ISSN:0975-9646

ONTOCOMM: Ontology Based E-Commerce Application using RDF

S.S. Dhenakaran[#], S.Yasodha[*]

 # Asst. Prof.of Computer Science
Alagappa University
Karaikudi, Tamilnadu, India

** Asso. Prof.of Computer Science*
Govt. Arts College(W),
Pudukkottai, Tamilnadu, India

***Abstract* : The combination of two powerful technologies - the Semantic Web and Web Mining - will probably bring the internet and even the intranet closer to human reasoning than we ever thought possible. The internet is simply viewed as one huge, distributed database just waiting to be made sense of. Preliminary work in transforming this huge corpus of text, images, sound and video is already available. There is still a long way to go until efficient algorithms for automatic conversion of traditional data into ontology will be found.**

In this paper, an ontology based RDF crawler design is proposed. This crawler design, named ONTOCOMM is an Ontology based E-Commerce application. This paper provides a review and the background of Resource Description Framework (RDF), an ontology language. RDF is used to generate the user details and at the administrator side JAVA is used to select and access the E-Commerce sites depending on the query passed. The performance of ONTOCOMM is evaluated by two evaluation metrics, the accessing time and precision. ONTOCOMM minimizes the accessing time and improves precision.

Keywords: **Ontology, Resource Description Framework, RDF Crawler, Semantic Web, Web Mining.**

I. INTRODUCTION

Over the last decade, there is an explosive growth of information available on the Web. Today, web browsers provide easy access to myriad sources of text and multimedia data. More than one billion pages are indexed by search engines, and finding the desired information is not an easy task. One of the fundamental issues regarding the efficiency of information gathering is "overload" [1]. The problem of information overload occurs when a large number of irrelevant documents may be considered to be relevant. The existing approaches of information retrieval and information filtering could be used to solve this problem [2] [6]. The problem, however, is that most of these approaches cannot explicitly interpret user profiles (e.g., the user feedback, or the user log data).

The human ability for information processing is limited on one hand, whilst otherwise the amount of available information in the Web increases exponentially, which leads to increasing information saturation [3]. In this context, it becomes more and more important to detect useful patterns in the Web, thus use it as a rich source for data mining [4]. Web intelligence (WI) is a new direction to push technology towards manipulating the meaning of Web data and creating a distributed intelligence that can actually get things done [5].

Traditionally, the knowledge engineers spend much time in the acquisition of knowledge from domain experts that is the "knowledge acquisition bottleneck" problem. Machine learning and evolutionary computing approaches have all been shown to have niches in which they perform well. In most cases, however, these approaches have either not had the impact of the expert systems produced by the usual knowledge engineering methods; or they have required significant domain expertise for the design of the algorithms, training set, etc [9]. Furthermore, the knowledge acquisition bottleneck in Web-based intelligent information systems becomes particularly difficult, because such systems must have a time-consuming and centralized domain knowledge engineering for ubiquitous information.

The research area of Semantic Web Mining is aimed at combining two fast developing fields of research: the Semantic Web and Web Mining.

The semantic web is a step towards Web intelligence. It is based on languages that make more of the semantic content of the webpage available in machine-readable formats for agent-based computing. One of the components of semantic Web techniques is to use ontology for marking up Web resources and assisting the generation and processing of semantic markup. This need forces the Web users to use more meaningful XML documents and more new semantic markup languages to describe information in ontologies. However, manual ontology construction remains a tedious, cumbersome task that can easily result in a bottleneck for Web intelligence.

The Semantic Web is an intelligent retrieval system, where the information is associated in a way that can be processed easily by machines, worldwide. At present, only human are able to understand the product information that is available online. The emerging semantic web technologies have the potential to extremely influence the further development of the Internet market.

