International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

# Personalized Web Search using Ontology and Page Ranking

**A. N. R. Latha Kumari**

Department of CSE, ANITS, Sangivalasa, India

**Abstract:** ***Web Users are always interested to search various types of information in the web. While searching the World Wide Web for particular Queries, users usually get irrelevant and redundant information causing a waste in user time. Providing more relevant information to the users based upon user's interests and needs from their behavior have become increasingly important.. Web mining is one of the techniques that could help the User's to retrieve data based upon their interests. In this paper, ontology is proposed to create user interest, and to provide more relevant information to the Query, Web pages are ranking using Page Ranking algorithms.***

**Keywords:** Ontology, Web crawler, Page Rank, Web Graph, Personalized search, Information Retrieval, User Profiles

## 1. Introduction

In recent decades, the amount of Web information has exploded rapidly. Gathering useful information from the Web, becomes challenging to Web users. The current Web searching systems cannot satisfy Web users, as they are mostly based on keyword-matching mechanisms and suffer from the problems of data mismatching and information overloading. Keyword-based searching systems can access the information quickly; however, the gathered information may possibly contain much useless and meaningless information. This is particularly referred as the fundamental issue in searching. Capturing user information needs through a given query is extremely difficult.

The capture of user information needs requires the understandings of user's personal interests and preferences. For this purpose, user profiles are created for user Interest domains. User profiles represent the concept models possessed by users when gathering web information. To create user profile or user concept model ontologies are utilized in personalized searching. Such ontologies are called ontological user profiles or personalized ontologies.

Search engines play an important role in searching web pages. The search engine gathers, analyzes, organizes the data from the internet and produces result to the user .The major components of a Search Engine are the Crawler, Indexer, Query Processor. A Web Crawler is a relatively simple, automated program, or script that methodically scans or ―crawls through internet pages to create an index of the data it is looking for. The Search Engine which uses this general web Crawler returns links. It may return millions of pages in response to a query and user interests. It is not possible for a user to preview all the returned result set. So search engine makes use of ranking algorithm to display the resultant pages in a ranked order using page ranking algorithms. The search engines on the Web need to be more efficient because there are extremely large number of Web pages as well queries submitted to it.

## 2. Personalized Ontologies

Creation of User profile acquisition techniques can be categorized into three groups: the interviewing, non-interviewing, and semi-interviewing techniques. The interviewing user profiles are entirely attained using manual approach; such as questionnaires, interviews. Users read documents and assign positive or negative judgments to the documents against given topics Based upon the assumption that users know their interests and preferences exactly. However, this kind of User profile acquisition mechanism is costly. The semi interviewing techniques involves limited user involvement. The users were given set of topics or domains from that user has to specify the interesting and non interesting topics. Based upon the feedback and user activity and behavior, user profiles were created. The non-interviewing techniques do not involve users directly but ascertain user interests instead. Such user profiles are usually acquired by observing and mining knowledge from user activity and behavior. In this paper semi interviewing method is used to create user profiles.

User Profiles can be represented using Ontologies. An ontology formally represents knowledge as a set of concepts within a domain, and the relationships between those concepts. It can be used to reason about the entities within that domain and may be used to describe the domain.

## 3. Developing an Ontology Includes

1) Defining classes(domain names) in the ontology.
2) Arranging the classes in a taxonomic (subclass–super class) hierarchy.
3) Defining relations and describing allowed values for these relations.

## 4. Formation of Personalized Ontology

The users may come from different backgrounds. Ontology contains three types of references: In this paper, they are encoded as the *is-a, part-of, Relate-to* relations in the Ontology. *is-a* are used for many semantic situations, including broadening the semantic extent of a subject and describing compound subjects and subjects subdivided by other topics.*Part-of* relation is used to define an action or an object. When object is used for an action, object becomes a part of that action (e.g., "a key is used for opening a locker

so key is *part-of* locker"); when object1 is used for another object 2, 1 becomes apart of 2 (e.g., "a wheel(object1) is used for a car(object2) so wheel is *part-of* car "). These cases can be encoded as the part-of relations. *Related-to* are for two subjects related in some manner other than by hierarchy. They are encoded as the related-to relations in our world knowledge base.

