ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

To Develop Web Personalization System Analysing User Browsing Behaviour

Sukhmeen kaur Hanjraw[1], Kuldeep Yadav[2] & Karamjeet kaur[3],

Ph.D Student, Department of CSE, Uttrakhand Technical University, Dehradun, India[1],

Associate Professor, Department of CSE, COER, Roorkee, India[2]

Assistant Professor, Department of CSE, Thapar University, Patiala, India[3]

ABSTRACT: Research on intelligent system can be traced back to the late 1950s, and a significant number of retrieval techniques have been invented during the period. Specially, intense evolution in this research field has occurred along with the birth and proliferation of the World Wide Web. The exponential progressions in web technologies have enabled users to experience enhanced delivery of personalized services & information through the integration of various existing technologies. In this instance the requirement for research activities in web management & enhancements by developing a standard, flexible but intelligent, adaptive and distributed framework for the support of heterogeneous infrastructure is obvious. The combination of web mining and semantic personalization plays a vital role in these circumstances.

KEYWORDS: WSD, NLP, Web Mining, Web Personalization, Semantic Technologies.

I. INTRODUCTION

Many applications of the World Wide Web need to discover the envisioned meaning of certain textual resources (e.g., data to be annotated, or keywords to be searched) in order to semantically describe the result. However, this vision is more complicated because current search engine focus only on retrieving the documents containing the user keywords, and lots of data that may carry the desired semantic information remains overdue. So in these circumstances Word Sense Disambiguation (WSD) & Neuro Linguistic Programming (NLP) plays an important role to retrieve valuable information from the web. Word sense disambiguation is the task of computationally determining which sense of a word is activated by its use in a particular context, considering sense as the representation of one of the possible meanings of a word, expressed on the basis of an electronic dictionary, a lexical knowledge base, ontology, or a certain application specific inventory.

In the last few years, a lot of attention has been devoted to improving web search and recommender systems through web mining. It allows sifting through large quantity of data for useful information. Web mining gives an innovative direction for scientific research and pushing web technology to making the meaningful information and exploits some data mining techniques to automatically mine valuable information from the World Wide Web. It makes an environment where the information available on the web can be semantically interpreted. It assembles more feature to built web personalize interaction and customizing a web site according to the requirements of users, obtaining advantage of the knowledge attained from the study of the user's browsing behavior.

In this work, we analyze data stored in web search engines' logs to discover usage patterns, and the aim is to enhance performance of search tools as well as to help users to find information on the web.

II. WEB MINING

Web mining technique gives set of standards used to extract important and relevant information from the web documents. It uses the data mining techniques for mining more relevant information through the use of certain

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

sophisticated algorithms. With the amount of data increases in the web every year, web mining is becoming an increasingly important area to transform this data into information. For this purpose, some of the existing data mining techniques such as association rule mining, clustering, statistical analysis, and classification [3] are used for effective web data analysis in addition to the new techniques proposed specially for web mining.

i. Semantic Personalization

The Semantic Personalization gives a common framework that allows information to be shared and reused across applications. It visualizes a globally interconnected network of machine process able information, made possible by the sharing of semantic data models, known as ontologies. The intense competition among Internet-based businesses to acquire new customers and hold the existing ones has made Web personalization a key portion of e-commerce.

The Figure 1 describes the current growth of web mining & semantic web that result is semantic personalization. Semantic personalization implies the delivery of dynamic and personalized content, such as textual elements, links, advertisement, product recommendations, etc., that are customized to needs or interests of a particular user or a segment of users [11]. The process of personalization involves data collection and preprocessing phase in which the information pertaining to user interests is obtained and preprocessed and a discovery phase in which user profiles are constructed from the data collected.

