Use Of Ontology And Web Usage Mining For Web-Page Recommendation

NazneenTarannum S.H. Rizvi, Prof. Ranjit R. Keole

Abstract— ***As the Web rapidly evolved into an immense repository of content and continues to grow at an exponential rate, human users discovered that they could no longer effectively identify the content of most interest to them or receive ambiguous results when they try to navigate through them. Several approaches are developed for improving our ability to find content. Basically not done by traditional recommendation systems, Web-page recommendation can automatically recommend Web-pages that are most interesting to a particular user based on the user's current Web navigation behavior. Also Semantic annotation helps assist automated (or computer-assisted) processing of content to better identify the real contents of pages. In this paper, we presented a framework for recommending better web-pages based on the queries fried by the users and thus provides a better search utility over google search engine using ontology and web usage mining…***

Index Terms— **Domain knowledge, Ontology, Semantic Web, Web-page recommendation, Web usage mining,**

I. INTRODUCTION

The World Wide Web is one of the most used interfaces to access remote data for both the commercial and non-commercial services. Due to the mushroom growth of usage of WWW by the users, these transactions are growing very quickly. Many experts forecast that the subsequent huge growth is forwarded in web information technology by adding semantics to web data, and will almost certainly consist of the semantic web.

Web semantic search is a key technology of the web database, since it is the major process through which the access content in the web data can be performed. Current web search technologies are fundamentally based on grouping of textual keyword search using ranking via the link structure of the web. Semantic Knowledge-Based as presented in showed how to abstract away from the raw real-world information step by step by means of semantic technologies. The framework triggered a knowledge exchange between the status monitoring agents but failed to apply the approach to more complex scenarios involving other agents. Semantic Knowledge-Based framework did not deal with the acoustic communication limitations associated to the underwater environment.

Conversely, present web semantic search does not permit semantic processing of web search queries, which analyzed based on both web search queries and web pages. The web pages with respect to "keyword" return precisely the semantically appropriate pages for a query. For the same reason, current standard web search does not allow complex web search queries that engage reasoning over the web.

Web usage mining is one of the frequent usage areas of web mining. The awareness of Web mining lies in analyzing user's behaviour on the web after exploring access logs and its popularity is increasing at a faster face especially in E-services areas. The applications in these web semantic search areas added its approval and made it as an inevitable part in computer and information sciences. Details like user log files demand for resources and maintain web servers, which is the core mining area of web usage. The semantic analysis gives the user browsing patterns utilized for target advertisement, development of web design, fulfilment of users and making market analysis. Most of the web service providers realized the fact behind it to retain their users.

Web-page recommender systems are one kind of recommender systems, which can automatically recommend Web-pages that are most interesting to a particular user based on the user's current Web navigation behaviour. Since a website is usually designed to show the index pages on the home page, the index pages take the role of guiding users to the content pages on the website through Web-page links, whereas with the index pages, a user usually has to navigate a number of Web-pages to reach the content page they are interested in. If the index pages of a website are not well designed, which is often the case, Web users will struggle to find useful pages and are very likely to leave the site. For a commercial website, this means losing potential customers. For an egovernment website, this will mean that the citizen's needs are not satisfied. Therefore, Web-page recommender systems have become increasingly valuable for helping Web users to find the most interesting and useful Web-pages on specific websites. Good Web-page recommendations can improve website usage and Web user satisfaction.

II. LITERATURE REVIEW

The related work and literature review covers the background, latest development of and related techniques for semantic-enhanced recommender systems.

Recommender Systems (RS) are information search tools that have been proposed to cope with the information-overload problem, i.e, the typical state of a consumer, having too much information to make a decision.

Recommender Systems can be either: rating-based (content-based or social/collaborative-based) – users explicitly express their preferences by giving binary or multiscale scores to items that they have already experienced, or feature-based (case-based, utility-based, knowledge-based and critiquing-based) – evaluating the

Manuscript received April, 2015.

NazneenTarannum S.H. Rizvi*, Department of Information Technology,S.G.B.University,H.V.P.M'sCOET, Amravati, India,*
Mobile No:9403307299.
Prof.R.R.Keole*,Department of Information Technology, S.G.B.University,H.V.P.M'sCOET,Amravati,India*
MobileNo:9823852893

match between a user's need and the set of options available .Recommendation systems attempt to predict items a user may be interested in, given some information about user's preferences and past behaviour, i.e., a user profile. Most existing recommender systems take advantage either of :

• collaborative filtering techniques, i.e., analyzing past actions and behaviour of all users in order to identify interesting associations between them or between the objects, which can be used to make recommendations to a single person (memory-based collaborative filtering and model-based collaborative filtering);

• content-based methods, i.e., recommending objects by analyzing the associations between user's past choices and descriptions of new objects;

• knowledge-based i.e., system suggest products based on inferences about user's needs and preferences ,

• hybrid filtering methods combining previous ones.

