Research Paper Computer Engineering

Personalized Web Search

Piyush Joshi Student Computer Engineering Noble Engineering College Junagad India

Dr. Vipul Vekariya HOD Computer Engineering Noble Engineering College Junagad India

ABSTRACT

Users are accessing the World Wide Web to find the information for their use. But due to the growing of web pages roughly by millions of web pages, it is very difficult for the user to find the relevant information. Because of some commercial search engine, users are able to find the relevant information from it. User enters the query, find the no. of results and identify relevant results with his context. But such a search engine process user's general request, it cannot identify the interest of individual users. So the personalization of web search results has to carry out so that it can able to process the request of users individually and gives the results on context of his interest. In this paper we have presented various methods to personalize web search results. Based on their process to personalize results, it has been divided into the three different methods. We have presented the competitive analysis of such search engines and measure the precision value for each. At last we have presented some potential direction on which there is a need to personalize web search results.

KEYWORDS component; Information retrieval, Personalized web search, User Profiling , Semantic web mining.

INTRODUCTION

Over recent years, the World Wide Web has become a new communication medium thorough with Web information access. This incorporates with informational, cultural, social and evidential values to be specific. With the existence of various Search Engines e.g. Google, Yahoo and many more, the users are tend to use them for retrieving their desired information Web pages and. Although today's search engines can meet a user's general request, they cannot distinguishes different users' specific needs well. For example if some user X is a programmer and he/she searches for term as "python". Search engine retrieves result related to both python as language and python as category of snake. As the same time if any user Y who is interested to know about python as category of snake he/she would also get both the result on search term as python. Thus, Personalization of Web search has to carry out retrieval for each user incorporating his/her interests. Web users normally issue keyword queries to the Web search engines to fetch relevant information on a number of topics. As the users may have diverse backgrounds and Different expectations for a given query, some search engines try to personalize their results to better match the overall interests and preferences of an individual user. This task involves two major challenges.

- The search engines need to be able to effectively identify the user interests and build a user profile for every individual user.
- Once such a user profile is available, the search Engines need to re-rank the results in a way that matches the interests of a given user

In short it requires to customize content of web sites and web search engine in such a way that particular user can find information that is more relevant to him/her. Process of customizing content of web sites or web search engine based on user preference is called web personalization [1]. Web Personalization is used to retrieve the more relevant result based on user interest. Through the personalized system user can retrieve result which she/he wants from very large amount of data. There is no. of research done to provide the personalized system in respective fields, but again size of web data is increasing roughly it becomes these techniques less efficient. Personalized web search is an effective that provides specific results to different users when they submit their query [2]. How to obtain user's real-time information need is a main issue in personalized search. Again there is no. of research that has been done to provide personalized web search and re-rank the page ranking algorithm to display result based on user interest.

Personalized web search become an important tools for E-learning, E-Commerce, Scholar, etc. Usage of online material is playing an important role for users in self discovery learning process but it is difficult to find relevant pages from huge web pages on internet. So it is very crucial for developing personalized model for E-learning. Today many E-commerce web sites such as flipkart, amazon, ebbay, etc provide personalized web sites to direct user based on user preference. Generally they are using domain ontology which is reference ontology to discover knowledge from usage data to refine navigational patterns of particular user. Still it requires to personalized web search engines because if any user is frequently buying some product related to computer from flipkart. When that user will visit flipkart site he/she will get recommendation based on user preference. But if same user will search for term as "mouse" than again he/she will find some of the irrelevant information.

The paper is organized as follows: Section 2 gives background knowledge Semantic Web Mining. Section 3 describes some related works done to personalized web search. Section 4 describes comparative study of existing commercial search engines. Finally section 5 concludes paper.

SEMANTIC WEB MINING

Semantic web Mining Aims to combine Two Fast Developing Research area Semantic Web and Web Mining [1]. Semantic Web converts human readable data to machine readable form and different web mining techniques are applied to gain knowledge from that data.

Semantic web

To make sense on web data is more difficult because these data is largely unorganized and human understandable. Semantic web is to convert your data into well defined format in such a way that machine efficiently process can process your

data. If computer can understand meaning behind the data than it can give more relevant result based on user interest. So Semantic Web is very useful to develop standard and technology for both user and computer. Architecture of Semantic Web is shown in Fig.1.

