

Ontology based personalized web information gathering

B.Umamaheswari¹, Shikha Shrivastava²

¹CSE, JECRC, Sitapura, Jaipur

²IT, JECRC, Sitapura, Jaipur

Abstract. Ontology is a paradigmatic for describing the knowledge and is used to verbalize profile of user in personalized web information collection. User profiles can be characterized by extracting knowledge from either a global repository or local repository. The global analysis makes use of global repository and produce effectual performance. The local analysis digs out user behavior from user background information. The user background knowledge can be better observed and represented if we put together global and local analysis. In this paper, a personalized ontology model is used to generate user profiles. This model observes profile of user from both global repository and local repository. The search results are more personalized and have on- topic specificity. The relevance and efficiency measures produce more accurate results. Hence this model is preferred for web information.

Keywords: ontology, personalization, user profile, global repository, user background knowledge

I. INTRODUCTION

Over the last few decades the information on the web increased to a greater extent. So collecting useful information from the web has become mind-numbing task. It has faced the problem of information divergence and overload. In the case of divergence the relevant pages and pages of user's interest are not retrieved. In the overload the retrieved pages are not of user's interest. Recently many web extraction systems are flourishing with the motto to extract the information based on user personal interest. It changed the notion of keyword based search to concept based search. User profile containing background knowledge [4] [7] of the user came into existence. It helps the system to gather not only relevant information but also personalized result for user's search.

User Profile [7] acquiring and representing plays an eminent role in the personalization of web information gathering. User profile can be represented in information format and data format. The former deals with topics of interest according to user information needs. The latter deals with user registration form and portfolio of user. User profile can be acquired by describing it through vector space i.e. a set of keywords using machine learning techniques. This approach had some drawback. They are user profile are poorly interpreted. Effectiveness depends on number of labeled data set. It is difficult to differentiate non interesting topics from interesting ones.

To represent user profile, the background knowledge of user is necessary. The global analysis and local analysis have contribution in the innovation of user background knowledge. The global analysis [10] use global repository such as Wordnet, Wikipedia and LCSH system. It has effectual performance but the topic it covers is restricted by the feature of knowledge bases. The local analysis use local information of user in the browsing history or consider preference of user in user profile. It is ineffective and may contain noise and uncertain information because they obtained results is based on data mining techniques (classification techniques). So user background knowledge can be retrieved and better represented if we integrate global and local analysis.

Ontology [1] is used to simulate user profile so that search results are more accurate. Ontology [12] is a corpse of knowledge describing some domain, normally a common sense knowledge domain using a representation vocabulary. It plays an important role to enable web-based knowledge acquisition and delivery, sharing and reuse between applications. It provides a common understanding [2] of topics that can be communicated between human and application systems. It consists of a hierarchy of concepts with in a domain and describes each concept's vital properties through an attribute-value mechanism.

The Major contributions are

- * Information extraction depends on ontology.
- * User profile includes registration details collected about users.
- * Use of web crawler for global database generation.
- * User preference is maintained in local repository.
- * Search is more personalized for the user.

The presented work integrates personalized web information gathering based on ontology. The following section presents the research background. Proposed architecture, methodology, Results and conclusion are presented in the later section. References conclude the paper.

II. RESEARCH BACKGROUND

The following section covers various researches in the field of web information gathering, information retrieval and ontology.

Chin-Ang Wu et al. [1] proposes a framework with multidimensional association mining that incorporates with ontology of user preference, which contains substitute queries that correspond to frequently used queries in the query history record. The representative power and the user inclination of the substitute queries are derived and expressed in fuzzy linguistic terms. The construction of the ontology and how active mining mechanism is assisted are also described. The relation of ontology of the user preference to the profile of user in the venture database allows conveying new mining results to the explicit users automatically.

Fensel et al. [2] in their book discusses the role played by ontology in the area of knowledge management and electronic commerce. They further added that how arising

web standards such as RDF and XML are used as an underlying illustration languages for ontology.

N.Zhong [3] and Sieg [11] investigates representation and construction of ontology for Web intelligence. Three ontology groups are suggested by the author some of the research and development with respect to the three groups is presented and a survey of the major ontology languages is done and a multi-level process of automatic construction of the domain specific ontology is discussed.