In 2006, M. Eirinaki [1] proposed a semantic web personalization system, focusing on word sense disambiguation techniques which can be applied in order to semantically annotate the web site's content.

In 2006, A. Loizou [2] presents a semantics-based approach to Recommender Systems (RS), to exploit available contextual information about both the items to be recommended and the recommendation process, in an attempt to overcome some of the shortcomings of traditional RS implementations.

In 2007, B. Mobasher[3] presents an overview of Web personalization process viewed as an application of data mining requiring support for all the phases of a typical data mining cycle. These phases include data collection and preprocessing, pattern discovery and evaluation, and finally applying the discovered knowledge in real-time to mediate between the user and the Web.

In 2008, S. A. Rios [4] proposed a concept-based approach to add semantics into the mining process. The solution proposed, was applied to a real web site to produce offline enhancements of contents and structure.

In 2009, S. Salin [5] presents a framework for integrating semantic information with Web usage mining is presented. The frequent navigational patterns are extracted in the form of ontology instances instead of Web page addresses and the result is used for generating Web page recommendations to the visitor.

In 2011, C.Ramesh [6] proposed a novel framework integrating semantic information in the Web usage mining process. Sequential Pattern Mining technique is applied over the semantic space to discover the frequent sequential patterns.

In 2011, S. Grimm [7] discusses the development of a new information representation system embodied in ontology and the Semantic Web. The new system differs from other representation systems in that it is based on a more sophisticated semantic representation of information, aims to go well beyond the document level, and designed to be understood and processed by machine.

In 2012, Thi Thanh Sang Nguyen [8] presents a new framework for a semantic-enhanced Web-page recommender (WPR) system, and a suite of enabling techniques which include semantic network models of domain knowledge and Web usage knowledge, querying techniques, and Web-page recommendation strategies.The paper enables the system to automatically discover and construct the domain and Web usage knowledge bases, and to generate effective Webpage recommendations.

In 2012, V. Sitha Ramulu[9]presents an overview of the semantic web mining- Integration of domain knowledge in to web mining to form semantic web mining, the concepts of semantic web mining.

In 2014, Suresh Shirgave[10]propose semantically enriched Web Usage Mining method for Personalization (SWUMP), which is a combination of the fields of Web Usage Mining and Semantic Web. In this method, the undirected graph is derived from usage data with rich semantic information extracted from the Web pages and the Web site structure

In 2014, Thi Thanh Sang Nguyen[11] proposed a the conceptual prediction model to automatically generate a semantic network of the semantic Web usage knowledge, which is the integration of domain knowledge and Web usage knowledge.

III. DOMAIN ONTOLOGY FOR WEB-PAGE RECOMMENDATION

In the context of Web-page recommendation, the input data is Web logs that record user sessions on a daily basis. The user sessions include information about users' Webpage navigation activities. Each Web-page has a title, which contains the keywords that embrace the semantics of the Web-page. Based on these facts, we aim to discover domain knowledge from the titles of visited Web-pages at a website and represent the discovered knowledge in domain ontology to support effective Web-page recommendation.

A domain ontology is defined as a conceptual model that specifies the terms and relationships between them explicitly and formally, which in turn represent the domain knowledge for a specific domain . The three main components are listed as follows:

1) Domain terms (concepts),
2) Relationships between the terms (concepts), and
3) Features of the terms and relationships.

Ontologies are often implemented in a logic-based language, such as OWL/RDF, to become understandable to software agents or software systems. Therefore, ontologybased knowledge representation allows sharing and interchanging semantic information among Web systems over the Internet. It also enables the reuse of the domain knowledge, and reasoning the semantics of Web-pages from the existing facts . Furthermore, ontological representation of discovered knowledge from different sources can be easily integrated to support Web-page recommendation effectively. Depending on the purposes of ontologies, they can be designed as domain conceptualizations of various degrees of formality and can be in the form of concept schemes, taxonomies, conceptual data models, or general logical theories

A. Domain Ontology Model

Domain ontology can be obtained by manual or automatic construction approaches. Depending on the domain of interest in the system, we can reuse some existing ontologies or build a new ontology, and then integrate it with Web mining. Web logs in a Web personalization system. Ontology is a knowledge representation technology whose

implementation can be machine-understandable using the ontology language, such as OWL. Ontology defines the concepts and their associations in an application domain. In the context of Web-page recommendation, it is necessary to have an ontology that expresses the meaning of Web-pages for better understanding Web usage patterns and discovering frequently viewed domain terms for supporting more effective Web-page recommendations.

