Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 6 (2017) pp. 1683-1690

http://www.ripublication.com

Web Personalization through Semantic Annotation System

Sunny Sharma[1] and Vijay Rana[2]

Research Scholar1, Department of Computer Science & Engineering, Arni University Kathgarh, Indora, HP, India.

[2]Department of Computer Science & Engineering, Arni University Kathgarh, Indora, HP, India.

Abstract

In recent years, e-commerce website has become essential for human in daily life. However this vision is complication because current e-commerce systems are not able to find the expected product. So, there is need to personalize the search results. Extensive research has shown that Web result ranking can be significantly improved by considering personal behavioral signals such as past queries and user's hits for a website. To address this objective, a conceptual model is proposed for personalized search based on long-term behavioral signals.

Keywords: Web Personalization, SWM, Semantic Annotation, WUM, WCM, WSM.

1. INTRODUCTION

With the large amount of information available on the web, it is more complicated to extract desirable information from the existing knowledge system. At that instances semantic web and data mining techniques play an important role to mining valuable information from web. The term Semantic Web infers an Intelligent Web i.e. a meaningful web. It aims to make computers to understand the meaning of information on the web pages rather than merely presenting them to users [2]. The idea is to make World Wide Web (WWW) intelligent and machine readable by providing tools to find, exchange and interpret information to a limited extent by adding metadata.

Web mining exploits the data mining techniques to automatically extract useful information from the web and gave agreeable outcome to users. It utilizes three basic techniques content, structure and usage mining to extract meaningful information [7]

[10]. Content mining is a superlative tool for retrieving desired information a web page. Structure mining is the process of extracting knowledge from the interconnected hypertext document, tells the system how pages are interlinked with each other on the web and usage mining an imperative technique of automatically discovering and analysis the user interaction patterns with web servers.

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 manner 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. Semantic Personalization implies the use of semantic knowledge for creating customized experiences for visitors to a website. Instead of providing a single, broad experience, website personalization allows web masters to present information to visitors with unique experiences tailored to their needs, and intentions. The basic idea behind this work is to make innovative semantic personalization mining technology.

2. THE STATE OF ART: WEB MINING & SEMANTIC PERSONALIZATION

2.1 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 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.

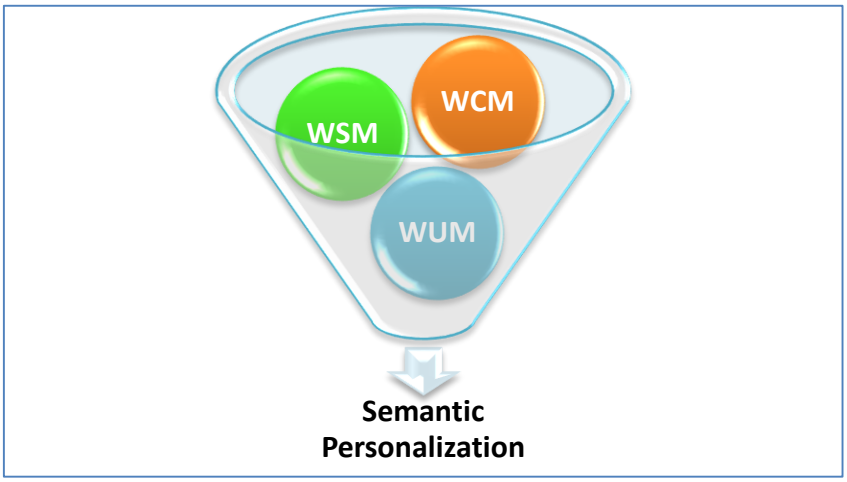

Figure 1: Hybrid Model of Semantic Personalization

It uses three basic techniques: web content, structure and usage mining. Web Content Mining (WCM) is a mechanism for retrieving meaningful information from the contents of web resources. Web Structure Mining (WSM) is the use of sophisticated algorithm to know the relation between interconnected web pages, and web Usage Mining (WUM) an imperative technique of automatically discovering and analysis the user interaction patterns with web servers.

