International Journal of Computer Applications (0975 – 8887)
Volume 124 – No.5, August 2015

A Semantic Web Search Personalization using Ontology

C.I. Arthi
VESIT
Mumbai, India

Kajal Jewani
VESIT
Mumbai, India

Pooja Nagdev
VESIT
Mumbai, India

Varkha Jewani
KC College
Mumbai, India

ABSTRACT
Web is the effective source of information gallery where it's retrieval is hampered by misunderstanding of user's query and user diversity. Hence we support shifting to the Semantic Web that in turn provides best search results and browsing facilities by providing machine readability of information on web taking advantage of ontologies. In order to obtain the better than best results, we have combined the results of personalized log based web search results and the results of semantic web and reranked those combined results. So that the user could get more relevant results than normal search engine. The merge of these techniques together could enhances search results when compared to normal web search methods.

Keywords
Semantic, Ontology, Personalization, Web Search, Rank.

1. INTRODUCTION
One criticism of search engines is that when queries are issued, most return the same results to users. In fact, the vast majority of queries to search engines are short and ambiguous. Different users may have completely different information needs and goals when using precisely the same query. For example, a biologist may query "mouse" to get information about rodents, while programmers may use the same query to find information about computer peripherals. When such a query is issued, search engines will return a list of documents that mix different topics, as shown in Table 1. It takes time for a user to choose which information he/she wants. On another query of "free mp3 download," although most users find websites to download free mp3s, their selections can diverge: one may choose the website www.yourmp3.net, while another may prefer the website www.seekasong.com. Personalized search is considered a solution to address these problems, since it can provide different search results based upon the preferences of users. Various personalization strategies, which include and have been proposed. However, they are far from optimal. personalization may lack effectiveness on some queries, and thus, there is no need of it for these queries; this has also been found by Teevan et al. [7]. For example, on the query "mouse" mentioned above, using personalization based on topical interests of users (for example, the one proposed by Chirita et al., we could achieve greater relevance for individual users than a common Web search. Beyond all doubt, the personalization is of great benefit to users in this case. Contrarily, for the query "Google," which is a typical navigational query as defined by Broder [9] and Lee et al. [10], almost all users consistently select a link to Google's homepage. Therefore, none of the personalized strategies could provide obvious benefits to users. In order to obtain the relevant results we have included the results of semantic web search engine. The target of the semantic web is to be "a web that understands machines", i.e the information in the web could be effectively accessed by the machines by understanding the content of the web. By this way web could take advantages of computers for information processing using ontologies. This definition is easily related to what already exists on the web: wrappers for extracting data from regularly structured pages, natural language analysis for extracting web page contents, indexing schemes, syndication facilities for broadcasting identified web resources. Much of this is painful and fragile: the semantic web should make it smart and robust.

2. PERSONALIZED WEB SEARCH
There have been several prior attempts on personalizing Web search. A comprehensive survey on personalized search can be found in [14]. In the following sections, we will summarize previous personalized search strategies, including personalized search based on content analysis, personalized search based on the hyperlink structure of the Web, and personalized search based on user groups.

In most of the above personalized search strategies, only the information provided by a user himself/herself is exploited to create user profiles. These are also some strategies that incorporate the preferences of a group of users to accomplish personalized search. In these approaches, search histories of users who have similar interests with a test user are used to refine the search. Collaborative filtering (CF) is a typical group-based personalization method and has been used in personalized search by Sugiyama et al. [12] and Sun et al. [1]. Sugiyama et al. [12] constructed user profiles based on a modified CF algorithm [13]. Sun et al. [11] proposed a novel method, named CubeSVD, to apply personalized Web search by analyzing correlations among users, queries, and web pages in click-through data. In this paper, we also introduce a method that incorporates click histories of group of users with similar topical affinities to personalize Web search.

Time, Cookie GUID, Query String, Browser GUID
{
(Position$_1$,URL$_1$), (Position$_2$,URL$_2$), ... (Position$_n$,URL$_n$)
}

Fig. 1 Representation of a log entry.

Search Query logs
Search Query logs consist of logs of searches made by users of search engines. They are usually collected at the search engine server. They typically consist of : user identity (ip address or anonymous id etc), search queries, corresponding clickthroughs made by the user and click information regarding it like the click time, no of clicks made etc. Some times the query logs are also captured on the client side i.e.,

on the user's computers. Clickthrough data/Query logs have been the most important source for capturing user context for user modeling. There has been some work in this connection some of which are described below. Sugiyama et. al used web browsing history in past N days for personalized search. They partition the browsing history data into three categories according to the time stamp, i.e., persistent data (before today), today data (today but before the current session) and current session data. They found that the performance of using web browsing history is competitive with that using relevance feedback. Speretta et. also used users search history to construct user profiles. Several other works have made use of past queries mined from the query logs to help the current searcher.

