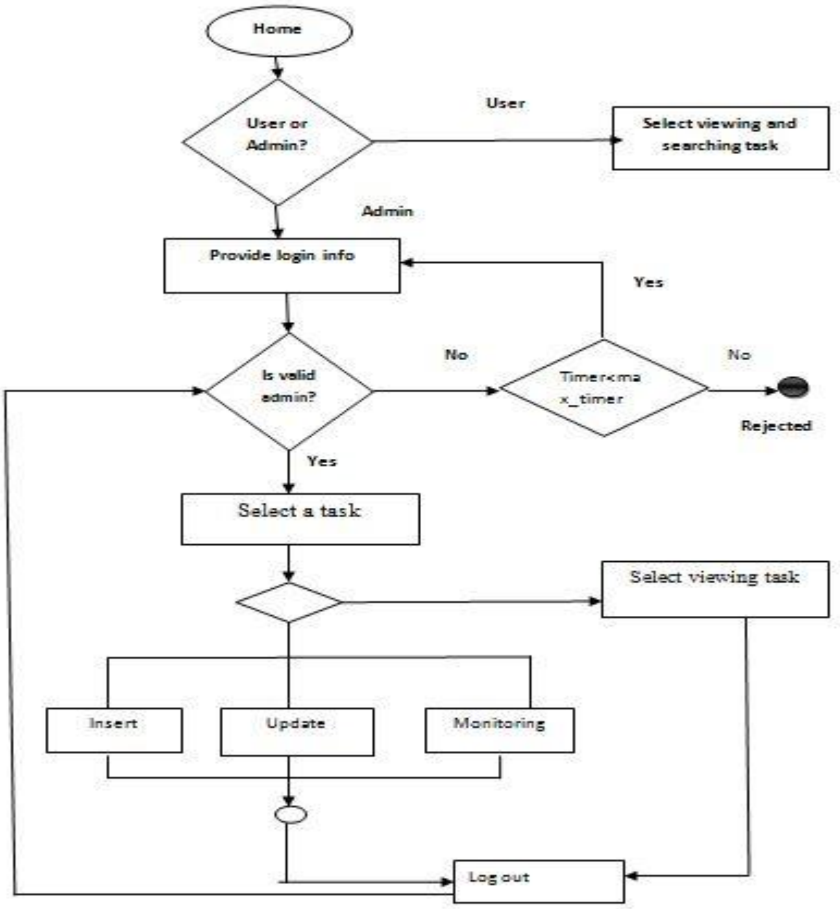


Figure 13: Activity Diagram of SRS

2.7 Architecture Diagram

Architecture Diagram of SRS designed for searching shows the different components of the system. We have populated the central database of the system by MySql database and populated it in the MySql Essential-5.0.67. The web server used is Xampp-1.7.3. This system designed by HTML, PHP, JavaScript, and CSS provides an interface with the user using both the servers. This interaction requires a network among the computer from which the personnel will provide their requirements.

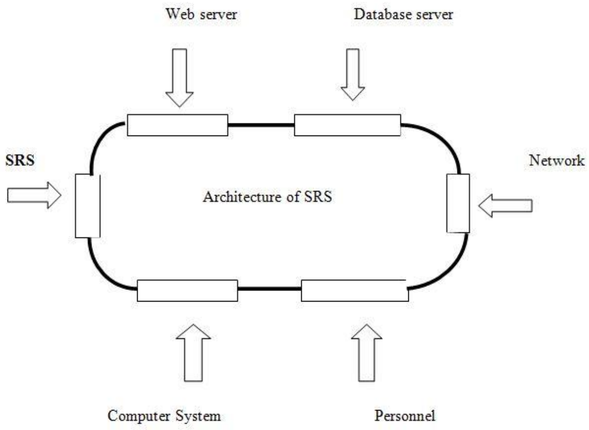


Figure 14: Architecture Diagram of SRS

III. TOOLS AND TECHNOLOGY

The latest demanding tools and technologies such as HTML, PHP, CSS, JavaScript, MySQL database, and Apache web server have been used to develop the SRS. More significantly, a phonetic algorithm namely Metaphone algorithm has been used for searching in SRS to ignore the spelling error in the searching keywords.

