Personalized Web Search and User Profile Mining using Ontology

B.Umamaheswari
Dept. of Computer Engineering
D.Y.P.I.E.T
Pune, India

Pramod Patil
Dept. of Computer Engineering
D.Y.P.I.E.T
Pune, India

ABSTRACT
Ontology is a replica for knowledge portrayal and is used to articulate profile of user in personalized web information extraction. Earlier user profile was created based on either global knowledge base or local repository. This model extracts global knowledge from web crawler and combines it with user preference to obtain user background knowledge. Finally specificity is calculated based on semantic relations is-A and part-Of. The search results are more personalized and accurate. Hence this model is suggested for web information extraction.

General Terms
Information Retrieval, Data Mining.

Keywords
Ontology, personalization, user profile, global repository, user background knowledge

1. INTRODUCTION
The World Wide Web (WWW) is the first widely exploited many-to-many data interchange medium and due to its immense growth it poses new requirements for extraction of useful data. It also visage mismatch and overload problem. Initially search is keyword based rather than concept based. Currently, the web has been intended for direct dispensation, but the next generation web, the semantic web as mentioned by Tim Berners-Lee aims at machine-processable information. The semantic web [1] will enable intellectual services such as information agents, search engines and information filter which offer greater functionality and interoperability than current stand-alone services. It will be only possible once if further levels of interoperability have been established.

User profile infers semantic meaning of queries and incarcerates user information requirements. It symbolizes the concept model held by the user when gathering information from the web. Ontology [2] is used to simulate user profile. User background knowledge plays an important role in representing user profile [3] and it is extracted from either global knowledge base or local repository. Here in this paper knowledge is mined both global and local repository for better search results.

Ontology learning [4] is a task of information extraction. The goal of ontology learning is to semi-automatically extract relevant concepts and relations from data sets to form Ontology. Global knowledge bases [5] were used by many existing models to learn ontology for web information gathering. Many model mined user background knowledge from user local information.

Ontology [6] aims to represent the knowledge contained within software application, enterprises and business procedures for a particular domain. It aims to reuse of the domain knowledge. It separates the domain knowledge from the current databases

2. ARCHITECTURE
C denotes crawler which is used to generate global repository. By giving url as input to crawler all type of pages (.html, .htm, .asp etc) from the corresponding website will be retrieved and stored in the repository. G denotes global repository which is used for global search. Topic will be given as input. Keyword based search is done and result is retrieved.

O denotes ontology database which is generated by acquiring html pages from global repository. The html pages are processed to acquire the schema by removing the tags from the html pages and stored in sql server. Support of each pages is calculated. Ontology database is used for local or ontology based search. Topic will be given as input and result is retrieved based on support value.

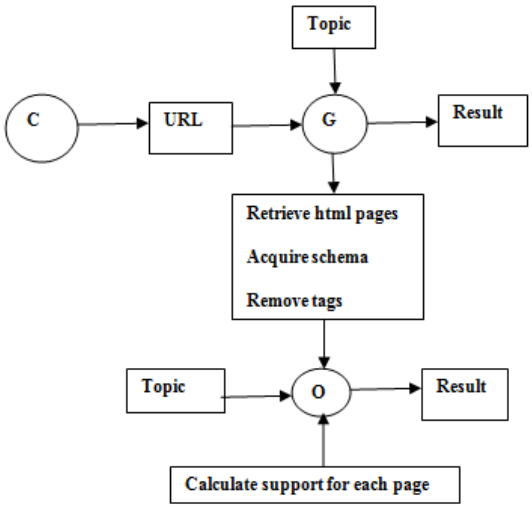


Figure 1: Architecture of proposed model

3. IMPLEMENTATION AND RESULTS
This section is providing performance results of ontology based search (mentioned as local search) compared to keyword based search (mentioned as global search). We are taking values for different number of searches. All searches are performed on certain type of domain. Here performance is calculated on basis of number of technical stuffs. Results are taken for domains such as movie, sports, tours and passion.

Any topic for the mentioned domain can be given as input to the proposed ontology model and the output was compared with keyword search. The global repository is constructed based on output of web crawler which runs in the background. The crawler is refreshed at regular intervals to update the global database. The information retrieved are first

preprocessed and then converted to human understandable format and stored in an SQL server database.

In order to make this ontology model [8] run more efficiently, only the newsworthy, communal, and location based topics were kept in the global repository. Then the user personalized ontology is built through user interaction. The aim is to show that the ontology based method produce more personalized results compared with keyword search. The global database is created with one lakh crawled pages covering different topics. The ontology database is created by feature extraction [10] (considering only html/htm).

