SEMANTIC NET BASED WEB SEARCHING USING ONTOLOGIES

PRIYANKA SHARMA, EAKANSH MANGLIK, ALOK KUMAR, DR. SANJEEV SHARMA

Abstract: Focused Crawling aims at matching the actual information required by the user by retrieving documents relevant to predefined topic. To intelligently fetch the most relevant results many models had been proposed by various researchers, however as the demand of the users keeps changing, the ways to extract desired information should be changed to meet such demands. This paper aims at achieving more relevant results that too in reduced time by making the use of knowledge present in semantic web. A model is proposed that breaks the traditional search into two steps, where first step refines the search keywords and second step retrieves the focused result in order of their relevancy.

Keywords: ontology; semantic; searching; internet; topic-map.

Introduction: Internet is the largest, multi-disciplinary source of knowledge over the planet. Due to its huge size and increasing search needs day-by-day, it's becoming difficult to search the appropriate result instantly. As the major portion of the web repository is unstructured, it's difficult to inference, relate or represent knowledge in multiple formats. Humans by nature, are good at image processing rather than numerical evaluation. We propose a model, which will transform the primitive searching, mostly into graphical searching with the help of semantic network. We'll use ontology, to store the relationship between the words and semantic network to represent this knowledge in the form of network (inter-connection of nodes and edges, where nodes will represent the word & edges will re represent the type-of relationship between them).

In this paper, we construct/update ontology in parallel at the time of crawling, to store sematic relationships between different words. A methodology of web searching using semantic network as the backbone, to support focused searching and idea of pruning semantic network to generate a specific view with the help of topic maps is influenced from ontology modeling.

Background:

Sematic Network: Semantic network is a graph where vertices represent concepts and edges represent relations between concepts. Semantic network S at the level of ontology, represents vocabulary that is especially helpful for focused searching and fast processing. There are following kinds of relations that can exist between the concepts that are used in a semantic network. [8]

- Synonym: means concept A express something as concept B.
- Antonym: means concept A expresses opposite as concept B.
- Meronym, holonym: means expressing part-of or has-part relation between two concepts.
- Hyponym, hypernym: inclusion of semantic range (context) between the concepts in both the directions.

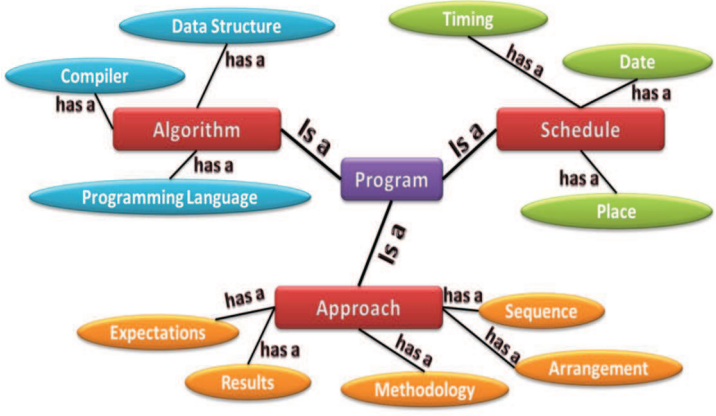


Fig. 1: Semantic Network

A semantic network, is a graph of the structure of meaning. Specifically, "it is a graphical notation for representing knowledge, in patterns of interconnected nodes and arcs". The nodes represent the concepts and the arcs are the interrelationship between every two nodes. It provides a convenient approach to visualize a knowledge base. Semantic network has been applied for many ontology development projects. It is believed that semantic network is the most appropriate representation method for capturing and encapsulating the massive amounts of semantic information in an intelligent environment. [4]

There are basically two kinds of semantic networks:

- Executable semantic net: can perform inferences and search for patterns and associations.
- Learning semantic net: build and extend their representation by acquiring knowledge.

Our proposed model, is a concept of "hybrid semantic networks" which contains both of the above properties.

Topic Map: Topic maps are syntactically standardized forms of semantic networks. They allow using topics (concepts), associations (relations) between topics including specification of roles of topic in various associations and occurrences (resources relevant to topic, i.e. context). By using topic map, we can create ontology of a topic, and at the time of searching, we can focus our results by choosing a specific topic map from stored ontology, that is most relevant to user's query. Topic Maps are similar to concept maps and mind maps in many respects, though only Topic Maps are standardized.