Gauch et al. [4] observed personalized ontology from Open Directory Project to understand user preference and user interest particularly on web search. The main reward of this approach is that system can function as a reference ontology generated from any topic that has associated textual content. They have placed reference ontology on topic level and associated Web pages from different search engines and the Open Directory Project.

Sakthi Priya T et al [5] proposed an efficient search engine which can be aligned better with user preferences. The efficiency is improved by making search result match the user interest. The system is evaluated by making the user to give the feedback. It is found to be more reliable and provide better response time compared to other search engines. Teena Skaria [6] proposed new personalization method by capturing user click through and mining the search results which provide more accurate result.

Li and Zhong [7] applied techniques like pattern recognition and association rule mining for knowledge innovation from user's local database and used the same for the construction of ontology. Xiaohui Tao et al. [8] projected an ontology model in lieu of user background knowledge and profile dependent web information extraction. The model erects user profile ontology by considering expert knowledge from the LCSH system and observing user background knowledge from user local repository.

Jiang., X and Tan, A.H [9] used web content mining techniques to observe semantic knowledge from domain based text document and used the same for Onto learn. P. Devisree and P. Revathi [10] proposed the different schemes that spawn user local repositories to match the depiction of an expert knowledge. R.Y.K. Lau et al. [12] demonstrate the function of extraction method to achieve adaptive eLearning using fuzzy domain ontology. The mechanism can automatically generate concept relation based on the texts deployed on online discussion forums.

Bhaganagare Ravishankar C et al. [13] performed survey on various research areas such as personalization, web mining, and information extraction. S.Vigneshwari and M.Aramudhan [14] constructed ontology by extracting documents from web and identified the concepts from them and used the concepts for building taxonomy. The taxonomy will be used for ontology construction.

Xiaohui Tao [15] proposed knowledge-based model which is a novel contribution to better understanding of Web personalization using ontology and user profiles, and enhance designs of personalized web information extraction systems. In reference [16] we made a survey of personalized ontology model and covered the basic terminologies related to ontology, personalization and web information gathering.

David Sanchez, Antonio Moreno [17] proposed a methodology to erect automatically ontology by considering the keyword available in the information gathered from the web. A normal search engine is used to extort concepts and

their relations from a set of web pages and display the result in a typical way.

III. PROPOSED ARCHITECTURE

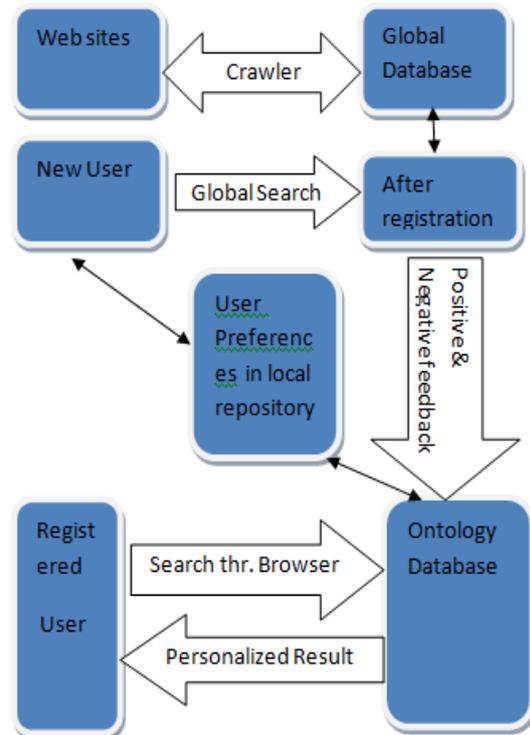


Fig. 1 Architecture of Proposed Model

In Fig: 1 the architecture of proposed model is explained.

From the websites with the help of web crawler pages will be retrieved and Global database will be generated. New user's preferences will be collected from his registration form and stored in local repository. User will make global search from global database and give his positive and negative feedback which along with local instance repository form the ontology database. The query given by the user will be directed to ontology database to get more personalized result.

