SOFTWARE ARCHITECTURE FOR SEMANTIC WEB MINING IN THE PARALLEL DISTRIBUTED ENVIRONMENT

A. Anny Leema, Phd Scholar
Karpagam University
annygeo_annygeo@yahoo.co.in
annyleema@gmail.com

&

Dr. Hemalatha M
Assistant Professor
Department of Computer Applications
Karpagam University
Pollachi Main Road, Eachanari (PO)
CBE-21
hema.bioinf@gmail.com

ABSTRACT: The web has become a major vehicle in performing research and education related activities for researchers and students. There is some tremendous amount of information and knowledge existing on the web which is to be discovered, shared and utilized. Web mining is the use of data mining technologies to automatically interact and discover information from web documents, which can be structured, unstructured or semi-structured form. Web mining is now a reality and the challenge is to carry out semantic web mining. It is about machine-understandable web pages to make the web more intelligent and able to provide services to the user. This means information on the web has to be mined so that the machine can understand the content. An enterprise framework was developed that uses semantic web mining, resource description framework, ontology and XML technology for an educational domain. Performing ontology maintenance - currently takes a lot of time and is very difficult. Recurring patterns can indicate relevant and/or necessary changes to the ontology. For decision making and to increase speed and efficiency it is intended to implement semantic web mining in parallel distributed environment in all tiers. It is also planned to establish an operational ontology base with more granularities to realize intelligent interaction between user and machine and act as better learning platform.

Keywords: Semantic Web Mining

1. INTRODUCTION

The World Wide Web provides plentiful contents for Web-based learning, but its hyperlink-based architecture connects Web resources for browsing freely rather than for effective learning. Therefore, how to effectively organize learning resources of various types to support e-learning in a semantic context becomes a challenge. 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. Semantic web services are built around universal standards for the interchange of semantic data, which makes it easy for programmers to combine data from different sources and services without losing meaning. Web services can be activated "behind the scenes" when a web browser makes a request to a web server, which then uses various web services to construct a more sophisticated reply than it would have been able to do on its own. Semantic web services can also be used by

ISBN: 97881 8371 362 7
automatic programs that run without any connection to a web browser. The semantic web is vitally dependant on a formal meaning for the constructs of its languages. The Semantic web is designed to let users make explicit statements about any resource, and maintain that data themselves in an open and distributed manner[3]. The process of building the Semantic Web is currently an area of high activity. Its structure has to be defined, and this structure then has to be filled with life. In order to make this task feasible, one should start with the simpler tasks first. The following steps show the direction where the Semantic Web is heading:
1. Providing a common syntax for machine understandable statements.
2. Establishing common vocabularies.
3. Agreeing on a logical language.
4. Using the language for exchanging proofs.

2. ONTOLOGIES:
Ontology comprises a set of knowledge terms, including the vocabulary, the semantic interconnections and some simple rules of inference and logic for some particular topic. Ontologies applied to the Web are creating the Semantic Web]. Ontologies provide the necessary armature around which knowledge bases should be built and set grounds for developing reusable Web-contents, Web-services and applications[2]. Ontologies facilitate knowledge sharing and reuse, i.e. a common understanding of various contents that reaches across people and applications. Technically, an ontology is a text-based piece of reference-knowledge, put somewhere on the Web for agents to consult it when necessary and represented using the syntax of an ontology representation language. There are several such languages around for representing ontologies[1] for an overview and comparison of them. It is important to understand that most of them are built on top of XML and RDF.

2.1 Performing Ontology Maintenance
Ontologies require maintenance as they evolve. Identifying recurring structures helps in revising ontology concepts. For instance, in a financial portal the concept of tax appears, but in real life taxes are often of different types and for financial experts it is difficult to work only with one big concept. So they decide to split this concept into more useful categories, such as: income tax, luxury tax, import tax, value added tax etc. This results in more meaningful modeling[4].

3. SEMANTIC WEB MINING
Semantic Web Mining aims at combining the two areas Semantic Web and Web Mining by using semantics to improve mining and using mining to create semantics[5]. Last but not least, these techniques can be used for mining the Semantic Web itself.

