

Web-Page Recommendation Based on Web Usage And Domain Knowledge

Mr. Nitin Y. Suryawanshi¹, Dipali V. Dose², Shubhada S. Bhandarkar³, Jayashri V. Chaudhari⁴, Harsha P. Patil⁵

^{1, 2, 3, 4, 5} Dept of Computer Engineering

^{1, 2, 3, 4, 5} SSBT College of Engineering and Technology Bambhori, Jalgaon

Abstract- In intelligent web system web-page prediction plays a big role. People uses websites for various purpose like for study, education and entertainment point of view and access websites for online shopping purpose. Each website contains multiple WebPages. The proposed system consists of two models to provide better web page recommendation. The first model use an ontology to represent the domain knowledge. The second model uses semantic network to represent domain terms, Web-pages, and the relations between them.

Keywords- Web-page recommendation, domain ontology, semantic network.

I. INTRODUCTION

The web recommender system is use to predict web page or pages that are visited from a given web-page of a website. Web-page recommendation is shown as links to most viewed pages at websites or related books and stories.

When a user searches a website, a order of visited Web-pages during a session (the period from starting, to existing the browser by the user) can be produced. This order is organized into a Web session $S = d_1d_2 \dots d_k$, where d_i is the page ID of the i th visited Web-page by the user.

The growth of World Wide Web means internet results in high demand. But it is necessary to provide better required information to a user. This paper proposes a novel method to efficiently provide better Web-page recommendation through semantic-enhancement by integrating the domain and Web usage knowledge of a website. Two new models are proposed to represent the domain knowledge. The first model uses an ontology to represent the domain knowledge. The second model uses one automatically generated semantic network to represent domain terms, Web-pages, and the relations between them.

II. LITERATURE SURVEY

The classical systems failed to recommends the newly added pages or the products to the visitors since these pages or products are not in the current common navigation profiles.

So, to overcome the New page Problem the common navigation profile can be extracted in terms of semantic meaning or information. For that particular purpose ontology should be used. But in classical systems ontology was not used. Another problem is clustering in the existing systems. In this clustering number of recommended pages get increased. Non Useful or Unrelated pages or links are recommended which user never prefers.

The comparison of two sequences to determine their similarity is one of the fundamental problems in pattern matching technique. The Longest Common Subsequence method generates a sequence or a list of recommended products to the user. It is useful for the latest trend shopping that is online shopping purpose but what about the other users who does not use such online shopping websites but they still want the personalization in their web page recommendation system. The Numerical or numbering features should be represented by single components of vectors representing items. These components hold the exact value of that feature. Unfortunately, classifiers of all types tend to take a long time to construct. For example, if wish to use decision trees, need one tree per user. Constructing a tree not only requires that look at all the item profiles, but the problem is have to consider many different predicates, that could involve complex combinations of features. Hence, this approach tends to be used only for relatively small problem sizes.

The navigation of the users is largely driven by the semantics. Every time while user is searching he actually aims at finding some information concerning a particular different subject. There are many methods to extract keywords that characterize the web content. Which should be the exact matching between the terms determines the similarity between documents. Previous systems use this approach to find the similarity between documents. But by using this approach, only the binary similarity is achieved. That means no actual semantic similarity is considered. Semantic similarity is very important as far as similarity between documents is considered. Many number of research approaches integrate other information sources. require more abstract representation which enables a more flexible and uniform document matching process. It uses the semantic web structures, such as

ontology. Early systems possess a very common problem of caching of web pages. When a web user searches for an already cached page, this action is not recorded in the web sites log.

III. PROPOSED SYSTEM

propose a mathematical programming model to improve the user navigation on a website while minimizing alterations to its current structure. Results from extensive tests conducted on a publicly available real data set indicate that our model not only significantly improves the user navigation with very few changes, but also can be effectively solved. In addition, we define two evaluation metrics and use them to assess the performance of the improved website using the real data set. Evaluation results confirm that the user navigation on the improved structure is indeed greatly enhanced. More interestingly, find that heavily disoriented users are more likely to benefit from the improved structure than the less disoriented users.

