ISSN No. 0976-5697

Volume 1, No. 3, Sept-Oct 2010

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Business Intelligence using Semantic Web

S. Yasodha*
Assistant Professor in Computer Science
Govt. Arts College for Women (A), Pudukkottai India
yasodha_karthik@yahoo.co.in.

Dr. S.S. Dhenakaran,
Asst. Professor in Computer Science
Alagappa University,
Karaikudi, Tamilnadu India

Abstract: Computers have become an essential thing in our day-to-day life. Currently, computers have changed from single, isolated devices to entry points in order to exchange information on the World Wide Web (WWW). Human beings are capable of using the web to carry out a task. However, a computer cannot accomplish the same task without the human direction, because web pages are designed to be read by the people, not by the machines. Usually, human beings interact with the web to gather information or to improve their business intelligence. The information is accessed by giving a related keyword to search. This keyword is used by the server to provide the relevant information. The server provides only limited information for the given keyword. But there exists a lot of information but it is not provided to the user due to the lack of words in the given keyword. So searching becomes intangible to fulfill the requirements of the user in some situation. This kind of searching is made with the presence of human beings. To overcome this problem and to reduce the human effort, a new method is employed in this paper. This new method is termed as "**SEMANTIC WEB**".

The Semantic Web is an extension of the WWW. This is used to carry out the task directly by the computers instead of by human beings. The computer generates the Machine-Processable Information. The Semantic Web provides the automated schemes for gaining more relevant information by analyzing the behavior of users who currently use the web. To store the information, the Semantic Web uses RDF (Resource Description Framework). RDF stores information in the form of XML. With the help of Semantic Web, the users can access the web and gather more information with less effort. In this paper two models are proposed, the User Model and the Query Model for providing service to the users and they are illustrated using an online transaction application.

Keywords: Automated Schemes, Business Intelligence, Extension of WWW, Intangible, Machine-Processable information, Semantic Web.

I. INTRODUCTION

In this Internet world, clients use websites for gathering information to acquire knowledge or to do transactions or to improve business. Users gain information by sending request to the server in the form of keywords and get the response.

This leads to a new challenging task that the keyword search becomes inefficient to retrieve data from a huge amount of data in the warehouse. This challenge can be overcome when the information becomes Machine-Processable Information. When the information is machine-processable, computers can automatically analyze and retrieve all the information needed by the user.

Machine-Processable Information can point the search engine to the relevant pages and can thus improve both precision and recall. For this kind of action, Semantic Web can be used.

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.

Semantic Web is

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.

Semantic Web uses RDF to store the information in the form of XML Language. The **Resource Description Framework (RDF)** is a standard for describing Web resources such as the title, author, date of modification, content and copyright information of a Web page. A web pages database is an application that allows users to create an RDF database in the browser without any knowledge necessary.

1. RDF is a framework for describing resources on the web.
2. RDF is designed to be read and understood by computers.
3. RDF is not designed for being displayed to people.
4. RDF is written in XML

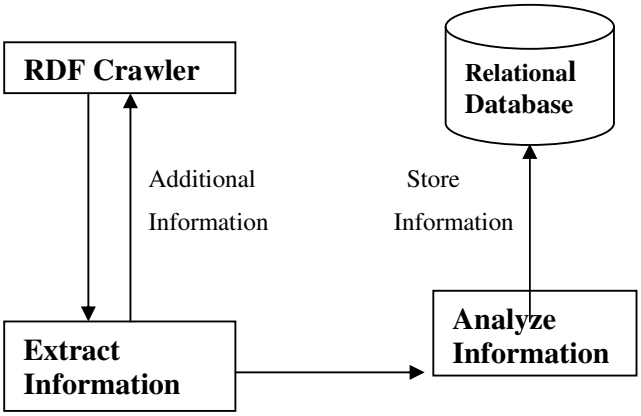

Figure-1 RDF Crawler Design

In Semantic Web, lot of information can be stored in RDF in an XML file format. This information can be extracted by the users depending upon their needs. This can be done by accepting the user request and providing response to the user by extracting the information from the RDF. Thus the information can be easily extracted by the Web.

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

Semantic Web is the advanced technique used by the web to fulfill the client's requirements. In this paper, some application areas are described for this new technology, and the ongoing works in the fields of knowledge management and Electronic Commerce are also focused. A detailed survey on the Semantic Web is also presented in this paper.

II. RELATED WORK

In paper [1], Two new and fast-developing domains are presented: Semantic Web and Data Mining. The authors suggest how these areas can be combined and they present three different approaches to semantic web mining: an approach to recurring pattern mining, a text classification algorithm called AdaBoost and a framework for generating better customized content on the web by using web mining combined with embedded ontologies.