Web log Clustering

Search Query logs consist of logs of searches made by users of search engines. They are usually collected at the search engine server. They typically consist of : user identity (ip address or anonymous id etc), search queries, corresponding clickthroughs made by the user and click information regarding it like the click time, no of clicks made etc. Sometimes the query logs are also captured on the client side i.e., on the user's computers. Clickthrough data/Query logs have been the most important source for capturing user context for user modeling.

Here in our Experimental web log we have obtained user viewed URL's, the specific id of the individual website and also user viewed time of corresponding website. Those details could be collected in the server side and then it could be clustered. The clustering could be done by filtering most commonly viewed URL's. And the URL's getting highest count is ranked first in the search results. And then the websites viewed mostly could be displayed first. The screenshot for web log clustering is given below.

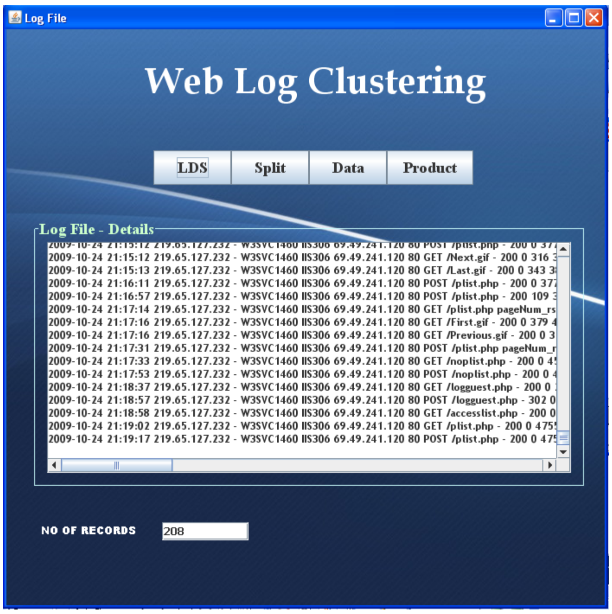


Fig.2.web log clustering

3. REQUIREMENTS FOR SEMANTIC WEB

In order to make search efficient, the results of queries need to be relevant and accurate. The results of irrelevant data can be avoided by a common framework called semantic web. The semantic web is a metamodel of machine understandable information and it generally described in the form of layers. Semantic web is layered and it is based on RDF(Resource Description Framework) XML provides syntax for content within documents with the goal of usability and data integration, whereas RDF defines the structure of data. RDF is designed to be a machine readable one, so that it cannot be displayed on web. RDF can describe properties, content and description of the terms (such as items) in the web pages. Computers can integrate the information from the web pages with the help of RDF.

Fig 3. Layered view of the semantic web

Uniform resource locators

Uniform resource locators (urls) are defined in [bm+98a] as "compact strings of characters for identifying an abstract or physical resource". they are com-monly used to address pages on the web (http://www.ideanest.com), specify ftp lo-cations (ftp://seng330@www.engr.uvic.ca/notes.txt) and name newsgroups (news://news.uvic.ca/uvic.engr.seng330). their ubiquitous acceptance stems from the fact that they are human readable and can be easily extended to encompass any number of protocols without compile-time changes to existing tools thanks to a consistent yet flexible syntax. Only new protocol handlers may need to be plugged in; the generic url parser never changes.

1) ***Ontologies***

In semantic web, the terms and the relationships could be defined by vocabularies. The vocabularies can be used to segregate relationships between constraints. The vocabularies can be simple with one or two relationship only or complex with plenty of terms. Those complex vocabulary relationships are termed as 'ontology'. Typically, the purpose of vocabularies on the semantic web is data integration and knowledge organization, further it canprovide results of the query according to its semantics. The vocabularies can be defined in standard format such as RDF ad RDF schemas, web ontology language (OWL) etc. by W3c. for example , in a shopping website an user can surf with terms like 'normal' or 'common'. Both queries should be reverted with same results. Hence the data can be integrated and defined by adding the extra information to RDF data. To make more detailed the terms 'common' can be related to terms such as 'usual'.

for example, expresses the following three relationships in RDF/XML format:
Figure 4,

```
<?xml version="1.0" encoding="utf-8"?>
        <rdf:Description (‘http://www.w3.org/author’>
        <author:name>”lee kong”</author:name>
        </rdf: Description>
        <rdf:Description  (‘http://www.w3.org/book/’”)>
        </rdf:Description>
        <authorOf:book>”ISBN9850432”
        </authorOf:book>
```

Fig 4. Merging an ontology Language with rdf schema

The RDF Description Model

RDF gives proper resource description. Resources have their own attributes. RDF defines a *resource* as any object by an Uniform Resource Identifier (URI). The attribues of resources are identified by types, and by their corresponding values. *Property-types* tells the relationships of values with those resources. In RDF, *values* can be text, numbers, or other resources. These attributes that points to the same resource is known as *description*.RDF is a syntax-independent structure for representing attributes and their own descriptions. RDF description is explained by following diagram.