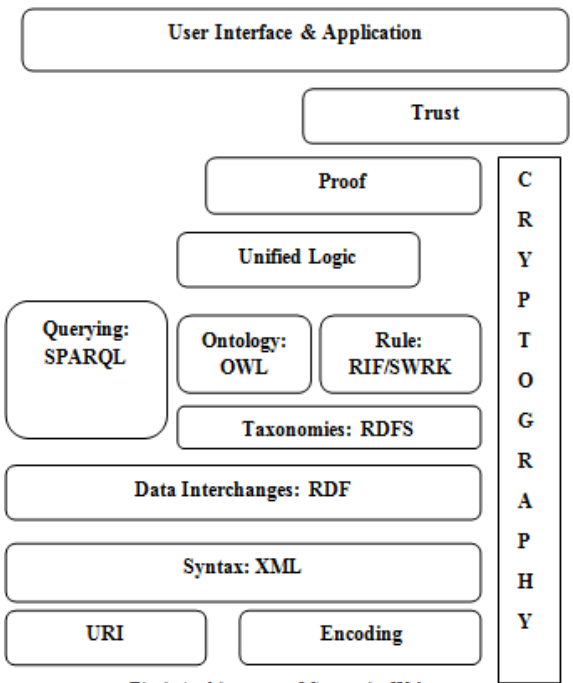


Fig.1 Architecture of Semantic Web

URI (uniform resource identifier) & Unicode:
URI is used to identify web resource. URI identification allows interaction with resource over a network using specific protocols like http or ftp.

Representation and constructs includes classes, properties and individuals. URIs is the fundamental benefit of semantic web technology. URIs provides users to know exactly what it is they are being referred.

Unicode is an encoding character sets and that allow all user languages can be used to read and write on the web by using standardized form.

XML & Namespace:
XML is a very powerful language used to transport and store data on the web. Its aims carry data, not to display data. The Tags in XML is user defined. XML schema is used to describe the structure of the XML document. XML schema also called as XSD XML Schema Definition.

XML Namespace is used to avoid conflict data or names.

RDF/RDFS Model:
Resource Description Frameworks (RDF): RDF is a framework for semantic web which is based on XML. RDF is XML based language used to describe resource with classes, properties and values assigned on the web. In web semantic RDF is used to describe the web resources. RDF contains the information about resources, such as the author, title, and modification date of a Web page. RDF is used for storing any other semantic data. RDF convert data into a form that machine can easily process. RDF Schema provides the framework to describe application specific classes and their properties

Ontology (OWL):
OWL is on the top of the RDF and XML based language. RDF is used to represent knowledge about things and their relationship but in some complex knowledge like "Only one person is allowed in project" this kind of data cannot be converted into knowledge using RDFS. It can be done by the standard language like OWL which used to represent complex knowledge about the things and their relationship.

Rules & Query:
RIF (Rule interchange Format) and SWRL (Semantic Web Rule Language) provide rules for the semantic web. SPARQL is query language for RDF. SPARQL used to querying the RDF data, RDF Schema and OWL Ontologies with knowledge. SPARQL based on RDF data model. The results of SPARQL queries retrieved in XML form.

Proof & Trust:
Proof layer is used to verify the results produced by the agents whether it is authenticate the agent behaviour. Trust layer is to provide a mechanism for trust and poise between information users (man or machine) and information sources

Web Mining
Web Mining is used to extract useful information from web data. All the data mining technique which is applied on web than it is called web mining. Web mining is very helpful to find out relative information, to established relationship between no. of page which has a same concept, to identify user behaviour etc [1]. Based on its function it can be divide into following three categories:

- Web Content Mining
- Web Structure Mining
- Web Usage Mining

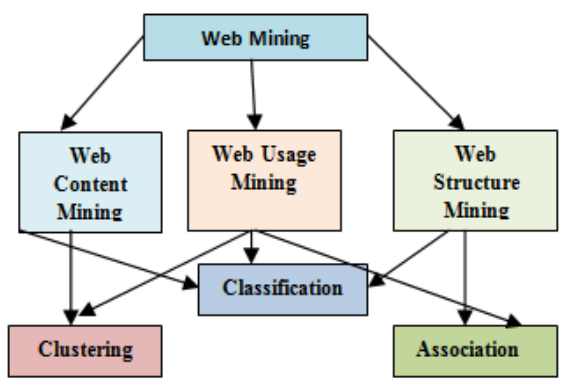


Fig. 2 Categories of Web Mining

Web Content Mining
It is a technique to retrieve useful information from huge amount of web data . Its aims to analyze content of web pages, pre-process data and find out useful information from raw data which is too structured. It is generally used for indexing to find out hidden web data.

Web Structure Mining
Web Structure Mining aims to describe the relationship between web pages and categorized web pages based on their similarity and generates useful information. Web Structure Mining is useful for Web crawler.