2.1.1 Web Content Mining

Web Content Mining (WCM) is the process of extracting desired information from a web page. It may be text, images, audio, and video. It is related to text mining and data mining [14] because much of content of the web is unstructured and most of the data mining techniques are used in web mining such as clustering, association, classification. There are two types of approaches utilized in web content mining: Agent based approach and database approach. Agent Based Approach finds relevant information and analyzes the discovered information using domain characteristics and user profiles. Database Approach to Web mining have concentrated on strategies for sorting out the semi-organized information on the Web into more organized information and utilizing standard database querying mechanisms and data mining techniques to analyze it.

2.1.2 Web Structure Mining

Web Structure Mining is the process of finding structured information from the Web. It tells the system how the web pages are interlinked with each other and relationships between the web pages. Based on the kind of structure information used, it can be further classified into two parts. First is Documents Structure that contains the content in a Web page that can be sorted out in a tree structured format, based on the various HTML and XML tags. And the second one is Hyperlink Structure that links a location in a web page to a different location, either within the same web page or on a different web page.

2.1.3 Web Usage Mining

Web Usage Mining is the process of customizing a website to the needs of specific users by automatic discovery of user access patterns from web servers. The user's access patterns can be collected in a log file which can be placed either at client side or server side. Other source of information include referrer log which tells from which page a user is visiting to a page. Analyzing such data can determine the life time value of customers. It is segmented into three parts: Data collection and pre-processing, Pattern discovery, and Pattern Analysis.

2.2 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. Web 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 [12]. 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.

3. 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. This part describes and highlights the work of eminent researchers.

Mathieu *et al*., [9] have proposed a semantic web search engine called Watson, which provides various functionalities to find 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 set of API tool consisting of high level elements for searching, exploring and retrieving semantic data from web. The search engine architecture includes a crawler, indexes and query mechanisms to these indexes.

Cilibrasi *et al.,* [11] have proposed Google Distance algorithm to find the semantic relatedness measure between two words based on Google page counts. It also could find relative frequency whenever two terms emerge on the web within the same documents. Finally, he uses the WordNet database as an objective against which to judge the performance of their method.

Fathy Samaret *et al*., [13] have proposed dynamic and Hybrid Model for Emotion Detection from text that are related with emotion detection in facebook posts, twitter messages, whatsapp, email, where the training examples are automatically modeled through hash tags and emotions contained. The basic idea behind this model is to retrieve valuable information from input sentences and align with the ontology base which assembled from simple ontologies. The intelligent information retrieved from the input sentence by using a triplet information extraction algorithm, and then the ontology matching process is applied with the ontology base.

Cooley *et al.,* [12] presented that the Web is providing a direct communication medium between the vendors of products and services, and their clients. He and his team also proposed several techniques in which the user preference is automatically

captured from Web usage data, by using data mining techniques. Specifically, they developed techniques for preprocessing of Web usage logs and clustering URL references into sets called user transactions, thus making the personalization process both automatic and dynamic by using data mining techniques.

4. PROBLEM DEFINITION

With the rapid growth in e-commerce society it is essential to propose some knowledge based techniques, which are able to understand the user needs online. However this task is tedious due to the dynamic nature of web. A major issue on the Web Mining is inability to predict the user web surfing behavior appropriately on web and gain new knowledge through this interaction. Recently the web mining communities have focused on classifying standards that evaluates user browsing behaviors on e-commerce websites but failed to enhance user's satisfaction towards ambiguous queries from different perspectives. This problem is more complicated when user search same queries in different ways. The current systems simulates the user browsing behavior on the basis of cookies and hits, means system only tries to store the cookies instead of understanding the meaning of query or expanding the query based on user intense.

For example, the user searches a query four hundred times in a day and also searches another query every day for one year. The current system only stores the behavior of user where one query is being searched four hundred times a day, which is based on number of hits that is not actual behavior of the user. The actual behavior of the user is where same query is search every day for one year.

5. METHODOLOGY

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. 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.

Behavior Evaluation and Web Personalization (BEWP)

BEWP system consists of User Identification, Query Processing, Context Similarity and Domain Analysis, and Pattern Determine and Analysis.