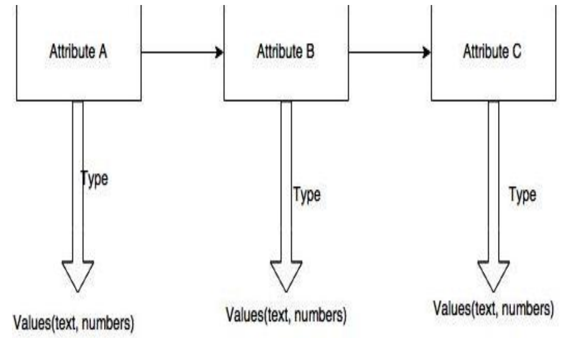


Fig 4. RDF Description model

The application and use of the RDF description model can be explained by following example. Consider the following statements:

1. "The owner of Acc no.20101 is Peter ladis"
2. "Peter ladis is the owner of Acc no. 20101"

The normal syntactic meaning of the statement same (Peter ladis owns the Acc no. 20101). The statement should be changed as machine readable. For machines these are simple combination of strings. Whereas humans can exactly extract meaning from these syntactic constructs. RDF attempts to provide a clear method of expressing semantics in a machine-readable structure by the use or properties, types, and values of attributes.

The RDF provides the methodology of representing the syntactic constructs in the form of directed graph or model in which the attributes are related by its types and values. Hence the resources or attributes are identified by its types and values. Given this representation, the descriptive model corresponding to the statement.

If the attributes needs more description or detailed information then the machine readable RDF descriptive model will be made very complex adding more information detail.

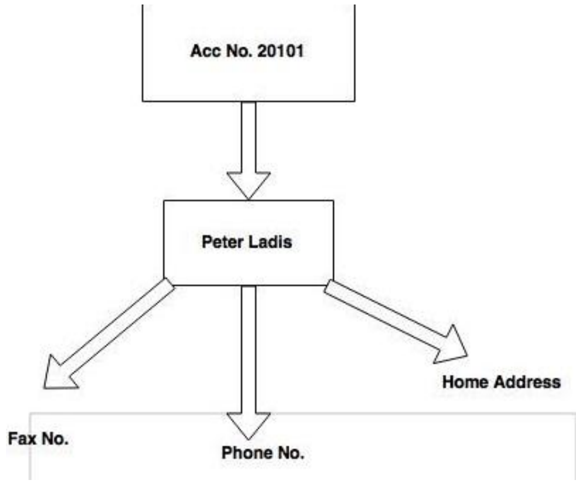
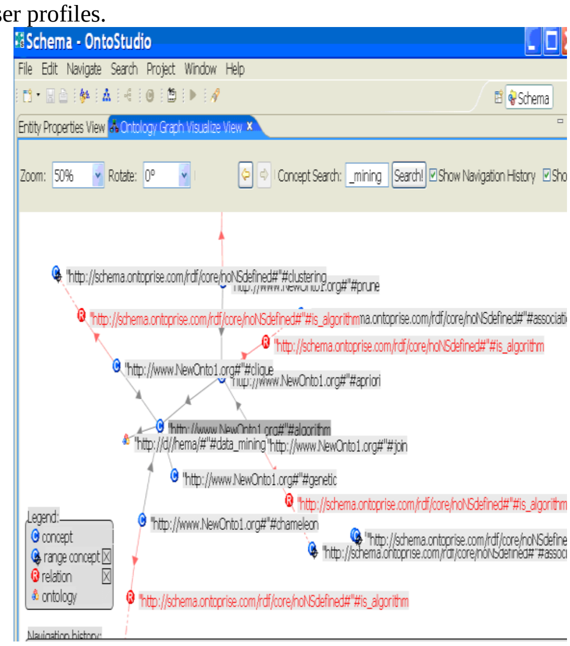


Fig 5. RDF Description model- example

When the statements are significantly complex or contains more details, the RDF description model also will be constructed so complex.

4. EXPERIMENTAL METHODOLOGY

According to the experimental methodology we have evaluated the semantic web search using ontologies and proposed a method which could give relevant results of users query, relevance to the user profiles. To do this, an approach for personalized search is implemented with resource of the semantic Web standards (RDF and OWL) to represent the information and the user profiles. The framework consists of the following phases: (1) building the log based clusters(users query) (2) generating the personalized results based on users query (3) clustering the documents which we obtain in the normal web search, (4) re-ranking the user's search results based on his/her profile, and (5) Combining the results of normal web search and personalized web search based on semantics. The experimental results show that the user context can be effectively used for improving the results of query obtained by re-ranking the search results based on the user profiles.