3.1 Importance and Phonetic Algorithm

A phonetic algorithm is an algorithm for indexing of words by their pronunciation. The main advantage of phonetic algorithm is to eliminate misspelling of words. When user search for a keyword than it may happen that he/she will place misspell of his desired keyword. To solve this problem we use Metaphone algorithm in this system which produce phonetic similarity of different alphabets or group of alphabets to avoid the loss of information. Here, users will give search keywords and actual data are matched by its phonetic similarity.

3.2 Metaphone Algorithm

Metaphone is a phonetic algorithm, published by Lawrence Philips in 1990, for indexing words by their English pronunciation. It fundamentally improves on the Soundex algorithm by using information about variations and inconsistencies in English spelling and pronunciation to produce a more accurate encoding, which does a better job of matching words and names which sound similar [5]. As with Soundex, similar sounding words should share the same keys. Metaphone is available as a built-in operator in a number of systems, including later versions of PHP.

Metaphone codes use the 16 consonant symbols 0BFHJKLMNPRSTWXY. The '0' represents "th" (as an ASCII approximation of Θ), 'X' represents "sh" or "ch", and the others represent their usual English pronunciations. The vowels AEIOU are also used, but only at the beginning of the code. This table summarizes most of the rules in the original implementation:

1. Drop duplicate adjacent letters, except for C.
2. If the word begins with 'KN', 'GN', 'PN', 'AE', 'WR', drop the first letter.
3. Drop 'B' if after 'M' at the end of the word.
4. 'C' transforms to 'X' if followed by 'IA' or 'H' (unless in latter case, it is part of '-SCH-', in which case it transforms to 'K'). 'C' transforms to 'S' if followed by 'I', 'E', or 'Y'. Otherwise, 'C' transforms to 'K'.
5. 'D' transforms to 'J' if followed by 'GE', 'GY', or 'GI'. Otherwise, 'D' transforms to 'T'.
6. Drop 'G' if followed by 'H' and 'H' is not at the end or before a vowel. Drop 'G' if followed by 'N' or 'NED' and is at the end.
7. 'G' transforms to 'J' if before 'I', 'E', or 'Y', and it is not in 'GG'. Otherwise, 'G' transforms to 'K'.
8. Drop 'H' if after vowel and not before a vowel.
9. 'CK' transforms to 'K'.
10. 'PH' transforms to 'F'.
11. 'Q' transforms to 'K'.
12. 'S' transforms to 'X' if followed by 'H', 'IO', or 'IA'.
13. 'T' transforms to 'X' if followed by 'IA' or 'IO'. 'TH' transforms to '0'. Drop 'T' if followed by 'CH'.
14. 'V' transforms to 'F'.
15. 'WH' transforms to 'W' if at the beginning. Drop 'W' if not followed by a vowel.
16. 'X' transforms to 'S' if at the beginning. Otherwise, 'X' transforms to 'KS'.
17. Drop 'Y' if not followed by a vowel.
18. 'Z' transforms to 'S'.

19. Drop all vowels unless it is the beginning [5].

3.3 Works on Metaphone Algorithm

- Metaphone Calculator [6].
- Doing a fuzzy match in MySQL. Soundex and Metaphone algorithm [7].
- Naushad UzZaman and Mumit Khan, “A Bangla Phonetic Encoding for Better Spelling Suggestions” [8].
- Chakkrit Snae and Michael Brückner, “Novel Phonetic Name Matching Algorithm with a Statistical Ontology for Analyzing Names Given in Accordance with Thai Astrology” [9].
- Chakkrit Snae, “A Comparison and Analysis of Name Matching Algorithms” [10].

