Automatic collaboration and analysis of semantic web information for electronic learning environment

Vishwanath Gupta and Shirish M. Dubey
Shri ram institute of technology Jabalpur, India.

ABSTRACT
Semantic web technology enables machine to read and process web pages without human intervene. Semantic web information will no longer only be intended for human readers, but also for processing by machines, enabling intelligent information services, personalized web sites, and semantically empowered search-engines. The web has become a major vehicle in performing research and education related activities for researches and students. With increasing student mobility and boom of the international student exchange programmes, a need arises for unifying and presenting information about academic study programmes on the Web. Publishing study programmes using semantic web technologies enables students to easily search and select study topic of their interest. In this paper we present an enterprise semantic framework for e-Learning system which improved the quality of web mining results but also enhanced the functions, services and the interoperability of e-Learning system. The system helps to find suitable semantic data related to students, faculties and courses for the clients. We have implemented semantic web mining in parallel distributed environment in all tiers for decision making, and also increased speed and efficiency of information retrieval.

© 2013 Elixir All rights reserved.

Introduction
Web mining [1] is the application of data mining technologies to automatically interact and discover information from web documents, which can be in structured, unstructured or semi-structured form. Semantic Web[2] which can be seen as a new generation of information infrastructure is a new distributed intelligent network platform based on semantic information processing. The basic idea of Semantic Web is that embed machine-readable, on behalf of certain types of knowledge mark in the Web message. So that the data on the Web is not only used to display, but also be understood by the machine so as to enhance the quality of the information services and explore a variety of new, intelligent information services. The Semantic Web is considered an evolving extension of the World Wide Web in which the semantics of information and services on the web is made explicit by adding metadata. Metadata provides the web content with descriptions, meaning and inter-relations. The Semantic Web is envisioned as a universal medium for data, information, and knowledge exchange.

With the ability of intelligent analyses, it can help people acquire appropriate information and discovery the latent semantic knowledge effectively. From the perspective of semantic web proposed by Berners- Lee as the core hierarchical of describing semantics, ontologies have promoted and facilitated interoperability among information systems, intelligent processing by agents and sharing and reusing of knowledge that can be communicated among people, machine, heterogeneous and widely spread application systems. Nowadays, semantic web and ontology have shown their usefulness in application areas such as intelligent information integration, information brokering and Natural Language processing.

The Web has become a major vehicle in performing research and education related activities for researches and students. With increasing student mobility and boom of the international student exchange programmes, a need arises for unifying and presenting information about academic study programmes on the Web. Publishing study programmes using semantic web technologies enables students to easily search and select study topic of their interest. The core technique of Semantic web mining is ontology. In computer science, ontology represents a set of precisely defined terms about a specific domain and accepted by this domain’s community. Ontology is an explicit specification of a conceptualization. XML allows the representation of semi-structured and hierarchal data containing not only the values of individual items but also the relationships between data items. Element tags and their nesting therein dictate the structure of an XML document. XML was designed to transport and store data. The RDF is a simple meta model for defining and exchanging information on the semantic web. The basis of a particular way of providing meaning for metadata is embodied in the model theory for RDF, the language at the base of the Semantic Web.

We have developed semantic web mining for an educational domain. For this we have developed an enterprise web framework that uses semantic web mining, Resource Description Framework, Ontology and XML technology. The system helps to find suitable semantic data related to students, faculties and courses for the clients and also improved speed and efficiency of information retrieval.
The paper is organized as follows: section II describes the overview of semantic web, ontology, Resource Description Language (RDF) and XML (eXtensible Markup Language). In section III we describe related works. Section IV presents our approach. In section V implementation work, simulation example and results are discussed. Section VI draws the conclusions and future works.

**Semantic Web Architecture , Ontology, Rdf, Xml**

Web mining can be generally defined as: Extract interested, useful patterns and implicit information from the WWW resources and behavior. In general, Web mining can be divided into three categories: Web content mining, Web structure mining and Web usage mining.

Web content mining is used to extract the text, image, or other information and knowledge component of the web content. Web content mining has two strategies: page text mining; process results for search engine query further to get more accurate and useful information.

Web structure mining is used to extract the network topology information, that is, the link between pages of information. Mine knowledge from the WWW organization and links. Which pages are linked to other pages? It extracts interested patterns from the access to records of Web

Semantic Web: The basic idea of Semantic Web[2] is that embed machine-readable, on behalf of certain types of knowledge mark in the Web message. So that the data on the Web is not only used to display, but also be understood by the machine so as to enhance the quality of the information services and

Explore a variety of new, intelligent information services. If the knowledge that reflect the link between data and application are embedded in a variety of different information sources in a user transparent manner, Web pages, database, procedures will be able to link up through the agent and each other collaborate.

**Semantic web mining architecture**

The bottom layer is the Unified Resource Identifiers (URIs) and Unicode[3]. The URI is an industry standard of representing entities, objects or concepts in the Semantic Web. Unicode is required by modern standards to represent a unique number for every character and for exchanging symbols. The next layer over URI and Unicode is XML and its namespaces.