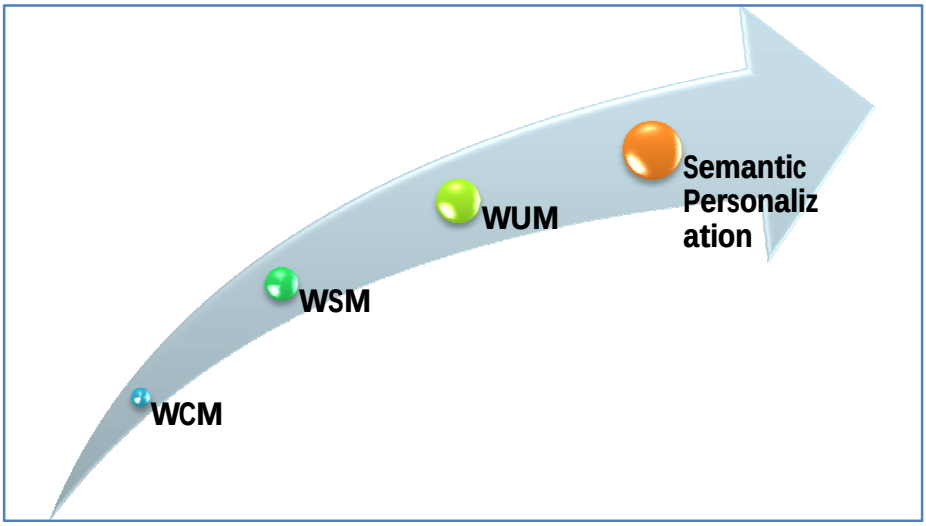

Figure 1: Hybrid Model of Semantic Personalization

III. LITERATURE REVIEW

Literature survey plays an imperative role in our research work. It is the documentation of a comprehensive review of particular theme, which holds the information of past and present development of the topic. Thus it motivates to develop innovative techniques and models. This work describes the work of eminent researchers and highlights the challenges, which still require to be addressed.

Sepliarskaia et.al in [1] describes the model for ontology matching over two educative content repositories. The basic idea behind this work is to automatically improve the efficiency of homogeneity resources for e-learning.

Rana and Singh et.al [10] have proposed a Semantic Web Mining interface, which is competent to handle heterogeneity issue and provide meaningful information in non-redundant way. Their work focused on finding the most ambiguous words and finding the relatedness measure with other important keywords in the query. It also considered removing

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

redundancy among keywords being matched, but performs little effort to enhance the social web and e-commerce activities

Mathieu et.al in [9] has proposed a semantic web search engine called Watson, which provides various functionalities to discover and locate ontologies and semantic data online. It provides new possibilities in terms of enlarging semantic applications used by the content of the semantic web by exploiting a with set of API tool consisting of high level elements for searching, exploring and retrieving semantic data from web.

Heidari et.al in [5] describes the need for combining the semantic web and e-commerce and evaluated the benefit of this combination. Their work describes the importance of semantic web technique and emphasized that the technology has the ability to extremely persuade the prospect improvement of the internet nation. They proposed innovative approach that support conventional web development and e-commerce to join the semantic web resulting.

IV. PROPOSED MODEL

The basic idea behind Behavior Evaluation and Web Personalization is to develop a standard consistent Web Mining approach for predicting the user browsing behaviors on web and to decrease the redundancy of accessed data. Our objective is to incorporate Web mining and semantic web in order to enhance the effectiveness of web personalization system. The view of the proposed System for the computation of user's behavior and for the reduction of the redundancy of accessed data is recommended below in figure 2.

i. Behavior Evaluation and Web Personalization (BBWPS)

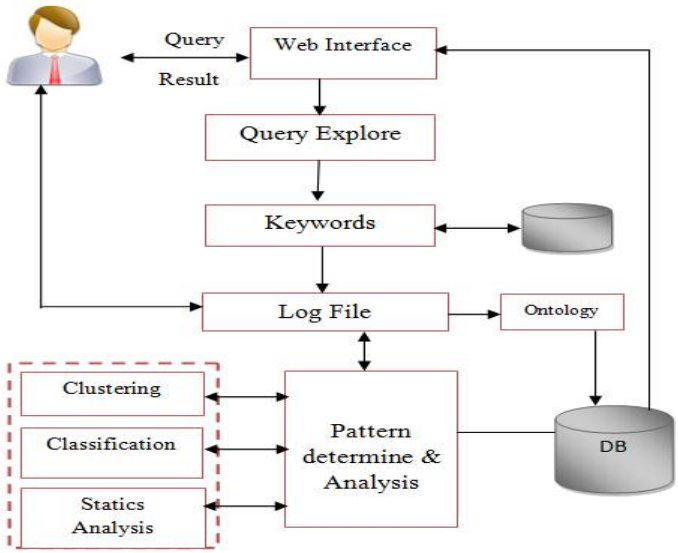

Figure 2: High Level View of Proposed Model

Content Module (CM):

