International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 11 | Nov -2017 www.irjet.net p-ISSN: 2395-0072

Web page recommendation using domain knowledge and web usage knowledge

Miss. Swapnali S.Jadhav[1], Prof. Sachin P.Patil[2]

[1]*Student, Dept.of computer science and Engineering, ADCET Ashta, Maharashtra, India.*
[2]*Professor, Dept.of computer science and Engineering, ADCET Ashta, Maharashtra, India.*

Abstract - *As, population increases use of World Wide Web goes an increasing for various purposes. People surfing websites for study purpose, education and entertainment point of view and access websites for online shopping purpose. But useful knowledge discovery and satisfactory knowledge representation is challenging one. Because each website contains multiple web pages it is time consuming and challenging task. So to overcome this challenge proposed system gives novel approach to provide webpage recommendation. The proposed system consists of three knowledge based models. To improve the performance further key information extraction algorithm is used and comparison takes place between results obtained by applying only three knowledge based models and models along with key information extraction algorithm. Experimental result shows that execution time required for traditional system is more and Precision is less as compared to proposed system.*

Key Words: **Extract, Webpage, Recommend, Semantic, Ontology, Conceptual.**

1. INTRODUCTION

The growth of World Wide Web means internet results in high demand. But there is a need to provide better required information to a user which is provided by web mining technique. Here the ultimate goal of any user-adaptive system is to provide users what they require. Most of the time users get search result by websites. Each website contains multiple WebPages. So as result it is complicated work. Session created for user surf result.

There are much more issues for web page recommendation like learn from available historical data. How to discover knowledge from that? So the framework is provided by novel models. A Novel model clears recommendation difficulty of web pages.

2. LITERATURE REVIEW

In[1] this, paper describes process of Web personalization viewed as an technique of data mining required to guide all the data mining cycle. The phases include collection of data then preprocessing after that pattern discovery and evaluation. Finally this discovered knowledge is used in real-time for user and web.Here semantic web technology is positive for recommendation. Along with that web mining plays important role which is extraction of web usage knowledge from webpage. This is used as semantic network analysis model [3].

Here 3 models are used first is ontological model for extracting domain terms. Next is semantic network analysis model for extracting relationship between domain terms and WebPages. And last is Conceptual prediction model for extracting web usage knowledge. By using these three models webpage recommendation is provided and result compared by using precision and recall [4].

3. MATHEMATICAL MODEL

Fig.1 shows mathematical model.

1. Let IQ be the set of input query.

2. Web Pages Extracted based on input IQ are W={w1, w2, w3,..., wn}.

3. T be the extracted terms Where T is the subset of IQ.

4. Calculate Relevant Score S= {S1, S2,..., Sn}.

5. It means relevance is Rw1, Rw2 ,Rw3,...., Rwn with IQ.

6. If score is less than T Check for V={Vw1, Vw2,...., Vwn} Where V=Vocabulary.

7. Output relevant information OUTPUT= {D1, D2,..., Dn} Where D is the Related documents.

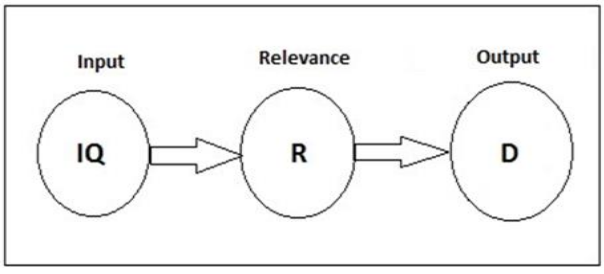


Fig-1: Mathematical Model

4. SYSTEM ARCHITECTURE

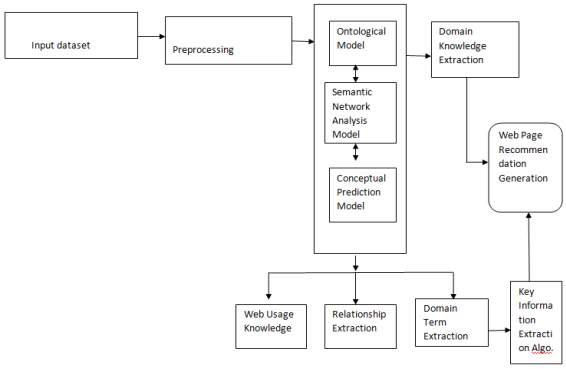


Fig-2: System architecture

4.1Input Dataset:-

User enters search query here then on that basis results are displayed. Then weblog session created means extraction of links takes place by processing on google links. Finally input dataset created as links or web pages.

