American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-7, Issue-2, pp-284-287
www.ajer.org
Research Paper Open Access

A Survey of Web Search from Web Documents Based On Semantic Ontology Technique

[1]s.Suma Singh, [2]g.Sangeetha
[1,2] *Assistant professor Department of Computer Science And Engineering, Valliammai Engineering College, Chennai ,India*
Corresponding author: s.Suma Singh

ABSTRACT: There exists a gap between Web mining and the effectiveness of using Web data. The main reason is that we cannot simply utilize and maintain the discovered knowledge using the traditional knowledge-based techniques due to the huge amount of discovered patterns, many noise in discovered patterns and even some useful patterns with uncertainties. In this paper we discuss ontology approaches for building a bridge between Web mining and the effectiveness of using Web data, which tend to automatically construct and maintain ontologies for representations, application and updating of discovered knowledge.
KEYWORDS: Semantic Web, Ontology, Web Mining,

Date of Submission: 14-02-2018 Date of acceptance: 03-03-2018

I. WEB MINING

Data mining turns data into knowledge. Web mining is the process of applying data mining techniques to extract and uncover knowledge from web documents and services. In other words web mining is discovering interesting and useful information from Web content and usage. The web is not a relation table it is a textual information and linkage structure. Web mining can be divided into three different types – Web usage mining, Web content mining and Web structure mining.

Web content mining is extracting and integration of useful data, information and knowledge from Web page content. There are many search engines like Lycos, Alta Vista, WebCrawler, Aliweb, MetaCrawler, and others to provide some useful retrieval documents to the users.

Web structure mining uses graph to analyze the node and link connection structure of a web site. According to the type of web structural data, web structure mining can be divided into two kinds:

1. Extracting patterns from hyperlinks in the web: a hyperlink is a structural component that connects the web page to a different location.
2. Mining the document structure: analysis of the tree-like structure of page structures to describe HTML or XML tag usage.

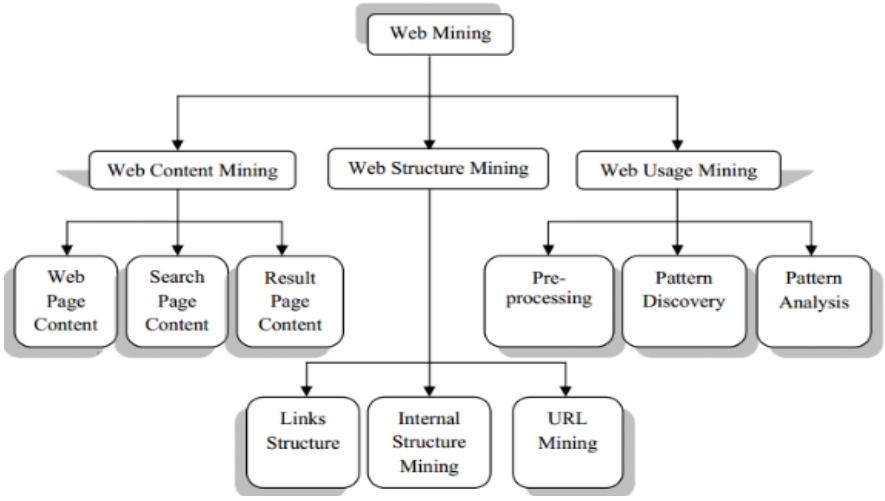

Fig1: Taxonomy of Web Mining.

Web structure mining terminology:

- web graph: directed graph representing web.
- node: web page in graph.
- edge: hyperlinks.
- in degree: number of links pointing to particular node.
- out degree: number of links generated from particular node.

Web Usage Mining applies data mining techniques to discover interesting usage patterns from Web data. In order to understand and serve better the needs of Web-based applications. Web Usage data identifies origin of Web users along with their browsing behavior , browsing history with a Web site. Typical example includes data like IP address, page reference and access time. The ability to track various kinds of business events and log them in application server logs.

However the Web data is not structured that is the web data may be semi or unstructured and the data exhibits heterogeneity the automated discovery of targeted information is a challenging research problems. These factors insist the researchers to develop more intelligent methods for information retrieval, such as intelligent web agents, data mining techniques and AI systems to provide a higher level of retrieval from semi-structured and unstructured data available on the web.

II. SEMANTIC SEARCH

Semantic search improves search accuracy by understanding user target meaning and the contextual meaning of terms to generate more relevant results. Author Seth Grimes lists "11 approaches that join semantics to search", and Hildebrand et al provide an overview that lists semantic search systems and identifies other uses of semantics in the search process.

