Journal of Theoretical and Applied Information Technology
20th March 2015. Vol.73 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EFFECTIVE FRAMEWORK TO IMPROVE THE EFFICIENCY OF SEMANTIC BASED SEARCH

[1]MERLIN ANN ROY, [2]VIGNESHWARI. S
[1,2] Department of Computer Science and Engineering ,Faculty of Computing, Sathyabama University, Chennai, Tamilnadu, India
E-mail: [1]mrlnann@gmail.com[2]vikiraju@gmail.com

ABSTRACT

The incredible progress in the size of data and with the great growth of the amount of web pages, outdated search engines are not suitable and not properly any longer. The search engine is the best significant device to determine any information in World Wide Web. Semantic Search Engine is innate of outdated search engine to solve the above problem. The Semantic Web is a postponement of the existing web where data is are given in fixed meaning. Semantic web tools have a vital role in improving web search, because it is functioning to produce machine readable data and semantic web technologies will not exchange traditional search engine

Keywords:*Semantic,Query,Keyword,Search Engine, Data,Relevant Document*

1. INTRODUCTION

The keyword search engine does not provide the relevant result because they do not know the meaning of the words and expressions used in the web pages. The incredible progress in the size of data and with the excessive development of the amount of web pages, traditional search engines are not suitable and not proper anymore.The search engine is an important tool to determine any information in World Wide Web. The Semantic Web is a postponement of the existing web where data are given in fixed meaning. Semantic web machineries have a vital role in improving web search, because it is functioning to produce machine readable data and semantic web technologies will not exchange traditional search engine. The keyword search engine like Google and Yahoo and the semantic search engine like Hakia, DuckDuckGo and Bing are selected to search. While comparing both of the search engines, the semantic engine result was shown better than keyword search engine.

Some pages contain hundreds of words just to attract the users. It shows only the advertisement of the page rather than giving the relevant result to the users. If a user gives a keyword in the search engine that it will suggest for so many pages according to the previous user search. But if the keyword is wrong, it does not g show up anything. This research work proposes a framework to resolve this problem called enhanced skyline sweep algorithm. The algorithm says that even if the particular keyword given by the user is wrong, the search engine is going to give the relevant result to the user

2. LITERATURE SURVEY

Coffman et al [1] says that researchers are comparing the performance of different keyword search techniques and their results was not up to the expectation level. The run time performance was poor and the execution times for various search techniques vary for different evaluations. They evaluated the run time performance of various keyword search techniques and they found that their performance was poor and they don't provide the relevant result to the users. They implemented an algorithm called relational keyword search technique and it does not show the document if the particular keyword is wrong

Jarunee et al [2],explains and proposes an effective move towards keyword query in relational database. Keyword search technique in the web cannot be applied directly to the databases as data which present in the internet is of different forms. That is in databases the information is seen as data tuples and relationships. Researchers propose a model called semantic graph model consist of database metadata, database values, user terms and their semantic connections. This paper proposes an active structure that agrees users to query a database using metadata and database terms and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

their chosen terms. To accomplish the process, it employs an approach called metadata search and the database needed accommodate the terms it uses a semantic graph

Baid et al [3], describes that, systems produced answers quickly for many queries, but the other side many others, they take a long time or sometimes fail to produce answers after exhausting memory. It concludes that this approach is successful in returning a combination of answers in a predictable amount of time. They proposed techniques to solve the above problem and their proposed system says that our systems give forms to distinguish the unfamiliar part of the answer. If the user is not getting what the user wants from the answers which are in the right then the person can check the forms on the left and then click to fill the form and then submit

Surajit et al [4],says that researchers investigates about the problem that occurs when the user searches for a database query on a SQL database SQL database suggests so many tuples that satisfies the given query. The problem is when too many tuples are there in the answer. It leads to many-answers problem. They proposed a ranking approach for the answers to database queries. The ranking was based on the IR models and also it did a survey on the quality and performance efficiency of the different ranking system.