3.4 Working Steps of Metaphone Algorithm

The main two steps of working of the Metaphone Algorithm are illustrated below:

Step-1:

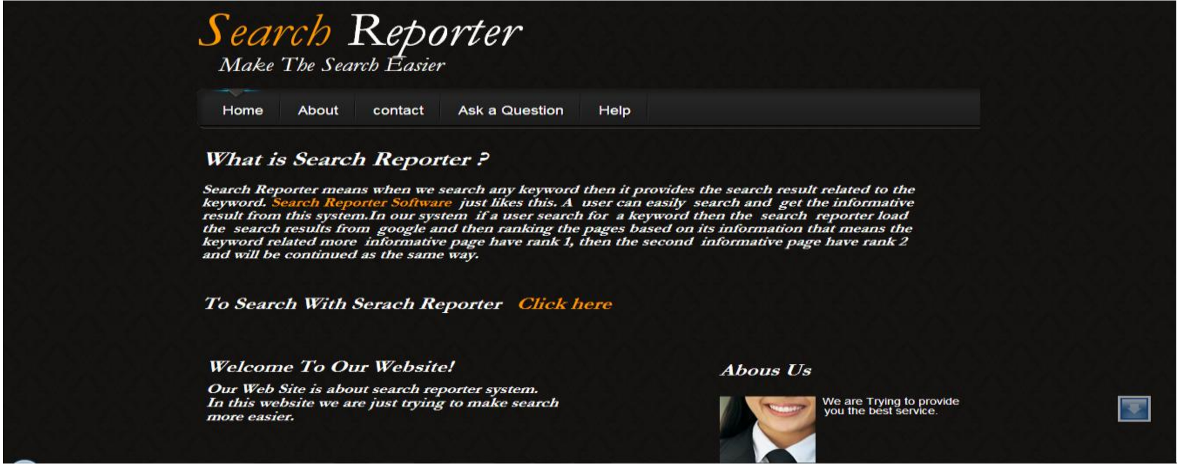
- Take the input for inserting in the database
- Select the words which may be searched
- Make Metaphone code for the searchable words
- Store Metaphone code in database

Step-2:

- Create Metaphone code for each of the words in the search key
- Retrieve data from database based on the code of the words
- Calculate percentage of matching with the keywords and the actual data in the database
- Display retrieved data in the ascending order of matched with the actual data
- Divide the whole data into several pages if it is required

IV. SNAPSHOTS OF SRS

The home page in the SRS looks like the following:



Search Reporter

Make The Search Easier

What is Search Reporter ?

Search Reporter means when we search any keyword then it provides the search result related to the keyword. Search Reporter Software just likes this. A user can easily search and get the informative result from this system.In our system if a user search for a keyword then the search reporter load the search results from google and then ranking the pages based on its information that means the keyword related more informative page have rank 1, then the second informative page have rank 2 and will be continued as the same way.

To Search With Serach Reporter Click here

Welcome To Our Website!

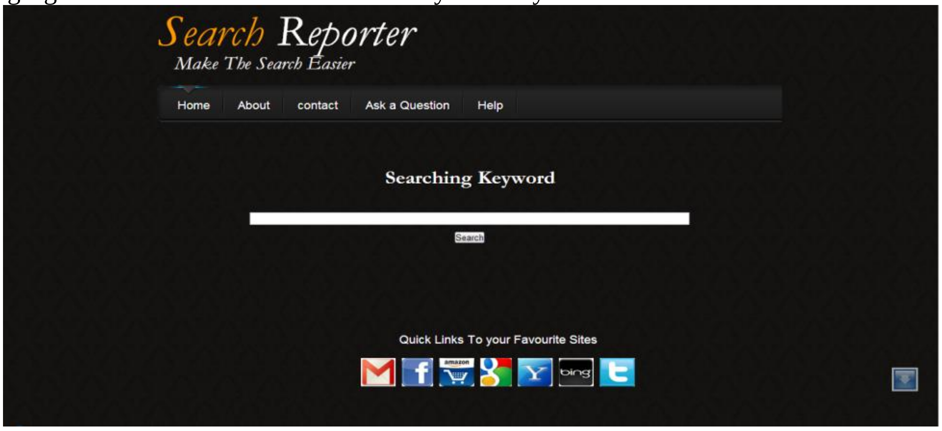
Our Web Site is about search reporter system. In this website we are just trying to make search more easier.

