A Novel Approach to Provide Web Page Recommendation Using Domain Knowledge and Web Usage Knowledge

[1] Ms.Priyaka Kolekar, [2] Prof.Suchita Wakhade,
[1][2] Department of Computer Engineering, Trinity college of Engineering and Research
Pune, India

Abstract: As, population increases use of World Wide Web goes on increasing for various purposes. People surfing websites for study purpose, education, entertainment point of view and also access websites for online shopping purpose. But required knowledge capturing is difficult and time consuming which is challenging one. So to overcome this challenge the proposed system gives a novel approach to provide a Web page recommendation. The proposed system consists of three knowledge based models. To improve performance in future key information extraction algorithm is used and comparison takes place between results obtained from applying only three knowledge based models and models along with key information extraction algorithm. Recommendation is given to page from weblog records. Experimental result shows that recommendation for webpage is better by using proposed system than existing system and execution time required for the proposed system is less as compared to existing system and accuracy of proposed system is more than existing system. The key information extraction algorithm is used to achieve better result.

Keywords: - Webpage, Recommendation, Ontological, semantic, Conceptual.

I. INTRODUCTION

Here, firstly to provide recommendation to the web pages of website three knowledge based models used. These are Ontological model, Semantic network model and conceptual prediction model respectively. First model is Ontological model in which titles of WebPages extracted by using link extraction tool. Then titles can be split into sequence of term by using term extraction algorithm. So here by using this model Domain terms can be obtained.

Second is a semantic network analysis model which is used to find out relationship between webpage and domain terms. So here terms sort in descending order then relationship is find out between web pages and terms

Third model is Conceptual Prediction Model which is used extract web usage knowledge by inlink and outlink property of CPM schema. This shows visitors count of webpages and next visited webpage of user by using inlink property. But on the basis of these three model recommendation to WebPages is not correct so to improve performance there is need to work on data present on WebPages. For that purpose key information extraction algorithm is used In which two word extraction algorithm is used and multiword extraction algorithm is implemented.

II. LITERATURE REVIEW

It describes process of Web personalization viewed as a technique of data mining required to guide all data mining cycle. The phases include collection of data , then preprocessing after that pattern discovery and evolution. Finally, discovered knowledge used in real time for user and web[1]

Semantic web technology is positive for recommendation example, Thi Thah Sag Nguyen. Along with that web mining plays important role which is extraction of web usage knowledge from web page which is used as semantic network analysis model.[2]

Here 3 models are used first s ontological model for abstracting domain terms next is semantic network analysis model for the abstracting the relationship between domain terms ad Web pages ad last is a The conceptual prediction model for extracting web usage knowledge. From using these three models a Web page recommendation is provided and result compared from using precision ad recall. [3]
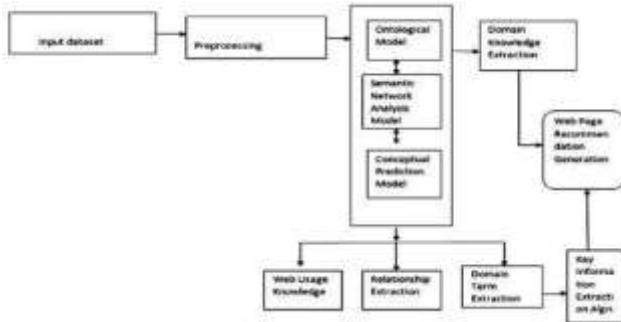
III. THE PROPOSED SYSTEM ARCHITECTURE

Fig 1:- Proposed System Architecture

1. Input Dataset:-

Here, user enters search query for e.g. Wireless sensor network. Then result display as links or snippets which are related to user search. Which are nothing but html pages .these collection of pages are weblog records called as Input dataset.

2. Preprocessing:-

Weblog records are the input files here which contain irrelevant data. So there is necessary to convert irrelevant data to relevant data. To achieve this objective preprocessing technique means stemming is used. Which removes stop words like' a','an','the' Finally, relevant data is output here.

