ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

A Novel Framework for Context Based Focused Search Engine

[1] Dr.K.Vaishali
[1] Prof. of CSE, Dept. of CSE, Jyothismathi Institute of Technology and Sciences, Karimnagar

Abstract: **- Context based Focused search engine is the information retrieval tool that narrows the search space by finding the documents relevant to user query only. The objective of a focused search engine is to return more specific results in response to a user query rather than large number of results. In this paper, the architecture for such a focused search engine has been proposed that works on the basis of the context of web pages and serve the user with contextually more related web pages semantically.**

Keywords**: Search Engine, Semantic Search, Context based Search.**

I. INTRODUCTION

User searches for information from WWW, related to his topic of interest by using various information retrieval tools. The most prevalent and often utilized instrument is a search engine. The real duty of a search engine is to keep up an archive of downloaded web pages and react to the client search inquiry with a rundown of coordinated web pages. As of now, there are around 50 billion web documents on the Web, put away on a huge number of web server around the world. There might be countless documents from client's inquiry perspective, a couple of progressively and a couple of less important.

An ordinary search engine returns a large number of coordinated documents in light of a question. The measure of data gave by such a search engine is too vast to experience and along these lines, prompts the issue of data pointless excess. This issue additionally exasperates on account of unpracticed clients, attempting to search the data from the web. In accordance with the acclaimed "8-second control", such clients take a gander at an initial couple of results and tend to dismiss. This conduct rouses the need to create devices with high accuracy that can give top 'k' important connects to web documents, rendering search engines and related advances a potential region of research. Accordingly, the testing errand for a search engine is to locate the pertinent documents identified with client search question. The answer for it has turned out as the engaged search engine.

The engaged search engine is a data recovery device that recovers the pertinent web pages write that the client question. Additionally, it focuses more on the quality or pertinence of the web pages rather on amount. In this paper, a novel engineering for the setting based centered search engine is planned, which show topic-particular documents from the Web, taking into an account the need of a client to search profoundly significant documents.

II. CONSTRUCTION OF A GENERAL SEARCH ENGINE

This section provides a brief review of general search engine architecture. A search engine generally comprises of the following components:

A) User Interface
B) Query Processor
C) Index Builder
D) Ranking Module
E) Crawler

The design is made out of two layers back-end process and front-end process. Back-end process is made out of crawler, web diagram developer, page rank adding machine and list manufacturer. The front-end process is made out of the UI, query processor and positioning module. The significant obligation of back-end process is to keep up the store of downloaded web documents and fabricate an upset list for quick recovery from the archive. The significant duty of frontend process is to acknowledge the client query and give a rundown of coordinated web documents in slipping request of their figured rank.

III. MOTIVATIONS FOR IMPROVEMENT

Keeping in mind the end goal to search the data utilizing a search engine, by and large, clients enter query keywords without giving the specific circumstance. A solitary query keyword has different meanings relying on

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

its utilization called logical faculties. This property of words having an alternate importance in various settings is called polysemy. General search engine design does not channel and rank documents in view of polysemy setting. From that point, the search engine searches and delivers the outcomes without thinking about the client's search setting and thus, restores an extensive number of documents which could possibly satisfy their prerequisites or instance, for a query keyword 'Server', a general search engine reacts with top outcomes identified with the PC server, for example, Server (Computing), File Server, Virtual Server, Server (Ubuntu) and Server Research Center and so forth. Though, word 'Server' can likewise allude to a server, court recreations, a software engineering host or a utensil. The general search engine does not give an interface to catch the correct or real setting of query keywords.

Data needs of a client along these lines can't be chosen simply on the premise of query keywords. In this manner, change in existing design is required to get the particular set of client's search and that of the web page Moreover as indicated by the client search patterns; just 40% of the clients go up-to third page of showed comes about. The real gathering of clients either changes the query, the search methodology or the data recovery instrument. It demonstrates that clients are more intrigued to get pertinent outcomes either in top 10 or 20 positions on the principal page.

III.I Contextual Sense: Meaning Scope and Examples

For the most part, query keywords go by clients have various meanings (as Server); accordingly, the data needs of client stay uncertain and can't be obviously reasoned from query keywords. This prompts numerous unwanted outcomes in top positions. For instance, in English 'mouse' is a pointing gadget in registering and rat somewhere else. Ebb and flow search engines can't find the setting based on just query keywords, subsequently, for a query keyword 'mouse' they react with a rundown of web pages containing documents identified with the mouse in a feeling of pointing gadget and rodents. It might happen that occasionally wanted substance are not found even in the top couple of results. Hence it's a testing assignment to make search engine distinguish distinctive relevant faculties of a word. The proposed setting based centered search engine engineering has presented another module called Context sensitive extractor (CSE) to expel equivocalness in query keywords because of polysemy property.

