Folksonomy based Personalized Search using ConceptNet

Rohith Krishna V
School of Information Technology and Engineering
VIT UNIVERSITY
Vellore, India

Sujatha R
School of Information Technology and Engineering
VIT UNIVERSITY
Vellore, India

***Abstract*— The amount of information on the web is increasing day by day at an exponential rate. There is a need for retrieving the necessary information required for the user. The requirement of the information varies from user to user (example: variation in the required information between a geek and geek's mother). So there is a need for the personalized search where the results should vary based on the user interests. This paper will discuss about the research going on in this particular area and also serve as a guide in designing a personalized search engine efficiently and effectively using the ConceptNet with a detailed architecture.**

Keywords- ConceptNet; Folksonomy; Tagging; WordNet;

I. INTRODUCTION

The number of digital resources is increasing in everyday lives; especially the Worldwide Web has become the largest source of the digital resources with the resources growing at exponential rate. In this situation there is a need for the user to identify the required resources from the vast available data. The search engines like Google [1], Yahoo [2] and a lot more are serving this purpose. But a problem with these search engines is that each search is considered independent which means that the same results will be provided for different users irrespective of the their level of interest. Consider for example a set of user's type java as a keyword for searching, of which some may require java tutorials and some may require java software and some requires java code but the results will be same for the users if they use Google or Yahoo. Thus the need for the personalized results which is based on their interests is becoming more important for the users.

The common vocabulary extracted from social tagging is called as Folksonomy. But considering folksonomy there are many problems such as ambiguity, inconsistency and lack of precession which results in tags that do not best describe certain web resources and the results will vary from what is expected thus violating our goal.

This paper will discuss on the various methods for personalization and also solutions to the above discussed problems thus providing an idea on what issues to be considered and what to follow for the designing a system for personalized search.

The rest of the paper is organized as follows. The section II will discuss the past and ongoing research related to this field and the section III will propose a system for better designing of personalized search with explanations along with the architecture. The section IV will contain the conclusion for this paper. The section V will discuss on the future work related to this area.

II. RELATED WORK

A lot of work is being carried out for the personalization of search based on the user interests and comparatively a small amount of work considering folksonomy for personalization.
There is a lot of research going on in order to personalize the search based on search history of the user, user profile etc... The problem with the search history is that the user may be using different systems and search history varies from machine to machine and when a profile [4] is used then if the user wanted resources in different domains it becomes difficult. These follows a Taxonomic approach.

We have already discussed about the problems with the taxonomic approach in the introduction. This session mainly concentrates on the folksonomy based personalized search. This section is further organized as follows:

A. Overcoming Folksonomy Disadvantages of ambiguity and inconsistency

Thus in order to overcome the disadvantages of folksonomy Udell (2005) [5] provided a method for vocabulary stabilization around the common choices in tags but there is a loss of some valuable information which does not serve our purpose. Kipp and cambell (2006) [6] used co-word analysis to derive patterns from tags used in social bookmarking delicious. This showed good results but there is a problem with the same word having different meanings. (For example "apple" may represent "fruit apple" or the "company apple"). Using WordNet [8] we identify similarity of various tags by measuring the relatedness of their concepts.

B. *Personalizing Search based on Folksonomy*

The delicious website[3] provide an interface for tagging where personalized results will be returned based on the search tag but the problem with this is, the search tag must exactly match with a tag in the folksonomy (Dataset) i.e. it does not consider the sematic aspect of searching. Peter and stock (2007) [7] proposed natural language processing methods to identify the entities referenced by the tags that could be useful for retrieval by processing the sentences provided as a query for the search. Heung-nam Kim et al. (2011) [9] Proposes a method which uses user-tag matrix, item-tag matrix and user-item matrices. Where the tag similarities and the item similarities are identified and based on the cosine similarity calculation of various tags and items. The results are personalized based on the tag used by the user and their similarity with the other tags is used for identifying the items.

C. *Page Ranking after Personalization*

ching-chieh kiu et al. (2010) [10] Proposed a method for ranking of pages for personalized search. The rank of a page resulted from a search engine is re-ranked based on the pre-defined user profile which includes user interests and top-level domains of webpages which considers only few domains and will not produce the results from other domains. Heung-nam Kim et al.(2011) [9] Proposed a method called Folksonomy Boosted Ranking for ranking of the resources according to the user interests which uses the normalized user-tag matrix and tag-item matrix and displays the results which more suits user interests.

III. FOLKSONOMY BASED PERSONALIZED SEARCH

We have seen the various solutions to the problems faced by the Folksonomy based personalized system. This section will propose a method which provides accurate results of personalization. The major issues that we have to consider are:

A. *Identifying concepts from the query*

The query given by the user should be processed before it is used for searching. The initial step that has to be performed is to identify the parts of speech of the given sentence and extract the subject, object, verb and adjective (if any present) of the given sentence and the concepts relating all these should be extracted using the ConceptNet [11]. This method of using ConceptNet for extracting the concepts serves better as the matching will be done based on the concepts identified and not the word matching so the results will be accurate.