![Image 1: Semantic web mining architecture](image)

Semantic web is made up of XML which is independent of any platforms, hardware or software. Wireless Access Protocol (WAP) is the standard for information services on wireless terminals and is based on some internet standards, such as XML, TCP/IP, and HTML. WAP used Wireless Mark-up Language (WML), which is based on XML. XML Schema Definition (XSD) is an alternative to the Documentation Type Definitions (DTDs), which are used to describe document structure and the legal building blocks used in an XML document. An XML document and its XML Schema Definition (XSD) have namespaces to differentiate between contexts. The Resource Description Framework (RDF) is a W3C standard for describing resources on the Web, such as document author, creation/modification date, document title, document content, and some related copyright information of a Web page. RDF is adding metadata to resources on the Web so that computers can understand the content of the subject other than just displaying what it contains. Since RDF follows the XML syntax, computers can exchange RDF documents just like the exchange of XML documents. RDF documents are not for humans to view, but are designed for computer applications to read and understand. It describes resources on the Web using simple statements consisting of subject, predicate and object. The subject can be any resource that has a URI. A property describes the resource to a value. OWL (Web Ontology Language)[14] is very similar to RDF, but it is a heavier language with greater machine-interpretability than RDF. It describes the exact nature of resources and relationships between them. Spanning across these standards is the all-time requirement of security through signatures and encryption.

**Ontology**

The core technique of Semantic web mining is ontology[4].In computer science, ontology represents a set of precisely defined terms about a specific domain and accepted by this domain’s community. Ontology is an explicit specification of a conceptualization. an ontology is a formal explicit description of concepts in a domain of discourse (classes (sometimes called concepts)), properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties)), and restrictions on slots (facets (sometimes called role restrictions)). An ontology together with a set of individual instances of classes constitutes a knowledge base. In reality, there is a fine line where the ontology ends and the knowledge base begins. Classes are the focus of most ontologies. Classes describe concepts in the domain.

**RDF**

The RDF[5] is a simple meta model for defining and exchanging information on the semantic web. The basis of a particular way of providing meaning for metadata is embodied in the model theory for RDF [5,6], the language at the base of the Semantic Web. In particular, RDF has a very limited collection of syntactic constructs, and these are treated in a very uniform manner in the semantics of RDF.

**XML**

The eXtensible Markup Language (XML)[6,7] has become a standard language for data representation and exchange. XML is a Standard, flexible syntax for data exchanging. With the continuous growth in XML data sources, the ability to manage collections of XML documents and discover knowledge from them for decision support becomes increasingly important. Mining of XML documents significantly differs from structured data mining and text mining. XML allows the representation of semi-structured and hierarchal data containing not only the values of individual items but also the relationships between data items. Element tags and their nesting therein dictate the structure of an XML document. XML was designed to transport and store data

**Related Work**

According to [8] semantic e-business is founded upon three active stream of research: Semantic Web technologies with ontologies, knowledge representation, and intelligent software agents. The purpose of this literature review is to study, analyze
and identify the opportunities and limitation of various techniques regarding the application of semantic technologies in educational domain.

Ideas presented in [9] explain the concept of using semantic web in educational application to overcome the educational domain problems like decision making on the basis of incomplete knowledge, due to excess of unstructured information, and difficulties in integration between different educational applications. Semantic technologies provide a possible solution to their problems by associating ontologies with unstructured information and to provide uniform access layer to heterogeneous data sources. The strength of this work is general examples to address problems and their solutions. It highlights the application of semantics in web services and with web 2.0. The limitations of the work are that the author does not introduce any new method for solving problems, and no practical approach has been adopted. Ontology mediation and exploring semantic web services may leads more comparative advantages to educational domain.

It is considered that object-oriented programming (OOP) is a mainstream paradigm to design software development. Due to difference between RDF data model and OOP, RDF data format triples[11] can not directly used by OOP languages such as Java. It is called Triple-to-Object impedance mismatch, just like Relational-to-Object impedance mismatch. Of course, the usual way is to implement Triple-Object mapping framework. SOF is proposed to realize direct mapping between object-oriented classes and Semantic Web classes. It embeds semantic descriptions in source code to convert objects to RDF data. It is independent of programming languages. JenaBean1 is another mapping framework which can perform the mapping between triples and objects. It mainly utilizes Java annotation. ActiveRDF [9,10] is an object-oriented API for managing RDF data that offers full manipulation and querying of RDF data. It does not rely on a schema and fully conforms to RDFS semantics. Different from Triple-Object mapping framework,

It is popular that N-tier architecture is used to build web applications. In [9], Martin introduces a few famous layers architecture such as Microsoft DNA, Java EE etc. Java EE 3-tier architecture is undoubtedly used greatly in enterprise applications. [6] elaborately discusses what Java EE layers are and what responsibilities each layer owns. And it also explains why this architecture is more suitable for constructing enterprise applications running on internet. The key of Java EE 3-tier architecture[12] is how to implement model-view-controller (MVC) pattern. There are two main types of Java web frameworks: one is based on request and response mechanism such as Struts2, SpringMVC etc, the other is based on event-driven, such as Tapestry, Wicket etc.