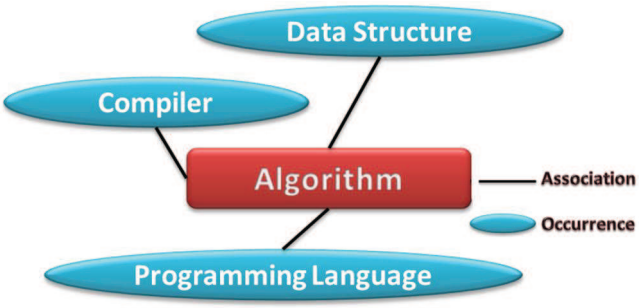


Fig. 2: Topic Map

However, with a little additional generalization, we can create a meta-model with potentially far wider application.

- topics, representing any concept, from people, countries, and organizations to software modules, individual files, and events,
- associations, representing hyper graph relationships between topics, and
- Occurrences, representing information resources relevant to a particular topic.

The semantic expressivity of Topic Maps is, in many ways, equivalent to that of Resource Description Framework (RDF), but the major differences are that Topic Maps (i) provide a higher level of semantic abstraction & (ii) allow n-ary relationships (hyper graphs) between any numbers of nodes, while RDF is limited to triplets.

Ontology: Ontologies are formal representation of knowledge within a domain. Ontologies provide a structural framework for organizing information and are used in artificial intelligence, semantic web, systems engineering, software engineering, biomedical informatics and information architecture as a form of knowledge representation about the world. In the context of knowledge sharing, we can define term ontology to mean a specification of a conceptualization i.e., ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. [6]

Related Work: "*Product Ontology Construction*

***from Engineering Documents*", J. M. Park, J. H. Nam, Q. P. Hu & H. W. Suh**: This paper describes the idea of developing the ontology in a bottom-up style manner and designs a specific format of ontology known as "DocOnto". The methodology is mainly composed of three phases such as defining ontology for terms in engineering documents, integrating the ontology with semantic networks for both a single document and a focused document group, and pruning the ontology for practically usage. In this approach, first-order logic (FOL) and semantic networks (SN) are used for formal and visual representation of ontology, and semantic mapping with similarity evaluation is used in integrating ontology.

"***Ontology driven Pre and Post Ranking based Information Retrieval in Web Search Engines*", Parul Gupta & A. K. Sharma:** This paper proposes, an ontology driven pre ranking of the documents with identical context and hence post ranking of the search results using keyword matching of the expanded query terms and document keywords in the pre-ranked search results. The idea makes the search more relevant and results in fast retrieval of documents. The extracted rank is stored in the context based index and hence the name is ranked context based index. The ontology repository can be used to extract the keywords in the domain and the sub domain of a particular context.

"***Concept based Focused Crawling using Ontology*", S.Thenmalar & T. V. Geetha:** The paper proposes and extends the idea of focused crawling by prioritizing the queue of URLs downloaded by the focused crawler. This idea is supported by designing a conceptualized vector which is obtained by combining concept vectors of individual pages associated with seed URLs. The conceptual rank is based on comparison between conceptual vectors at each depth, across depths and between the overall topics indicating seed concept vector.

"***Semantic Web to E-Learning Content*", T.Sheeba, S.Hameetha Begum & M. Justin Bernard:** The paper focuses on the idea of ontology development in a conceptualized manner so that it can serve as a basis which is highly suitable for establishing an appropriate educational technology system. This idea is supported by the thought that ontology is an important component in instructional design and its semantics can be applied to learning contents.

"***Using Topic Maps for web based education*", C. Dichev, D. Dicheva & L.Aroyo:** Topic Maps organizes and retrieves online information in an organized web-based manner. The paper considers Topic Maps to be a standard representation that encodes expert domains and instructional knowledge to build a conceptualized ontology. Also by using topic maps, further reuse, sharing and interoperability of knowledge structures gets developed.

Proposed Work: A critical analysis of the available search methodology reveals that there is one common thing on which every search technique relies and that is the matching process of keywords appearing in the web documents. This conventional processing leads to the most common problems of searching viz. Polysemy and Synonymy.