Let us see this with an example. Suppose the keywords that are extracted from the query are “money, Theft, gun”. Then the concept that is identified from the concept net will be “Robbery” (based on the data it seems like that). Thus the search will be done by searching for the keyword robbery which will yield accurate Results. More details about the ConceptNet will be found in [11]. Let us now take an example from the fig.1. The query if contains “bank, wallet, ATM” then the concept that is identified from the ConceptNet is “money”. This can be seen in the figure which is the root node of the three concepts.

Figure 1. A Sample Example of ConceptNet

B. *Personalizing the search based on user profile*

The next important issue is searching for the identified concepts. The issue here is that the search should be personalized i.e. it has to be based on the user profile. Here the user profile is based on the tags used by the user for the resources. Now first the identified concepts should be searched in the folksonomy dataset. If a concept is novel i.e. if a concept which has been identified is not present in the dataset then it can be identified by making analogies with the known concepts.

Another problem here is to search a word with different meanings. For example consider “Eat, chips” here the ConceptNet[11] will return the keyword as Potato chips than micro-chips which has more relevance to situation so this is solved by using ConceptNet.

Now considering user based personalization it is better to incorporate two methods [9][10] i.e. identifying the top level domain of interests of the users based on/her his dataset and also using his/her tag similarities used for various resources(items). For example consider apple. The top-level domain of the user will mention what type of domain he is interested in whether “fruit apple” or “company apple”.

C. Page Ranking

After identifying the personalized results it is also important that the relevant interested results should come first as everyone will be looking for the interested resources in the top 10 results. Thus in order to provide the interested results page ranking using Folksonomy Boosted Rank [9] will serve the purpose better as it will consider the user profile from the tags tagged by the user and returns the returns based on his profile i.e. based on the users interests the ranking of resources(items) will be done.

D. Maintaining stable and consistent Folksonomy Vocabulary

The vocabulary should be stable and consistent. In order to serve the purpose using WordNet or ConceptNet [11] will be better. The advantage of ConceptNet over WordNet is its integrated Natural language Processing Engine [11]. Thus using ConceptNet for storing the concepts along with the tags will maintain a unambiguous and stable vocabulary, as the concepts are stable and are well structured i.e. linked with the other concepts using the relations. The ConceptNet has 20 relations based on which it is related (or linked) with other concepts. The figure 2 will show the relations between the concepts eliminating the three k –line relations omitted.

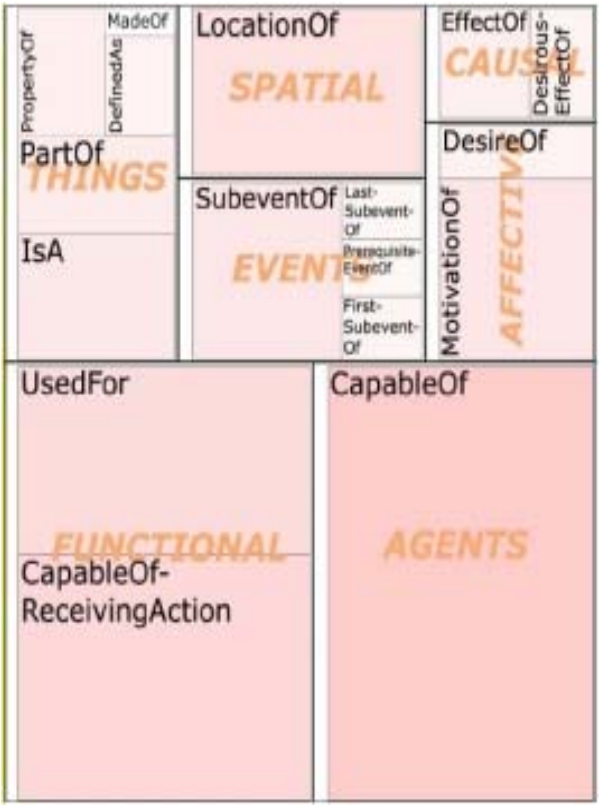

Figure 2. A Treemap of ConceptNet Relational Ontology (with the three K-line relations omitted)