4.2Preprocessing:-

The preprocessing is phase in system architecture as shown in fig 2. Weblog records are the input files here which contain irrelevant data. While preprocessing is used to process on irrelevant data, the cleaned weblog records are input to the ontological model parser. Finally, this preprocessing results in user file, semantic objects. Means finally it converts weblog records into required format.

4.3Model Constructions:-

4.3.1 Ontological model:-

4.3.1.1 Analysis: - Once URLs of visited WebPages extracted from the weblog, title is extracted using web crawler. The Web crawler is used to retrieve Web content based on URLs. The goal of Web crawling is to explore a collection of linked web documents by fetching WebPages at many levels based on one URL. The crawler is designed to look for the TITLE tags in Web documents [4] and to retrieve the corresponding values as the titles. Then the output of the crawler is the set of titles of the all viewed WebPages of the given website.

- Collecting Web-pages:-

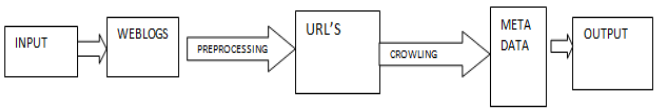


Fig-3: Collection of WebPages

The titles of Web-pages are based on the TITLE tag on the HTML documents of Web-pages. Fig 3 shows steps for collection of web pages.

4.3.1.2 Conceptualization: - There is process performs as below:-

- Extracting domain terms-

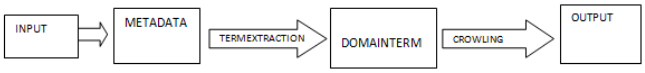


Fig-4: Extracting domain terms

Here term extraction algorithm used to obtain terms from webpage title. Fig 4 shows steps of extracting domain terms. Steps for term extraction algorithm are as follows.

Input: D (A set of titles)

Output: TS (A set of term sequences)

Process:

i. Set D = null

// Token extraction

ii. **For each** title in set D

a. Remove invalid words e.g. "&", ")".

b. Remove stop words, e.g. "an", "and", "for".

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

C = t1t2 … tn; ti (i = [1...n]) a token, n: the sequence length.

In this way ontological model is constructed.

4.3.1.3 Implementation

In this way these analysis and conceptualization applied to each webpage to extract domain terms from title of web pages. In implementation stage the DomainOntoWP [4] is used to describe domain term T.D1 is the page id input for the system. We get domain terms related to TITLE of visited web pages [4].

Input: D1(PageID)

O(DomainOntoWP)

Output:-T(Domain terms)

4.3.2. Semantic network analysis model:-

Here, we present the second model, i.e. a new semantic network of a website, which is a kind of knowledge which represents relations including the collocations of domain terms.

Evaluation:-The returned domain terms sorted in descending order of their occurrence weights based on the fact that the more times a domain term occurs on WebPages, the more likely the term has been viewed. Based on the algorithm is described in logics notation as follows.

- TermNetWP[4], is defined as a tuples:
  O auto := <T, L, D, R>

Where,

- T = set of domain terms and the corresponding occurrences,
- D= {d1,...dn} no of WebPages
- R = is a set of relations between domain term t and Web-page d.

Algorithm:-

To find out relation between web pages and domain terms TermNetWP [4] algorithm used. Input to the TermNetWP is Domain terms extracted during preprocessing of system. Here relation is obtained among domain terms and WebPages.

4.3.3. Conceptual prediction Model (CPM):-

But on the basis of these models only we can't correctly recommend webpage so the third model introduced is Conceptual Prediction. Here we get Web usage knowledge. Here we refer to this semantic network as TermNavNet. This model motivated by the Markov models, which is a kind of well-known probabilistic model. They have remarkable features of events prediction by learning sequential data patterns. Here the idea of Markov models is very useful i.e. FWAP, for the Web-page recommender system.