1. DOMAIN ONTOLOGY

In the connection of Web-page proposal, the info information is Web logs that record client sessions on a day by day premise. The client sessions incorporate data about clients' Webpage route exercises. Every Web-page has a title, which contains the magic words that grasp the semantics of the Web-page. In view of these certainties, mean to find space learning from the titles of went to Web-pages at a web-website and speak to the found information in an area philosophy to help powerful Web-page proposal. A space philosophy is characterized as a theoretical model that points out the terms and connections between them expressly and formally, which thus speak to the area information for a particular space. The three principle segments are recorded as takes after: Domain terms (ideas), Relationships between the terms (ideas), and Features of the terms and connections.

Algorithm for Domain Ontology

Step1-Gather The Domain Terms: gather the Web log record from the Web server of the site for a time of time (no less than seven days), run a preprocessing unit to dissect the Web log document and produce a list of URLs of Web-pages that were gotten to by clients, run a product specialists to slither all the Webpages in the URL rundown to concentrate the titles, and apply a calculation to concentrate terms from the recovered titles.

Step2-Define the Terms:In this steps the Domain Concept Will be defined based on extracted terms. In this step, the area

ideas will be characterized for the given site in light of the removed terms. show the MS web webpage as a sample. This site concentrates on the application programming, for example, MS Office, Windows Operating System, and Database.

Step3-Define Taxonomic And Non Taxonomic Relations: There are three conceivable ways to add to the taxonomic connection boats, for example, a top-down advancement methodology begins from the most general ideas in the area and after that distinguishes the ensuing specialization of the general concepts, a base up improvement process begins from the most particular ideas as the leave hubs in the concept various leveled structure tree structure, then gatherings these most particular ideas into more general ideas, a crossover advancement procedure is the mix of the topdown and base up metho0dologies. recognize the center concepts in the area first and afterward sum up and practice them fittingly. The nontaxonomic connections can be the connection boat sorts utilized as a part of a social database aside from the connections between a superset and a subset, for example, referencing toward oneself, 1-M and M-N connections. In the MS webwebsite case, the fundamental sorts of nontaxonomic connections are recorded as beneath. The "gives" connection depicts the M:N connection send between idea Manufacturer and ideas Product, Solution, Support, and News. The 'is Provided' connection is the converse of the "gives" connection. The "has" connection portrays the M:N relationship between idea Application and ideas Product, Solution, Support, and News. The "is Applied For" relation is the opposite of the "has" connection. The "has Page" connection portrays the M:N connection deliver between an idea, for example, Application and product.

2. SEMANTIC NET OF A WEBSITE

The second model, i.e. new semantic net of a website, is generated which is a graph of concepts representing domain terms, Web-pages, and relations enclosing the collocations of domain terms, and the associations between domain terms and Webpages. Initially, the domain terms are gathered from the Web-page titles based on the assumption that a generally composed Webpage ought to have an instructive title; then the relations among these terms are extracted from the accompanying two perspectives.

Step 1- Collection Of Visited Web-Pages Titles: To collect the titles, collect the Web log le from the Web server of the website for a certain length of time, run a pre-processing unit that examines the Web log le thus producing a list of URLs of Webpages that were approached by users, and run a software agent to crawl all the Web-pages in the list to extract the titles.

Step 2-Extraction Of Sequence Terms: The algorithm used in the domain ontology construction is applied to extract the terms from the retrieved titles. The extracted terms are presented in the order as they appear in each title, namely they are collected as sequence terms.

Step 3- Build The Semantic Net Semnetweb: In SemNetWeb, each node represents a term in the extracted sequence terms and the order of sequence determines the from- Instance and to Instance relations of a term between other terms.

Step 4- Implementation of automatic construction of SemNetWeb: The SemNetWeb is incorporated in RDF to enable reusability and liability of the domain term network by other parts of a Web-page recommendation system. The table shows the algorithm to automatically construct a SemNetWeb FVTP (Frequently Viewed Term Patterns) FWAP (Frequent Web Access Patterns) CPM (Conceptual Prediction Model).

IV. RESULTS AND ANALYSIS

This section presents the detailed discussion about the results which are obtained from the experimentation. The experimentation is done on the proposed approach using synthetic dataset and the results are evaluated with the precision, applicability and hit ratio.