E. Architecture

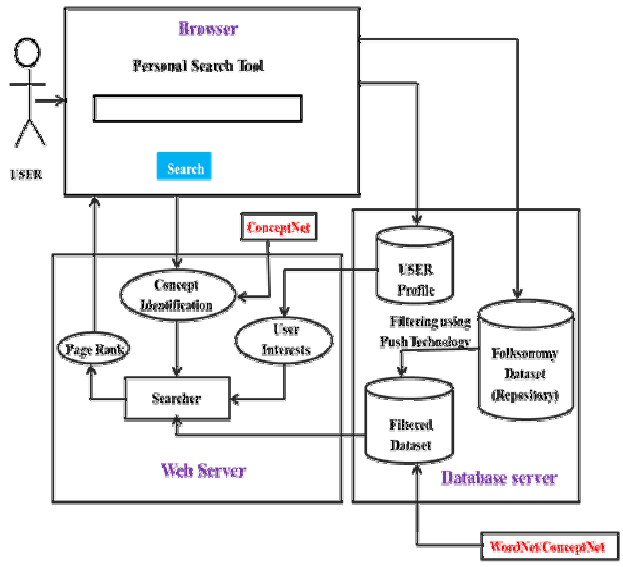

Figure 3. Architecture Diagram for Folksonomy based Search

The fig.3 represents the architecture of the system designed. The browser will provide the interface for the user for tagging as well as for searching a resource on the web. When users tag a resource on the web the data will be stored in a repository called folksonomy dataset. Push technology is used for the fast production of results i.e. whenever the repository is updated the filtered dataset will be notified about the update and hence it will update the data from the folksonomy dataset by filtering the tags using ConceptNet or WordNet. Thus stable and consistent vocabulary is maintained in the filtered dataset. If pull technology is used instead of push technology the data will be extracted by the filtered dataset every time it is called which is unnecessary. Now considering the query the user enters in the search box of the interface provided by the browser, the concepts are extracted from them using the ConceptNet. The user interests are identified based on the user profile. Thus based on the user interests, categories and the folksonomy vocabulary the searcher performs the job of searching for resources on the web and after that finally the page rank algorithm will rank the resources based on the user interests and will return to the browser which will display the results to the user.

IV. CONCLUSION

The system designed for the personalized search tries to suppress some problems faced by the folksonomy such as ambiguity and inconsistency using the ConceptNet. Because of the reduced problems the accuracy of the results produced

by the system will be improved, by identifying the concepts using the ConceptNet and then searching based on the user profile. Thus ConceptNet best serves the purpose of semantic search supporting Personalization.

V. FUTURE WORK

The system designed now needs an explicit user intervention i.e. a user should tag resources based on which his profile is created and personalization is done. Instead of this personalizing based on his/her historic data over the web will serve better as there is no need for the user to understand the concepts of tagging and Folksonomy as the personalization is done implicitly. Another work is to integrate the folksonomy structure in the taxonomy which will overcome some of the disadvantages of taxonomy such as updating the taxonomy and adding extra tags to the taxonomy, also some of the disadvantages of folksonomy such as unstructured and unambiguous tags. Thus the integration may help in maintaining a common vocabulary which is consistent.

REFERENCES

[1] www.google.com (Google)
[2] Search.yahoo.com (Yahoo)
[3] www.delicious.com (Delicious)
[4] Mehmet S Aktas, Mehmet A. Nacar, "An application of Personalized Page Rank Vectors: Personalized search Engine"
[5] J Trant, "Studying Social Tagging and Folksonomy: A review and Framework," Journal of Digital Information, Vol 10, No 1 (2009)
[6] Kipp, M.E., & Campbell, D.G., "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices" In American Society of Information Science & Technology Annual Conference 2006,Austin, Texas (US),3-8 November 2006, unpublished
[7] Stock, W.G., "Folksonomies and science communication. A mash-up of professional science databases and Web 2.0 services," Information Services & Use, 27, 97-103 (2007)
[8] Ted Petersen,Siddharth Patwardhan & Jason Michelizzi 'WordNet::Similarity – Measuring the Relatedness of Concepts," 2004 (references)
[9] Heung-Nam Kim, Majdi Rawashdeh, Abdullah Alghamdi, Abdulmotaleb El-Saddik, " Folksonomy-based personalized search and ranking in social media services," Inf. Syst. 37(1): 61-76 (2012).
[10] Mehmet S. Aktas, Mehmet A. Nacar, and Filippo Menczer, "An Application of Personalized PageRank Vectors: Personalized Search Engine," (references).
[11] Hugo Liu and Push singh : A Practical commonsense Reasoning Tool Kit, Media Laboratory, MIT University (references).

R.Sujatha received her B.E. degree in computer science from Madras Uninversity, in 2001, the M.E. degree in computer science from Anna University in 2009 and currently pursuing the Ph.D. degree in Vellore Institute of Technology, Vellore. She was a lecturer and currently assistant professor, with School of Information Technology and Engineering in Vellore Institute of Technology, Vellore. Her area of research interest include Data mining, Image Processing and Management of Information systems.

AUTHORS PROFILE

Rohith Krishna V, is currently pursuing his final year of MS(Software Engineering) at VIT University and is interested in the area of Data Mining mainly personalizing the search.