The Web usage knowledge can be discovered from Web usage data through unsupervised learning processes, such as sequential pattern mining techniques, but without the semantics of Web-pages, the discovered knowledge are limited in supporting Web-page recommendation, such as no alleviation to the "new page" problem. Domain ontology is really useful to enhance a Web-page recommendation process by adding semantics to Web-pages, but how to build effective domain ontology for Web-page recommendations is always a big challenge. The study presented in this chapter builds domain ontology of Web-pages of a website that can be used to interpret the semantics of Web-pages. This chapter proposes a domain ontology model that represents the domain concepts, Web-pages, and the relations among them for a given website to support semantic-enhanced Web-page recommendation and also presents a novel method to build such domain ontology for a website.

In the context of Web-page recommendation, we build the domain ontology of Web-pages of a given website based on the visited Web-pages to represent the domain concepts (general domain terms), the relationships between the concepts with constraints, the instances of concepts (specific domain terms), Web-pages, and the links between Web-pages and specific domain terms.

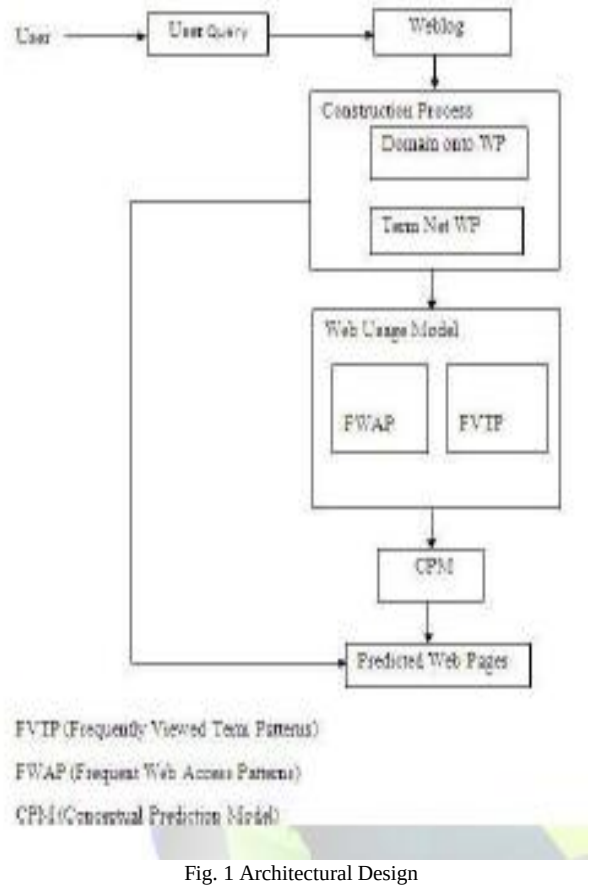

Fig. 1 Architectural Design

Definition 3.1 (Domain ontology model of Webpages-DomainOntoWP) A domain ontology structure of a website is defined as a four-tuples: Oman:= < C, D, PMAN, A>, where C represents terms extracted from the Web-page titles within the given website, D represents the Web-pages of the website, PMAN represents properties defined in the ontology, and A represents axioms, such as, an instantiation axiom assigning an instance to a class, an assertion axiom assigning two instances by means of a property, a domain axiom for a property and a class, and a range axiom for a property and a class. In details, C, D, and PMAN are further divided into sets:

C= C ∪Tman comprises a set of general domain terms (concepts) C, and a set of specific domain terms (instances of the concepts) Tman, D= SemPage ∪D comprises class SemPage which represents Web-page instances, and a set of Web-pages D, PMAN = Rman ∪Aman comprises a set Rman of the relations between terms (Rc) and the relations between terms and Web-pages (Rp), and a set of attributes Aman defined in the ontology. In particular, Rc will be specified depending on the application domain. Rp= haspage ⊔ isAbout , where the 'hasPage' relation states that a domain term may have some Web-pages, and the 'isAbout' relation is the inverse of the 'hasPage' relation. That means each domain concept class has the 'hasPage' object property referring to class SemPage, and class SemPage has the 'isAbout' object property referring to the domain concept classes.