In paper [2], the authors provide a discussion on a technical solution which is aimed at improving the semantics. The CORPORUM tool set that is developed for this task exists for a set of programs that can fulfill a variety of tasks, either as 'stand-alone', or augmenting each other. The aim of the semantic web is not only to enhance the precision and recall of search, but also to enable the use of logical reasoning on web contents.

In paper [3] the authors argue that even more important than semantics is pragmatics; that is, to really enhance web usability it is critical to capture and react to aspects of the end use context. Most centrally, to make the web truly responsive to human needs, they need to understand the "users" of the web and their purposes for using it. In this paper they elaborate this argument in the context of e-learning systems. They propose an approach to the design of e-learning systems which they call the ecological approach. Moving from the open web to repositories of learning objects, they show how the ecological approach shows promise not only to allow information about learners' actual interactions with learning objects to be naturally captured but also to allow it to be used in a multitude of ways to support learners and teachers in achieving their goals. The approach involves attaching models of learners to the learning objects they interact with, and then mining these models for patterns that are useful for various purposes. The ecological approach turns out to be highly suited for e-learning applications. It also has interesting implications for e-learning research, and perhaps even directions for semantic web research.

In the past few years, there have been many attempts at "breaking the syntax barrier" on the web. A number of them rely on the semantic information in text corpora that is implicitly exploited by statistical methods. Some methods also analyze the structural characteristics of data; they profit from standardized syntax like XML. In this paper, they concentrate on markup and mining approaches that refer to an explicit conceptualization of entities in the respective domain. These relate the syntactic tokens to background knowledge represented in a model with formal semantics. When they use the term "semantic", they thus have in mind a formal logical model to represent knowledge.

The aim of the paper [4] is to give an overview of where the two areas of Semantic Web and Web Mining meet today. In their survey, the authors first describe the current state of the two areas and then discuss, using an example, their combination, thereby outlining future research topics. They also provide references to typical approaches. Most of them have not been developed explicitly to close the gap between the Semantic Web and Web Mining, but they fit naturally into this scheme.

The challenge of the Semantic Web Mining technologies in the e-Learning domain can relate to the provision of personalized experiences for the users. Particularly, these applications can take into consideration the individual needs and requirements of learners. In paper [5], the authors propose a framework for personalized e-Learning based on aggregate usage profiles and a domain ontology. They have distinguished two stages in the whole process, one is offline tasks that includes data preparation, ontology creation and usage mining and the other is online tasks that concerns the production of recommendations.

Providing semantic web personalization needs to tackle the technical issues on how to define web access activities, discover hierarchical relationships from web access activities,

transform them into ontology automatically, and deduce personalized usage knowledge from the ontology. The paper [6] proposes a web usage mining approach for semantic web personalization. The proposed approach first incorporates fuzzy logic into Formal Concept Analysis [7] to mine client-side web usage data for automatic ontology generation, and then applies fuzzy approximate reasoning [8] to generate personalized usage knowledge from the ontology.

With the explosive growth of information available on the World Wide Web, it has become more difficult to access relevant information from the Web. One possible approach to solve this problem is web personalization [9]. Web usage mining [10], which aims to discover interesting and frequent user access patterns from web usage data, can be used to model past web access behavior of users. The acquired model can then be used for analyzing and predicting the future user access behavior. Semantic Web [11] provides a common framework that allows data to be shared and reused across application, enterprise and community boundaries. In Semantic Web environment, user access behavior models can be shared as ontology. Agent software can then utilize the ontology to provide personalized user services such as recommendation and search. Ontology has become an important component for Semantic Web, as it allows the description of the semantics of web content. Various techniques such as Natural Language Processing (NLP) [12], association rules [13], hierarchical clustering [14] and Formal Concept Analysis (FCA) [15] have been investigated for ontology generation. However, majority of these works have focused on generating concept hierarchy for building ontology from text documents. Recently, semantic web personalization [16-18] has become an active research area. However, the current research works create ontology manually and are unable to deal with temporal access behavior. Further, most of them investigated the problem mainly for a specific domain such as e-learning [16,17].

Several other researchers have recognized the value in bringing semantics to email. For example, the Information Lens system1 lets users send semi structured email messages and filter those messages using production rules. Users can send to a special mailbox called “anyone,” and anyone can choose to receive messages from this mailbox based on production rules. This flips the nature of widely broadcast emails on its head. Instead of starting with receiving all emails and whittling down based on filtering rules, the user starts with an empty inbox and pulls in emails of interest. This is similar to the RSS subscription model. As RSS feeds contain more semantic information, the semantic subscription model exemplified by Information Lens might become more commonplace.