- Experimental set up and dataset description : The proposed web page recommendation approach is implemented in Java (jdk 1.6) with I3 processor of 2GB RAM. Here, the synthetic dataset is generated as like the same format of real datasets and the performance of the proposed approach is evaluated with the evaluation metrics. The generated synthetic dataset is divided into two parts such as, Training dataset (It is used for building the pattern tree model and test dataset (It is used for testing the web recommendation approach).
- Evaluation metrics For evaluating the proposed approach, we have used three measures such as precision, applicability and hit ratio [12, 8]. The formal definition of these three measures are given as, Where, C+ is Number of correct recommendations. I- is Number of incorrect recommendations.
- Performance of the web page recommendation algorithm The proposed web page recommendation approach is analyzed with the help of precision, applicability and hit ratio. Here, the testing dataset is given to the tree model constructed with the help of training data. Subsequently, the precision is computed based on the result obtained for

the test dataset Here, the results are taken for the proposed approach.

V. CONCLUSION

Ontology based learning and domain knowledge extraction is used to perform better enhancement in web page recommendation system. A number of Web-page recommendation strategies have been proposed to predict next Webpage requests of users through querying the knowledge bases. The experimental results are promising and are indicative of the usefulness of the proposed models.

For the future work, a key information extraction algorithm will be developed to Compare with the term extraction method in this work, and will perform intense comparisons with the existing semantic Webpage recommendation systems.

REFERENCES

- [1] B. Liu, B. Mobasher, and O. Nasraoui, "Web usage mining", in *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, B. Liu, Ed. Berlin, Germany: SpringerVerlag, 2011, pp. 527603.
- [2] International Journal for Research in Applied Science and Engineering Technology, "An Intelligent Web System by Integrating Domain and Web Usage Knowledge" on (Volume:3, Issue:3)march 2015,ISSN:2321-9653.
- [3] Loizou and S. Dasmahapatra, "Recommender systems for the semantic Web", in *Proc. ECAI, trento,Italy2006*.
- [4] 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. 4375.
- [5] B. Mobasher, Data mining for web personalization, in *The Adaptive Web*, vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdi, Eds. Berlin, Germany: Springer-Verlag, 2007, pp. 90135
- [6] G. Stumme, A. Hotho, and B. Berendt, Usage mining for and on the Semantic Web, in *Data Mining: Next Generation Challenges and Future Directions*. Menlo Park, CA, USA: AAAI/MIT Press, 2004, pp. 461480.
- [7] H. Dai and B. Mobasher, Integrating semantic knowledge with web usage mining for personalization, in *Web Mining: Applications and Techniques*, A.Scime, Ed. Hershey, PA, USA: IGI Global, 2005, pp. 205232.
- [8] S. A. Rios and J. D. Velasquez, Semantic Web usage mining by a concept-based approach for online web site enhancements, in *Proc. WI-IAT08, Sydney, NSW, in Australia*, pp. 234241.

- [9] G. Stumme, A. Hotho, and B. Berendt, “Semantic Web mining: State of the art and future directions,” *J. Web Semant.*, vol. 4, no. 2, pp. 124–143, Jun. 2006.
- [10] B. Zhou, S. C. Hui, and A. C. M. Fong, “CS-Mine: An efficient WAP-tree mining for Web access patterns,” in *Proc. Advanced Web Technologies and Applications*. vol. 3007. Berlin, Germany, 2004, pp. 523–532.
- [11] J. Borges and M. Levene, “Generating dynamic higher-order Markov models in Web usage mining,” in *Proc. PKDD*, Porto, Portugal, 2005, pp. 34–45.
- [12] C. I. Ezeife and Y. Lu, “Mining Web log sequential patterns with position coded pre-order linked WAP-tree,” *Data Min. Knowl. Disc.*, vol. 10, no. 1, pp. 5–38, 2005.
- [13] B. Zhou, S. C. Hui, and A. C. M. Fong, “Efficient sequential access pattern mining for web recommendations,” *Int. J. Knowl.-Based Intell. Eng. Syst.*, vol.10, no. 2, pp. 155–168, Mar. 2006.