Global Journal of Computer Science and Technology: E
Network, Web & Security
Volume 14 Issue 2 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

# Design and Implementation of Information Retrieval using Ontology

By Dr. V. S. Dhaka & Aditi Sharma
*Jaipur National University , India*

*Abstract-* An approach is proposed that can be used to make these arch adaptive according to each user's need using ontology .Our approach is distinct because it allows each user to perform more fine-grained search by capturing changes of each user's preferences without any user effort. Such a method is not performed in typical search engines.

*Keywords:* *ontology, fine-grained, user's preferences, search engines.*

*GJCST-E Classification :* *H.3.3*


*Strictly as per the compliance and regulations of:*

# Design and Implementation of Information Retrieval using Ontology

Dr. V. S. Dhaka [α] & Aditi Sharma [σ]

*Abstract-* An approach is proposed that can be used to make these arch adaptive according to each user's need using ontology .Our approach is distinct because it allows each user to perform more fine-grained search by capturing changes of each user's preferences without any user effort. Such a method is not performed in typical search engines.

*Keywords: ontology, fine-grained, user's preferences, search engines.*

## I. Introduction

Over the past few years, the World Wide Web (WWW) has become the largest and most admired means of communication and dissemination of information. Users often feel disoriented and lost in this information overload which continues to develop. Therefore, the ultimate need nowaday is that of predicting the user needs in order to improve the usability and approaches to make the search adapting to satisfy the user requirement user retention of a web site. We propose. The approaches conversed here are derived from ontology and active user profile. The presented approach aims to effectively personalize search results according to each user's information need by accurately identifying the user context, updating user profile timely, recommending documents according to similar users and by reorganizing the information satisfying the needs. The Web pages are customized according to the characteristics (interests, the social category, the context) of an individual. Personalization technology enables the lively insertion, customization or hint of content in any format that is pertinent to the individual user, based on the user's implicit actions and inclinations.

### *a) Dynamic User Profile and Ontology*

Personalization needs user profile and to construct a user profile, some basis of information concerning the user required to be collected. This information may be collected explicitly and implicitly. Explicit profile creation is not preferred as it puts an additional saddle on the user. Additional issues related to explicit profile creation are the user may not accurately report their interests; the profile, so created, remains inert while the user's interests may keep altering interests may keep altering over time. Hence, the user needs to update the profile. Implicit profile building based on observations of the user's actions describes model considers the frequency of visits to a page, the amount of time spent on the page, how recently a page was visited and whether or not the page was book marked. A set of m finite number of users is termed as U. An ith user(ui)is indicated as a person who poses the question /query to search engine through web browser.

NewUserisauserwhoposesthequeryfirsttimeusingtheemployedsearchengine. Newuserset

$$NU \subseteq U;$$

OldUseristheuserwhohascreatedthequerypreviouslyonthesearchengine.

$$\text{Hence} \quad OU \subseteq U;$$

ActiveUser(denotedasa)istheuserwhoiscurrentlyworking;solivelyuser,attime,iseitherafreshuseroranoldusr

$$ui \in U\{ui:1 \leq i \leq m\}$$

$$\text{and} \quad U = OU \cup NU$$

Query Topic (denoted as QT) is a search query that comprises of one or more keywords/ terms. extent/ dimension of query are number of terms present in it. New Query is a query created by the user firstly. Old Query is a query that has previously been searched by a user. W(u,j)is weight given to the jthquery topic for the user u.

Context is the description of a user's aim / need for information reclamation. In this chapter, context is implicitly defined which are update do over time to reproduce changes in user interests/requirements. Contexts are extracted from Word Net in terms of concepts.

## II. The Proposed Approach

In this information age, it is a deplorable state that, despite the information overload, we fail regularly to identify relevant information. In particular, in the field of education, several terabytes of content related to various educational institutions such as universities, colleges are downloaded from the Internet every week, and the demand for these resources is still rising. But this is not satisfactory in terms of access to information that the generic search engine in terms of overtime on bad links and relevance links. There can be many reasons, the most important in terms of lack of

*Author α σ : Department of Computer Science Jaipur National University , Jaipur. e-mail: vijaypal.dhaka@gmail.com, ksaditi2@gmail.com*

recognition of context and semantics of the user query to get the required results.

To address these critical issues of information retrieval, the proposed system is designed. The proposed system retrieves semantically relevant results for the user account application semantics and context of the request. The semantics of the query is analyzed using the following procedures:

- The user's request is first analyzed and syntactically by the analysis.
- The synsets related to key words in the query are retrieved .
- The keywords of the ontology of domain are collected to form the refined query.

The results obtained in the proposed approach are more relevant by adopting the following procedure:

- "The refined queries which are entries in the search engine are formed on the basis of the semantic analysis on user request.
- "The Web links retrieved for all the refined queries newly formed are again classified according to the information specific to a domain.

The low-level design of our proposed system is demonstrated as follows:

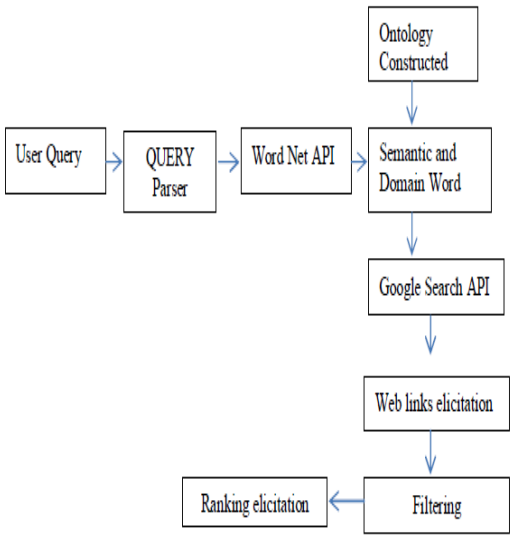


*Figure 1 :* Low Level Design of Proposed System

*a) High Level Design*

The three major components of the proposed System are as follows Ontology Construction

- Refined query formation
- Ranking of retrieved links

*i. Ontology Construction*

Elementary knowledge that the main body of this component description forms, institutional construction of suggestion. Other spheres of learning and organizations of related concept, gathered from various websites and other origins, such as Word only. These concepts centralized in a stratified form in the foundation territory related keyword of ontology. These key words are used to train the purification inquiry.

*ii. Refined Query Formation*

Improvement, to provide better search result, uses this module the inquiry that is assigned by the user. In this part, the inquiry analysis that is assigned by the user, the speech recognition part of inquiry words and expressions. Then, about in keyword the retrieval of synonym collection in the inquiry contains. The key words territory, the semantic query related extract completes from the main body. This step will cause the more semantic related words the restoration of quantity. Then is used in the open country training purification inquiry these key words. These inquiry fine inquiries, the key words expand, have the related semantics of involving.

*iii. Modules*

- Query parsing
- Synsets retrieval
- Keywords extraction
- Refined query formation

*b) Filtering and Ranking of Retrieved Links*

The collaborative filtering is a universal Web technology produces the personalized suggestion. Example of the use includes iTunes, Netflix Corporation in Amazon, lastfm, Stumble Upon Corporation, with Delicious.

Collaborative filtering is a technology utilized chiefly to predict individuals' inclinations. The initiative of collaborative filtering has its basis in information filtering, which leads a reader's pick by filtering a large amount of information and obtaining inclinations collaboratively based on inclinations shared by like readers.

Collaborative filtering works by first sifting through an individual's inclinations or purchase history to find a group of individuals, or a 'neighborhood', with similar inclinations or purchase histories, and then envisaging what else the individual will like, based on the collective inclinations or purchase histories of other individuals in the neighborhood. The predicted inclinations can then be used to make product or service recommendations to the individual.

*i. Strengths*

- Intuitive, easy to comprehend and implement.
- No data structure assumptions.

*ii. Weaknesses*

- Requires a large sample to make meaningful recommendations.
- Erroneous recommendations result when close neighbors don't exist.

- Direct insights into the drivers of the exhibited inclinations are difficult to derive.
- Does not unswervingly use product or item content information and customer profile or behavior information for making recommendations.

*iii. Modules:*

- Retrieval of web links
- Filtering and Ranking of retrieved links

## III. Results

*a) Parsing of The Input Query*

The query given by the user is parsed by m n sea of query parser and the output is:

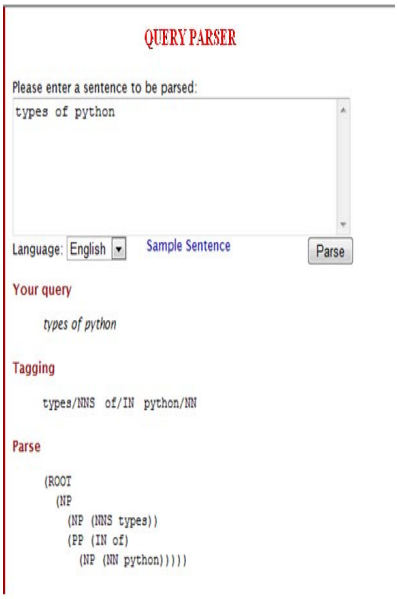


*Figure 2 :* Query Parser Results Screenshot For Query

*b) Retrieval of Synsets from Wordnet*

Now the related synsets for the words present in the query are retrieved from the wordnet.

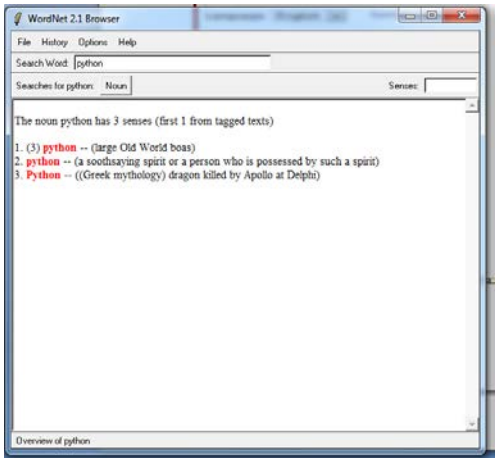


*Figure 3 :* Wordnet Results Screenshot for Query

*c) Extraction of Domain Keywords from Ontology:*

The domain keywords that are semantically related to the words in the query are extracted from ontology.

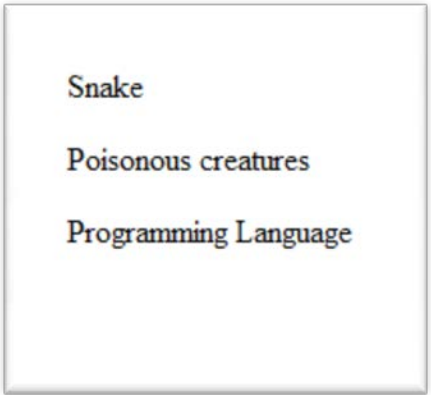


*Figure 4 :* Jena Api Results Screenshot for Query

The query 'Python' will be expanded with" programming language ",for the users fascinated in computer programming language, and with" snake ',for the users fascinated in "wild life". To get the appropriate context of query topic, the Word Net is used to retrieve appropriate context using the following algorithm and the user profile is updated accordingly.

*d) Web Links Retrieved:*

*i. User Query:*

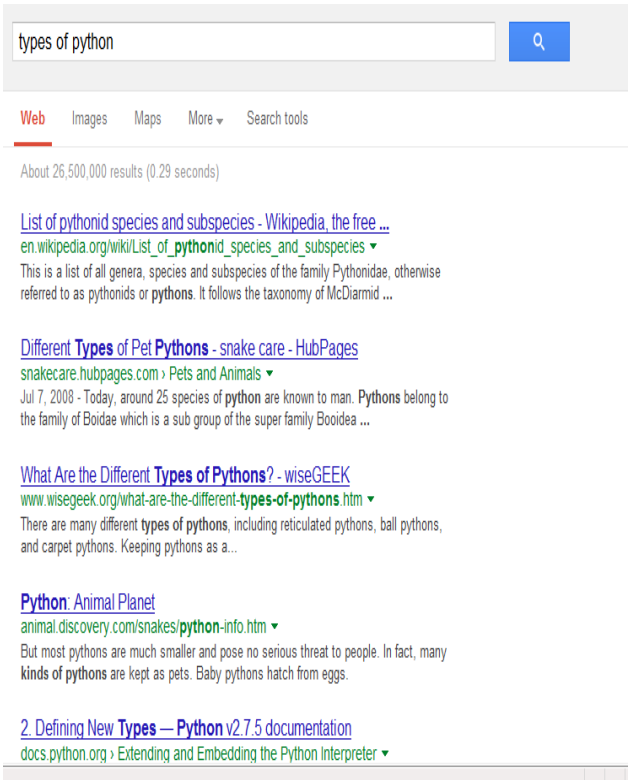


*Figure 5 :* Google Results Screenshot for Query

*ii. With Refined Query*

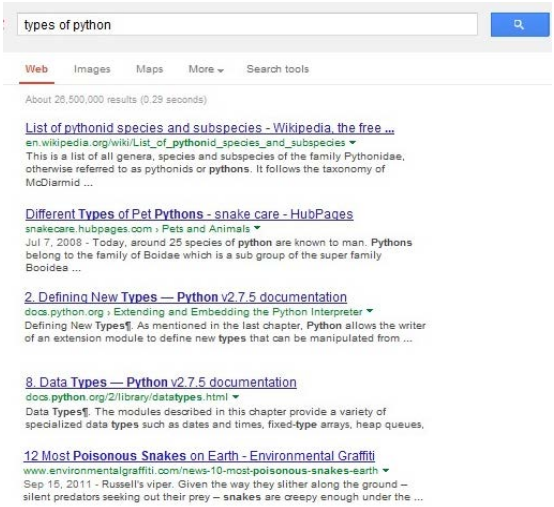


*Figure 6 :* Refined Query Results Screenshot

## IV. Summary and Conclusion

The design and implementation of the proposed approach using Dynamic User profile and Ontology.

The experiments designed are first discussed, followed by the experiment frame work and environment. The overview of the proposed system. In addition ,it gives details of the query parser tool and implemented for query expansion using ontology and re-ranking of documents with using user profile. Evaluation of Context aware applications is quite difficult as they depend on context. The contexts or situations of interest depends on user to user and can't be generalized.

## References Références Referencias


1. Abecker A., Bernardi A. Hinkelmann K. Kühn, O. &Sintek M. (1998): Toward a Technology for Organizational Memories, IEEE Intelligent Systems, 13, May/June, p.40-48.
2. BerthierRibeiro-Neto, Ricardo Baeza-Yates. Modern Information Retrieval. ACM Press, New York, 1999.
3. Borghoff U. M. &Pareschi R. Eds. (1998) Information Technology for Knowledge Management. Springer Verlag.
4. S. Dumais. Improving the retrieval of information from external sources. Behavior Research Methods, Instruments, and Computers, 23(2):229–236, 1991.
5. Newell A. (1982) The Knowledge Level. Artificial Intelligence, 18, p. 87-127.
6. Paralic, J., Sabol, T. & Mach, M. (2002) A System to support E-Democracy. Proc. of the First International Conference EGOV 2002, Aix-en-Provence, France, LNCS 2456, LNCS 2456, Electronic Government, R. Traunmuller, K. Lenk (Eds.), Springer Verlag.
7. Tiwana A. (2000) The Knowledge Management Toolkit. Prentice Hall.
8. Scott C. Deerwester, Susan T. Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. Indexing by latent semantic analysis. Journal of the American Society of Information Science, 41(6):391–407, 1990.
9. S. Adali, K. S. Candan, S. Chen, K. Erol, and V. S. Subrahmanian, "Advanced Video Information System: Data Structures and Query Processing," ACM-Springer Multimedia Systems Journal, vol. 4, pp. 172-186, 1996.
10. B. Arons, "SpeechSkimmer: Interactively Skimming Recorded Speech," in Proc. of ACM Symposium on User Interface Software and Technology, pp. 187-196, Nov 1993.
11. G. Aslan and D. McLeod, "Semantic Heterogeneity Resolution in Federated Database by Metadata Implantation and Stepwise Evolution," The VLDB Journal, the International Journal on Very Large Databases, vol. 18, no. 2, Oct 1999.
12. R. Baeza and B. Neto, Modern Information Retrieval, ACM Press New York, Addison Wesley, 1999.
13. M. Bunge, Treatise on basic Philosophy, Ontology I: The Furniture of the World, vol. 3, Reidel Publishing Co., Boston, 1977.
14. S. Gibbs, C. Breitender, and D. Tsichritzis, "Data Modeling of Time based Media," in Proc. of ACM SIGMOD, pp. 91-102, 1994, Minneapolis, USA.
15. J. Gonzalo, F. Verdejo, I. Chugur, and J. Cigarran, "Indexing with WordNetSynsets can Improve Text Retrieval," in Proc. of the Coling-ACL'98 Workshop: Usage of WordNet in Natural Language Processing Systems, pp. 38-44, August 1998.
16. T. R. Gruber, "A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition," An International Journal of Knowledge Acquisition for Knowledge-based Systems, vol. 5, no. 2, June 1993.
17. N. Guarino, C. Masolo, and G. Vetere, "OntoSeek: Content-based Access to the Web," IEEE Intelligent Systems, vol. 14, no. 3, pp. 70-80, 1999.
18. G. Hauptmann, "Speech Recognition in the Informedia Digital Video Library: Uses and Limitations," in Proc. of the Seventh IEEE International Conference on Tools with AI, Washington, DC, Nov 1995.
19. R. Hjelsvold and R. Midstraum, "Modeling and Querying Video Data," in Proc. of the Twentieth International Conference on Very Large Databases (VLDB'94), pp. 686-694, Santiago, Chile, 1994.
20. L. Khan and D. McLeod, "Audio Structuring and Personalized R Retrieval Using Ontologies," in Proc. of IEEE Advances in Digital Libraries, Library of Congress, pp. 116-126, Bethesda, MD, May 2000.
21. L. Khan and D. McLeod, "Effective Retrieval of Audio Information from Annotated Text Using

Ontologies," in Proc. of ACM SIGKDD Workshop on Multimedia Data Mining, Boston, MA, pp. 37-45, August 2000.
22. Bhogal J, Macfarlane A and Smith P, "A review of ontology based query expansion", Science Direct. Information Processing and Management, vol. 43, 2007, pp. 866-886.
23. Mittal, N., Nayak, R. Govil, M.C. ;Jain, K.C. "A Hybrid Approach of Personalized Web Information Retrieval", Web Intelligence and Intelligent Agent Technology (WI-IAT), Vol.1 , 2010, pp. 308 – 313.

This page is intentionally left blank