Web Usage Mining
Web Usage Mining is used to determine user behaviour by analyzing the browsing and navigational pattern of user. Web Usage Mining is a very useful in the Web Personalization System.

REALETED WORKS
Lots of work has been done to personalize web search results. In this section, we describe existing personalized web search techniques. Personalized web search has been achieved in context like personalized web search using keyword list, topical categories, collaborative filtering, Query expansion, session

based approach, etc. we also presents some existing work that personalize web search results by expanding user's search query.

Xiyuan Wu [2] has been proposed hybrid way to personalized web search. This approach combines content analysis and collaborative filtering technique to modify search result. User interest could be done with help of user's long term interest and with the help of collaborative filter it observes the interest of other user in same query and provides recommendation to user. Users send queries at the entrance of the tool. First, it will get the search results referring to the user's query. Then this model have computed the interest value of each searching item based on the user's interest model and the recommendation value of each item based on collaborative filtering.

Zheng Lu at, proposed a framework to fetch different user search goals for a query by clustering the feedback sessions. Feedback sessions were constructed from user click-through data. It can efficiently reflect the information needs of users. Second, It has been proposed a novel approach to generate pseudo-documents to better represent the feedback sessions for clustering. And finally, a new criterion "Classified Average Precision (CAP)" has been introduced to evaluate the performance of inferring user search goals.

John Garofalakis [3] has introduced a framework that generates the Cluster of User profile having same interest by analyzing log of users that have previously clicked on particular page. In [4] Open Directory Project (ODP) has been used as reference ontology to find out the semantic of the key words. From the history of user search it generates the concept based on concept and their relation to keywords users has put inside the cluster in that there are users which has same interest. It uses Google Search API and retrieves the result based on user interest which semantically annotate in user profile.

In [5], a novel query expansion algorithm has been proposed that was based on a model of personalized web search system. The new system served as a middleware between a user and a search engine and it was set up on the client machine. It could learn a user's preference implicitly and then generate the user profile from user's search log. When the user inputs search query, more personalized expansion words were generated by the proposed algorithm, and then these words together with the search keywords were passed to a popular search engine such as Baidu or Google. These expanded words could help a search engine to retrieve information for a user according to his/her implicit search intentions.

In [6], proposes a method for constructing an Enhanced User Profile by using user's previous browsing history and use it in domain knowledge. This Enhanced User Profile can be used for improving the performance of personalized web search results. The Enhanced User Profile specifically used for suggesting relevant pages to the particular user.

Zhicheng Dou at el, proposed three algorithms to personalize web search results. Generally personalized web search techniques were categorised in three group. It has been proposed an algorithm for all such methods to evaluate effectiveness of personalization on web search results.

Xuehua Shen at el [7], proposed a decision theoretic framework and developed techniques to identify user context for personalized web search. It has developed an intelligent client-side web search agent (UCAIR) that can perform eager implicit feedback. UCAIR was a browser plug-in that act as proxy for web search engine. This way, the captured user information always resided on the computer that the user was using, thus the user did not needed to release any information to the outside. Client-side personalization also allows the system to easily observe a lot of user information that might not be easily available to a server. Furthermore, performing personalized search on the client-side was more scalable than on the server side, since the overhead of computation and storage is distributed among clients.

Jie Yu [10] proposed a technique to provide short-term query context from web snippets to provide semantic background to user's behaviour and indentify related concept for the query. Context snap has been built from user search behaviour it built the context snap. Finally evolution of user context is considered by identifying forgetting factor to merge the independent user context snap in a user session.

[11] Proposed a RankBox, an adaptive ranking system to personalized web search. Based on user opinion on current query search it Re-Rank the result. This ranking algorithm learns from user feedback and replaces the current ranking algorithm with new machine learning based ranking technique. By analyzing user feedback RankBox learns to determine preference of user.

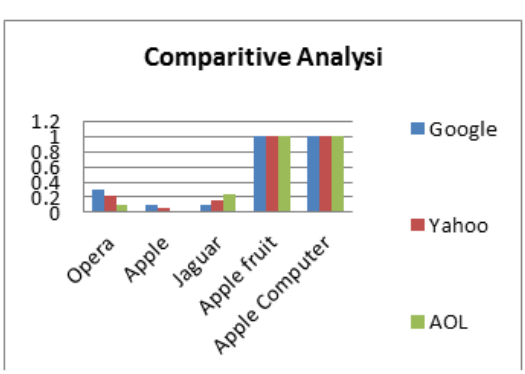
COMPARITIVE STUDY AND POTENTIAL DIRECTION OF RESEARCH

We have reviewed no. of research papers that personalize web search results. As per concern of user context, personalization is very critical part of today's commercial search engines due to the irrelevance in retrieved results and ambiguities in user search query. Table 1 shows the comparative analysis of some commercial available. As a part of comparative analysis we have all the search engines based on their previously defined categories like Google and Yahoo uses the hyperlink structure to retrieve the results. AOL search engine is use ODP as knowledge to personalize the web search results. UCAIR uses the content analysis techniques to personalize web search engines.