Semantic search tools includes various factors like context of search, location, intent, and variation of words, synonyms, generalized queries and specialized queries. Even concept matching and natural language queries are used to provide relevant search results. Major web search engines like Google and Bing incorporate some elements of semantic search. Semantic search is one which not only searches data but also the conceptual meaning of the given keyword data . Unlike other search algorithms, semantic search is based on the context essence, intent and concept of the searched keyword phrase. location, synonyms of a term, current trends, word variations and other natural language elements as part of the search. Semantic search concepts are derived from various search algorithms and methodologies, including keyword-to-concept mapping, graph patterns and fuzzy logic.

Semantic Web data are formatted according to Resource Description Framework (RDF), a triple/graph-based way to represent information. Furthermore, Web Ontologies described in RDF Vocabulary Description Language (RDFS) and the Web Ontology Language (OWL) provide shared Concepts, that is., classes and properties, for describing domain entities and thus enabling semantic interoperability of different applications. Semantic interoperability depends on reusing or extending existing Ontologies when developing new applications. Therefore, Ontologies search becomes a fundamental service for application developers.

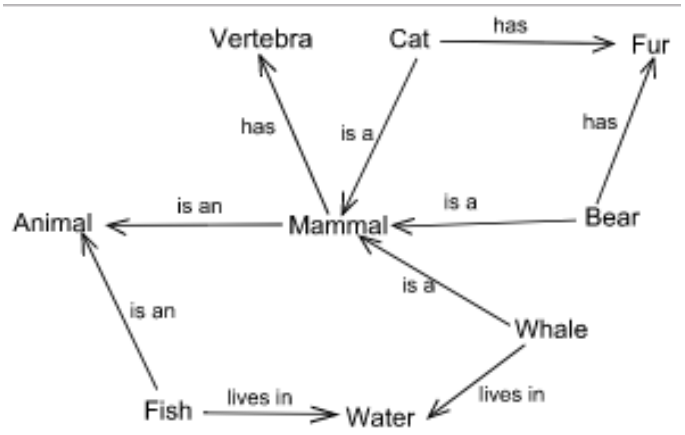

Fig2: Semantic Graph Representation Of Mammal.

III. ONTOLOGY

In the environment of computer and information sciences, an ontology defines a set of primitives to model a domain of knowledge . The primitives usually include classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the primitives are information about their meaning and constraints on their logically reliable application . Ontologies are in general represented with help of a language. The languages of ontologies are similar to first-order logic than languages used to model databases. On this grounds, ontologies are supposed to be at the "semantic" level, whereas database schema are models of data at the "logical" and "physical" level. Due to their independence from lower level data models, ontologies are used for integrating heterogeneous databases, enabling interoperability among different systems, and specifying interfaces to independent, knowledge-based services. In the available technology pile of the Semantic Web standards [1], ontologies are called out as an explicit layer. There are now standard languages and a various commercial and open source tools for creating and working with ontologies.

The Web Ontology Language (OWL) is a known as the family of knowledge representation languages for authoring ontologies. Ontologies are a standard form to describe taxonomies and classification networks, basically ontology defines the structure of knowledge for a variety of domains: the nouns represent classes of objects and the verbs represent relations between the objects. Ontologies are intended to represent information on the web and are expected to evolve constantly and continuously. Similarly, ontologies are typically far more flexible as they are meant to represent information on the Internet coming from all sorts of heterogeneous data sources.

The OWL languages characterize a standard semantics. They are built upon the World Wide Web Consortium's (W3C) XML standard for objects called the Resource Description Framework (RDF). OWL and RDF have involved remarkable change in the field of academic, medical and commercial interest of information retrieval.

Here we discuss different ontology ranking algorithm:

- AKTive Rank Algorithm

AKTiveRank is an experimental and outstanding system for ranking ontologies based on a number of measures. These measures evaluate the ontology in terms of how well it represents the concepts of interest. The targeted query submitted by the user to the search engine is obtained by AKTiveRank to recognize the concepts that match the user's request. The ranking measures used by AKTiveRank will be based on the representation of concepts and their neighbourhoods.

- Content-based Ontology Rank Algorithm

The content-based ontology ranking algorithm uses a list of ontologies from a search engine. The knowledge engineer rank the retrieved ontologies based on the term given by the user. The terms that is the concept label of the ontology is matched with the terms extracted from a WordNet. It is done related to the domain of knowledge recognized by the knowledge engineer‟s original search terms. Each ontology is then ranked according to how many of these new terms match class labels within them. The class match score (CMS) is used for ranking.

