Secure and Intelligent Decision Making in Semantic Web Mining

Ankita Jain
M. Tech Student

Technocrats Institute of
Technology, Bhopal

Prof. Ilyas Khan
Associate Professor

Technocrats Institute of
Technology, Bhopal

Dr. Bhupendra Verma
Director-PG Courses

Technocrats Institute of
Technology, Bhopal

ABSTRACT
With the huge amount of information available online the World Wide Web is a fertile area for data mining research. The Web has become a major vehicle in performing research and education related activities for researches and students. Web mining is the use of data mining technologies to automatically interact and discover information from web documents, which can be in structured, unstructured or semi-structured form. We present an enterprise framework regarding semantic web mining in distance learning, which can be used to not only improve the quality of web mining results but also enhances the functions and services and the interoperability of long distance educational information systems and standards in the educational field. For on line distance education system we propose an Ontology-based approach to share online data and retrieve all relevant data about students and their courses. Thus semantic web ontology help build better web mining analysis in educational institute and web mining in-turns helps contract basis more powerful ontology in distance learning. Since the majority of the online data considered as private data we need various mechanism for privacy preservation and control over the online presence data. We propose privacy protection in semantic web mining using role back access control.

Keywords
Semantic web mining, Association Mining, Ontology, OWL, RDF, XML,RBAC.

1. INTRODUCTION
In the short span of its existence, the World Wide Web has resulted in a revolution in the way information is transferred between computer applications. The goal of data mining is to extract or mine" knowledge from large amounts of data". Data mining[1] involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods. Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction. Data mining can be performed on data represented in quantitative, textual, or multimedia forms. Web mining is the use of data mining technologies to automatically interact and discover information from web documents, which can be in structured, unstructured or semi- structured form.

The Semantic Web [2] is designed to let users make explicit statements about any resource, and maintain that data themselves in an open and distributed manner. To enhance efficiency of information retrieval, several web mining techniques have been proposed including methods deriving from data analysis and conceptual analysis. With the ability of intelligent analyses, it can help people acquire appropriate information and discovery the latent semantic knowledge

effectively. Nowadays, semantic web and ontology have shown their usefulness in application areas such as intelligent information integration, information brokering and Natural Language processing.

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.

2. THE SEMANTIC MINING METHOD FOR EDUCATION DOMAIN ONTOLOGY, RDF, XML

2.1 Ontology
The core technique of Semantic web mining is ontology[3].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[6] 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[8]. Classes describe concepts in the domain. For example, a class of wines represents all wines. Specific wines are instances of this class. The Bordeaux wine in the glass in front of you while you read this document is an instance of the class of Bordeaux wines. A class can have subclasses that represent concepts that are more specific than the superclass. For example, we can divide the class of all wines into red, white, and rosé wines. Alternatively, we can divide a class of all wines into sparkling and nonsparkling wines. In practical terms, developing an ontology includes:

- defining classes in the ontology,
- arranging the classes in a taxonomic (subclass–superclass) hierarchy,
- defining slots and describing allowed values for these slots,
- filling in the values for slots for instances.

2.2 RDF
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 [4], the language at the base of the

Semantic Web. . The RDF Schema[5] (RDFS) draft specifies a small upper ontology on top of RDF, but it too is a work in progress and has never been officially published. Finally, an RDF Model Theory that formally defines the semantics of RDF and RDFS constructs is also under development. Consequently, while the foundations of RDF are fairly solid and well understood, the more advanced features (that nonetheless belong to the object layer of the IMI Reference Model) are still very much a moving target 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. The RDF thesis requires that no other syntactic constructs are to be used and that the uniform semantic treatment of syntactic constructs cannot be changed, only augmented.

2.3 XML

The eXtensible Markup Language (XML) has become a standard language for data representation and exchange. XML[9] is a Standard, flexible syntax for data exchanging Regular, structured data. Database content of all kinds: Inventory, billing, orders etc. It has small typed values and irregular, unstructured text. It can consists of documents of all kinds: Transcripts, books, legal briefs etc. 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. XML tags are not predefined. You must define your own tags XML documents[10] form a tree structure that starts at "the root" and branches to "the leaves".
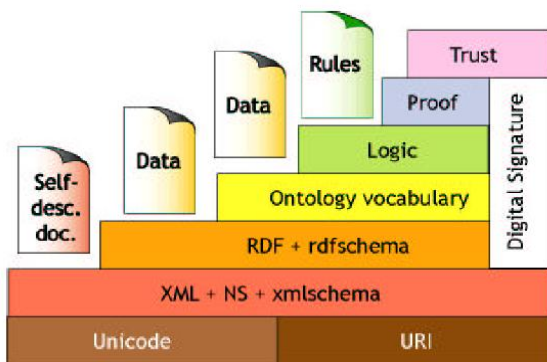
3. THE SEMANTIC WEB MINING

Figure 1: Semantic web mining architecture

The bottom layer is the Unified Resource Identifiers (URIs) and Unicode[12]. 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. 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[13] 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.