3.1 Layers of the Semantic Web

On the first two layers, a common syntax is provided. Uniform resource identifiers (URIs) provide a standard way to refer to entities, while Unicode is a standard for exchanging symbols. The Extensible Markup Language (XML) fixes a notation for describing labeled trees, and XML Schema allows the definition of grammars for valid XML documents. XML documents can refer to different namespaces to make explicit the context (and therefore meaning) of different tags. The formalizations on these two layers are nowadays widely accepted, and the number of XML documents is increasing rapidly. While XML is one step in the right direction, it only formalizes the structure of a document and not its content. The Resource Description Framework (RDF) can be seen as the first layer where information becomes machine-understandable: According to the W3C recommendation4, RDF “is a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web.” RDF documents consist of three types of entities: resources, properties, and statements. Resources may be Web pages, parts or collections of Web pages, or any (real-world) objects which are not directly part of
the WWW. In RDF, resources are always addressed by URIs. Properties are specific attributes, characteristics, or relations describing resources. A resource together with a property having a value for that resource forms an RDF statement[4]. A value is a literal, a resource, or another statement. Statements can thus be considered as object–attribute–value triples.

4. ONTOLOGY BASED MODEL:

Ontologies are theories that use a specific vocabulary to describe entities, classes, properties and functions related to a certain view of the world. The use of ontology, translated into an active information system component, leads to ontology-driven information systems and, in the specific case of EDUCATION, leads to what we call ontology-driven learning systems. Intelligent, high-level services like information brokers, search agents, information filters, intelligent information integration and knowledge management, are what the users want from the Semantic Web. They are possible only if a number of ontologies populate the Web, enabling semantic interoperability between the agents and the applications on the Semantic Web, i.e. semantic mappings between terms within the data, which requires content analysis. One specific kind of ontology is necessary to enable high-level Semantic Web services- ontologies of services themselves. These ontologies should include a machine-readable description of services (as to how they run), the consequences of using the service (e.g., the fee) and an explicit representation of the service logic (e.g., automatic invocation of another service). Services have their properties, capabilities, interfaces and effects, all of which must be encoded in an unambiguous, machine understandable form, to enable agents to recognize the services and invoke them automatically. Course sequencing generally starts with the student entity component that receives the learning contents, while the student’s behavior is being observed. The instructor sends queries to the learning resources to search for learning content that is appropriate for the student entity component. The ontological knowledge is added to the learning resources as a resource for contextual learning and it may be searched by means of queries. The student’s performance is measured by the evaluation component and the result is stored in the student records database. The data in it can be used by the instructor component to locate a new content. Searching learning resources and sequencing a course can be done using a knowledge base of learning resources and a delivery component. To implement the knowledge base, first of all, the learning resources have to be described by means of metadata. The metadata consists of the contextual knowledge of the learning resources, i.e., ontology in our model. It contains the general representation of the structural knowledge on specific domains, such as computer science, mathematics, biology and so on[6]. The ontology can be used for adaptive learning to retrieve the context of a course and to structure the contents. Also the metadata actually consists of the framing description of each learning object of a subject, i.e., the modularized content, which is linked to the concept of the ontology.

5. ONTOLOGICAL APPROACH TO METADATA

Metadata standards pose different levels of representation granularity. To reach a finer-grained level, where components in a resource are represented and correlated, knowledge schemas play an important role in in-depth representation and more refined user access. The key advantage of an ontological representation within the realm of learning objects is its ability to handle different granularities. In order to describe learning resources at the collection level (e.g. web site) and further describe each of the components (e.g. interactive applet, image), relationships must be identified when the data are input. Only by having description at the component level will specific learning objects able to be retrieved by users[6]. The following system helps to find suitable semantic data related to students, faculties and courses for the clients.

![Ontology Based Educational System](image)

6. CONCLUSION AND FUTURE WORK

Thus we have developed semantic web mining for educational domain with the
help of enterprise web framework that comprises semantic web mining, RDF, Ontology and XML technology. This system is implemented with more granularities and thus implemented in the parallel distributed environment. It acts as better learning platform too.

7. REFERENCES