Scope

The word 'relevant sense' has been given distinctive definitions and degrees by various researchers. The setting is alluded to as data about an area, characters of individuals and questions in the nearness of the client and the adjustments in them [1]. Dark colouredet al [2] has included the season of day, season or temperature and so on to area and characters to get the unique circumstance. Another researcher [3] has included client's passionate state, the focal point of consideration regarding the area and close-by personalities, to consider as a setting. Dey and Aowd [4] has characterized the setting as any data that can be utilized to describe the circumstance of a client. Anders and Marius [5] have thought about the arrangement of natural states and its settings for the client, to get the set of client subordinate applications and administrations, for example, data recovery. In the region of data recovery, the setting depends on client's current exercises, for example, went by pages and messages, past search inquiries and so forth [6]. Some have considered the client's understood conduct with some metadata data [7] to recognize the ebb and flow search require. Another bit of work has given the possibility of programmed development of client profile [8]. Work done in [9] has attempted to evacuate the uncertainty utilizing query logs, navigate history inside a session, client profile and so forth.

Motivating Examples

Google is the most prevalent search engine which files the greatest number of publically index documents from the web among all other accessible search engines. It keeps up the word vocabulary and returns equivalent word words match in order arrange because of a given query keyword [10]. For example, because of the query keyword Student, the Google search engine restores a rundown of equivalent words matches as appeared in

Figure 1: Google results for keyword 'Student'

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

The results that Google returns in response to various query keywords are generally the synonyms word pairs and do not show the contextual senses of the keywords. It does not show the polysemy based contextual senses of the keywords. Similarly, various words that are having multiple meanings in different contexts when passed to Google, provides only the synonyms pair. Few such examples of words along with their different meanings and synonyms word pairs provided by Google. For instance, Google provides options 'Bata, Batman, Bata India and Batman Games'; whereas, the multiple meanings of 'Bat' are 'a kind of mammal and cricket bat'. None of the different meanings of 'Bat' is covered by Google.

III.II IMPROVEMENT IN EXISTING ARCHITECTURE

Thus the basic architecture of a search engine needs to be improved to determine and resolve the Polysemy based context of the query keywords and retrieve more relevant documents at higher positions. In order to achieve the same, some changes are proposed in the general architecture of search engine. Modifications are proposed in the query processor, ranking module and user interface of the front-end process layer. The back-end process layer has been divided into two sub-layers: middle layer and bottom layer. No modifications are proposed for the bottom sub layer whereas the changes have been done in index builder module and one extra module named back link extractor has been added at middle sub layer. The detailed proposed modified architecture of context based focused search engine is discussed in the following section.

IV. PROPOSED MODIFIED ARCHITECTURE OF SEARCH ENGINE

The proposed modified architecture of search engine is shown in Figure 2. It is referred as Context Based Focused Search Engine. It works on different contextual interpretation of the keywords, entered by the user in form of query and the words present in the web documents. The architecture consists of the three layers:

A) Top Layer
B) Middle Layer
C) Bottom Layer

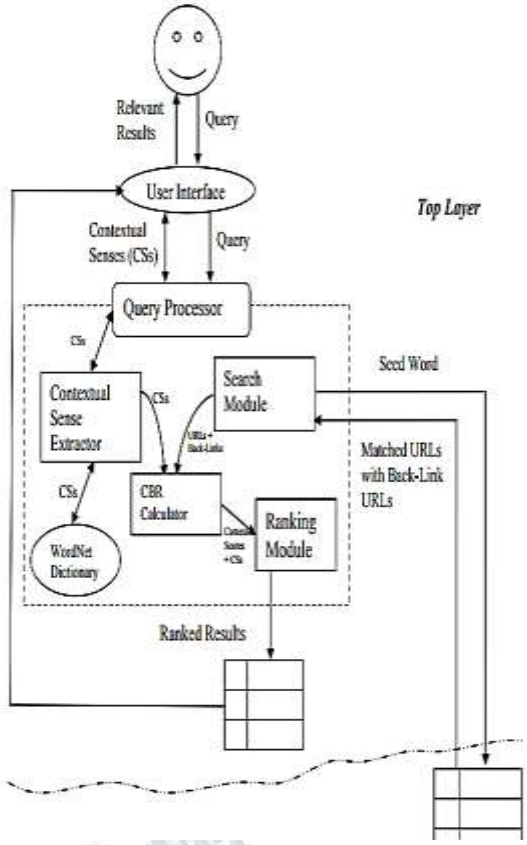

Figure 2: High level architecture of proposed context based focused search engine