Abous Us

We are Trying to provide you the best service.

Figure 15: Hope Page of SRS

The following figure shows the form to search for keywords by the users:



Search Reporter

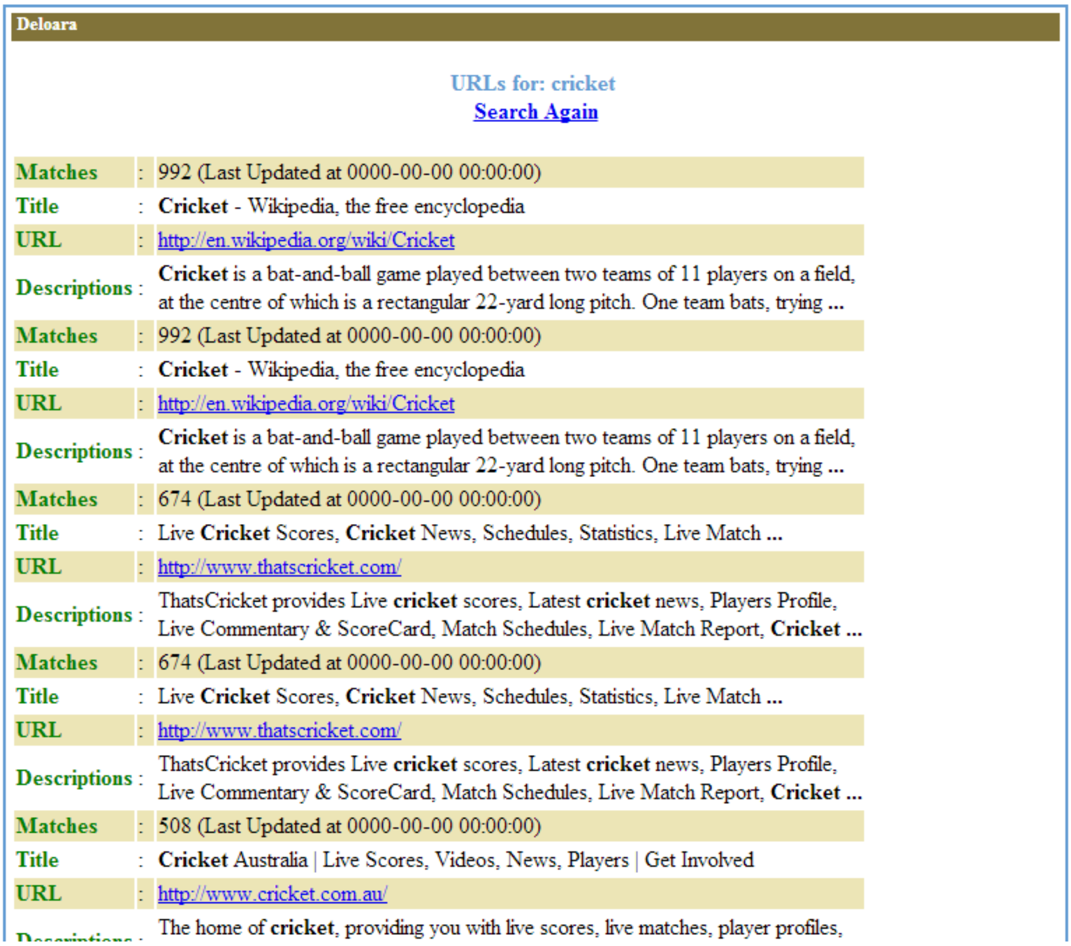
Make The Search Easier

Searching Keyword

Quick Links To your Favourite Sites

Figure 16: Searching Form

Suppose that the user has searched for “Cricket” in this SRS which is connected to a search engine such as Google. In searching, he/she can mistype the input keywords but the Metaphone algorithm used in SRS performs the related results as shown below:



Deloara

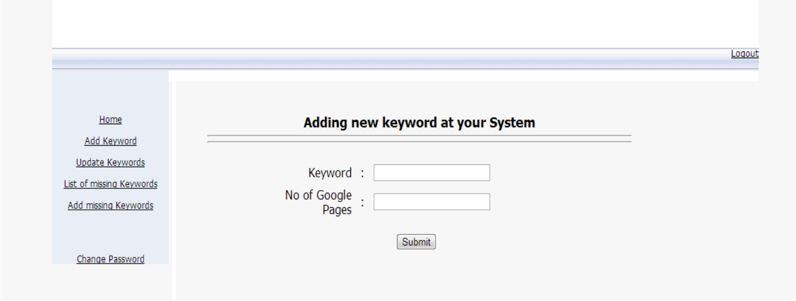
URLs for: cricket

Search Again

Matches	:	992 (Last Updated at 0000-00-00 00:00:00)
Title	:	**Cricket** - Wikipedia, the free encyclopedia
URL	:	http://en.wikipedia.org/wiki/Cricket
Descriptions	:	**Cricket** is a bat-and-ball game played between two teams of 11 players on a field, at the centre of which is a rectangular 22-yard long pitch. One team bats, trying ...
Matches	:	992 (Last Updated at 0000-00-00 00:00:00)
Title	:	**Cricket** - Wikipedia, the free encyclopedia
URL	:	http://en.wikipedia.org/wiki/Cricket
Descriptions	:	**Cricket** is a bat-and-ball game played between two teams of 11 players on a field, at the centre of which is a rectangular 22-yard long pitch. One team bats, trying ...
Matches	:	674 (Last Updated at 0000-00-00 00:00:00)
Title	:	Live **Cricket** Scores, **Cricket** News, Schedules, Statistics, Live Match ...
URL	:	http://www.thatscricket.com/
Descriptions	:	ThatsCricket provides Live **cricket** scores, Latest **cricket** news, Players Profile, Live Commentary & ScoreCard, Match Schedules, Live Match Report, **Cricket** ...
Matches	:	674 (Last Updated at 0000-00-00 00:00:00)
Title	:	Live **Cricket** Scores, **Cricket** News, Schedules, Statistics, Live Match ...
URL	:	http://www.thatscricket.com/
Descriptions	:	ThatsCricket provides Live **cricket** scores, Latest **cricket** news, Players Profile, Live Commentary & ScoreCard, Match Schedules, Live Match Report, **Cricket** ...
Matches	:	508 (Last Updated at 0000-00-00 00:00:00)
Title	:	**Cricket** Australia | Live Scores, Videos, News, Players | Get Involved
URL	:	http://www.cricket.com.au/
Descriptions	:	The home of **cricket**, providing you with live scores, live matches, player profiles,

Figure 17: Result Page

After successful login to the SRS, admin can insert keyword by using this page:



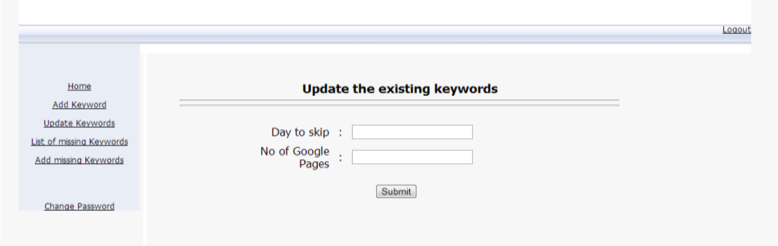
Adding new keyword at your System

Keyword :

No of Google Pages :

Figure 18: Keyword Inserting Form

Then the admin can update keywords by the following page:



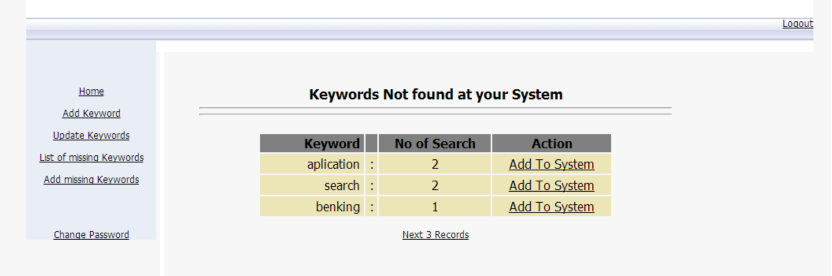
Update the existing keywords

Day to skip :