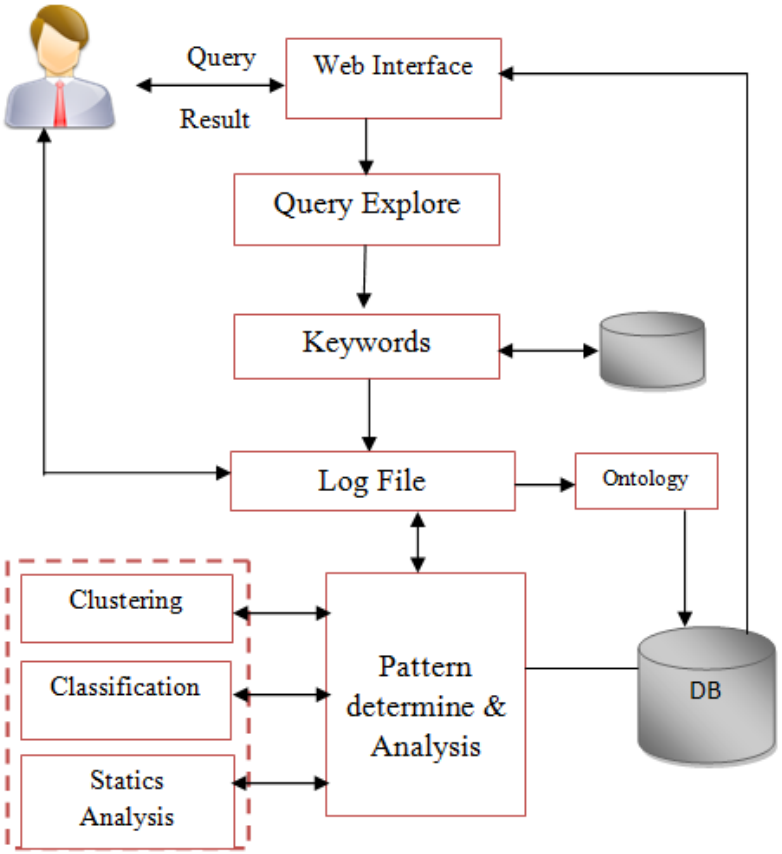

Figure 2: High Level View of Proposed Model

The Web master can identify a user by cookies, IP address, or exploit 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, whereas Login provides better privacy, accuracy and consistency. After user identification the next step is query processing. In query processing phase a set of keywords is given as an input for preprocessing, which describes the user information needs. Our mean to focus on finding the relevant meaning that describes the user's behavior and prevents from the repetition of accessed data. 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. After the preprocessing, Context Similarity and Domain analysis

aim to discover 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 context similarity and domain analysis.

At the final, the 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. At the end, the perception is obtained and the desired result provides to the user.

CONCLUSION

In this work we have proposed a new simple model for personalized search based on personal behavioral signals such as past queries and user's hits for a website that matches or outperforms the state-of-the-art for this task. We describe a general architecture for automatic Web personalization based on the proposed modules, and discuss solutions to the problems of data redundancy, knowledge extraction, and making recommendations based on the extracted knowledge.

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.

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

[10] R. Cooley, B. Mobasher, and J. Srivastava, (1997), Web Mining: Information and Pattern Discovery on the World Wide Web, IEEE Xplore Digital Library, pp 1-10.

[11] Rudi L. Cilibrasi & Paul M.B. Vitanyi, (2007), The Google Similarity Distance. IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 3, pp 370-383.

[12] Robert Cooley and Jaideep Srivastava, (2000), Automatic Personalization Based on Web Usage Mining, Magazine Communications of the ACM, pp 1-20.

[13] Samar Fathy, Nahla El-Haggar and Mohamed H. Haggag (2017), A Hybrid Model for Emotion Detection from Text, International Journal of Information Retrieval Research, Vol 7, N0.1, pp (31-37).

[14] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, (1996), From Data Mining to Knowledge Discovery in Databases, American Association for Artificial Intelligence, pp 1-18

[15] Yuahgui Lei, Victoria Uren and Enrico Motta, (2006), SemSearch: A Search Engine for the Semantic Web, Proceeding in 15th International Conference on Managing Knowledge in World of Network-EKAW'06, PP 238-245.

[16] Zhong N and Liu J, (2002), "In Search of the Wisdom Web", Journal of Computer-IEEE Computer Society, Vol 35, No 11, pp 27-31.