3 Model Constructions:-
3.1 Ontological model:-

Here extraction of titles of web pages takes place then these titles can be split into domain terms using term extraction algorithm as below:-

1. Input: D (A set of the titles)

2. Output: TS(A set of term)

3. Process:

i) Set D = null

ii) For each the title in set D

a) Remove unwanted symbols e.g."&"

b) Remove stop words, e.g. "a", "ad", "for"

c) Split words in the the TITLE into a sequence of tokens, called C

$$C = t1t2 \ldots tn; ti (i = [1..n]) \text{ a token, n: the sequence length.}$$

In this way ontological model is constructed.

3.2 Semantic network analysis model:-

In this model firstly domain terms obtained in ontological model are relisted in descending order of occurrence from each webpage. Then each domain term belongs to which web page is find out means relationship between domain term and web pages in graphical form along with occurrence of domain terms. Based on the algorithm is describes in logic notation as follows.

Tuple: Oauto := <T, D, R>
Where
a) T = set of domain terms and the corresponding occurrences,

b) D= {d1,…dn} no of Web pages

c) R = is a set of the relations between domain term and Web-page.

3.3 The conceptual prediction model (CPM):-

But on the basis of these models only we can't correctly recommend web page so the third model introduced is Conceptual Prediction. Here we get Web usage knowledge. Inlink and outlink property used here. It gives visitor count of each web page and user's previously visited web page and next visited webpage So it gives web usage knowledge by visitors count and inlink and outlink of each user.

4. A web page Recommendation:-

Finally, web page recommended using ontological model, Semantic analysis model and Conceptual Prediction model.

5. Key Information Extraction Algorithm:-

Finally, this algorithm is used for improving a Web page recommendation performance in more extends.

Algorithm:-
*1. Two-word extraction:-Input***:** A corpus *L* in any language.

Step 1: Collect bigram frequencies for *L* from using proximity database *THE DB*.
*Step 2***:** Then all 4-grams(*a,b,c,d*) in *L*, remove, count for *b,c ,in THE DB* if
- $mi(b,c) < mi(a,b) - k$ or
- $mi(b,c) < mi(\text{c, d}) - k$.

Step 3: For all entries (*b,c*) in *THE DB*, add (*b,c*) to a list *T* if: - $C(b,c) > minCount$
Output: The list *T* of two-word candidate from L

2. Multi-word Extraction:-
Input: A list *T* of two-word candidates
Step 1: Collect all possible substrings involving *c ,in the DB*.
Step 2: Then proximity database update and Remove each entry in *THE DB* that has frequency $< minFreq$.
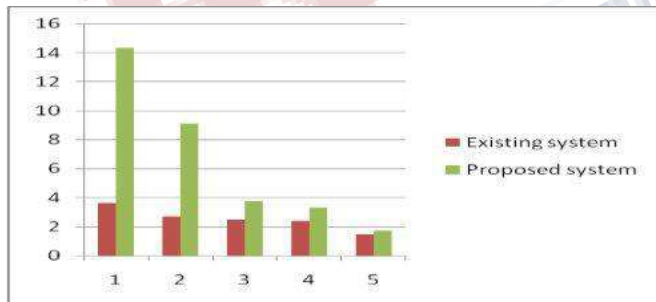*Step3:* Add Extracting multiword key for each candidate c,in T
Output: The list *E* of extracting multi-word key

IV. RESULT ANALYSIS

Result analysis shows comparison between existing system and proposed system for recommendation, execution time and accuracy.
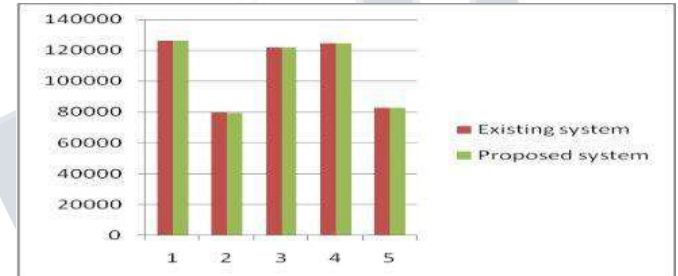
1. Recommendation:-

Graph1:-Comparison between existing system and Proposed system for Recommendation.