3.1 Global Search

Any user without login can perform global search. The search is done on the entire database and result is shown. So it takes more time and search is keyword based. The figure below shows the global search performed by any user for the search word manali.

3.2 User Registration and Login

The user has to register for giving his preferences. Here four domains are considered. They are movies, sports, passion and tours. For each of four domain four preference are filled by the user in the registration form. The preferences will be recorded as local repository of the user. For making local search user has to login. Only registered user can login. The user preference will be retrieved whenever the user register him and do the login procedure.

3.3 Absolute Local Search

Absolute local search is ontology based search. Absolute local search is based on semantic relation is-A. Here all the user preference will be considered. Search result will be more personalized.

3.4 Relative Local Search

Relative local search is based on part-of semantic relation. Here not all user preference will be considered. Only part of them will be considered. Search result will be less personalized.

3.5 Calculating F1 Measure

The evaluation scheme, F1 Measure, is calculated by:

$$F1 = \frac{2 * precision * recall}{precision + recall}$$

The F1 measure is calculated by giving equal magnitude to precision and recall. The performance will be better if F1 measure values are greater. The F1 measure of local search is more than global search. The experimental results shown in the table below proves that ontology model gives better results.

Table 1. F1 Measure

Id	Topic	Precision	Recall	F1	Precision	Recall	F1
		Global Search			Local Search		
1	cricket	0.74	0.54	**0.62**	0.73	0.70	**0.72**
2	football	0.76	0.59	**0.66**	0.77	0.73	**0.75**
3	manali	0.72	0.52	**0.59**	0.74	0.69	**0.71**
4	kerala	0.69	0.48	**0.57**	0.76	0.74	**0.74**
5	salwar	0.74	0.54	**0.62**	0.69	0.65	**0.68**
6	jeans	0.68	0.48	**0.56**	0.73	0.71	**0.72**
7	salman khan	0.76	0.59	**0.66**	0.77	0.72	**0.75**
8	aamir khan	0.72	0.52	**0.59**	0.74	0.69	**0.71**
9	ipl matches	0.69	0.48	**0.57**	0.76	0.72	**0.74**
10	goa	0.76	0.59	**0.66**	0.69	0.66	**0.68**

3.6 Calculating Specificity

Input: A personalized ontology [4,5] O(T) := {tax^s, rel};a Coefficient a between (0,1).
Output: *spea(s)* applied to specificity.

Step.1:Set k=1, get the set of leaves So from tax^s, for s_o € S_o Assign *spea(*So) = k;
Step.2: Get S' which is the set of leaves in case we remove the nodes So and the related edges from tax^s;
Step.3: If (S' ==Φ)then return;// the terminal condition;
Step.4: foreach s' € S' do
Step.5; if(is A(s') ==Φ)) then *spea* ^{1}(S') = k;
Step.6 else *spea* ^{1} (S') = a * min { *spea* ^{1} (S) s € is A(s')};
Step.7 if (part of(s') ==Φ) then spe_a^2 (S') = k;
Step.8: else $spe_a^2(S') = \frac{\sum separtof(s')spea(s)}{|partof(s')|}$

Step.9: = min (*spea* ^{1} (S'), *spea* ^{2}(S'));
Step.10: end
Step.11: k = k * a, So = So U S', go to step.2.

The is A(s') and part of(s') are two functions in the algorithm satisfying is A(s') ∩ part of(s') =Θ. The is A(s') € tax^s returns a set of subjects s € tax^s that satisfy tax(s -> s') = True and type(s ->s') =is A. The part Of(s') returns a set of subjects s € tax^s that satisfy tax(s ->s') = True and type(s ->s') = part Of.

Table 2. Value of Specificity (belief)

Page id	Url	Description	Title	Domain	Belief
450	~/site/movie/…	Movie related	Aishwarya Rai	movie	2.7856
451	~/site/movie/…	Movie related	Aishwarya Rai	movie	2.5672
452	~/site/movie/…	Movie related	Aishwarya Rai	movie	2.6578
453	~/site/movie/…	Movie related	Aishwarya Rai	movie	2.7243
454	~/site/movie/…	Movie related	Aishwarya Rai	movie	2.1343
455	~/site/movie/…	Movie related	Aishwarya Rai	movie	2.2345

The specificity of a search is the proportion of results that matches the query string. However specificity (belief) is calculated and the search results are retrieved based on belief value. The performance of the proposed system is also analyzed by specificity. The following table shows the experimental result for search carried on various domains. From the result we conclude that ontology based model shows better performance.