IV. METHODOLOGY

4.1 Global database creation

The entire topic regarding a particular area for e.g. (Tennis) is extracted from global repository. A web crawler will be used and the corresponding pages and all their links will be extracted in an automated fashion. A web crawler is a software agent or program that browses the web in a disciplined way. It begins with a set of URLs (here related to Tennis) to visit. Once it visits these URLs, it captures all the hyperlinks in that page and adds them to the list of URLs to visit and store the content extracted in a database.

4.2 User preference extraction

New user will have preferences to be submitted while registration and give positive and negative feedback based on search results. The feedback and the user preferences are combined to build the ontology database.

4.3 Ontology Construction

It is the process of extracting knowledge in the form of concepts and relation from web documents that are normally unstructured or semi structured.

Here the approach used to ascertain and select representative concepts, their relation and corresponding websites for a domain and construct the final ontology is described. It mainly consists of feature extraction and taxonomy building. The algorithm is based on investigating a enormous number of web sites in order to capture useful concepts for a domain by considering the primary keyword's vicinity. The assumption is that words that are next to the specified keyword are closely correlated.

The entrant concepts are evaluated in order to choose the most sufficient ones by accomplishing a statistical analysis [17]. The selected classes are finally added to the ontology. For each candidate, the corresponding websites from where the concepts are extracted are stored, and the procedure is repeated recursively so that new terms can be found and hierarchy of concepts is constructed. The taxonomy of terms obtained from above steps can be the base for obtaining more complex ontological relations between concepts, or it can be applied to perform a search process for finding information or a classification scheme from a document collection.

The feature extraction phase is the first stage in ontology construction. It elaborates the concepts from the keywords present in the global database. The concepts are the building block in the construction of the ontology. The concepts are extracted from the keywords in such way that, there will be relation between the other nodes. The main function in the feature extraction phase is the mutual information [14]. The mutual information gives the relation between two keywords under observation. Mutual information relates two entities by their semantic dependency. The concepts extracted are then used to build the taxonomy that constructs the ontology.

V. RESULTS AND DISCUSSION

Any topic can be given as input to the proposed ontology model and the output was profile of user consisting of positive documents i.e. related documents and negative documents i.e. unrelated documents. The global repository is constructed based on output of web crawler which runs in the background. The crawler is refreshed at regular intervals to update the global database. The information retrieve are first preprocessed and then converted to human understandable format and stored in an SQL server database.

In order to make this ontology model [8] run more efficiently, only the newsworthy, communal, and location based topics were kept in the global repository. Then the user personalized ontology is built through user interaction. The beneficial and negative topics are selected by the user for ontology construction, following the explanation and elucidation related with the topics chosen. Each personalized ontology model retrieved about 18 affirmative and 25 negative documents.

The Performance measures in web information gathering based on ontology are precision and recall. Precision is the portion of retrieved pages that are relevant. Recall is the portion of relevant pages that are retrieved. The search results are considered for calculating the precision and recall.

Precision = (no. of relevant items retrieved) / (no. of retrieved items)

Recall = (no. of relevant items retrieved) / (no. of relevant items)

Consider 30 pages are retrieved by applying search technique. Out of which only 20 pages are found to be relevant and also there are 40 pages which are relevant but not retrieved then precision is calculated as 20/30 and recall is calculated as 20/60

The aim is to show that the ontology based method produce more personalized results compared with keyword search. The global database is created with one lakh crawled (considering only html/htm) pages covering different topics. The advantages and disadvantage of keyword search is analyzed by calculating precision and recall values.

The other evaluation scheme, F1 Measure, is calculated by:

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

The F1 measure is calculated by giving equal magnitude to precision and recall. The performance will be better if F1 measure values are greater. The experimental results shown in the table below proves that ontology model gives better results.

TABLE I. COMPARISON OF ONTOLOGY MODEL WITH KEYWORD SEARCH MODEL

d	Topic	Precision	Recall	F1	Precision	Recall	F1
1	cricket	0.74	0.54	0.62	0.73	0.70	0.72
2	football	0.76	0.59	0.66	0.77	0.73	0.75

VI. CONCLUSION AND FUTURE SCOPE

In this paper, an ontology based sculpt is used for representing background knowledge of user in personalized web information extraction. The ontology model erects user profile by deriving global knowledge from the global database and perceiving user background knowledge from user local repositories. This model provides a solution by combining global and local knowledge in a unique model. The findings can be applied to the building of information extraction systems. The model also contributed [16] in Personalization, Information Retrieval System, Intelligence System, Recommendation Systems, and Information Acquisition Systems.