RBAC model

An advanced access control framework, "Role Based Access Control Model," was introduced in 1996 [27]. The RBAC model has been widely discussed and further developed since then. In 2001, NIST proposed a consensus model for RBAC, based on these two models. The model was further refined within the RBAC community and has been adopted by the American National Standards Institute(ANSI), International Committee for Information Technology Standards (ANSI/INCITS) as ANSI INCITS 359-2004 [4].

The Core RBAC model describes the basic concepts for the RBAC. These basic sets are:

1. Users(U)
A user is a human being or an autonomous agent.
2. Roles (R)
A role is a job function within an organization with some associated permissions to protected resources.
3. Permissions(P) (includes Objects (Obj) and Operations (OP))
A permission is an authorized right to access or perform operations on protected resources. Objects are the protected resources or tasks in the system.
An operation is an executed action performed by a user.
4. Sessions (S)
When a user logs in the system, he or she establishes a session during which the user activates some roles that are authorized to him or her.

4. SEMANTIC WEB ARCHITECTURE FOR TRAINING DOMAIN

Algorithm 1
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<Ii

```
            Newnode=node.root_path_array[Ii-node.distance_t-
            _max]
      Else
            Newnode=max(Ai)
      Endif
   Replace node with new node
 Endfor
 Remove duplication from R'
End


Algorithm 2
Input : Primitive rules set R
Output Generalized rules set R'
R' <- 0
N= | R |
   For I=0 to N-1 do
      r <- ri
      M <- |r|
            For j=0 to M-1 do
                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 include 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) have 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 [15]. 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.
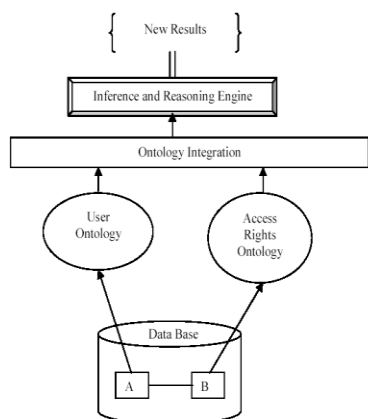

Figure 2: An architecture for semantic web mining and educational system

A: Students, Lecturers, Courses and other staff related tables
B: Role Based Access Control and Security Policy Solutions Tables

Web mining takes aspects of data mining and text mining and brings them together in the context of the world biggest information resource. The WWW web mining finds the info from web sources, may be from data warehouse and from own communal database. There is tremendous amount of information and knowledge existing on the web and waiting to be discovered, shared and utilized. Students, faculties, researchers require a lot of information about education and research activities.

Data mining technology has been playing an important role in education related field for many years. Web mining is the use of data mining techniques to automatically discover and extract information from the web documents which can be structured, unstructured or semi structured from. XML has become very popular for representing semi structured data and a standard for data exchange over the web. The data based on XML is self described; it can be exchanged and handled without internal description. The core technique of semantic web mining is ontology. Ontology represents a set of precisely defined terms about a specific domain and accepted by this domain's community, ontology is an explicitly specification of a conceptualization. The RDF is a simple meta model for defining and exchanging information on the semantic web.

Example Ontology for online training
```
<?xml version="1.0"?>
<rdf:RDF              xmlns:Ontologyowl="http://www.owl-
ontologies.com/Ontology1243411901.owl#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
   <Ontologyowl:Student rdf:about="it02">
     <Ontologyowl:studyMemberOf>
         <Ontologyowl:Project rdf:about="proj52"/>
     </Ontologyowl:studyMemberOf>
   </Ontologyowl:Student>
   <Ontologyowl:ClassOfCourse rdf:about="class21"/>
</rdf:RDF>
```

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.

5. IMPLEMENTATION

We have used Protégé to create Ontologies for student, courses. We have used Dot Net platform and C# programming language for implementation. We have used data from mits college database. In experiment, we built a computer 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 14 classes and 34 properties. The faculty consisted of 12 classes and 27 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

Student		
Eno	**Name**	**Grade**
1	Ashish	A
2	Vijay	B
3	Rahul	A

Table 2

Facuty		
ID	**Name**	**Sub**