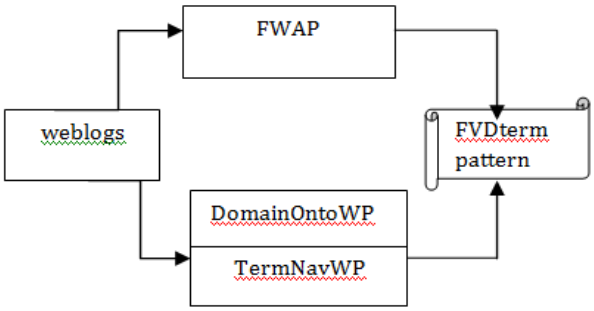


Fig-5: Frequently Viewed Dterm pattern

For better webpage recommendation we have to combine Domain knowledge model and semantic network together with web usage knowledge that can be discovered from web log files. To obtain frequently viewed term pattern i.e. semantic knowledge, we combine frequent web access patterns with DomainOntoWP or TermNetWP [5].Fig 5 shows frequently viewed term pattern.

Conceptual prediction model is developed to automatically generate weighted semantic network of frequently viewed domain terms with the weight being the possibility of the alteration between two adjacent terms based on FVTP. We refer to this semantic network as TermNavNet. Fig 6 illustrates conversion of FVTP into TermNavNet using CPM that act as a formatter.

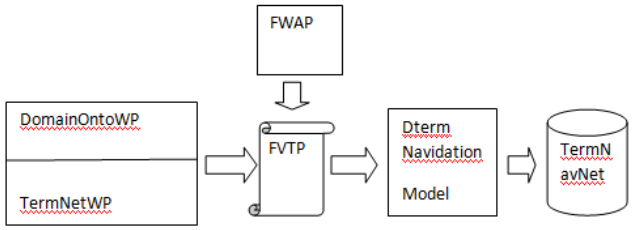


Fig-6: Web usage knowledge extraction

Algorithm for TermNavNet[4] Construction:-

Input: P (FWAP)

Output: M (WPNavNet)

So finally from this we get web usage knowledge.

4.4 Web page Recommendation:-

Finally using:-

1. Domain terms by using ontological model

2. By using semantic network analysis model Domain terms and relationship between WebPages

3. By using CPM web usage knowledge is obtained.

Finally recommendation is given to webpage.

4.5 Key Information Extraction Algorithm:-

Finally, this algorithm is used for improving webpage recommendation performance in more extend.

Algorithm:-

4.5.1. Two-word extraction:-

Input: A corpus *L* in any language.

Step 1: Collect bigram frequencies for *L* by using proximity database *DB*.

Step 2: Then all 4-grams *(a,b,c,d)* in *L*, remove count for *b,c* in *DB* if

$$- mi(b,c) < mi(a,b) - k$$

or

- $mi(b,c) < mi(c, d) - k$.

Step 3: For all entries (b,c) in *DB*, add (b,c) to a list *T* if:

- $C(b,c) > minCount$

Output: The list *T* of two-word candidate from L .

4.5.2. Multi-word extraction algorithm:-

Input: A list *T* of two-word candidates

Step 1: Collect all possible substrings involving *c* in *DB*.

Step 2: Then proximity database update and

Remove each entry in *DB* that has frequency < *minFreq*.

Step3: Add Extracted multiword key for each candidate c in T

Output: The list *E* of extracted multi-word key.

5. RESULT ANALYSIS

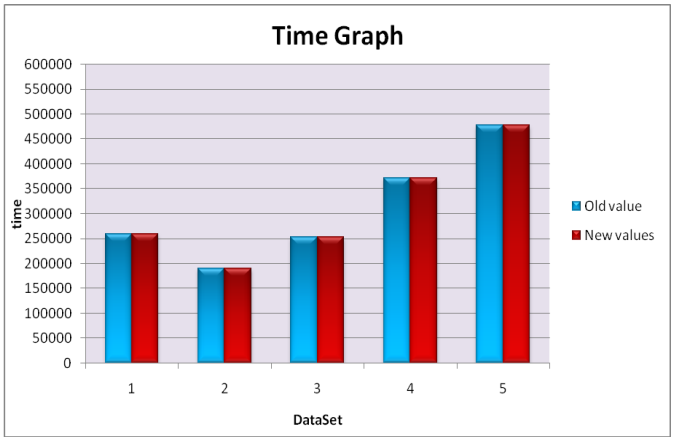
Analysis is made using two parameters, execution time and precision. We compare previous execution time and new execution time. Previous execution time is time require for recommendation using semantic network .New execution time is time require for recommendation using key information extraction model. Next we compared precision. Precision is calculated using following formula. Precision=TP/TP+FP.