Usage of automated tool for saving the profile of user in ontological format. This helps the user to view the personalized ontology. Integration of the global knowledge base like LCSH with the application that enable the integration of all of the related classes so that the user can give any type of string in the search process.

VII. REFERENCES

- [1] Chin-Ang Wu, Wen-Yang Lin, and Chuan-Chun Wu, "An Active Association Mining Framework with user preference ontology", *International Journal of Fuzzy Systems*, vol. 12, no.2, pp. 125-135, 2010.

- [2] Fensel, Dieter; Michael L. Brodie., “Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce”, *Second Edition, December, 2003. Springer-Verlag.*
- [3] N. Zhong, “Representation and Construction of Ontologies for Web Intelligence”, *Int’l J. Foundation of Computer Science, vol. 13, no. 4, pp. 555-570, 2002.*
- [4] S. Gauch, J. Chaffee, and A. Pretschner, “Ontology-Based Personalized Search and Browsing”, *Web Intelligence and Agent Systems, vol. 1, nos. 3/4, pp. 219-234, 2003.*
- [5] Sakthi Priya T, Revathy P, Pradeesh T, Rene Robin C.R, “Design and Development of an ontology based personal web search engine”, *International Conference on Communication, Computing and Security, Procedia Technology 6 pp. 299-306, 2012.*
- [6] Teena Skaria, Prof. T. Kalaikumar, Dr.S.Karthik, “A Cluster based multidimensional ontology mining for personalized search”, *International Journal of Electronics and Computer Science Engineering, vol. 1, no.3, pp.1390-1396, 2006.*
- [7] Y. Li and N. Zhong, “Web Mining Model and Its Applications for Information Gathering,” *Knowledge-Based Systems, vol. 17, pp. 207-217, 2004.*
- [8] Xiaohui Tao, Yuefeng Li, and Ning Zhong, Senior Member, IEEE, “A Personalized ontology model for Information gathering” *IEEE Transactions on Knowledge and data engineering. vol 23, No 4, April 2011.*
- [9] Jiang., X and Tan,A.H., (2005), “Mining Ontological Knowledge from Domain-Specific Text Documents”, *Proc. Fifth IEEE Int’l Conf. Data Mining (ICDM ’05), pp. 665-668.*
- [10] P. Devisree and P. Revathi, “ Web Search Personalization with Ontological User Profiles and World Knowledge Base” *International Conference on Computing and Control Engineering (ICCCE 2012), ISBN 978-1-4675-2248-9.*
- [11] A.Sieg, B. Mobasher, and R. Burke, “Web Search Personalization with Ontological User Profiles”, *Proc. 16th ACM Conf. Information and Knowledge Management (CIKM ’07), pp. 525-534, 2007.*
- [12] R.Y.K. Lau, D. Song, Y. Li, C.H. Cheung, and J.X. Hao, “Towards a Fuzzy Domain Ontology Extraction Method for Adaptive e-Learning”, *IEEE Trans. Knowledge and Data Eng., vol. 21, no. 6, pp.800-813, June 2009.*
- [13] Bhaganagare Ravishankar C, Dharmadhikari Dipa D, “ Web Personalization Using Ontology: A Survey”, *ISSN : 2278-0661 Volume 1, Issue 3(May-June 2012), pp 37-45 .*
- [14] S.Vigneshwari, M. Aramudhan, “A Technique to ontology Mining for Semantic Web Information Extraction”, *ISSN 1450-216X vol.94 no.1, 2013, pp. 49-60.*
- [15] Xiaohui Tao, “A Knowledge-based Model Using Ontologies for Personalized Web Information Gathering”, 1570-1263/10/\$17.
- [16] B.Umamaheswari, Pramod Patil, “Personalized Ontology Model – A Survey”, *International Conference on Hybrid Intelligent System, 978-1-4673-5116-4/12/31.00 2012IEEE.*
- [17] David Sanchez, Antonio Moreno, “Creating Ontologies from web documents”, <http://deim.urv.cat/~itaka/Publicacions/ccia04dsr.pdf>