Web mining is a new technology that has emerged as a popular area in the field of Web intelligence. Currently Web mining could be viewed as the use of data mining techniques to automatically retrieve, extract, generalize, and analyze information [7]. It is obvious that data mining techniques ([1] [3] [4]) can be used for Web mining. Web mining, however, is very different from data mining in that

the former is based on Web-related data sources, such as semi-structured documents (HTML, or XML), log, services, and user profiles, and the latter is based on more standard databases. Web mining is used to automatically discover and extract information from Web-related data sources such as documents, log, services, and user profiles. Although standard data mining methods may be applied for mining on the Web, many specific algorithms need to be developed and applied for various purposes of Web based information processing in multiple Web resources, effectively and efficiently.

Ontologies play an important role in supporting knowledge-based applications in the semantic web [11]. They are the backbone of semantic Web, a semantic-aware version of the World Wide Web [12]. Ontology has been widely applied to information retrieval, artificial intelligence, knowledge network and knowledge management [8]. Ontologies ensure efficient retrieval by enabling inferences based on domain knowledge, which is gathered during the construction of knowledge base [10]. The success of the semantic Web depends strongly on the proliferation of ontologies [13]. Ontologies represent a key aspect for the integration of information coming from different sources, for supporting collaboration within virtual communities, for improving information retrieval, and more generally, it is important for reasoning on available knowledge [14]. Thus with the growth of semantic web and the knowledge management systems in the corporate world and other organizations, Ontologies will be vital tools for knowledge sharing. Ontologies are at present, hand-crafted. Thus the challenge is to learn and construct ontologies [15].

The RDF (Resource Description Framework) data model has gained popularity in many domains as a representation format for heterogeneous structured data on the Web. In addition, the growing popularity of knowledge sharing communities such as Wikipedia and the advances in automatic information extraction have contributed to the presence of large general-purpose RDF knowledge bases [16].

In this paper, an ontology based RDF crawler design named ONTOCOMM is proposed. ONTOCOMM is an E-Commerce application which makes use of RDF databases. It is designed with the objective of improving precision and minimizing accessing time.

II. BASIC DEFINITIONS

A. *Semantic Web*

Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in co-operation. The semantic web will provide intelligent access to heterogeneous, distributed information enabling software products to mediate between user needs and the information source available.

Semantic web is a vision of information that is understandable by computers, so that they can perform more of the tedious work involved in finding, sharing and combining information on the web. It comprises the standards and tools of Extensible Markup Language (XML), XML Schema, Resource Description Framework (RDF), RDF Schema and Web Ontology Language (OWL) that are organized in the stack. This provides service in E-Commerce and Business-to-Business applications.

Semantic Web is a future where Web information has exact meaning and the web information can be easily understood and processed by computers. RDF is also responsible for integrating the information from the web.

B. *Semantic web mining*

The human ability for information processing is limited on the one hand, whilst otherwise the amount of available information of the Web increases exponentially, which leads to increasing information saturation[3]. In this context, it becomes more and more important to detect useful patterns in the Web, thus use it as a rich source for data mining [4].

The existing keyword search is inefficient as it provides enormous unwanted information. Ts overcome this difficulty, semantics can be provided for search keywords so as to retrieve more meaningful information.

The research area of Semantic Web Mining is aimed at combining two fast developing fields of research: the Semantic Web and Web Mining. The idea is to improve, on the one hand, the results of Web Mining by exploiting the new semantic structures in the Web; and to make use of Web Mining, on the other hand, for building up the Semantic Web. These two fields address the current challenges of the World Wide Web (WWW): turning unstructured data into machine-understandable data using Semantic Web tools.

As the Semantic Web enhances the first generation of the WWW with formal semantics, it offers a good basis to enrich Web Mining: The types of (hyper)links are now described explicitly, allowing the knowledge engineer to gain deeper insights in Web structure mining; and the contents of the pages come along with a formal semantics, to apply mining techniques which require more structured input.