The existing work needs to include the following issues–

- The search engine must cover the two major aspects of information retrieval i.e. Context (Meaning) and Domain (Range of meanings) simultaneously. [6]
- Context refers to the actual meaning of word to be searched by a user. For example, the term bank as their query may refer to a financial center, river bank or a collection of objects. [11]
- Domain refers to multiple meanings of the same word, such as car and automobile both come under the category of vehicle class. A standard Boolean engine is not able to return semantically related documents whose keywords are synonyms of the keywords specified in the original query. [3]

This paper works in following phases to solve/reduce the above problems:

a) *Develop/Update ontology in XML, at the time of Crawling, in terms of documents.*

b) *Integrating this ontology with the semantic nets so that a focused document group can be created.*

c) *Pruning this semantic network, according to the user's query using the concept of topic maps for practical usage in searching.*

To realize the above discussed processes and to make searching process more effective in terms of relevancy and speed, we have break the traditional techniques of searching into following steps:

Develop/Update a Cognitive Ontology: This phase builds a cogitative ontology of web-documents (structured form) which provides a parallel support to web-searching. This ontology also has the facility of timely updates that keeps its refresh rate high, suitable for future references. [5]

The Architecture of Onto-Net Builder works on following steps:

Step1: Crawler downloads the web pages which gets stored in a web repository, it also provides an unique-ID to each web page which makes their maintenance easier.

Step2: Pre-processing of downloaded documents is conducted (Extraction of keywords, URLs, removal of stop words, etc.) at this step.

Step3: Multiple contexts of each extracted keywords

are generated from the thesarus. Eg: BANK can have multiple contexts like Financial instituition, River bank, Reservoir etc.

Step4: Keywords, extracted domains, related words are given as input to the tool named as ONTO-NET BUILDER, which generates an appropriate semantic network for each keyword representing its all possible contexts and meanings

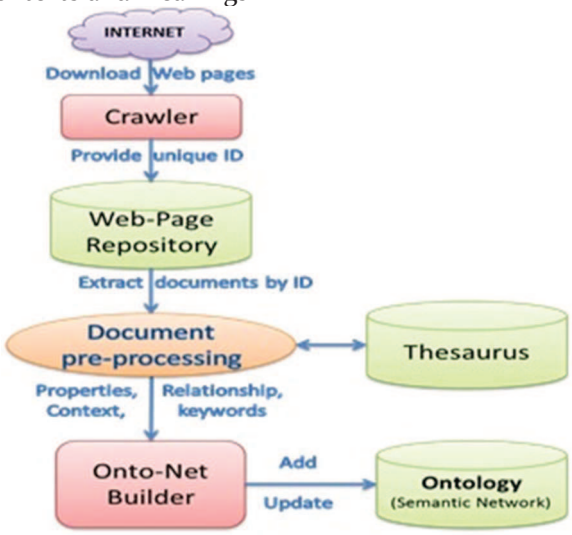


Fig. 3 Architecture of Onto-Net Builder

Web-Searching based on cognitive ontology: Architecture of searching defines how the searching process is to be performed, what all preprocessing, processing and references are required in the searching at different levels.

The diagram below depicts the architecture of "searching". It consists of following steps:

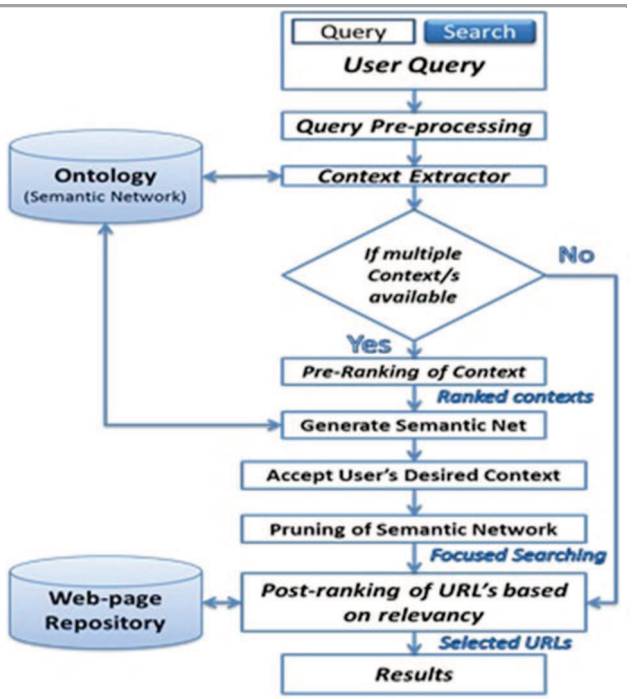
Step 1: As the user enters any search query, keywords are extracted from it using query preprocessing.