Dataset	Existing system	Proposed system
1	3.6	14.31
2	2.7	9.12
3	2.5	3.76
4	2.4	3.3
5	1.5	1.71

Table 1:-Comparison between existing system and proposed system for Recommendation
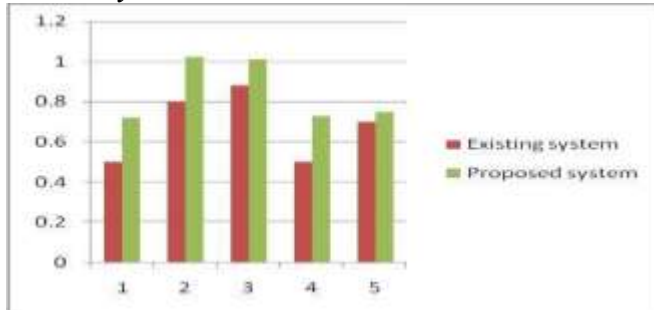
2. Execution time:-

Graph2:-Comparison between existing system and proposed system for Execution time.

As shown in graph 2 comparison between existing system and proposed system for Execution time in milliseconds. The execution time required for proposed system is less as compared to existing system shown in table. Average difference is about 50 to 55 ms so result of comparison shown below in table.

Dataset	Existing system	Proposed system
1	126252	126162
2	79541	79510
3	122015	121973
4	124483	124462
5	82817	82731

Table2:-Comparison between existing system and Proposed system for Execution time.

3. Accuracy:-

Graph3:-Comparison between existing system and proposed system for Accuracy.

As shown in graph 3 comparisons between existing system and proposed system for accuracy in percentage. The accuracy for proposed system is more as compared to existing system shown in table. Average increase in accuracy is about 15\% to 20% so result of comparison shown below in table.

Dataset	Existing system	Proposed system
1	0.5	0.72
2	0.8	1.02
3	0.88	1.01
4	0.5	0.73
5	0.7	0.75

Table 3:- Comparison between existing system and proposed system for Accuracy

V. CONCLUSION

Finally, this paper provides better a Web page recommendation using methods which are the first model is the ontological model that for domain term extraction. Next means second is semantic network analysis model, for extracting the relationship between domain terms and web pages then third model is nothing but a conceptual prediction model for extracting web usage knowledge. On the basis of these three models Recommendation, execution time and accuracy calculated for existing system. In order to improve performance of webpage recommendation Key information extraction algorithm is used and Recommendation, execution time and accuracy calculated for proposed system. Experimental result shows that recommendation for webpage is better by using proposed system than existing system and execution time required for the proposed system is less as compared to existing system and accuracy of proposed system is more than existing system. This can be achieving using key information extraction algorithm.

VI. FUTURE SCOPE

In future experimental comparison takes place of key information extraction algorithm with differentiate methods of recommendation of Web pages ad differentiate various types of methods used to for providing a Web page recommendation using differentiate technique to extract domain terms.

REFERENCES

[1] Modraj Bhavsar "Web page recommendation using web mining", Int. Journal of Engineering Research and Applications Vol. 4, Issue 7(Version 2), July 2014, pp.201-206

[2] Chhavi Rana Dept. of CSE, "Trends in Web Mining for Personalization" University Institute of Engineering and Technology, MD University, Rohtak, Haryana, India.

[3] Bamshad Mobasher, "Data Mining forWeb Personalization Center for Web Intelligence" School of Computer Science, Telecommunication, and Information Systems DePaul University, Chicago, Illinois.

[4] Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu, "Web-Page Recommendation Based on Web Usage and Domain Knowledge".

[5] International Journal of Advance Research in Research Article / Survey Paper / Case Study Available online at: www.ijarcsms.com A Preliminary Review of Web-Page Recommendation in Information Retrieval Using Domain Knowledge and Web Usage Mining by Nazneen Tarannum S.H. Rizvi1 M.E. Computer Science &Information Technology Dept H.V.P.M's C.O.E.T

Amravati, India.

[6] Sciences Thi Thanh Sang Nguyen,” SEMANTICENHANCEDWEB-PAGERECOMMENDER SYSTEMS“A dissertation submitted for the degree of Doctor of Philosophy in Computing Faculty of Engineering and Information Technology UNIVERSITY OFTECHNOLOGY, SYDNEY Australia December, 2012.