- OntoRank Algorithm

The OntoRank algorithm apply the link analyze method for ranking . Here two concepts are measured as a reference relationship "if and only if" a relationship exists between the two classes in a relation set . The

reference relations used in this algorithm are directional and transitive. It evaluates the importance of ontology and does not consider the user query as an effective factor in ranking the results.

- Ontology Structure Rank Algorithm.

Ontology Structure Rank algorithm called Ontology Structure Ranking (OS_RANK) ranks the ontologies based on its semantic relation and structure. The overall ranking criteria are based on the three ranking scores:

- Ranking based on class name
- Ranking based on semantic relation
- Ranking based on ontology structure.

The search engine retrieves the ontology from the user targeted query and these measures are applied to them . The user can decide the weights of the ranking measure according to the needs and importance of their applications

- SIF Rank Algorithm

Semantic-aware Importance Flooding (SIF RANK) retrieves the OWL ontology and converts them into directed graph. The importance of a node is calculated with the iteration fix point computation in each graph. It is based on the nine kinds of patterns and semantically treated correct. This computation reaches the maximum number of iterations and the normalization is done to neglect the nodes which are not semantically linked.

IV. CONCLUSION

Semantic web search is based on knowledge representation which contains a large number of ontologies. Nowadays the increasing demand for user interested web pages has triggered a growing number of usable ontology in web. Alike to the web page searching and ranking , ontology searching results are also to be ranked. The ranking method increases the scope of the knowledge searching in ontology-driven searches. This paper gives an overview of different types of ranking algorithm their methodology. It is used to help the researchers to select the most suitable algorithm for their application based on the efficiency.

REFERENCES

[1]. How to Use Search Engine Optimization Techniques to Increase Website Visibility JOHN B. KILLORAN, IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION, VOL. 56, NO. 1, MARCH 2013.

[2]. WebCap: Inferring the user's Interests based on a Real-Time Implicit Feedback. Nesrine zemrili, Information system department,978-1-4673-2430-4/12-2012.

[3]. A Collaborative Decentralized Approach to Web Search Athanasios Papagelis and Christos Zaroliagis, *Member, IEEE* IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 42, NO. 5, SEPTEMBER 2012.

[4]. A Web Search Engine-Based Approach to Measure Semantic Similarity between Words. Danushka Bollegala, Yutaka Matsuo, and Mitsuru Ishizuka, Member, IEEE IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 7, JULY 2011.

[5]. Correspondence Falcons Concept Search: A Practical Search Engine for Web Ontology Yuzhong Qu and Gong Cheng IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 41, NO. 4, JULY 2011.

[6]. One Size Does Not Fit All: Towards User & Query Dependent Ranking For Web Databases Aditya Telang, Chengkai Li, Sharma Chakravarthy *Department of Computer Science and Engineering, University of Texas at Arlington* July 16, 2009

[7]. Long-Term Cross-Session Relevance Feedback Using Virtual Features Peng-Yeng Yin, Bir Bhanu, Fellow, IEEE, Kuang-Cheng Chang, and Anlei Dong, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 20, NO. 3, MARCH 2008.

[8]. Automated Ranking of Database Query Results, Sanjay Agrawal, Surajit Chaudhuri, Gautam Das,Microsoft Research., Aristides Gionis Computer Science Dept, Stanford University, Proceedings of the 2003 CIDR Conference.

[9]. An Efficient k-M,eans Clustering Algorithm: Analysis and Implementation Tapas Kanungo, Senior Member, IEEE, David M. Mount, Member, IEEE, Nathan S. Netanyahu, Member, IEEE, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, Senior Member, IEEE, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 7, JULY 2002

[10]. An introduction through information retrieval Power point slides. Jian-Yun Nie ,University of Montreal,Canada.

[11]. A New Algorithm for Inferring User Search Goals with Feedback Sessions Zheng Lu, Student Member, IEEE, Hongyuan Zha, Xiaokang Yang, Senior Member, IEEE, Weiyao Lin, Member, IEEE, and Zhaohui Zheng

s.Suma Singh.“ A Survey of Web Search from Web Documents Based On Semantic Ontology Technique” American Journal of Engineering Research (AJER), vol. 7, no. 2, 2018, pp. 284-287.