V. THE WORKING OF THE PROPOSED SEARCH ENGINE:

In this manner, in the proposed design the documents are put away in a storehouse and an inverted index is maintained. The query processor takes the client query and searches the inverted index to get the rundown of coordinated documents. The different contextual faculties of the query keywords are shown to the client and solicited to indicate the context from the search. At that point the coordinated documents are positioned in that contextual faculties and corresponding documents from the store are shown to the client as the final outcomes. For instance, query for keyword 'Server' is appeared in Figure 3. The inverted index is maintained in sequential request. The different contextual faculties of server are then shown to the client, for example, Waiter, Host, Server (Utensil), Server (Court Game) and so forth where, client has chosen the sense as 'host' and then the corresponding pertinent positioned documents i.e. D1, D5, D6, D11 will be shown as the consequences of search.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

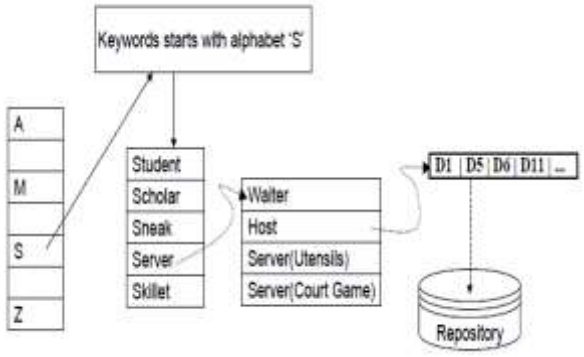

Figure 3: Example for the proposed architecture

VI. CONCLUSION

In this paper, the changed engineering for centered search is recommended that thinks about the different contextual faculties of the words to recognize the correct need of the client. The search is engaged, as just the web documents identified with the client chose sense are shown in the outcome list. The different contextual faculties of words are extricated. The reasonable information structure, the context-based index is intended to store the contextual faculties. The context-based relevance evaluation and ranking of web documents empower to show just the pertinent archive at higher positions i.e. prior to the outcome list.

REFERENCES

[1]. Schilit B.N., Theimer M.M., "Disseminating Active Map Information to Mobile Hosts", IEEE Network, vol. 8, issue 5, pp. 22-32, Sep. 1994.

[2].Brown P.J., Bovey J.D., Chen X., "Context-Aware Applications: From the Laboratory to the Marketplace", IEEE Personal Communications, vol. 4, issue 5, pp. 58-64, 1997.

[3].Ryan N., Pascoe J., Morse D., "Enhanced Reality Fieldwork: the Context-Aware Archaeological Assistant", Computer Applications in Archaeology, Proc. of 25th anniversary conference, University of Birmingham, April 1997.

[4].Dey A.K., Abowd G.D., "Towards a better understanding of context and context awareness", Proc. of the first International Symposium on Handheld and Ubiquitous Computing, Karlsruhe, Germany, pp. 304-307, 1999.

[5].Internet World Stats Usage and Population Statistics, available at, www.internetworldstats.com/emarketing.htm, accessed on 10th Aug. 2013.

[6].Teevan J., Dumais S.T., Horvitz E., "Personalizing search via automated analysis of interests and activities", SIGIR'05, Proc. of the 28th annual international ACM SIGIR conference on research and development in information retrieval, pp. 449-456, 2005.

[7].Belkin N.J., Muresan G. and Zhang X.-M., "Using user's context for IR personalization." Proc. of the ACM/SIGIR Workshop on Information Retrieval in Context, ACM Press, Sheffield, UK, pp. 1-3, July 25-29, 2004.

[8].Pitkow J., Schütze H., Cass T., Cooley, R., Turnbull D., Edmonds A., Adar E., Breuel T., "Personalized Search", Communications of ACM, vol. 45, issue 9, pp. 50-55, Sep. 2002.

[9].Bai J., Nie J.-Y., Bouchard H., Cao G., "Using Query Contexts in Information Retrieval", SIGIR'07, Proc. of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, Amsterdam, Netherlands, July 23–27, pp. 15- 22, 2007.

[10].Page Lawrence, Brin Sergey, Motwani Rajeev, Winograd Terry, "The PageRank Citation Ranking: Bringing Order to the Web", Technical Report, Stanford University InfoLab, 1999.