Kacholia et al [5], analyzed that researchers found that the problem for the graphical, structured textual data is extracting best answer trees from a data graph. XML and HTML data can be characterized as graphs by using entities as nodes and relationships as edges. To achieve this elasticity, they create a novel search frontier prioritization technique and this technique is centered on spreading activation.

Ritu et al [6], proposed a new approach semantic search engine which will answer the intelligent queries and also more efficiently and accurately. They used XML Meta tags to search the information. The XML page contains built in and user defined tags. The proposed approach proves that it takes less time to answer the queries. Using W3C compliant tools helps the system to work on any platform

Duygu et al [7], evaluates search performance of various search engines by allowing each query to run in a keyword based search engine as well as semantic based search engine. For both keyword-based search engines and the semantic based search engine semantic search engine performance were low

Thanh et al [8], offered a generic methodology for mapping queries. He converts user language into logical language. He also offered certain instantiation, that converts keyword queries into DL queries by using the information that is presented in KB.

David et al [9], predicted that semantic knowledge has repeatedly been engaged to apply relational database reliability. It also proposals the chance to convert a query into a semantically equivalent query which is more efficient. This paper explains a meaning based transformation technique that uses constraints and semantic integrity to reduce the cost of query processing

Latha et al [10], surveyed on the web search engine that are developed by different authors and they confirmed that no search engine gives answers properly and seamlessly modern means up-to-date

Anusree et al [11], surveyed about the semantic based search engine to extract the gifted features of various semantic search engines and also it says about the explanation of some of the better semantic search engines.

Anuradha et al [12], experimented with the approaches and features of some of the semantic search engines and they give a detail about the various advantages and techniques of some of the best semantic search engines. And the difference between the semantic search engines and traditional search

Madhu et al [13], says that retrieving relevant information from the search engine is tough. To solve the above problem the semantic search engine plays a vital role in computer system. A survey is done on the generations of search engines and advantages, features of the various search engines and also survey is done on the role in the web

Junaidah et al [14], mentioned that the traditional search engine does not provide the relevant information because it does not know the meaning, but the semantic search engines are meaning based search engine and it can overcome the above problem. This paper gives a brief about the traditional search engine and keyword search engine

Joel et al [15], said that, however a number of techniques have been implemented and proposed those all had a lack of standardization for system evaluation. This paper gives an empirical evaluation of the performance of the relational keyword search systems. They concluded with the results like many existing search techniques are not giving a good performance and also discover the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

relationship between execution time and factors that mottled in earlier calculations

Vanitha Muthusamy et al[16], The network enviourment where the server cluster is running contain multiple usrs and the data may be scattered in different data centers.The cloud computing platform shared by multiple users are not logically isolated and the data for the multiple users is stored in the physical enviournment.These equipments are not implemented properly and the data is not secured.It proposes a method to identify the data of the customer individually

Isabella J. et al[17], IDF(Information Document Frequency) is used to find the importance of a word in the document.It is a main factor used in text mining and information retreival.The paper explains about the functions of the inverse document frequency to deal with the unstructured text

S Gowri et al[18], The algorithm used is users clustering which is the another form of reverse factor algorithm and the drawbacks of the algorithm is pattern length not extended and data loss occurs

S Vigneshwari et al[19],The semantic similarity algorithm is used and the concept is based on ranking and the relationships can be measured using the tool called ontology.The two descriptors are textual and parameter descriptors.The drawback is it is not applicable for multiple users

3. MATERIALS AND METHODS

The semantic search greatly advances the query related data exactness, and in covering the user intention. There's no rejecting the control and reputation of the Google search engine. By using semantic search engine we will ensure that it results in more relevant and smart results. There are four methods for semantic search. And the method differs that is based on the semantic search engine .First method uses contextual analysis to help to disambiguate queries. Second is reasoning and third is natural language reasoning and the fourth is ontology search.

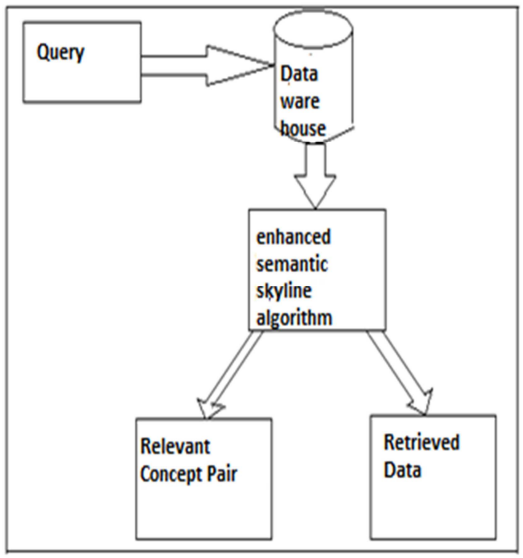

Figure 1: Proposed System Architecture

The above Fig 1 says that when the user gives a particular query in the semantic search engine it will extract the relevant result and gives to the user. If the particular query is wrong the result is not going to show to the user . So the skyline sweep algorithm helps to give the relevant result,by key combination process.This can be achieved even if the particular query is wrong. The process can be completed by courtesy of a genuine query load instead of a larger load with queries Runtime performance is intolerable for utmost search mechanisms. Memory ingesting is also unnecessary for many search techniques. Our proposed system is efficient to search data in various search engines and it is easy to execute with different patterns.

Searching keywords in databases is complex task than search in files. Information Retrieval (IR) process search keywords from text files and it is very important that querying keyword to the relational databases. Structure Query Language can be used for retrieving the data from the relational database and it is used to find relevant data or records from the database. For supporting effective and efficient queries there is a natural demand for database. The algorithm describes the frame-zword which takes keywords and the input ask and develops relevant records with the help of skyline sweeping algorithm and it is used to style the keyword search both efficiently and effectively.

To connect with server user must give their username and password, then only they can able to connect the server. If the user already exits directly can login into the server else user must register their details such as username, password

Journal of Theoretical and Applied Information Technology
20th March 2015. Vol.73 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and Email id, into the server. The server will create the account for the entire user to maintain upload and download rate. The name will be set as user id. . Logging is usually used to enter a specific page. *Example: Create* node and set name, port for that node. Nodes are created and displayed

3.1 Algorithm for KeyCombination using enhanced Skyline Sweep algorithm

Stept 1: Input: M, e, y, W

Stept 2: Output :H ∈ M,with |H| =f and low k-regret ratio

Stept 3: Let H= {k1∈ Mβ, where k1 is point in M with highest k1.0.

Stept 4: while |H| <f do

Stept 5: Let v be first k ∈ M . Initialize w= (0,1,2..n , 0)

Stept 6: For all k ∈ M |H do

Stept 7: for all i from 1 to W do

Stept 8: M|H| {k}is partitioned into M_0,...., M_j-2

Stept 9: Let max_regret (p) be result of Linear Program 3 with input k, H, M_0,...,M_{j-1}

Stept 10: If max_regret (k) has all xj>0 then

Stept 11: If max_ regret (k)>max_regret (v) then

Stept 12: Let v be k and w be $\bar{w}$ from linear program 3

Stept 13: End if

Stept 14: Break inner loop and go to next point.

Stept 15: End if

Stept 16: End for

Stept 17: End for

Stept 18: Let B= {v} and w

Stept 19: For all k ∈ M\H do

Stept 20: If k.w ≥ v.w then

Stept 21: Let B=B U {k}.

Stept 22: End if

Stept 23: End for

Stept 24: Let b be best member of B with heuristic of choice

Stept 25: Let H=H U {b}

Stept 26: End while

Stept 27: return H

The advantage of my proposed system is the performance is good and accurate

4. RESULTS AND DISCUSSIONS

Table 1: Number Of Relevant Documents Retrieved.

Query	Key word	Key word Relevant	Keyword Irrelevant	Retrieved Data
Satellite	5	3	2	10
Disambiguation	4	2	3	8
Computer	5	4	1	7
Server	3	2	3	8

Admin maintains the user information. And he can upload the file to search, the user. The file upload should be completed, so that the user can able to search the required file. And then admin can check the user information. Suppose here one file is searched that related all information is stored into admin. Searching information means when the user searched the file and timing everything stored in the admin. Finally admin check what file we are uploaded. For example admin upload the files into the database .And then check the uploaded files

Query processing means what we are searching that is passed by the query. Admin uploaded all files are stored in the database. User search in a database where is available the requested keyword. Suppose the requested file is available in a database that is passed to the user.

Journal of Theoretical and Applied Information Technology
20th March 2015. Vol.73 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Suppose the user give one keyword depends upon the keyword all related lines are displayed. In that line from user get what are the data we need. This file searching and execution details is stored in the database. Whenever need this we can able to view these details. *Example*: User searches the Query (keyword) in database. User gets that query related output.

Recommended module meant to suppose now we give any keyword wrongly, that word automatically going to mapped correct keyword. And then displayed what are the keyword mapped related that word. Suppose we give any wrong keyword that related all correct words going to mapped and displayed. Here we used "Skyline sweep algorithm" for automatically checked that correct keyword. *Example*:The user gives the wrong query. Key combination will give the correct output

Top rank meant most of the files viewed by a user that is called top ranked. Those files come first. After then only comes the user searching a keyword. So now we can easily understand which files are mostly viewed by the user. That ranking is displayed in the chart. *Example*: User searches the keyword. The keyword already viewed by the user, that keyword displayed in first .The advantages are reduced time consumption during retrieval, Efficient to search data in various search engines, Easy to execute in a realistic manner.

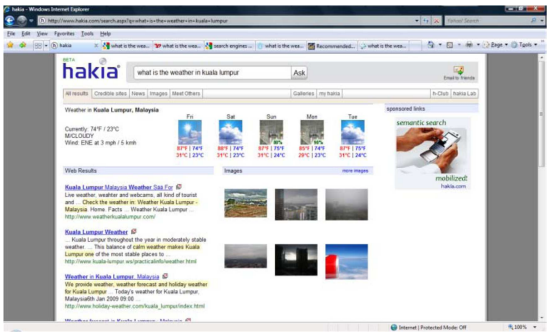

Figure 2: Hakia Search Engine

Fig 2 shows a sample search in Hakia search engine Relevancy ratio is calculated by the recall (equation 1), where red represents the relevant documents and rtd represents the retrieved documents. Table 2 shows that Hakia gives the more relevant documents and the recall ratio is more for Hakia.

$$recall = \frac{\{red\} \cap \{rtd\}}{\{rtd\}} \qquad (1)$$

Table 2: Number of Relevant Documents Retrieved For both Keyword and Semantic Search Engine

Query	Recall on Keyword search engine		Recall on Semantic Search Engine	
	Google	Yahoo	Hakia	Bing
1	0.6	0.7	0.9	0.8
2	0.7	0.6	0.7	0.9
3	0.4	0.4	0.9	0.6
4	0.2	0.3	0.8	0.7

5. CONCLUSION AND FUTURE WORK

Searching the internet today is a encounter and it is projected that approximately partial of the complex questions go unanswered .Semantic search has the power to enhance the traditional web search. Whether a search engine can meet all these conditions still remain a question .We proposed a framework using an enhanced skyline sweep algorithm to overcome this problem. In which the process can be completed by the courtesy a genuine query load instead of a larger load with queries Runtime performance is intolerable for utmost search mechanisms. Memory ingesting is also unnecessary for many search techniques in our experimental results, And our proposed system is efficient to search data in various search engine and it is easy to execute with different patterns.