Step 2: These keywords are passed to context extractor, to find various contexts available for that keyword.

Step 3: If multiple contexts are available, all of them are pre ranked and loaded on the user screen in the form of semantic network and user selects the desired context, just by a single click on the desired node.

Step 4: As per user's selection, semantic network is pruned to topic map and qualifying URLs, that are based on selection criteria are retrieved from the web repository.

Step 5: The final component ranks the URLs and documents given by the pruning of semantic network and the focused results are displayed.



Pseudo Code of Semantic Net based Web Searching using Ontologies

Begin
Foreach document in web-repository {
Provide unique ID to document; Extract keywords;
Find multiple contexts for keyword using thesaurus;
Find synonyms, related words & properties for context;
Provide keyword, contexts, synonyms & properties to Onto-Net Builder;
Add/update Ontology;
}
Foreach keyword in search query {
Retrieve context from ontology;
If multiple context available
Pre-ranking of contexts based on relevancy;
Generate semantic net;
Accept desired context from user;
Pruning of the semantic networks;
Post-ranking of URLs based on relevancy;
\Display output results;
}
End

Implementation: Semantic Network for each word is stored in the form of ontology and ontologies are stored in XML. Description of ontology is given in the Fig. 5. Each word is described using its context and each of the respective context have related topics stored under them. E.g.: CURRENT has contexts {latest, electric-current, river-current} and each specific context is described, using a related topic/s.

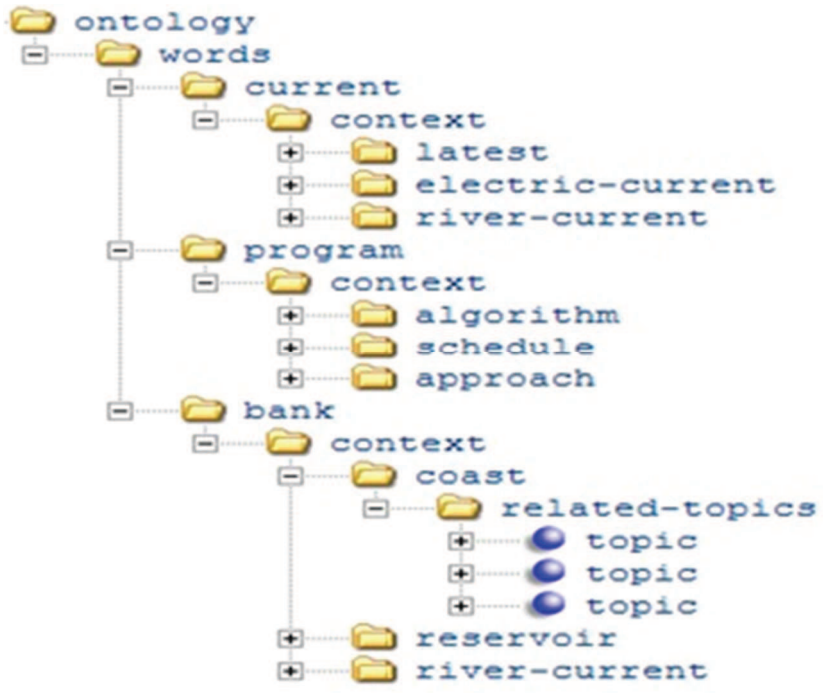


Fig 5: XML representation of Ontology

URL: http:// Electric_current.html. TITLE: Electric current KEYWORDS: electric, current, charge, ampere. DISCRIPTION: Electric current is a flow of electric charge through a medium. The SI unit for measuring the rate of flow of electric charge is the ampere.

Fig 6: Query page of the simulated Search Engine for word "Bank"

