Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 5, May 2016, pg.231 – 234

Efficient Keyword Query Routing for Search Engines

Viral Patel[1], Hitul Patel[2]
[1]M.E in C.E, Swaminarayan College of Engineering & Technology (S.C.E.T., Kalol), India
[2]HOD of ME(CE) in Swaminarayan College of Engineering & Technology (S.C.E.T., Kalol), India
patelvj0209@gmail.com; hitulce@gmail.com

ABSTRACT - Increasing data on web has resulted in increasing challenge for keyword query routing. Top k routing is an approach used for keyword query search. Keyword query routing work efficiently for one or two keyword but its efficiency drastically reduces while handling multiple keywords. Our approach is to increase the efficiency even for multiple keyword querying. We are going to follow keyword-element relationship which could show the relation among keywords and data elements referring them. We are going to adopt link state approach for routing.
Keywords— Keyword query, keyword search, keyword query routing

I. INTRODUCTION

Data mining can be defined as the process designed to explore huge amount of data also called “big data” by methods like artificial intelligence, machine learning and database system.[6] Web mining can be defined as the data mining technique used to discover and fetch out information from the web documents.
Web is a collection of countless documents presented in a distributed manner on different databases. Day by day plenty of data is being added which has created a challenge for its efficient retrieval. The main reason behind the need of keyword query routing is the large number of data which could be exploited by the structured languages. There are several solutions proposed by which the relevant results can be found but only from the solo database. So this approach could not work for web search engine results as there are tons of linked data sources. Our aim is to develop routing plans that can compute multiple sources as well as multiple keyword. For this we are providing below approach: We are going to analyze keyword query routing technique for the large number of linked data resources. And finding the optimum path for the keyword through sources. Hence we can reduce the cost of searching to beneficial level.

We have preferred top-K routing algorithm for the keyword query searching and link state for routing.

It is that the link state approach far better and efficient than the existing approach for routing. So by using link state approach, we can get more efficiency than the existing solution. Moreover, our approach is also to optimize the text by filtering out the articles from the text document.

II. Introduction to Keyword Query Routing

Researchers are constantly working on finding the approach to reduce the time to fetch the results from multiple databases. As said, there are existing solution which only works only for one or couple of keywords, however for multiple keyword input, the time taken is about few seconds to minutes also. This is absolutely no acceptable solution for multiple keyword input. There are millions and millions of RDF triples linked, but it is not going to show the same triple again even if it is having similar kind of meaning. One demonstration was done on publicly available 150 sources on web depicted that valid plan(precision at 1 of 0.92) which are highly relevant (mean reciprocal rank of 0.89) computed in 1 second on an average.[1]

III. Work involved in Query routing

There are mainly two kind of work. There are

1) Keyword search

As far as the keyword search is concerned. There are two more types of technique available in keyword search. They are:

a) Schema based approach

Schema based approaches are executed on top of off-the-shelf. In this approach, keywords are mapped on database.

b) Schema agnostic approach

Schema agnostic approaches are executed on data directly. The results are evaluated on the bases of data graph. The graph structure used in it is steiner graph.

2) Database selection [1][4]

It is necessary to find the appropriate database as the aim is to fetch out and display such result which user is requesting. So the main aspect of selection of relevant database is by considering the keyword relationships. For example, M-KS is a database selection technique which does by using matrix.

IV. Keyword routing system architecture

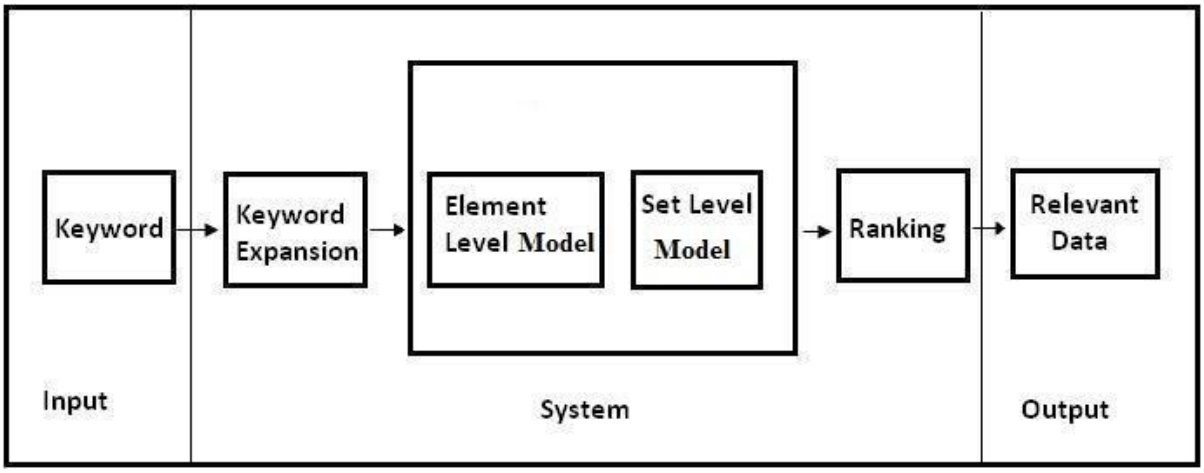

Fig.1 keyword routing system architecture

V. Introduction to Smith-Waterman

Below is the listing of the phases of keyword routing system shown in the above figure:

1) Keyword (input)

This is the input string which the user wants to query to find the sources of his/her wish, which is treated as keyword

2) Keyword expansion

This is the second phase after the keyword is input. In case of multiple keywords, each word is separately treated and the whole text is tokenized and stemmed.