More recently, MailsMore2 lets users annotate an e-mail’s content with Resource Description Framework (RDF) triples and automatically includes RDF triples based on standard email headers such as the “To,” “From,” “Subject,” and body fields. This can be used for semantic filtering and filing of emails. The Mangrove system3–5 takes this idea further. It allows not only structured email content but also semantic email processes. Users can script email clients with declarative workflows that automatically aggregate information obtained from many email responses, automatically resend emails to people who haven’t responded, or analyze the semantic content of incoming email messages and respond accordingly. Most relevantly, Microsoft Exchange 2003 lets administrators create query-based distribution groups, which are essentially mailing lists whose recipients are based on a Lightweight Directory Access Protocol (LDAP). This alleviates much of the administrative work required to maintain a mailing list. However, because only an administrator can create the mailing lists, users can not send SEA mail, and the information upon which the lists are based is not under users’ control. None of this application’s functionality is available to the users and very little to the administrators. In fact, users can’t see that a distribution group is query based: each query based distribution group has a name, so to an outsider looks like a regular mailing list.

The research area of Semantic Web Mining is aimed at combining two fast developing fields of research: the Semantic Web and Web Mining. These two fields address the current challenges of the World Wide Web (WWW): turning unstructured data into machine-understandable data using Semantic Web tools, and automatically extract knowledge hidden in the vast amounts of Web data using Web Mining tools [19]. Semantic Web Mining is a convergence of these two fields, where the tools of the Semantic Web can be used to improve Web Mining and vice versa. For example, in the vast quantities of data, Web Mining can discover semantic structures to build semantics for the Semantic Web.

Similarly, semantic structures can improve the task of mining by allowing the algorithms to operate on certain semantic levels or choose appropriate levels of abstraction. Semantic Web Mining can then be thought of as Semantic (Web Mining) or (Semantic Web) Mining to cover the spectrum of topics [19].

The paper [20] summarizes the different characteristics of web data, the basic components of web mining and its different types, and their current states of the art. The reason for considering web mining, a separate field from data mining, is explained. The limitations of some of the existing web mining methods and tools are enunciated, and the significance of soft computing (comprising fuzzy logic (FL), artificial neural networks (ANNs), genetic algorithms (GAs), and rough sets (RSs) highlighted. A survey of the existing literature on “soft web mining” is provided along with the commercially available systems. The prospective areas of web mining where the application of soft computing needs immediate attention are outlined with justification. Scope for future research in developing “soft web mining” systems is explained. An extensive bibliography is also provided.

In paper [21], the authors highlight the challenges faced in the semantic web. The paper [22] discusses what kind of entity the proposed Semantic Web (SW) is, in terms of the relationship of natural language structure to knowledge representation (KR). It argues that there are three distinct views on the issue: first, that the SW is basically a renaming of the traditional AI knowledge representation task, with all the problems and challenges of that task. If that is the case, as many believe, then there is no particular reason to expect progress in this new form of presentation, as all the traditional problems of logic and representation reappear and it will be no more successful outside the narrow scientific domains where

KR seems to work even though the formal ontology movement has brought some benefits. The paper contains some discussion on the relationship of current SW doctrine to representation issues covered by traditional AI, and also discusses issues of how far SW proposals are able to deal with difficult relationships in parts of concrete science.

Secondly, there is a view that the SW will be the World Wide Web with its constituent documents annotated so as to yield their content or meaning structure more directly. This view of the SW makes natural language processing central as the procedural bridge from texts to KR, usually via a form of automated Information Extraction. This view is discussed in detail and it is argued that this is in fact the only way of justifying the structures used as KR for the SW. There is a third view, possibly Berners-Lee's own, that the SW is the foundation of a system of web processes and services, but it is argued that this ignores the whole history of the web as a textual system, and gives no better guarantee of agreed meanings for terms than the other two approaches.

III. METHODOLOGY

Proposed Method

The aim of the proposed method is to develop a web to process the client's request by the machine itself and to reduce human effort.

The summary of this method is as follows: In the internet, there exists more number of websites and are used for various purposes. Some websites are used for purchasing goods and some are used for sales. Based on its purpose, its usability will be differentiated.

For the machine to process by itself, it needs some requirements. One such requirement is to store the user's information and to give response by considering the user behavior. To store the information, the Semantic Web contains RDF. RDF stores information in the form of XML.

In this paper, a method has been proposed considering the online transaction application. For providing service to the user, the Semantic Web requires 2 models:

1. User Model
2. Query Model

These two models are stored in RDF in XML File Format.

The User model contains the details about the users who use the web. The details include the customer's personal details and also the transaction details whereas the Query model contains the criteria to provide service to the user based on the details in the User Model.

Semantic web is useful to manage the customer requirements by itself without any human interaction. So it is necessary to maintain all the details that are required by the computer to satisfy the customer requirements. These details are maintained in the RDF, including the customer's contact information, his previous transactions in the website and so on. And also the RDF can be updated regularly to satisfy all the needs.

When the customer enters into the web, the User Model is executed to identify the user and then execute the Query Model to provide the service. When the transaction is completed, the information also gets updated in the RDF for future references.

