2013 International Conference on Applied Computing, Computer Science, and Computer Engineering

Knowledge-Based Data Mining Using Semantic Web

Sumaiya Kabir[a,b], Shamim Ripon[a]*, Mamunur Rahman[b] and Tanjim Rahman[b]

[a]*Department of Computer Science and Engineering, East West University, Bangladesh*
[b]*Dept. of CSE, Patuakhali Science and Technology University, Bangladesh*

Abstract

Semantic web offers a smarter web service which synchronizes and arranges all the data over web in a disciplined manner. In data mining over web, the accuracy of selecting necessary data according to user demand and pick them for output is considered as a major challenging task over the years. This paper proposes an approach to mapping data over the web 3.0 through ontology and access the required data via an intelligent agent. The agent provides all the searched data related to user query from which user can find desired information. When the user does not have sufficient search parameter, knowledge can be perceived from the information provided by the agent. The derivation of such unknown knowledge from the existing can be achieved by semantic web mining. We present an intelligent agent-based web mining model where users' query is being searched by following existing traditional way, e.g. by Google. The intelligent agent checks the searched data and derives only those are the semantically related to users search parameter. A work-in-progress case study of *University Faculty Information* presented to examine the effectiveness of the proposed model.

Keywords: Semantic Web; Web Mining; Intelligent Agent;

* Corresponding author. Tel.: +88-1928-891-978.
E-mail address: dshr@ewubd.edu.

1. Introduction

Ongoing rapid progress and extensive application of the internet, there is a massive amount of information distributed on the web. The conventional string based search often failed to hit the relevant pages and feedbacks a lot of irrelevant pages from user request. A common problem for a user is that "Everything is on the web, but we just cannot find what we need" [1] is partially true as most of the data over the web is scattered, unstructured, often inconsistent and insufficient. Data sets are not interlinked with each other which makes mining even more difficult to manage.

Discovering unknown knowledge is almost impossible in web2.0, as no relationship is established among data sets making traditional web mining result almost unsatisfactory. For an improved mining, people are now facing toward web3.0. Here, information is presented in a well-defined and structured manner and enable machines and human to work cooperatively. Data in the semantic web is interlinked among each other through ontology which makes effective discovery, mechanization and assimilation possible. These data are machine readable and can be shared and processed by automated tools as well as people.

The semantic web network is a layered architecture [2][3] consists of various levels. In this layered architecture, RDF [4] (Resource Description Framework) and RDF Schema provides a semantic model used to describe the information on the Web and its type. RDF query language SPARQL [5] can be used to query any RDF-based data (i.e., including statements involving RDFS and OWL [6]). The ontology vocabulary layer defines shared knowledge and describes the semantic relationships between various kinds of information. Ontology is considered as the backbone [7][8] for the semantic web architecture as it provides a machine-processable semantics and a sharable domain which can facilitate communication between people and different applications.

The Semantic Web is based on a vision of enriching the Web by machine-processable information. For instance, today's search engines are already quite powerful, but still too often return excessively large or inadequate lists of hits. Machine-processable information can point the search engine to the relevant pages and can thus improve both precision and recall.

Data mining is a process to extract useful and interesting knowledge from large amount of data. Web Mining aims at discovering insights about the meaning of Web resources and their usage. Given the primarily syntactical nature of the data being mined, the discovery of meaning is impossible based on these data only. Therefore, formalizations of the semantics of Web sites and navigation behavior are becoming more and more common. Semantic Web Mining combines Semantic Web and Web Mining. The nature of most data on the Web is so unstructured that they can only be understood by humans, but the amount of data is so huge that they can only be processed efficiently by machines. The Semantic Web addresses the first part of this challenge by trying to make the data (also) machine understandable, while Web Mining addresses the second part by (semi-)automatically extracting the useful knowledge hidden in these data, and making it available as an aggregation of manageable proportions. Instead of data mining semantic web enables knowledge mining over web.

Intelligent agent [9] facility enables users to find desired results for all possible related terms with respect to requirements. This paper focuses on how an agent detects all possible entities from ontology during web mining [10] related to a user query request on its own in an automated manner which enables the user to discover unknown knowledge.

In the rest of the paper, in Section 2, we first illustrate our proposed model of semantic web mining and show the steps of how the model can be used. In the following section, we briefly describe our work-in-progress case study of semantic web-based representation to *University Faculty Information*. We briefly describe how an intelligent agent can be used to acquire unknown knowledge with the support of ontology. Finally, we conclude our paper by summarizing our work and outlining our future plan.