IV. THE MINING PROCESS FOR PROPOSED SYSTEM

The proposed system gives a novel method to efficiently provide better Web-page recommendation through semantic enhancement by integrating the domain and Web usage knowledge of a website. In this dissertation, we are implementing the system on the basis of already available data of Microsoft, to provide analytics of the dissertation. Using the current visited Web-page (referred to as a state) and k previously visited pages (the previous k states), the Web-page(s) that will be visited in the next navigation step can be predicted. With the help of Domain Ontology Construction, we will collect the terms available in metadata of the web-pages and then depending upon the metadata, they will be categorized in order i.e. define the concepts and after this we will define taxonomic and non-taxonomic relationships between words. By considering above model, semantic knowledge representation model of web usage of website for webpage recommendation will be considered. With the help of this model, the pages can be predicted.

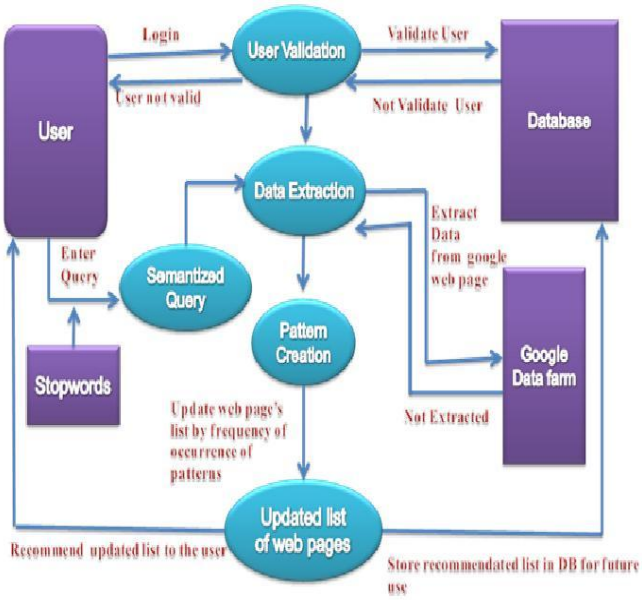

fig: Data Flow Diagram

The flow of work in WRS is illustrated by step by step explanation as follows:

1. Firstly User signs up in WRS to create a username and password.
2. On successive logins user provides his username and password.
3. User enters a query in WRS for accessing information.
4. Every query received by WRS interface is passed to Google API for extracting websites containing relevant information from the Google. At the same time every query is scanned to analyze whether the same query was fired by the user previously. If so, the already recommendated web-pages are provided to users.
5. The web pages extracted from the Google API are given to the Recommendation system where the CPM Algorithm will be executed and the resultant web pages will be displayed by their priority based on frequency of their domain terms
6. All information accessed through WRS i.e. all the terms searched for or websites visited are maintained in the browsing history list dedicated for the user. Admin keeps an eye on every query provided by the user and the information accessed from the response generated by WRS. Admin records this information in browsing history list maintained for the user and later mines it to draw useful knowledge about user interests.
7. Once the user session is over, Admin mines the browsing history list, classifies user interests and applies clustering techniques to draw useful knowledge about the user interests.

V. PERFORMANCE EVALUTION

The performance of Web-page recommendation strategies is measured in terms of two major performance metrics:*Precision* and *Satisfaction*. In order to calculate these two metrics, we introduce two definitions: Support and Web-page recommendation rules, as follows:

Definition (Support). *Given a set _ of WAS and a set P = {P1, P2 . . . Pn} of frequent (contiguous) Web access sequences over Δ, the support of each Pi €P is defined as:* $\sigma(Pi) = |\{S \in \Delta : Pi \subseteq S\}|$ / $|\Delta|$ *, where S is a WAS.*

Support is used to remove infrequent Web-pages and discover FWAP from WAS. This is accomplished by setting a Minimum Support (*MinSup*) and using it as a threshold to checkWAS. TheWeb access sequences whose Support values are greater than, or equal to *MinSup* are considered as FWAP. The smaller MinSup is set, the more FWAP are discovered.