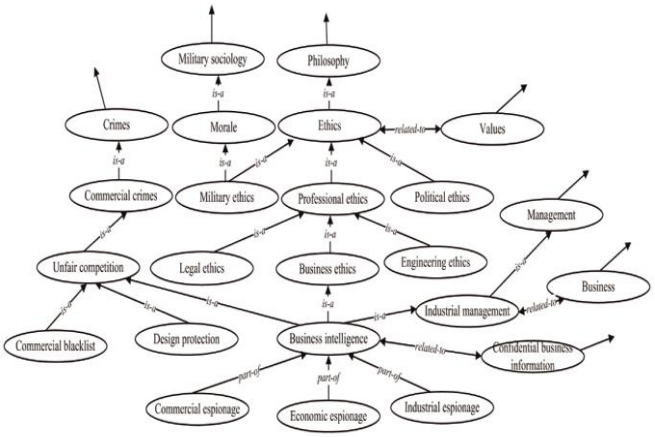


**Figure 1.1:** Ontology with *is-a, part-of, related-to*

## 5. Ontology based Web Crawler

In Ontology Based Web Crawler the web pages are first checked for validity (i.e. of the type html, php, jsp etc). If it is valid then it is parsed and the parsed content is matched with the ontology. If the page is relevant it is indexed otherwise it is not considered. The algorithm is as follows:

1) Get the seed URL.
2) If the web page is valid i.e(if it is of the defined type html, php, jsp etc.) then it is added to queue.
3) Parse the content of web page.
4) Get the response from the server if it is ok then read the protégé file of ontology. And match the content of web page with the terms of ontology. Return millions of pages in response to a query.
5) Add's the web page to index and caches file to a folder. With the help of cache and index searching can be done.
6) Search engine makes use of Page ranking algorithm to display the resultant pages in a ranked order .

## 6. Page Rank algorithm

The Page Rank algorithm is used to rank web pages. The page rank depends upon link structure of the web pages. Link structure of web may be viewed as directed graph and it is called as web graph. In web graph G(V,E), Nodes(V) represent web pages and Edges(E) associated with hyperlinks. Page Rank importance of web page by simply counting the number of pages that are linking to it. These links are called as back links(fig 1.1). If a back link comes from an "important" page(highest page rank), then that back link is given a higher weighting than those back links comes from non-important pages.

*Assume page A has pages p1...pn which point to it (i.e., back links). The PageRank of a page A PR(A) is given as follows:*
***PR(A) = (1-d) + d (PR(p1)/C(p1) + ... + PR(pn)/C(pn))***

- PR(A) is the Page Rank of page A,
- PR(Pi) is the Page Rank of pages Pi which link to page A,
- C(Pi) is the number of outgoing links on page Pi,
- d is a damping factor which can be set between 0 and 1.

The following steps explain the method for implementing Page Rank Algorithm.
Step 1: Initialize the rank value of each page by 1/n. Where n is total no. of pages to be ranked. Suppose we represent these n pages by an Array of n elements. Then A[i] = 1/n where 0≤ i< n

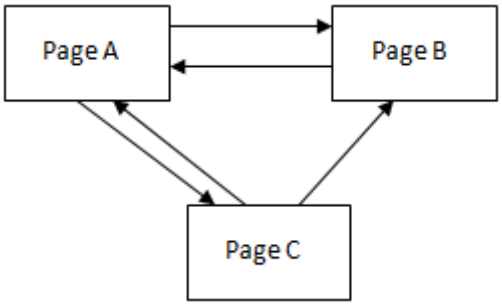


**Figure 1.2:** Web Graph with three pages

An Example of back link is page A is back link of B and C. and B is back link of A. C is back link of A and B