No of Google Pages :

Figure 19: Keyword Updating Form

And the admin can see the missing keywords to add them in the SRS by this page:



Keywords Not found at your System

Keyword		No of Search	Action
aplication	:	2	Add To System
search	:	2	Add To System
benking	:	1	Add To System

Figure 18: Monitoring Missing Keyword and Then Inserting Them

V. CONCLUSION

Analyzing the above descriptions of the Search Reporter System along with a search engine like Google it can be concluded that the SRS is highly effective, efficient and user-friendly for the fulfillment of the user's requirement. Users can be highly benefited using the SRS. With the rapid enhancement of modern technology people want beneficial information within a short moment of time. Web search engine can provide that required information but it shows many more necessary and less necessary URLs about the searching keyword elapsing a more time. Hence the SRS aims to reduce the less necessary and totally unnecessary URLs from a list of URLs resulted from a search engine after analyzing some factors for the same reducing the searching time. To get full benefits of modern web technology using the SRS, it can be integrated with other search engines.

Future Plan:

In the world nothing is free from error. So it is very common that the SRS may contain error. The SRS fully dependent on any search engine like Google. In future the following features will be integrated with the SRS:

- Searching based on search analyzer
- Mailing facility
- Chatting facility

REFERENCES

[1] http://en.wikipedia.org/wiki/Web_search_engine

[2] http://en.wikipedia.org/wiki/Web_crawler.

[3] http://www.brightplanet.com/2012/11/deep-web-search-engines-vs-web-harvest-engines-finding-intel-in-a-growing-internet/

[4] http://en.wikipedia.org/wiki/Information_flow_diagram

[5] http://en.wikipedia.org/wiki/Metaphone

[6] http://www.vbforums.com/showthread.php?655230-Metaphone-Calculator

[7] http://theunderweb.com/doing-a-fuzzy-match-of-names-in-mysql-soundex-and-metaphone-algorithms.html.

[8] Naushad UzZaman and Mumit Khan, "A Bangla Phonetic Encoding for Better Spelling Suggestions", Proc. of 7th International Conference on Computer and Information Technology (ICCIT 2004), pp. 76-80.

[9] Chakkrit Snae and Michael Brückner, "Novel Phonetic Name Matching Algorithm with a Statistical Ontology for Analyzing Names Given in Accordance with Thai Astrology", Issues in Informing Science and Information Technology, 2009, V.6, pp. 497-515.

[10] Chakkrit Snae, "A Comparison and Analysis of Name Matching Algorithms", World Academy of Science, Engineering and Technology, 2007, Issue 1.

Md. Palash Uddin (palash_cse@hstu.ac.bd) received his B.Sc. degree in Computer Science and Engineering from Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh in 2013. His main working interest is based on artificial intelligence, bioinformatics, algorithm analysis, database structure analysis, software engineering, theory of computation etc. Currently he is working as a lecturer in Dept. of Computer Science and Information Technology in Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh. Previously, he was a lecturer in department of Computer Science and Engineering at Central Women's University, Dhaka, Bangladesh. He has research publications in various fields of Computer Science and Engineering.

Mst. Deloara Khushi received his B.Sc. degree in Computer Science and Engineering from Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh in 2013. Her main working interest is based on communication theory and computer algorithms. Currently she is working as a lecturer in Dept. of Computer Science and Engineering in Bangladesh University of Business and Technology, Dhaka, Bangladesh

Fahmida Akter received his B.Sc. degree in Computer Science and Engineering from Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh in 2013. Her main working interest is based on bioinformatics and data mining. Currently she is working as a lecturer in Dept. of Computer Science and Engineering in East Delta University, Chittagong, Bangladesh

Md. Fazle Rabbi received his B.Sc. degree in Computer Science and Engineering from Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh in 2008. His main working interest is based on bioinformatics, data structures and algorithm etc. Currently he is working as an assistant professor in Dept. of Computer Science & Information Technology in Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh