IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 3, Ver. VIII (May-Jun. 2014), PP 52-55
www.iosrjournals.org

Ranking Algorithm of Web Documents using Ontology

[1]Gurdeep Kaur, [2]Poonam Nandal
[1,2]Dept. of Cse, FET, MRIU Faridabad, Haryana

Abstract: *Internet has become the most important part of everyone's life where the large amount of information is stored in the form of text, audio, video etc. The Web information retrieval is a technique which helps to match the query provided by the user on the web search engine with the documents stored on the web. The next generation web called the Semantic Web will help the user to retrieve the useful data that is stored on the web in the form of ontology and make the data visible to the user which is hidden behind the web. The aim of the proposed ranking algorithm is to provide users the result set of relevant data .For this purpose, we have stored the information of the web documents in the form of ontology which is an explicit specification of a conceptualization and will help us to retrieve the relevant information for the user stored in web documents. The proposed ranking algorithm in conjunction with search engine will display the relevant web pages that best match the user query on the top of the result set provided by the search engine.*

Keywords: *Semantic Web, ranking algorithm, ontology, relevant data.*

I. Introduction

The World Wide Web is a system of interlinked hypertext documents which help people to exchange their information using the internet that connects billion of computers in all over the world through multiple networks. Using internet, one can communicate to others and retrieve useful information which may help them solving their problems. Web browser helps the user to access the internet using web search engine which provides all kind of information according to the user's query by accessing the large amount of data stored on the web. The data which appears on the web pages of internet contains the irrelevant pages information which is not useful to the user and the relevant information is still hidden behind the web. The semantic web will enable the user to access the useful information that doesn't appear during the normal search. The information stored on the semantic web will be used by machines not only for display purposes but also for automation, integration and reuse of applications. In this paper, a Ranking Algorithm is used to provide user the relevant result set that best matches the user query. The proposed algorithm will match the keyword with the keywords of web documents stored in ontology and web pages and gives the result accordingly. In section 2, we have explained the different ranking strategy and algorithm proposed by the other authors. In section 3, we have explained the proposed ranking algorithm which is designed to give the relevant information to the user based on the highest ranking. In section 4, a structural diagram is given to explain the method of proposed algorithm from user's query to the result set. In section 5, performance evaluation is given and it is explained why our ranking method is providing better and accurate result as compare to the other methods.

II. Related Work

The author S. Qiao et. al [14] proposed a SimRank Algorithm , which is based on a simple and graph-model that works on similarity measure of objects . According to this paper, the two objects are similar if they matches the other object. For example if the web pages contains the three objects with the name mohit, anil, rohit then they all are similar to other object because all are male. Similarly if there are some objects like red, white, green, black they are similar to the other object named color. The proposed algorithm uses the basic SimRank equation, bipartite SimRank, Naive method and vector -based method to rank the objects according to the similarity between them.

N. Preethi et. al [13] proposed a CARE algorithm for semantic web search engine where it has used the textual case based reasoning and Relation -based page ranking to provide the user the relevant result set according to the query entered by the user . The semantic web helps to access the knowledge database stored on the web and matches the query to provide the needed information. The proposed algorithm used by author is based upon Page Rank algorithm which was first used by Google to browses the World Wide Web and to rank the important websites in their search engine results. It basically tells about the importance of the websites.

The author L. Fabrizio et.al [1] proposed a relation-based page rank algorithm using Semantic Web search engines to provide the ordered result set to the user according to the keywords entered by them on the search engine. It uses the user interface which allows the user to enter the query on the search engine. It accesses the page database and builds the unordered result set including all pages containing the keyword or query entered by user. The proposed algorithm analyzes the user query and constructs the query sub-graph. It also

constructs the page sub-graph for each page in the result set and computes the page spanning forest by considering the number of edges or relations in the database stored in the form of ontology. The algorithm compute the page score for each page and build the ordered result set to retrieve the needed information .In ordered result set, pages that best fit the user query are displayed first.

III. Proposed Ranking Algorithm

Large amount of data is stored on the internet in the form of text, images, audio, video etc. The problem is how to access the needed data stored on the web. The search techniques designed as a client-server communication method where the client sends its query to the server and the server receives the query sends the answer to the client related to the query. If the server doesn't contain any information related to the query then a message contains no record found will be send to the user. Using the same approach we have proposed a ranking algorithm of web documents using Ontology which will provide the result set according to the query given by the user in search engine, the relevant pages to the query will be on the top of the result set and irrelevant page will be on the bottom of the result set. The algorithm uses the concept-relation based technique using graph model to store the data in the form of ontology and to access the data by considering the relationships between the concepts/keywords. The notations used in the algorithm are given below:

G(C,R) -> Ontology graph G where C denotes for set of concepts and R denotes for set of relations between the concepts
$C = \{c_1, c_2, c_3 \ldots\ldots c_n\}$ -> set of concepts taken as vertices of the ontology graph
$R = \{R_{ij} | i=1,2,3,\ldots\ldots n, j=1,2,3\ldots\ldots n, j>i\}$ -> set of relations taken as edges of the Ontology graph
$R_{ij} = \{r_{ij(1)}, r_{ij(2)}, \ldots\ldots r_{ij(n)}\}$ -> set of relations between concepts C_i and C_j in G(C,R)
P(O,P) -> joint probability of a page sub-graph w.r.t. Ontology
P(O,Q)-> Joint probability of a query sub-graph w.r.t. Ontology
N_{ij} -> Number of relations linking concepts i and j in Ontology
D_{ij} -> Number of relations linking concepts i and j in page sub-graph
d_{ij} -> Number of relations linking concepts i and j in query subgraph
$P(r_{ij},p)=T_{ij}=D_{ij}/N_{ij}$ -> Relation probability for each edge of pag-e w.r.t ontology
$P(r_{ij},q)=T_{ij}=D_{ij}/N_{ij}$ -> Relation probability for each edge of query w.r.t. ontology
P(Q,P)=(P(O,P)*P(O,Q) -> Joint probability of each query and page w.r.t. ontology
The algorithm 1, algorithm 2 and algorithm 3 is used to compute the result set is given below:

Algorithm 1: Compute relational probability and joint probability of a page sub-graph w.r.t. ontology

1: if (C_i==0, C_j==0)
2: No record found
4: Else
5: if(i==j) then Set (C_{ij}==0)
6: for(i=1;i<N;i++)
7: for (j=1;j<N;j+;+)
8: N= N_{ij} ; // Number of relations exist in the ontology graph linking concepts i and j
9: for (i=1;i<D;i++)
10: for(j=1;j<D;j++)
11: D = D_{ij} ; //Number of relations exist in the page sub-graph linking concepts i and j
13: $P(r_{ij},p)=T_{ij}=D_{ij}/N_{ij}$ for all i,j=1--n
16: $P(O,P)= {}^{n}\prod_{i=1,j=1}(P(r_{ij},p))$; //Joint probability of page w.r.t. ontology
17: End

Algorithm 2: Compute relational probability and joint probability of a query sub-graph w.r.t. ontology

1: if (C_i==0,C_j==0)
2: then
3: No record found
4: Else
5: if(i==j) then Set (C_{ij}==0)
6: for(i=1;i<N;i++)
7: for (j=1;j<N;j+;+)
8: N = N_{ij} ; // Number of relations exist in the ontology graph linking concepts i and j
9: for (i=1;i<d;i++)
10: for(j=1;j<d;j++)
11: d = d_{ij} ; //Number of relations exist in the query sub-graph linking concepts i and j

12: {
13: $P(r_{ij},q)=T_{ij}=d_{ij}/N_{ij}$ for all i,j=1—n
14: T_{ij}++ ;
15: }
16: $P(O,Q)= {}^{n}\prod_{i=1,j=1}(P(r_{ij},q))$;//joint probability of query w.r.t. ontology
17: End

Algorithm 3: Compute the Joint probability of page sub-graph and query sub-graph w.r.t. ontology

1: P(Q,P)= P(O,P)*P(O,Q);
2: End;

IV. Methodology

Designing of the Ranking Algorithm for Web Documents using Ontology needs the information of web documents stored in the form of Ontology which is the structural form of data containing useful information about a particular domain. The Proposed algorithm is designed to access the information stored in the ontology and page database to provide the relevant information to the user according to the query entered by them on the search engine. Methodology of the Proposed Ranking algorithm works in the steps given below:

- The whole information about a particular domain is stored in the structural form of Ontology using OWL(Ontology Web Language).
- The pages are stored on the internet containing different information about a particular domain individually.
- The user interface allows the user to enter the query on search engine and provide result accordingly.
- The Proposed algorithm computes the joint probability of page with respect to ontology by matching concepts and relationships between concepts. Similarly, the joint probability of query with respect to ontology by matching concepts and relationships between concepts are calculated.
- Finally, the joint probability of both page and query is calculated with respect to ontology and the result set that best matches the user query will be displayed first.

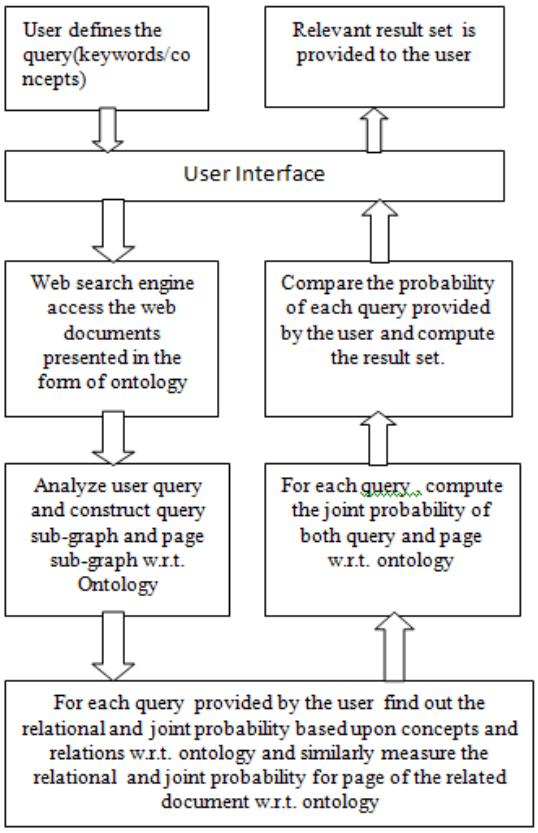

Fig1. Workflow from query definition to the results

V. Performance Evaluation

We have computed the performance of the proposed ranking algorithm using ontology to the baseline keyword based- approach, we have calculated the probability measurement of web document stored in the form

of ontology which is shown in Table 1. After analyzing the performance, we found that our proposed algorithm provides the result set in the order which is close to the human approach rating. The proposed algorithm provides the relevant pages on the top of the result set, as our algorithm ranking the documents according to the query given by the user by considering not only keywords but also relationships between concepts/keywords.

Table 1. Performance Evaluation By Different Methods

Number of pages	Keyword based approach	Human based approach	Our Ranking algorithm
Page1	0.42	0.76	0.66
Page2	0.32	0.66	0.76
Page3	0.54	0.52	0.50

The ranking calculated using the above methods is explained below:
Keyword based approach → 3>1>2
Human The probability → 1>2>3
Our ranking algorithm → 2>1>3

Hence, we have concluded that the variance between the keyword based approach and human based approach is much more than the variance between human based approach and our ranking algorithm .So, it is proven that our ranking algorithm is providing the better and accurate result set containing relevant information.

VI. Conclusion And The Future Work

The Proposed Ranking Algorithm is providing the relevant result set to the users of the internet according to the query or keyword specified by them. The algorithm proposed is considering the joint probability of web page and query with respect to the Ontology to consider the concepts and relationships that exists between the concepts. This algorithm will also help the user in terms of both time complexity and accuracy. The next generation web, the semantic web will access the database of web documents which are stored in the form of ontology and match the keyword entered by the user to provide useful information and to help extracting information which is hidden behind the web. The Ranking strategy used by the algorithm will show the result set that best matches the user query on the top of the web pages using the search engine and the other information related to the query will be displayed on the bottom of the web pages. We will also try to make our approach more scalable by making the ontology construction an automatic process.

References

[1] A.Kemafor, A. Maduko, and A. Sheth. "SemRank: Ranking Complex Relationships Search Results on the Semantic Web." Proceeding WWW '05 Proceedings of the 14th International Conference on World Wide Web. 117-27, 2005.

[2] A. Pisharody and H.E. Michel, "Search Engine Technique Using Keyword Relations," Proc. Int'l Conf. Artificial Intelligence (ICAI '05), pp. 300-306, 2005.

[3] B. Aleman-Meza, C. Halaschek, "A Context-Aware Semantic Association Ranking," Proc. First Int'l Workshop Semantic Web and Databases (SWDB '03), pp. 33-50, 2003.

[4] C. Junghoo, H. Garcia-Molina, and L. Page, "Efficient Crawling through URL Ordering," Computer Networks and ISDN Systems, vol. 30, no. 1, pp. 161-172, 1998.

[5] Cohen, S., J. Mamou, "A Semantic Search Engine." Proc. of Proc. 29th Int'l Conf. Very Large Data Bases. 45-56, 2003.

[6] D. Finin, A. Joshi, R. Pan, "Swoogle: A Search and Metadata Engine for the Semantic Web." Proc. 13th ACM Int'l Conf. Information and Knowledge Management (CIKM '04). 652-59. 32 , 2004.

[7] Eyal, Oren, Knud Hinnerk Moller, Simon Scerri, Siegfried Handschuh, and Michael Sintek. What Are Semantic Annotations? Science Foundation Ireland under Grants No. SFI/02/CE1/I131 and SFI/04/BR/CS0694 and by the European Commission under the Nepomuk Project FP6-027705.

[8] H.Liang, "Describing the Semantic Relation of the Deep Web Query Interfaces Using Ontology Extended LAV", JOURNAL OF SOFTWARE, VOL. 5, NO. 1, JANUARY 2009

[9] H. Zhao, W.Meng, "Fully Automatic Wrapper Generation For Search Engines", Dept. of Computer Science SUNY at Binghamton Binghamton, NY 13902, USA ,2011

[10] J. Trinkunas, "A graph oriented model for ontology transformation into conceptual data model", Information Systems Research Laboratory, Vilnius Gediminas Technical University Olegas Vasilecas Lithuania ,2007

[11] K. Javed and M. Hardas, "Hierarchial course Knowledge Representation Using Onotlogies", , CSE Kent State university,2007

[12] L. Fabrizio,, A. Sanna, and C. Demartini. "A Relation-Based Page Rank Algorithm for Semantic Web Search Engines." IEEE Transactions on Knowledge and Data Engineering 21.1, 123-36, (2009)

[13] N.Preethi , and T.Devi, "Semantic web Case and Relation (CARE) based Page Rank Algorithm for Search Engines", (IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012)

[14] S. Qiao , L. Tianrui "SimRank: A Page Rank Approach based on Similarity Measure", (IEEE Intelligent Systems and Knowledge Engineering (ISKE), 2010 International Conference