For Fig 1.2 page A has two outgoing link , Page B has one outgoing link , page C has two outgoing link (i.e. C(A) = 2 , C(B) = 1 and C(c)=2).Assume the initial Page ranks of all pages is one.PR(A)=1,PR(B)=1, PR(C)=1. And damping factor (d)=0.85.
PR(A)=(1-d)+d(PR(B)/C(B)+PR(C)/C(C))
PR(B)=(1-d)+d(PR(A)/C(A)+PR(C)/C(C))
PR(C)=(1-d)+d(PR(A)/C(A))
**First iteration:**
PR(A)=(1-0.85)+0.85(1/1+1/2)=2.5
PR(B) )=(1-0.85)+0.85(1/2+1/2)=1
PR(C)=(1-0.85)+0.85(1/2)=0.57
**Second Iteration:**
PR(A)=(1-0.85)+0.85(1/1+0.57/2)=2.592
PR(B)=(1-0.85)+0.85(2.5/2+0.57/2)=1.454
PR(C)=(1-0.85)+0.85(2.5/2)=1.212

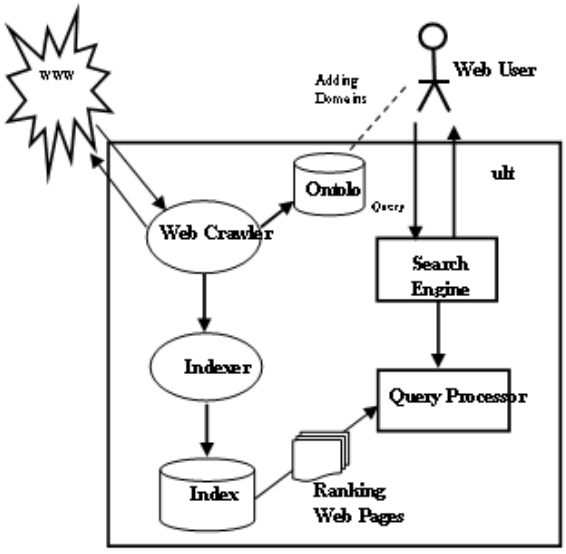


**Figure 1.3:** Architecture of Personalized Search

## 7. Conclusion

The main aim of our paper is to retrieve relevant Web pages and discards the irrelevant ones. We have developed an ontology to represent user profiles and ontology based crawler which retrieves Web pages according to user interests. The returned Result pages are ranked using page rank algorithm to calculate a relevance of web pages for the given query and discards the irrelevant Web pages. In this we have use the concept of Ontology which provides the meaning of terms and relationship between them. We believe that our crawler will not only be helpful in exploiting fewer web pages such that only relevant pages are retrieved but also will be an important component for the future Semantic Web which is going to become very popular in the years to come. Hence, such an improved crawler suggested by us in this paper can help in applications areas like Social Networking Portal, Online Library for Books Information etc. and can add to the benefits of them in their respective fields.

## References

[1] Raman Kumar Goyal1, Vikas Gupta2, Vipul Sharma3, Pardeep Mittal4, ―*Ontology Based Web Retrieval*‖, 1Lecturer (Information Technology), RIEIT, Railmajra, 2AP (CSE), RIEIT, Railmajra,3Student, UIET, Panjab University, Chandigarh, 4AP (CSE), BFCET, Bathinda.
[2] Felix Van de Maele, ―*Ontology-Based Crawler for the Semantic Web*‖, Faculty of Science, Department of Applied Computer Science, Vrije Universiteit Brussel, May 2006
[3] Marc Ehrig, Alexander Maedche, ―*Ontology Focused*
[4] Raymond Kosala & Hendrik Blockeel. Web Mining Research: A Survey. ACM SIGKDD, July 2000.
[5] An Intelligent Model for Redesigning Websites using Web Mining Techniques
[6] J. Hou and Y. Zhang, "Effectively Finding Relevant Web Pages from Linkage Information", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 15, No. 4, 2003.

.