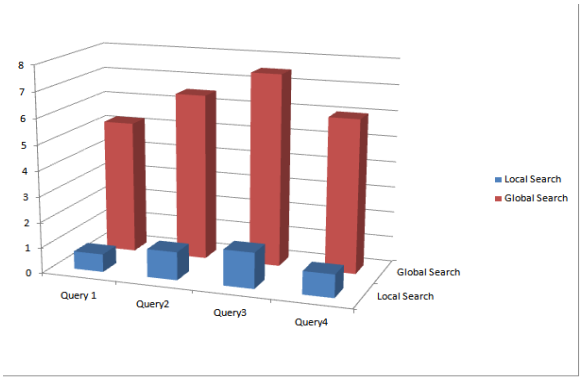
3.7 Comparison of Global vs. Local Search

The accuracy and relevance of both global and local search are determined. For any query global search takes more time to return search result. The local search returns result in less time. Hence ontology based web extraction is promising in terms of accuracy, relevance and results are obtained in less time.

Table 3. Comparison of Global vs. Local Search

S. No	Domain	Topic	Search Type	Req Time	Retrieved Result	No of Relevant Result
1	Tours	Manali	Global	7 min	16	3
2	Tours	Manali	Absolute local	40 sec	5	5
3	Tours	Manali	Relative local	40 sec	20	5
4	Sports	IPL matches	Global	10 min	13	3
5	Sports	IPL matches	Absolute local	40 sec	2	2
6	Sports	IPL matches	Relative local	1 min	80	20

Finally a graph is plotted showing comparison of global and local search by taking x coordinate as number of queries and y as time in seconds.



X: coordinate show the number of queries fired by user

Y: Required time in minutes

Figure 2: Comparison of Global vs. Local Search

4. CONCLUSION

A personalized ontology model is proposed to combine global knowledge and local knowledge for specific domain. We consider domain such as fashion, sports, movies and tours. The model extracts global knowledge from a web crawler and constructs ontology database through feature extraction. User preferences are also considered. Finally personalized search is performed from ontology database. Support for each page is also calculated for more personalized results.

Usage of automated tool for saving the profile of user in ontological format. This helps the user to view the personalized ontology. Integration of the global knowledge base like LCSH with the application that enable the integration of all of the related classes so that the user can give any type of string in the search process.

5. REFERENCES

[1] Stefan Decker. et. al "The Semantic Web: The Roles of XML and RDF" http://computer.org/internet/ 1089-7801/00/$10.00 c 2000 IEEE

[2] Fensel, Dieter; Michael L. Brodie., "Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce", Second Edition, December, 2003. Springer-Verlag.

[3] Y. Li and N. Zhong, "Web Mining Model and Its Applications for Information Gathering," Knowledge-Based Systems, vol. 17, pp. 207-217, 2004.

[4] Xiaohui Tao, "A Knowledge-based Model Using Ontologies for Personalized Web Information Gathering," 1570-1263/10/$17.00.

[5] P.Devisree and P. Revathi, "Web Search Personalization with Ontological User Profiles and World Knowledge Base" International Conference on Computing and Control Engineering (ICCCE 2012), ISBN 978-1-4675-2248-9.

[6] Bhaganagare Ravishankar C, Dharmadhikari Dipa D "Web Personalization Using Ontology: A Survey" ISSN : 2278-0661 Volume 1, Issue 3(May-June 2012), PP 37-45.

[7] B.Hima Bindu, I. Sandhya and Dr. P. Raja Prakasha Rao "Novel computation model for emphasizing global and local knowledge using Ontology" ISSN NO : 66023127 IJART, Volume 3, Issue 1 August 2012), PP 29-34.

[8] Sakthi Priya T, Revathy P, Pradeesh T, Rene Robin C.R, "Design and Development of an ontology based personal web search engine" International Conference on Communication, Computing and Security (ICCCS-2012], Procedia Technology6 (2012), 299-306.

[9] B.Umamaheswari, Pramod Patil, "Personalized Ontology Model - A Survey" International Conference on Hybrid Intelligent System 978-1-4673-5116-4/12/31.00 2012 IEEE.

[10] S.Vigneshwari, M. Aramudhan, "A Technique to ontology Mining for Semantic Web Information Extraction", ISSN 1450-216X vol.94 no.1, 2013, pp. 49-60.