[7] Jose Borges1 and Mark Levene, ”Generating Dynamic Higher-Order Markov Models in Web Usage Mining” School of Engineering, University of Porto, R. Dr. Roberto Frias.

[8] Abraham, A.,“ i-Miner: A Web Usage Mining Framework Using Hierarchical Intelligent Systems”, The IEEEInternational Conference on Fuzzy Systems, IEEE Press, 2003, pp. 1129-1134.

[9] Adomavicius, G., Sankaranarayanan, R., Sen, S., Tuzhilin, A.,“Incorporating contextual information in recommender systems using a multidimensional approach”, ACM Transactions on Information Systems, Vol. 23, No. 1, pp. 103-145, 2005.

[10] Milan Lathia, Gridalogy and the University of Illinois at Urbana-Champaign,“A Business Perspective on the Semantic Web”Adaptive Information: Improving Business through Semantic Interoperability, By Jeffrey T. Pollock and Ralph Hodgson.

[11] Cynthia Changxin Wang**, Dongbai Xue”, Using Domain Ontology in a Semantic Blogging System for Construction Professionals**” Volume 13, Number S1, October 2008.

[12] Daniel Fett, Ralf Küsters, and Guido Schmitz”, An Expressive Model for the Web Infrastructure:Definition and Application to the BrowserID SSO System University of Trier, Germany.

[13] Arundhati Patil, Prof. Supriya Sarkar Computer Dept” Personalized Web Page Recommendation Using Ontology” ,in Volume**:** 3 Issue: 7 .

[14] H. Dai and B. Mobasher, "Integrating Semantic Knowledgewith Web Usage Mining for Personalization",

in Web Mining:Applications and Techniques, A. Scime, Ed. Hershey, PA,SA: IGI Global, 2005, pp. 276 - 306.

[15] FENG XIA1, (Senior Member, IEEE), ZHEN CHEN1, WEI WANG1,JING LI1, AND LAURENCE T. YANG2,”, MVCWalker: Random Walk-Based Most Valuable Collaborators Recommendation Exploiting Academic Factors”, publication 8 September 2014 current version 30 October 2014.

[16] S. Perugini, M. A. Gonçalves, and E. A. Fox, ``Recommender systems research: A connection-centric survey," *J. Intell. Inf. Syst.*, vol. 23, no. 2, pp. 107_143, 2004.

[17] Shangguang Wang, Member, IEEE, Zibin Zheng, Member, IEEE, Zhengping Wu, Member, IEEE,Michael R. Lyu, Fellow, IEEE, and Fangchun Yang, Member, IEEE,” Reputation Measurement and MaliciousFeedback Rating Prevention in WebService Recommendation Systems”, IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 8, NO. 5, SEPTEMBER/OCTOBER 2015.

[18] J. Caverlee, L. Liu, and S. Webb, “Socialtrust: Tamper-resilienttrust establishment in online communities,” in Proc. 8th ACM/IEEE-CS Joint Conf. Digit. Libr., 2008, pp. 104–114.

[19] S. Ries and E. Aitenbichler, “Limiting sybil attacks on bayesiantrust models in open SOA environments,” in Proc. Symp. Workshops Ubiquitous, Auton. Trusted Comput., 2009, pp. 178–183.

[20] Teevan, J., Cutrell, E., Fisher, D., Drucker, S. M., Ramos, G., Andre, P., and Hu, C. "Visual Snippets: Summarizing Web Pages for Search and Revisitation." 27th International Conference on Human Factors in Computing Systems. Boston, MA: ACM, 2009. 2023-2032.

[21] Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu,” Web-Page Recommendation Based on WebUsage and Domain Knowledge”,VOL. 26, NO. 10, OCTOBER 2014.

Miss.Priyanka Kolekar Received B.E degree in Information technology from Shivaji university, Kolhapur in 2009.Where she is currently perusing Masters degree in Computer Engineering from Trinity college of engg and research,Pune under savitribai Phule university, Pune .Her research interest include Data mining and information retrieval.