Table 1: Comparative Analysis

	Google	Yahoo	AOL	UCAIR
Opera	0.3	0.23	0.1	0.367
Apple	0.1	0.067	0	0.5
Jaguar	0.1	0.1667	0.2333	0.2
Apple fruit	1	1	1	0.93
Apple Computer	1	1	1	1

Fig. 3 Comparative Analysis



We have observed top 30 results of each search engines and entered some queries (ambiguous and unambiguous both) in search different search. The value presented in Table 1 is precision i.e. out of 30 results, how many no. of results are relevant. For example in query opera the context of user is "music" but Google were able to find the 10 results from 30 results so precision of Google for query opera in context of music is 0.3.

The no. of researches is still going on to personalized web search engines. Because of the some of the issues that are still remains in current search engine. Followings are represented potential direction to do the research on such issues.

- It is very difficult to identify short term interest from previous search history.
- Conflict in personalized web search result when users are accessing same machine (When user profile is generated from client side data).
- Threaten privacy between users.
- Link analysis algorithm unable to identify interest of individual users.
- Client side analysis of user profile is not always feasible.
- Sources for user profiling are unstructured and diverse in format.

CONCLUSION

By reviewing various paper on personalized web search, we conclude that personalized web search addressed many solutions and algorithms for personalization. We have presented various methods like content based analysis, hyperlink structure analysis and user group analysis and there were lots of research has been done in content based analysis and user group or by combining both of them. But from Comparative study we have found that existing processing technique is very hard for any web search engine to identify the changing behaviour of users. Still there will be many directions towards the research that was able to predict the current interest of users.

REFERENCES

[1]. Stumme, Gerd, Andreas Hotho, and Bettina Berendt. "Semantic web mining: State of the art and future directions." Web semantics: Science, services and agents on the world wide web 4.2 (2006): 124-143. | [2]. Wu, X., Fu, Y., Tian, S., Zheng, Q., & Tian, F. (2012, May). A hybrid approach to personalized web search. In Computer Supported Cooperative Work in Design (CSCWD), 2012 IEEE 16th International Conference on (pp. 214-220). IEEE. | [3].Garofalakis, John, Theodoula Giannakoudi, and Agoritsa Vopi. "Personalized web search by constructing semantic clusters of user profiles." Knowledge-Based Intelligent Information and Engineering Systems. Springer Berlin Heidelberg, 2008. | [4]. Zhu, Z., Xu, J., Ren, X., Tian, Y., & Li, L. (2007 Expansion Based on a, October). Query Personalized Web Search Model. In Semantics, Knowledge and Grid, Third International Conference on (pp. 128-133). IEEE. | [5.]Kumar, Rakesh, and Aditi Sharan. "Personalized web search using browsing history and domain knowledge." Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on. IEEE, 2014. | [6]. Liu, Fang, Clement Yu, and Weiyi Meng. "Personalized web search by mapping user queries to categories." Proceedings of the eleventh international conference on Information and knowledge management. ACM, 2002. | [7]. Shen, Xuehua, Bin Tan, and ChengXiang Zhai. "Implicit user modeling for personalized search." Proceedings of the 14th ACM international conference on Information and knowledge management. ACM, 2005. | [8]. Linden, Greg, Brent Smith, and Jeremy York. "Amazon.com recommendations: Item-to-item collaborative filtering." Internet Computing, IEEE 7.1 (2003): 76-80. | [9]. Sieg, Ahu, Bamshad Mobasher, and Robin Burke. "Web search personalization with ontological user profiles." Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. ACM, 2007. | [10]. Yu, Jie, and Fangfang Liu. "Mining user context based on interactive computing for personalized Web search." Computer Engineering and Technology (ICCET), 2010 2nd International Conference on. Vol. 2. IEEE, 2010. | [11]. Chen, Na, and Viktor K. Prasanna. "Rankbox: An adaptive ranking system for mining complex semantic relationships using user feedback." Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on. IEEE, 2012. | [12]. Page, Lawrence, et al. "The PageRank citation ranking: Bringing order to the web." (1999).