Where TP=True positive and FP=false positive.

Following Table1 and graph1 shows comparison of traditional system and new system for execution time. Execution time is less for new system as compared to traditional system. Execution time is measured in milliseconds.

Execution time		
Data Set	Old value	New values
1	259228	259141
2	190613	190470
3	252463	252322
4	371419	371321
5	477262	477118

Table-1: Execution time for traditional system and new system in millisecond.

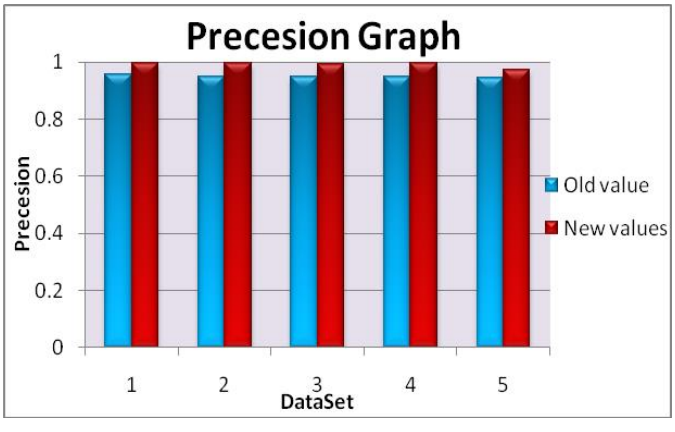


Graph-1: Comparison of traditional system and new system for execution time.

Following table2 and graph 2 shows comparison of traditional and new system for Precision. Precision of traditional system is less as compared to new implemented system.

Precision value		
Data Set	Old value	New values
1	0.9545454	0.995283008
2	0.95	0.994791687
3	0.9444444	0.974193573
4	0.95	0.994505465
5	0.95	0.994186044

Table-2: Precision values for Traditional system and new system.



Graph-2: Comparison of traditional system and new system for precision.

6. CONCLUSION

Finally this paper provide better webpage recommendation using methods which are the first model is the ontological model that can be semi-automatically constructed namely DomainOntoWP for domain term extraction. Next means second is semantic network analysis model, namely TermNetWP for extracting relationship between domain terms and web pages and the third model is nothing but conceptual prediction model is also proposed to namely TermNavNet for extracting web usage knowledge. To improve performance Key information extraction algorithm is implemented for webpage recommendation.

7. FUTURE SCOPE

In future experimental comparison takes place of key information extraction algorithm with different methods of recommendation of WebPages and different various types of methods used for providing webpage recommendation using different 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, March2012.

[3] Anwar Alhenshiri Dalhousie University," Investigating Features in Support of Web Tools for Information Gathering", 2014 ,47th Hawaii International Conference on System Science.

[4] Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu, "Web-Page Recommendation Based on Web Usage and Domain Knowledge" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 10, OCTOBER 2014.

[5] S.H.Rizvi, Ranjit R Keole," A Preliminary Review of Web-Page Recommendation in Information Retrieval Using Domain Knowledge and WebUsage Mining" International Research of Advance Research in Computer science andManagement studies,Vol. 3,Issue 1,January 2015.

[6] Arundhati Patil, Prof. Supriya Sarkar," Personalized Web Page Recommendation Using Ontology", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 3 Issue: 7,July 2015.

[7] Nguyen,Thi Thanh Sang, "semantic-enhanced web-page recommender systems" in Active Media Technology, vol. 5820, Australia December, 2012.

[8] Daniel Fett, Ralf Kusters, and Guido Schmitz", An Expressive Model for the Web Infrastructure: De_nition and Application to the Browser ID", SSO System University of Trier, Germany, September/October 2015.

[9] 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 Je_rey T. Pollock and Ralph Hodgson,August 2005.

[10] A. Harth, M. Janik, and S. Staab, "Semantic Web architecture," in Handbook of Semantic Web Technologies, J. Domingue, D. Fensel, and J. A. Hendler, Eds. Berlin, Germany: Springer-Verlag, 2011, pp. 43–75.