Definition (Web-page recommendation rules). *Let* $S = S_1S_2 \ldots S_kS_{k+1} \ldots sn\ (n \geq 2)$ *be a WAS. For each prefix sequence* $S_{prefix} = S_1S_2 \ldots S_k\ (k \leq n - 1)$*, a Web-page recommendation rule is defined as a set of recommended Web-pages generated by a Web-page recommendation strategy, denoted as RR = {r1, r2, . . . , rM}, where ri (i = [1 . . .M]) is a recommended Web-page.*

A Web-page recommendation rule is deemed as *correct*, and/or *satisfied*, or *empty* based on the following conditions:

1) If *sk*+1 €*RR*,*RR* is *correct*.
2) If ∃ *si* €*RR*$(k + 1 \leq i \leq n)$, *RR* is *satisfied*.
3) If $M = 0$, *RR* is *empty*.

VI. CONCLUSION AND FUTURE RESEARCH

Dominion Ontology Semantic Search matches the semantic content with the user given query. The web search results more suitable to the user query are extracted after the syntactic and semantic evaluation during context analysis in structured dominion ontology. This paper has presented a new method to offer better Web-page recommendations through semantic enhancement by three new knowledge representation models. Two new models have been proposed for representation of domain knowledge of a website. One is an ontology-based model which can be semi-automatically constructed, namely DomainOntoWP, and the other is a semantic network of Web-pages, which can be automatically constructed, namely TermNetWP. A conceptual prediction model is also proposed to integrate the Web usage and domain knowledge to form a weighted semantic network of frequently viewed terms, namely TermNavNet. A number of Web-page recommendation strategies have been proposed to predict next Web-page requests of users through querying the knowledge bases. The experimental results are promising and are indicative of the usefulness of the proposed models.Compared with one of the most advanced Web usage mining method, i.e. PLWAP-Mine, the proposed method can substantially enhance the performance of Web-page recommendation in terms of precision and satisfaction. Moreimportnatly, this method is able to alleviate the "new-page"problem mentioned in the introduction because it based on not only the Web usage knowledge, but also the semantics of Web-pages.

Future work will focus on further experiments with different combinations of the system's functionalities, further contextualization possibilities from the Semantic Web Mining area, and an evaluation of the proposed approach with respect to learning support and to open-corpus learning

ACKNOWLEDGMENT

My thanks to the Guide, Prof. R.R.Keole and Principal Dr.A.B.Marathe, who provided me constructive and positive feedback during the preparation of this paper

REFERENCES

[1] S. Grimm, A. Abecker, J. Völker, and R. Studer, "Ontologies and the semantic Web," in Handbook of Semantic Web Technologies,J. Domingue, D. Fensel, and J. A. Hendler, Eds. Berlin, Germany: Springer, 2011, pp. 507–580.

[2] 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. 205–232

[3] Y. Yalan, Z. Jinlong, and Y. Mi, "Ontology modeling for contract: Using OWL to express semantic relations," in Proc. EDOC'06, Hong Kong, China, pp. 409–412.

[4] 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

[5] 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.

[6] A. Loizou and S. Dasmahapatra, "Recommender systems for the semantic Web," in Proc. ECAI, Trento, Italy, 2006.

[7] S. A. Rios and J. D. Velasquez, "Semantic Web usage mining by a concept-based approach for off-line web site enhancements," in Proc. WI-IAT'08, Sydney, NSW, Australia, pp. 234–241.

[8] S. T. T. Nguyen, "Efficient Web usage mining process for sequential patterns," in Proc. IIWAS, Kuala Lumpur, Malaysia, 2009,pp. 465–469.

[9] Modraj Bhavsar Int. Journal of Engineering Research and Applications "Web Page Recommendation Using Web Mining", ISSN: 2248-9622, Vol. 4, Issue 7(Version 2), July 2014, pp.201-206.

[10] Thi Thanh Sang Nguyen "Semantic-enhanced web-page recommender systems" December, 2012.

NazneenTarannum S.H. Rizvi, received the B.E.degree in Information Technology from H.V.P.M's College Of Engineering And Technology, Amravati in 2012. She is currently persuing Master's Degree in Computer Science and Information Technology from H.V.P.M's College of Engineering And Technology, Amravati.

Prof. Ranjit R. Keole, received the B.E.and M.E degree in Computer Science from Prof. Ram Megha Institute of Technology, Badnera in 1992 and 2008, respectively. His field of specialisation is web Mining. He is currently working as Associate Professor at H.V.P.M's college of Engineering and Technology,Amravati.