The user can be identified by the cookies, IP address, and login authentication to identify the user. Common uses for cookies are authentication, storing of site preferences, shopping cart items, and server session identification. One of the most common privacy issues involves website cookies. So the user might delete the cookie. The cookie may be harmless but the privacy problem emerges when an unprincipled user gets hold of this cookie that divulges information regarding the site that you entered sensitive facts. Instead of using only cookie, **BBWPS** is identifying the user with hybrid system having login facility also. Login provides better accuracy and consistency.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Processing Query (PQ):
After user identification the next step is query processing. In query processing phase a set of key words is given as an input to PQ, which describes the user information needs. PQ uniquely contributes a different and novel algorithm that focus on finding the relevant meaning that describes the user's desires. The algorithm tries to find the user behavior and prevents from the repetition of accessed data. Thus PQ is an intelligent module as it improves the probability of success by finding the appropriate results. It performs several tasks to achieve the preprocessing phase: Tokenization and part of speech. Tokenization is the process of splitting up a query string into a set of tokens or words and Part of Speech is the process of evaluating a string of symbols, either in natural language or in computer languages, conforming to the rules of a formal grammar.

Algorithm 1: Intelligent Context Selection
Step 1; import WordNet, POS
Step 2: User Input t = "$Query$"
Step 3: Output = ambiguous keywords
Step 4: Tokenization text = $t.split(\text{"}Query\text{"})$
Step 5: $pos.tag(text)$
$[(tk_1, WP), (tk_2, VB), (tk_3, AD), (tk_n, NN)]$
Step 6: Ignore the stop words
$stop = stopwords.words('text')//call\ stopword$
The query string is an unstructured form of information and in order to convert it to proper input form i.e. structured input, automatic procedures are applied.

Domain Evaluation Module (DEM)

DEM discovers correspondences among semantically related entities of ontology and determines the set of synonyms having different names and structures. It finds the most appropriate meaning of ambiguous words according to the context in which it occur. Available literature reflects that probability model & page rank algorithms have been used to resolve the issues relating to query mapping. Probability model is based on probability of relevant & non relevant results while Page Rank computes the back links of web pages. Both these algorithms neither address the ambiguous queries nor do these compute the sentence and context related meanings of words thus found, therefore the motivation to propose DEM.

Algorithm 2. Domain Evaluation Module (DEM)
Step 1: input S_S to PQ
$input: source\ synset: S_S$
$output: target\ synset: T_S$
Step 2: call DEM( )
2.1: compare S_S & T_S
2.1.1 $if\ \#\ of\ elements\ in\ S_S < \#\ no\ of\ elements\ in\ T_S$
$then\ S_S\ subsumes\ T_S$
2.1.2 $if\ \#\ of\ elements\ in\ S_S = \#\ of\ elements\ in\ T_S$
$then\ S_S\ equivalent\ T_S$
2.1.3 $else$
$S_S\ plugs\ in\ T_S$
Step 3: Calculate probability score & $find\ the\ words\ with\ highest\ probable\ score$
$MAW = \sum_{P_S} Log(S_S) + Log(T_S)$
Step 4: return (MAW)
The algorithm returns MAW i.e. the words with the highest probability score thereby reducing the redundancy.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Pattern Reorganization Module (PRM)

In this phase system is trying to analysis the log file contents with the help of data mining techniques like clustering, classification and statistical analysis. It aims to determine the opinion of a user with respect to overall contextual polarity or intelligential reaction to a query. Then clustering and filtering techniques is used to access the desirable results. In this phase if the perception is obtained then the desired result provides to the user.

Algorithm 3: Pattern Reorganization Module (PRM)

$\boldsymbol{Step\ 1}$: $\boldsymbol{take\ t\quad e\ query\ (DEM)}$
$\boldsymbol{SteP\ 2}$: $\boldsymbol{Find\ t\quad e\ web\ pages\ from\ t\quad e\ index(w1, w2, w3 \ldots\ldots\ldots wn)}$
$\boldsymbol{Step\ 3}$: $\boldsymbol{calculate\ t\quad e\ page\ rank(PR)}$

$$(1-d) + d[\frac{PR(T1)}{C(T2)} + \cdots + \frac{PR(Tn)}{C(Tn)}]$$

$\boldsymbol{Step4}$: $\boldsymbol{Clustering}$ & $\boldsymbol{Classification}$
$\boldsymbol{Step\ 5}$: $\boldsymbol{most\ probable}$ Pattern $\boldsymbol{wit\quad accending\ order}$

V. RESULTS

Our system designed & implemented on open source general-purpose scripting language PHP, including Apache, MySQL, PhpMyAdmin and Xdebug. The modules are being implemented with Knowledge based technique and integrated with WordNet 3.0, one of the popular databases for the NLP domain. We have used EditPlus as text editor and window 8 workstation with dual quad-core 3.00 GHz Intel Xeon processors.
The internet user cannot get the appropriate result quickly because millions of information which are relevant and irrelevant is coming. To find the desired information among the huge result is difficult for ordinary user and expert IT professional as well. The performance of search engines is improving day by day. Some of the search engines are using semantic web technology either partially or as much as possible but not fully.