1	Shankar	DS
2	Rakesh	JAVA
3	Mukesh	CSO

User input the query words in ontology as expansion words and its performance can be showed through precision and recall ratios that are calculated from experimental results. Through 10 times different information requests, we compute recall and precision ratios. The experimental results are as follows.
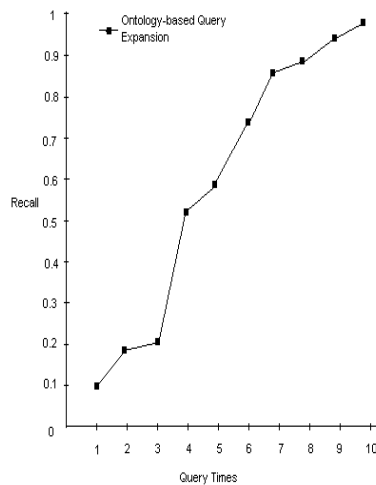

Figure 3: Recall ratio of query methods

From experimental results, we can clearly see that the semantic query expression method based on ontology makes good use of semantic relations and logical reasoning in ontology to expand initial query and better conveys user's willingness.
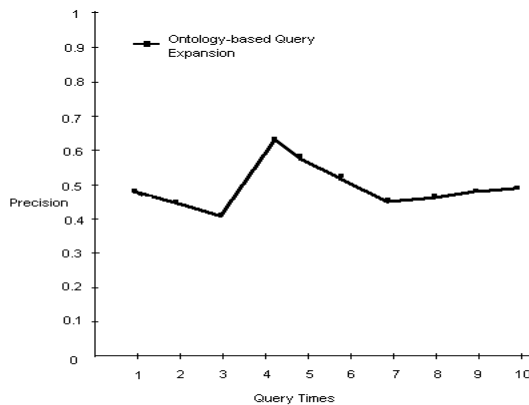

Figure 4: Precision ratio of query methods

Experimental results shows that this method can improve the precision and recall ratios of web information retrieval.

6 .CONCLUSION AND FUTURE WORK
Data mining technology has been playing an important role in education related field for many years. Web mining is the use of data mining techniques to automatically discover and extract information from the web documents which can be structured, unstructured or semi structured from. XML has become very popular for representing semi structured data and a standard for data exchange over the web. The data based on XML is self described; it can be exchanged and handled without internal description. The core technique of semantic web mining is ontology. Ontology represents a set of precisely defined terms about a specific domain and accepted by this domain's community, ontology is an explicitly specification of a conceptualization. The RDF is a simple meta model for defining and exchanging information on 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.

In our future work the main task is to fulfill the domain ontology and establish an operational ontology base with more granularities to realize intelligent interaction between user and machine and cooperation among computers. We are planning to implement semantic web mining in parallel distributed environment in all tiers for decision making, and increasing speed and efficiency.

7. ACKNOWLEDGMENTS

8. REFERENCES
[1] Ian Horrocks, Peter F. Patel Schneider, Three Theses of Representation in the Semantic Web, ACM, 2003

[2] Stefan Decker, Frank van Harmelen, Jeen Broekstra,Michael Erdmann, Dieter Fensel, Ian Horrocks , Michel Klein, Sergey Melnik ,The Semantic Web - on the respective Roles of XML and RDF, IEEE 2006

[3] The semantic web: Roles of XML and RDF, Stefan Decker And Sergey Melnik, Frank Van Harmelen, Dieter Fensel, And Michel Klein Jeen Broekstra Michael Erdmann Ian Horrocks, IEEE Internet Computing, October 2000, vol. 15, nr. 3, pgs. 63--74.

[4] Nigel Shadbolt and Wendy Hall, Tim Berners-Lee, The Semantic Web Revisited, Internet Computing, 2006

[5] Kiavash Bahreini and Atilla Elici,SDISSASA: A multiagent-Based web mining via semantic access to Web resources in Enterprise Architecture,IEEE 2008

[6] Mao-sang Lin, Hui Zhang, Zhang Gua Yo, An ontology for supporting data mining process, IMACS CESA, 2006

[7] Pance Panov, Larisa N. Soldatova, OntoDM: An Ontology of Data Mining,IEEE International Conference on Data Mining Workshops, 2008

[8] Zhang Yofeng, Wu Jinhong, Wang Cuibo, Automatic Competitive Intelligence Collection Based On Semantic web mining IEEE 2007

[9] Chang-xin Song, Ke Ma ,Applications of Data Mining in the Education resource based on XML, International Conference on Advanced Computer Theory and Engineering, 2008

[10] Kiavash Bahereini and Atilla Elci ,A New Software Architecture for J2EE enterprise environments via semantic access to web mining by distributed intelligent software agents, IEEE international software and Applications conference, 2008

[11] Weider D. Yu, Soumya R. Jonnalagadda ,Semantic web and mining in healthcare, IEEE 2006.

[12] Categorization and Clustering of XML Documents, Ludovic Denoyer and Patrick Gallinari Report on the XML Mining Track at INEX 2007.

[13] An Immune-based Model for Web Data Mining Wang Feng Xuwei Li Zhu Hong (© 2005 IEEE Xplore

[14] An Evolutionary Algorithm to Optimize Web Document Retrieval André L. Vizine1, Leandro N. de Castro1 & Ricardo R. Gudwin2 @ 2005 IEEE.

[15] Jung-Won Lee, Kiho Lee, Won Kim; Preparations for Semantics- Based XML Mining in @2001 IEEE Xplore

[16] A Flexible Structured-based Representation for XML Document Mining Anne-Marie Vercoustre, Mounir Fegas, Saba Gul, and Yves Lechevallier; rXiv:cs/0607012v1 [cs.IR] 5 Jul 2006.