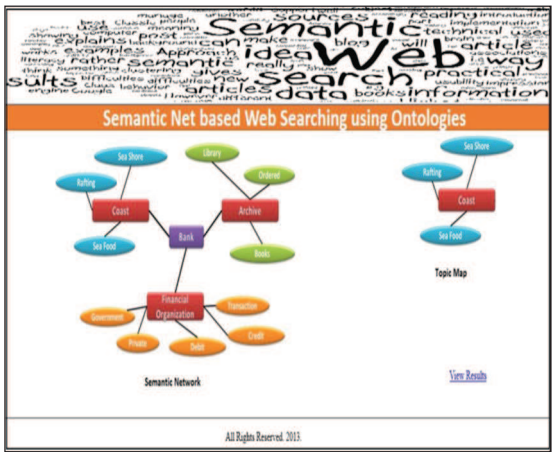
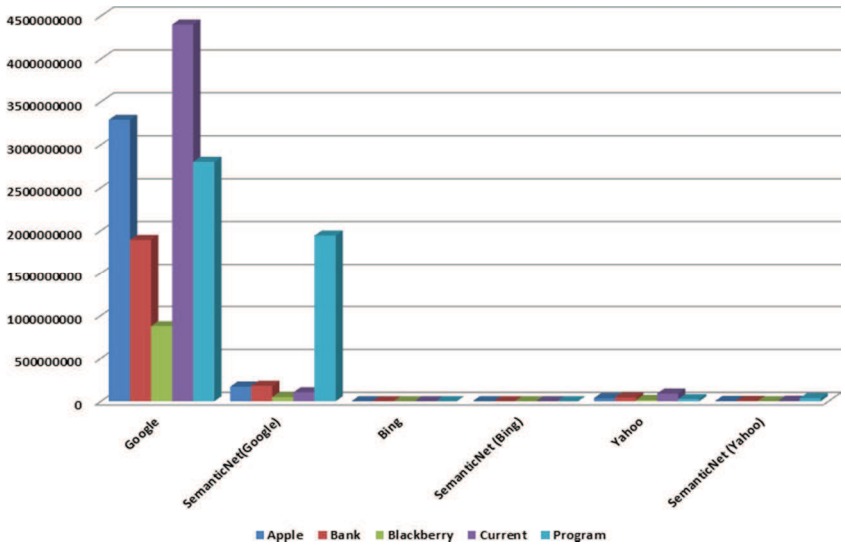


Fig 7: Semantic Net for the word "Bank" and selected Topic Map "Coast"

Performance Analysis: We have analyzed the same query over various search engines and the results are displayed in the form of bar-graph. It has been observed that in comparison to other search engines the results are more specific and less in number. It has been concluded by viewing the graph that the search becomes more focused.



Conclusion: Semantic representation of the document plays an important role in supporting the task of document classification and identification. The proposed technique improves the performance of the searching system in terms of accuracy and efficiency for retrieving more, appropriate documents as per the user's requirements. As the documents are stored contextually the relevant URLs are made available to the user in short span of time. Also it uses available contextual information and ontologies to rank the underlying documents as well as the search results. The use of contextual information results in better ranking of the documents and hence results in higher quality of the retrieved results. The use of topic maps for the pruning of semantic nets makes the searching process accurate and focused.

References:

1. Eakansh Manglik, Priyanka Sharma, Paramjeet Rawat and Nidhi Tyagi "Ontology based Context Synonymy Web Searching", ISCON, 2013.
2. Parul Gupta and A. K. Sharma, "Ontology driven Pre and Post Ranking based Information Retrieval in Web Search Engines", IJCSE, 2012.
3. Nidhi Tyagi, Rahul Rishi and R.P. Aggarwal, "Context based Web Indexing for Storage of Relevant Web Pages", IJCA, 2012.
4. Nidhi Tyagi, Rahul Rishi and R.P. Aggarwal, "Semantic Structure Representation of HTML Document Suitable for Semantic Document Retrieval", IJCA, 2012.
5. Nidhi Tyagi, Rahul Rishi and R.P. Aggarwal, "Contextual Ontology: A Storage Tool for Extracting Context from Web Pages", IJCA, 2012.
6. S.Thenmalar and T. V. Geetha, "Concept based Focused Crawling using Ontology", IJCA, 2011.
7. Parul Gupta and A. K. Sharma, "Context based Indexing in Search Engines using Ontology", IJCA, 2010.
8. J. M. Park, J. H. Nam, Q. P. Hu and H. W. Suh, "Product Ontology Construction from Engineering Documents", IEEE, 2008.
9. L. Weihua, "Ontology supported intelligent information agent", IEEE Symp, on Intelligent Systems, 2002.
10. Steve Lawrence, "Context in Web Search", IEEE Data Engineering Bulletin, 2000.
11. XML http://www.w3schools.com/xml.

* * *

Meerut/ Student/ SHU/
erpriyanka24@gmail.com,eakanshmanglik@live.com,
alokbondcs@gmail.com,
Dean Academics & Research/ JPIET/ dean.ar@jpiet.com,