RE-RANKING STATERGY

In the proposed methodology, we first download search results from the normal web search engine. Then, the ranking of clicked URLs in a log entry is performed by a selected personalization algorithm. Once again the entire query results are reranked.To be specific, the reranking menthodology of normal web search is available already technique to improve query results. Here we propose, First, reranking of personalized click based user profiles query's results. Those results will be T_*personalized*

Second, generate the results of semantic web search. Those result items will be T_*semantic.*

Third, combine the two rank lists T _*personalised* and T_*semnatic* and once again reranking of those combined results will be performed.

5. CONCLUSION

In this paper, we first download search results from the normal web search engine. Then, the ranking of clicked URLs in a log entry is performed by a selected personalization algorithm. Once again the entire query results are reranked. To be specific, the reranking menthodology of normal web search is available already technique to improve query results. we proposed, First, reranking of personalized click based user profiles query's results. Those results will be T_*personalized*

Second, generate the results of semantic web search. Those result items will be T_*semantic.*

Third, combine the two rank lists T_*personalised* and T_*semnatic* and once again reranking of those combined results will be performed.

Altogether the purpose of all these steps is to obtain relevant result of the given query. Here we are trying to obtain exact matching results by referring users profile details and by their log based clusters.

Also semantic web search methodology is the known and best one for obtaining semantic oriented results. For example, consider the query 'mouse'.A software engineer always or mostly search for the term mouse which is a device. He mostly don t go for the query 'mouse' which is a mouse. So here the keyword based query matching technology fails. In semantic web the query would be machine understandable by the use of ontologies which deals with the relationship of vocabularies. Hence semantic web will the better one when compared to common keyword based searching. We improved the results better than the best by combining the results of normal web search results and user profile based personalized results and finally reranking those combined results.

Hence we could improve the relevancy of results of searching query.

6. REFERENCES

[1] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, "Analysis of a Very Large Web Search Engine Query Log," ACM SIGIR Forum, vol. 33, no. 1, pp. 6-12, 1999.

[2] B.J. Jansen, A. Spink, and T. Saracevic, "Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web," Information Processing and Management, vol. 36, no. 2, pp. 207-227, 2000.

[3] J. Trajkova and S. Gauch, "Improving Ontology-Based User Profiles," Proc. Recherche d'Information Assiste´e par Ordinateur (RIAO '04), pp. 380-389, 2004.

[4] M. Speretta and S. Gauch, "Personalized Search Based on User Search Histories," Proc. IEEE/WIC/ACM Int'l Conf. Web Intelligence (WI '05), pp. 622-628, 2005.

[5] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," technical report, Computer Science Dept., Stanford Univ., 1998.

[6] J. Pitkow, H. Schutze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, and T. Breuel, "Personalized Search," Comm. ACM, vol. 45, no. 9, pp. 50-55, 2002.

[7] J. Teevan, S.T. Dumais, and E. Horvitz, "Beyond the Commons: Investigating the Value of Personalizing Web Search," Proc. Workshop New Technologies for Personalized Information Access (PIA), 2005.

[8] A. Pretschner and S. Gauch, "Ontology Based Personalized Search," Proc. 11th IEEE Int'l Conf. Tools with Artificial Intelligence (ICTAI '99), pp. 391-398, 1999.

[9] A. Broder, "A Taxonomy of Web Search," ACM SIGIR Forum, vol. 36, no. 2, pp. 3-10, 2002.

[10] U. Lee, Z. Liu, and J. Cho, "Automatic Identification of User Goals in Web Search," Proc. 14th Int'l World Wide Web Conf. (WWW '05), pp. 391-400, 2005.

[11] J.-T. Sun, H.-J. Zeng, H. Liu, Y. Lu, and Z. Chen, "CubeSVD: A Novel Approach to Personalized Web Search," Proc. 14th Int'l World Wide Web Conf. (WWW '05), pp. 382-390, 2005.

[12] K. Sugiyama, K. Hatano, and M. Yoshikawa, "Adaptive Web Search Based on User Profile Constructed without Any Effort from Users," Proc. 13th Int'l World Wide Web Conf. (WWW '04), pp. 675-684, 2004.

[13] J.S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proc. 14th Conf. Uncertainty in Artificial Intelligence (UAI '98), pp. 43-52, 1998.

[14] J.-R. Wen, Z. Dou, and R. Song, "Personalized Web Search," Encyclopedia of Database Systems, 2009.