- *Semantic Web Content and Structure Mining*

In the Semantic Web, content and structure are strongly intertwined. Therefore, the distinction between content and structure mining vanishes. However, the distribution of the semantic annotations may provide additional implicit knowledge. An important group of techniques which can easily be adapted to semantic Web content / structure mining are the approaches discussed as *Relational Data Mining* (formerly called *Inductive Logic Programming* (*ILP*)). Relational Data Mining looks for patterns that involve multiple relations in a relational database. It comprises techniques for Semantic Web Mining like classification, regression, clustering, and association analysis. It is quite straightforward to transform the algorithms so that they are able to deal with data described in RDF or by ontologies.

There are two big scientific challenges in this attempt. The first is the size of the data to be processed (i.e. the scalability of the algorithms), and the second is the fact that the data are distributed over the Semantic Web, as there is no central database server. Scalability has always been a major concern for ILP algorithms. With the expected

growth of the Semantic Web, this problem increases as well. Therefore, the performance of the mining algorithms has to be improved, e.g. by sampling.

As for the problem of distributed data, it is a challenging research topic to develop algorithms which can perform the mining in a distributed manner, so that only (intermediate) results have to be transmitted, and not whole datasets.

- *Semantic Web Usage Mining*

Usage mining can also be enhanced further if the semantics are contained explicitly in the pages by referring to concepts of an ontology. Semantic Web usage mining can for instance be performed on log files which register the user behavior in terms of an ontology. A system for creating such semantic log files from a knowledge portal has been developed at the AIFB. These log files can then be mined, for instance to cluster users with similar interests in order to provide personalized views on the ontology.

C. Ontology

An ontology is a formal specification of a conceptualization, that is, an abstract and simplified view of the world that we wish to represent, described in a language that is equipped with formal semantics. In knowledge representation, an ontology is a description of the concepts and relationships in an application domain. Depending on the users of this ontology, such a description must be understandable by humans and/or by software agents. In many other fields – such as in information systems and databases, and in software engineering – an ontology would be called a conceptual schema.

An ontology defines the terms used to describe and represent an area of knowledge, explicitly. Ontologies are used by people, databases, and applications that need to share domain information. Ontologies include computer usable definitions of basic concepts in the domain and the relationships among them. They encode knowledge in a domain and also knowledge that spans domains [W3C, 2001].

Ontology plays main role in accessing the relevant documents in the semantic web. Moreover it helps in getting the accurate information. Ontology is the heart of the semantic web applications and its metadata representation.

An ontology consists of classes, relationships and attributes. The classes in an ontology are general things (in the many domains of interest). Usually the names of classes are nouns. The properties (or attributes) are those the things may have.

D. Resource description Framework (RDF)

The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. It has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax formats. The W3C published a specification of RDF's data model and XML syntax as a Recommendation in 1999.

RDF is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data to be changed. RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple"). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.

E. RDF Crawler

RDF Crawler is a stand-alone application, which is given URIs and builds an RDF database from it (or increments an existing database).

Ontology servers and other tools dealing with meta information sometimes need to retrieve facts describing resources on the Web. The current standard of making statements about Web resources is RDF (Resource Description Framework), and there are a few more standards which build on top of the RDF, e.g. RDFS and OIL. Therefore we may need a utility to download RDF information from all over the Internet. This utility will be henceforth called **RDF Crawler**. It is a tool which downloads interconnected fragments of RDF from the Internet and builds a knowledge base from this data. At every phase of RDF crawling we maintain a list of URIs to be retrieved as well as URI filtering conditions (e.g. depth, URI syntax), which we observe as we iteratively download resources containing RDF. To enable embedding in other tools, RDF Crawler provides a high-level programmable interface (Java API). RDF Crawler utility is just a wrapper around this API - either a console application, or a windows application or a servlet.

III. METHODOLOGY

Proposed Method

There exists a large number of E-Commerce websites which are used for various purposes. The aim of the proposed method is to design an RDF crawler for E-Commerce application domain that reduces the accessing time of users who visit some websites frequently. The proposed method stores the users' access information in RDF files and display the previous access details. In the subsequent accesses, users are also taken directly to the previous page accessed without having to visit so many pages.