Evaluation Parameters
The section illustrates the evolutionary result of the BBWPS using some golden standard benchmarks with existing techniques. In order to analysis the overall quality of BBWPS, we exploited the standard precision and recall to analysis mapping results. These standards attempt to measure the amount of relevant and irrelevant information by evaluating the quantity of the obtained information.

Precision
Precision measures the ratio of probability of relevant information's extracted to the entire of irrelevant and relevant records retrieval. It is commonly described as a percentage. For example, RD is the list of important documents extracted and URD are the number of irrelevant documents extracted (see equation (i)).

$$\textbf{Precision} = \frac{\textbf{Number of RD}}{\textbf{RD+URD}} \times \mathbf{100} \qquad \text{(i)}$$

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Table 1: highlights list of 10 queries with their precision and recall measurer that describes the relevant and non-relevant results.

Table 1: List of Queries

Query No.	Queries	Number of link evaluated	More relevant	Less relevant	Irrelevant
Q.1	We sat along the bank of the Tevere river	5	4	1	0
Q.2	Book stays in London	5	3	2	0
Q.3	Total interest in last month	5	3	1	1
Q.4	They will definitely join our party	5	3	2	0
Q.5	Your flying planes with giant can be possible	5	2	1	2
Q.6	Apple	6	4	1	1
Q.7	Party	6	3	1	2
Q.8	Java	6	3	2	1
Q.9	Bank	6	5	1	0
Q.10	Interest	6	6	0	0

VI. CONCLUSION

We believe there is great potential for the studying user behaviour while surfing the Internet. However, researchers must first be able to characterize the tasks in which users engage on the Web and understand how the current features of Web surfing are used during different Web tasks. In this work we proposed a semantic personalization model, which is the process of creating customized experiences for visitors to a website. Instead of providing a single, broad experience, website personalization allows web masters to present visitors with unique experiences tailored to their needs, desires, and intentions.

REFERENCES

[1] Anna Sepliarskaia, Filip Radlinski, and Maarten de Rijke, (2017), Simple personalized Search Based on Long-Term Behavioral Signals , Springer European Conference on Information Retrieval 2017, pp 95-107
[2] Gerd Stumme, Andreas Hotho, Bettina Berendit, (2006), Semantic Web Mining, Elsevier, vol 4, pp 124-143.
[3] Jaideep Srivastava, Robert Cooley, (2000), Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, ACM, Vol 1, pp 1-12.
[4] Karun Bakshi, David R, (2005), Semantic Web Applications, *Proceedings* of the 29th ACM Symposium on Theory of Computing, pp 654-660.
[5] Karim Heidari, (2009), The Impact of Semantic Web on E-Commerce, World Academy of Science, Engineering and Technology", Vol 25, pp 303-306.
[6] Kenneth L, DIk L L and Wang-C L, (2010), Personalized Web Search with Location Preferences, IEEE Xplore Digital Library, pp 1-12.
[7] Kavita Sharma, Gulshan Shrivastava, and Vikas Kumar, (2011), Web Mining: Today and Tomorrow, IEEE Xplore Digital Library, pp 1-5.
[8] Li Ding, Tim Finin and Anupam Joshi, (2004), Swoogle: A Semantic Web Search and Metadata Engine, Proceeding of the Thirteenth ACM International Conference on Information and Knowledge Management, pp 652-659.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

[9] Mathieu Aquin and Enrico Motta, (2011), Watson. More than a Semantic Web search engine, INRIA, 2011- ACM Digital Library, pp 55-63.

[10] Vijay Rana, Singh G, (2014) "Analysis of Web Mining Technology and Their Impact on Semantic Web", CIPECH -2014, DOI: 10.1109/CIPECH.2014.7019035, IEEE Xplore, pp 5-11.

[11] Jia Hu and Philip K. Chan, (2008), Personalized Web Search by Using Learned User Profiles in Re-ranking, Fifth ACM International Conference on Web Search and Data Mining ,pp 1-14