3) Element level model

In this phase, the tokenized keyword is pointed to find out the elements where such keywords are present.

4) Set level Model

Different elements form a group (set) of the relevant and relative elements. This can also be called grouping of elements.

5) Ranking

Ranking refers to the arrangement of result stack according to its relevance.

6) Relevant data (output)

Exporting the final results fetched after the whole process of keyword querying of the input to user back.

VI. Method

Step 1 : Clean your text (remove punctuations and stop words).
Step 2 : Tokenize the text.
Step 3 : Stem the tokens.
Step 4 : Tokenize the text.
Step 5 : Find the TF(term frequency) for each unique stemmed token present.
Step 6 : Rank the stemmed tokens (keywords) using TF*IDF (IDF - Inverse Document Frequency)Build a network of keywords (repeat until all keywords and docs are being referred)
Step 7 : Create a single database with attributes based on graph
Step 8 : Display the results

VII. Flow Diagram

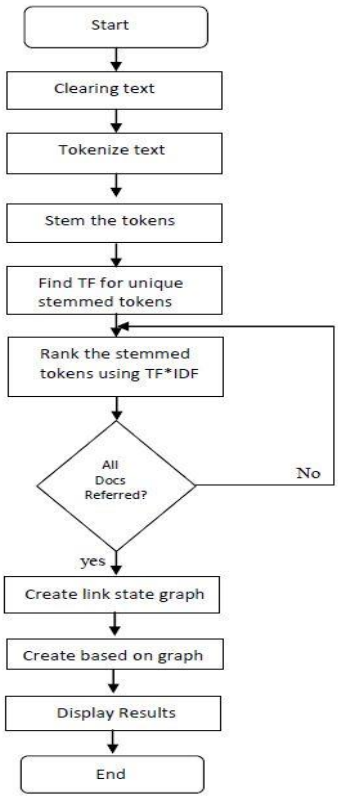

Fig 2 Flow chart

VIII. Keyword searching and retrieval steps

1. Find keywords
2. Ranking/Indexing of user's query results.
3. Elaborating representation and storage of information
4. Classification of documents (i.e. Pre-defined groups)
5. Clustering of documents (i.e. Automatically creates clusters)

IX. Experiments and evaluation

Tools used for our experiments are netbeans, eclipse, java(programming language). Our proposed algorithm is PPRJ with modification. Moreover, we have also filtered the articles from the document so that the data source become less so even this phase can help us to increase the efficiency. We have implemented the keyword extraction online as well as offline. As far as online is concerned, figure 6.1(c) show the demonstration of keyword extraction from the amazon.in's link in which we have extracted the keyword "furniture". Here, as the keyword is furniture, first of all we have collected all the data of the concerned url page, and treated it as an xml. Then we have filtered all the articles and made it light weight. We have calculated TF and IDF.
The approach for graph and routing we have adopted is link state approach which is far better approach than the existing one, which has resulted in the decreased delay time and increase in efficiency. The effort is even made to treat the offline file, that is the handy file present in our computer and the keyword extraction can be done here also.

CONCLUSION

Keyword query routing can greatly help to find the required source. By considering multilevel graph and applying link state approach and modifying the existing algorithm, more optimum path can be developed so reduce the routing cost through the nodes after input query. Moreover, the advantage is, we are not compromising with the quality of the results, but the time delay to fetch the results from the source is reduced. One more advantage is, it works efficiently even if the input are multiple keywords by the link state approach for routing the source nodes.

ACKNOWLEDGEMENT

The authors would like to thank the reviewers for their precious comments and suggestions that contributed to the expansion of this work.

REFERENCES

[1] "Keyword Query Routing" by Thanh Tran and Lei Zhang. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 2, FEBRUARY 2014
[2]Ms. Pawar Prajakta Bhagwat1 Mr. Niranjan L. Bhale, *ME Student, Dept of CE, HOD, IT Dept, MCERC, Nashik, India. MCERC, Nashik, India. "*Efficient Keyword Query Routing for Search Engines*"* Pawar Prajakta Bhagwat et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (1) , 2015, 434-437
[3]Prachi Karale, computer department vishwabharti college of engineering and Natikar S B Assistant professor , vishwabharti college of engineering. " *A Survey on Keyword Query Routing*"
[4] Mrs. Suwarna Gothane, Srujana.P M.E , Assitant professor, CMRTC , M.Tech, CMRTC "Approaches for Keyword Query Routing" *Mrs. Suwarna Gothane Int. Journal of Engineering Research and Applications* www.ijera.com *ISSN : 2248-9622, Vol. 4, Issue 10(Part -1), October 2014, pp.16-22*