ONTOCOMM works in two phases:

1. The first phase is the store phase. The store phase stores facts as ontology in the form of RDF templates.
2. The second phase is the search phase. This phase searches for the previous access log of users by validating the user-id.

During the first phase, the server analyzes the user's login particulars to identify some significant attributes. These details are stored in the RDF file.

During the second phase when the user enters the web, the server accepts the user's request and extracts the values of significant attributes. These values of attributes are used as the key for matching of facts.

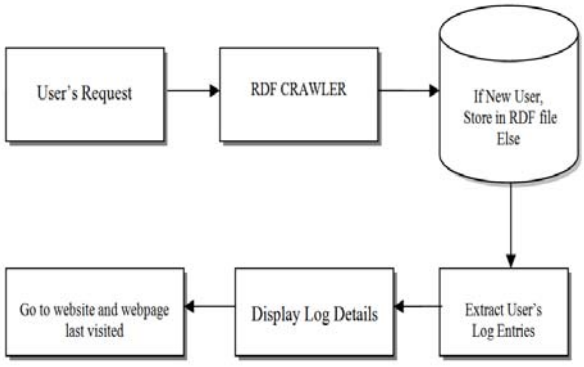

Figure-1 ONTOCOMM Design

The algorithm and RDF description of ONTOCOMM are given below:

The store-and-search Algorithm

First phase: The store phase
Begin
The Server Accepts the User's Request
Send the request to the RDF crawler
RDF crawler analyzes the User Behavior
Check the RDF file.
If the behavior already exists
then go to Search_Phase
Else
Store the user behavior in the RDF file.
End If
End

Second phase : The search phase
Begin
Extract log details from the RDF file
Display log details.
Go to the website and webpage last visited.
Update the RDF File with the current details.
End

The RDF file descriptions are given below:

```
The Store Phase:
<?XML Version="1.0"?>
<!—Store phase-->
<login>
<!-Store Login particulars of the User>
<entry>
<UserID> [User's ID]</UserID>
<Password>[PASSWORD]</Password>
</entry>
</login >
<!Accept user's request and send to RDF crawler>
<Fact table>
<!—Store fact table entries-->
<entry>
<Fact>[WEBSITE]</ Fact >
<Webpage>[ WEBPAGE]</ Webpage >
<Date>[ DATE]</ Date >
<Time>[ TIME]<./ Time >
<Item >[ PURCHASED_ITEM] </ Item >
<Price>[PRICE] <./ Price>
</entry>
</Fact table>
```

```
The Search Phase :
<?XML Version="1.0"?>
<!—Accept User's request as search keywords-->
<entry>
<Keyword><USER'S REQUEST></ Keyword >
<!—If User's request matches with the values of
stored attributes then extract facts-->
<Output>
<Fact>[WEBSITE]</ Fact >
<Webpage>[ WEBPAGE]</ Webpage >
<Date>[ DATE]</ Date >
<Time>[ TIME]<./ Time >
<Item >[ PURCHASED_ITEM] </ Item >
<Price>[PRICE] <./ Price>
</Output>
<!—Go to the Website and Webpage last visited>
</entry>
```

IV. EXPERIMENT SETUP

In a large application, if the user wants to retrieve relevant information for an application, it would become intangible when the user sends the request through Keywords. So in this paper, a Semantic Web Mining model is used to provide the necessary service to the users and fulfill their requirements. It can be applied in many fields such as Web Store, Web Shopping, E-Commerce, E-Jobs, E-Banking and so on. In this paper, an E-Commerce application is considered.

Consider a large Website for E-Commerce which is used by several users for electronic purchase. In such websites, it is the necessary task to store the information about the arrival and departure of the user and also the log maintenance. In the proposed design, this is achieved using RDF. Using the RDF file, the web can automatically analyze the user's log entries and display their previous access details and take them to the website and webpage last visited.

The proposed design is tested with 50 users who visited the website *amazon.com* for the purchase of different category of books, 50 users who visited *ebay.in* for online shopping and 50 users who visited *adidas.com* for the purchase of sports items. The accessing time for each item is measured and the precision of retrieval is evaluated. Their accessing time in subsequent visits is very much reduced as compared to their first visit.

V. SYSTEM PERFORMANCE EVALUATION

The effectiveness of the proposed design is evaluated using the performance metrics, Accessing time and Precision. Precision is the ratio of relevant information retrieved to the total retrieved.

$$\text{Precision} = \frac{\text{Number of relevant retrieved}}{\text{Total number retrieved}}$$

The proposed design is implemented in JAVA with the log entries stored in RDF files. The performance evaluated by using the two metrics, Accessing time and Precision are tabulated below:

Table 1: Performance of ONTOCOMM

Website	Accessing Time (in seconds)	Precision (in percentage)
amazon.com	0.094	96.1
ebay.in	0.168	92.7
adidas.com	0.189	90.4

V. CONCLUSION

The purpose of this paper was not to give an extensive coverage of semantic web mining, but rather to give a general overview of the possibilities that this area opens to future research and applications. We have shown how to use ontology in the E-Commerce domain using RDF. In this paper an ontology based RDF crawler design was proposed and tested using the E-Commerce websites. This may be further extended to other domains like E-Learning, E-Banking and health care.

REFERENCES

[1] N. R. Jennings, K. Sycara and M. Wooldridge, A Roadmap of agent research and development, *Autonomous Agents and Multi-Agent Systems*, 1998, **1(1):** 7-38.

[2] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, Addison Wesley, 1999.

[3] M.-S. Chen, J. Han, and P. S. Yu, Data mining: an overview from a database perspective, *IEEE Transactions on Knowledge and Data Engineering*, 1996, **8(6):**866-883.

[4] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R.Uthrusamy, eds., *Advances in knowledge discovery and data mining*, Menlo Park, California : AAAI Press/ The MIT Press, 1996.

[5] E. A. Feigenbaum, How the "what" becomes "how", *Communications of the ACM*, 1996, **39(5):** 97-104.

[6] D. A. Grossman and O. Frieder, *Information retrieval algorithms and heuristics*, Kluwer Academic Publishers, Boston, 1998.

[7] S. K. Pal and V. Talwar, Web mining in soft computing framework: relevance, state of the art and future directions, *IEEE Transactions on Neural Networks*, 2002, **13(5):** 1163-1177.

[8] Lei, Wang; Kuanjiu, Zhou & Peng, Qiu. Automatic domain ontology construction. *J. China Soc. Sci. Tech. Inf.*, 2010, **29**(1), 45-52.

[9] J. Hendler and E. A. Feigenbaum, Knowledge is power: the semantic Web vision, in N. Zhong et al., eds., *Web Intelligence: research and development*, LNAI 2198, Springer-Verlag, 2001.

[10] Saravanan, Ravindran & Raman. Improving legal information retrieval using an ontological framework. *Arti. Intell. Law*, 2009, **17**(2), 101-24.

[11] Jiang & Tan. CRCTOL: A semantic-based domain ontology learning system. *J. Amer. Soc. Inf. Sci. Technol*., **61**(1), 150-68..

[12] De Nicola, Missikoff & Navigli. A software engineering approach to ontology building. *Information Systems*, 2009, **34**(2), 258-75..

[13] Lin & Hanqing. Research on semi-automatic construction of domain ontology based on statistical NLP technique. *J. China Soc. Sci. Tech. Inf.*, 2009, **28**(2), 201-07.

[14] Gaeta, Orciuoli & Ritrovato. Advanced ontology management system for personalised e-learning. *knowledge-based Systems*, 2009, **22**(4), 292-301.S.

[15] Bettina Berendt , Andreas Hotho , and Gerd Stumme Towards Semantic Web Mining, *Springer-Verlog*, 2002.

[16] Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Ieong. Diversifying search results. In Proceedings of the Second ACM International Conference on Web Search and Data Mining, WSDM '09, pages 5–14, New York, NY, USA, 2009. ACM.