REFERENCES

[1] J. Coffman and A.C. Weaver, "An Empirical Performance Evaluation of Relational Keyword Search Systems,"Technical Report CS-2011-07, Univ. of Virginia, *IEEE transaction on knowledge and data engineering,* Vol. 26, No. 1,January 2014

[2] Jarunee Saelee, Veera Boonjing,"A metadata data search approach to keyword query in relational databases", *InternationalJournal of ComputerApplications*, May 2013,pp. 140-149

[3] A. Baid, I. Rae, J. Li, A. Doan, and J. Naughton, "Toward Scalable KeywordSearch over Relational Data,"*University of Wisconsin, Madison fbaid, ian, jxli, anhai*, naughtong@cs.wisc.edu

[4] Surajit Chaudhuri, Gautam Das, "Probabilistic Ranking of Database Query Results," *Microsoft Research One Microsoft Way Redmond, WA 98053 USA* {surajitc, gautamd}@microsoft.com

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[5] V. Kacholia, S. Pandit, S. Chakrabarti, S. Sudarshan, R. Desai, and H. Karambelkar, "Bidirectional Expansion for Keyword Search on Graph Databases," *Indian Institute of Technology, Bombay*

[6] Ritu Khatri, Kanwalvir Singh Dhindsa, Vishal Khatri "Investigation and Approach Of New Analysis Of Intelligent Semantic Web Search Engine" *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-1, Issue-1, April 2012

[7] Duygu Tümer, Mohammad Ahmed Shah, Yıltan Bitirim "An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia" *Fourth International Conference on Internet Monitoring and Protection*, 2009

[8] Thanh Tran, Philipp Cimiano, Sebastian Rudolph and Rudi Studer "Ontology-based Interpretation of Keywords for Semantic Search" *Institute AIFB, University ät Karlsruhe, Germany*

[9] W. David Haseman, University of Wisconsin-Milwaukee, daveh@uwm.eduTung-Ching Lin, Nationa Sun Yat-Sen University, Taiwan, tclin@mis.nsysu.edu.twDerek L. Nazareth, University of Wisconsin-Milwaukee, derek@uwm.edu "An Intelligent Approach to Semantic Query Processing"

[10] S. Latha Shanmuga Vadivu1, M. Rajaram2, and S. N. Sivanandam3 "A Survey on Semantic Web Mining Based eb Search Engines"ARPN Journal of Engineering and Applied Sciences VOL. 6, NO. 10, OCTOBER 2011

[11] Anusree.ramachandran, R.Sujatha School of Information Technology and Engineering, VITUniversity" Semantic search engine: A survey" *Int. J. Comp. Tech.* Appl., Vol 2 (6), 1806-1811

[12] G .Sudeepthi , G. Anuradha ,Prof. M.Surendra Prasad Babu "A Survey on Semantic Web Search Engine" *IJCSI International Journal of Computer Science* Issues, Vol. 9, Issue 2, No 1, March 2012

[13] G.Madhu1 and Dr.A.Govardhan2 Dr.T.V.Rajinikanth3 "Intelligent Semantic Web Search Engines: A Brief Survey" *International journal of Web & Semantic Technology (IJWesT)* Vol.2, No.1, January 2011

[14] Junaidah,Mohamed Kassim and Mahathir Rahmany" Introduction to Semantic Search Engine" *International Conference on Electrical Engineering and Informatics*5-7 August 2009, Selangor, Malaysia

[15] [Joel Coffman, Alfred C. Weaver "An Empirical Performance Evaluation of Relational Keyword Search Systems" *Department of Computer Science, University of Virginia Charlottesville, VA, USA*

[16] Vanitha Muthusamy,Kavitha C."Secured Data Detection in Cloud Based Multi-Tenant Database Architecture"*International Journal on Information Sciences And Computing,*Vol.6, No.2, july 2012

[17] Isabella J.,Suresh R.M."Application of Feature Extraction Technique to Unstructured Data" *National Journal on Advances in Computing and Management",*Vol.3,No.2,october 2012

[18] S.Gowri,G.S.AnandhaMala,G.Divya."Enhanci ng the Digital Data Retreival System Using Novel Techniques" Journal of Theoretical and Applied Information Technology, Vol. 66 No.2, 20th August 2014

[19] S.Vigneshwari, DR. M. Aramudhan " A Technique to User Profiling Ontology Mining and Relationship Ranking" *Journal of Theoretical and Applied Information Technology"* Vol. 58 No.3, 31st December 2013