**Proposed Work**

**Algorithm**

Input: Data set R, Attribute set Ai
Output: data set R'

R' -> R
For I=1 to n do
  Max (Ai) = the deepest node in the attribute set Ai
  If Max(Ai), Distance_to_max < Li
    Newnode=node.root_path_array[Li-node.distance_to_max]
  Else if generalized node of length r consists of r (r ≥ 1) nodes in the tree
    Newnode=Ai
  Else
    Newnode=max(Ai)
Endif
Replace node with new node
Endfor
Remove duplication from R'
N=| R |
For I=0 to N-1 do
  r <- ri
  M <- |r|
  If ri inconsistent with rule rn E then
    Restore the dropped condition aj
Endif
Endfor
Included in rule r
If rule r is not logically included in a rule r' E MRULE then
  MRULE <- r U MRULE
Endif
End

Several standards such as the Resource Description Framework (RDF) and Web Ontology Language (OWL)[13] ave been developed to realize the layer cake of the Semantic Web. From the viewpoint of end users, expressing semantics about students, faculties and their relationships has garnered considerable interest. We have created ontology for Training Institute of the Semantic Web. It is essentially a vocabulary for describing students, courses and faculties In this paper, we propose a Web mining approach for the Semantic Web for Training Institute. The approach uses a search engine and the traditional web as an information resource to produce semantically rich information. In particular, we examine one student or lecturer and extract its subjects and related information from the Web.

**RDF Example**

```xml
<rdf:RDF
 xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
 xmlns:rdfs=http://www.w3.org/2000/01/rdf-schema#>
 <rdfs:Class rdf:ID="Person">
   <rdfs:comment>Person Class</rdfs:comment>
   <rdfs:subClassOf
     rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Resource"/>
 </rdfs:Class>
 <rdfs:Class rdf:ID="Student">     <rdfs:comment>Student Class</rdfs:comment>
     <rdfs:subClassOf
       rdf:resource="#Person"/>
 </rdfs:Class>
 <rdfs:Class rdf:ID="Teacher">     <rdfs:comment>Teacher Class</rdfs:comment>
     <rdfs:subClassOf
       rdf:resource="#Person"/>
 </rdfs:Class>
 <rdfs:Class rdf:ID="Course">     <rdfs:comment>Course Class</rdfs:comment>
     <rdfs:subClassOf
       rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Resource"/>
 </rdfs:Class>
 <rdfs:Property rdf:ID="teacher"/>
 <rdfs:comment>Teacher of a course</rdfs:comment>
 <students><rdf:Seq>
   <Student rdf:id="er">
     <name>Rajesh Kumart</name>
   </Student>
   <Student rdf:id="gl">
     <name>Santosh Thakur</name>
   </Student>
 </rdf:Seq>
 <Student rdf:id="js">
   <name>Rajesh Kumart</name>
 </Student>
</students>
</rdf:RDF>
```
Implemention

We have used Protage[14] to create Ontologies for faculty, students, and courses. We have used Java technology, Jena and Java programming language for implementation. We have used our college web site as database. In experiment, we built student, faculty and courses ontology and input into the model. To test the developed method, we have created ontology for students and faculties from education domain. The student ontology was created from total 45 classes and 89 properties. The faculty consisted of 21 classes and 34 properties. Some concept pairs were chosen from each ontology and the relationships between them. The results are shown in table 1 and table 2.

Table 1

<table>
<thead>
<tr>
<th>Enrol</th>
<th>Name</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Santosh Thakur</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>Ajay Chudhary</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>Rakesh Singh</td>
<td>A</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Atul Rai</td>
<td>DS</td>
</tr>
<tr>
<td>2</td>
<td>Ritesh Sharma</td>
<td>JAVA</td>
</tr>
<tr>
<td>3</td>
<td>Satish Rane</td>
<td>CSE</td>
</tr>
</tbody>
</table>

Conclusion And Future Work

The Web contains huge amounts of data intended for clients. This information is mainly stored in hypertext form. The Semantic Web is considered an evolving extension of the World Wide Web[16] in which the semantics of information and services on the web is made explicit by adding metadata. Metadata provides the web content with descriptions, meaning and inter-relations. The Semantic Web is envisioned as a universal medium for data, information, and knowledge exchange. We propose a framework for personalised e-Learning[17,18] based on aggregate usage profiles and domain ontology. We have distinguished two stages in the whole process, one of offline tasks that includes data preparation, ontology creation and usage mining and one of online tasks that include data preparation and also increased speed and efficiency of information retrieval.

In our future work the main task is to provide security using role back access control. We are planning to implement semantic web mining in parallel distributed environment[19] in all tiers for decision making, and increasing speed and efficiency.

References

[4] Wang Jain, Li Zhuo, Research and Realization of long Distance Education platform based on Web Mining, IEEE 2009