The User Model in the RDF file is as follows:

<?XML Version="1.0"?>
<!—Customer Transaction, Last Updated ***[DATE]***-->
<Personal>
<!—Personal Details of the User-->
<entry>
<Name>***[NAME OF THE USER]***</Name>
<VotersID> ***[VOTERS ID]***</VotersID>
<Age> ***[AGE]***</Age>
<Gender>***[Male/Female]***</Gender>
<Address>***[DOOR NO, STREET NAME]*** </Address>
<City>***[CITY NAME]*** </City>
<Pin>***[PIN CODE]***</Pin>
<Telephone>***[TELEPHONE NUMBER]***</Telephone>
</entry>
</Personal>
<Transaction>
<!—User's Transaction Details-->
<entry>
<Date>***[DD-MM-YYYY]***</Date>
<Website>***[WEBSITE ADDRESS]***</Website>
<UserName>***[USERNAME]***</UserName>
<UserID>***[USER ID]***<./UserID>
<Category>***[CATEGORY OF THE ITEM]*** </Category>
<Amount>***[AMOUNT]***<./Amount>
<TransactionID>***[TRANSID]***</ TransactionID>
<Completed>***[Yes/No]***</ Completed>
<Count>***[NO. OF TIMES]***</Count>
</entry>
</Transaction>

The Query Model in the RDF file is as follows:

<?XML Version="1.0"?>
<entry>
<Output>
<Date>***[DD-MM-YYYY]***</Date>
<Criteria>***[CONDITION]***</Criteria>
<Query>***[QUERY]***</Query>
<RequiredField>***[REQUIRED]***</RequiredField>
<Result>***[RESULT]***</Result>
</Output>
</entry>

The proposed method is based on the following algorithm which works based on the given conditions:

A. Algorithm

Begin
The Server Accepts the Client Request
Analyze the User Behavior
Check the RDF's User Model
If the behavior already exists then
    Use the RDF's Query Model
Else
    Store the user behavior in RDF
    Create the Query Model for it.
End If
Provide service to the user based on the User Model and the Query Model.
Update the RDF XML File with the current details.

End

B. Explanation of the Algorithm

The proposed algorithm for providing service to the customer works as follows:

First, the web accepts the client's request. With the help of this request, the server analyzes the existing user behavior. Then it matches the behavior with the User Model. Then it proceeds further, if it exists. Else it creates the new entry for the new user in the User Model.

Then it finds the criteria in the Query model for the user. Based on the Query model, it provides service to the user.

IV. EXPERIMENTAL RESULTS

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

Consider a large Website for the **Web Store**, which can be used by several users for several purposes such as Online booking, Online Banking, Sales and so on. In such case, it is the necessary task to store the information about the arrival and departure of the user and also the transaction details like a history or log maintenance in the web through the RDF.

Using the RDF file, the web can automatically analyze the entering user, and provide the information based on the user's previous transactions. The details are stored in the RDF file.

The details are registered from the user's entry into the website to their exit. The User model also registers whether the entered user performed the transaction successfully or not. If the user did not complete the transaction, then the reason for this action is analyzed and stored. The reason may be the user wants only to visit the website or the user cancelled the transaction at the last moment due the lack of amount. These details are registered in the RDF XML file in the User model.

Based on this User Model, the web provides service to the current user. If the current user completed his previous transaction successfully, the web gives higher priority for that user by giving much preference. If the current user cancelled the transaction, it gives common preference. For those users who just visit the website for many times, the web gives much less preference. Thus based on the transaction category, the web provides its service to the user.

Thus the implementation of the Semantic Web technique is done successfully.

V. CONCLUSION

The aim of the paper is to make the web to provide service automatically through the Semantic Web. The uses of Semantic Web have been illustrated through an online transaction application.

Even though the Semantic Web performs better, it needs little user's help to create the RDF file. But its performance is better since it provides service by itself without human effort. The RDF pointer acts as a pointer to information about the things. The beauty about this is that we don't have to describe the things ourselves. The RDF application will sort it out for us.

Thus the Semantic Web will reduce the human effort and so it becomes more user-friendly and a fast information-provider.

VI. REFERENCES

[1] "Semantic Web Mining" Diana Cerbu* Socrates-Erasmus Student, Faculty of Computer Science, Iasi, Romania Konstanz, February 21, 2008.

[2] "CORPORUM: a workbench for the Semantic Web"-R.H.P. Engels, B.A.Bremdal and R. Jones, Halden, Norway, July 31,2001.

[3] "The Ecological Approach to the Design of E-Learning Environments: Purpose-based Capture and Use of Information About Learners" Gord McCalla, JIME Special Issue on the Educational Semantic Web, May 2004.

[4] "Semantic Web Mining State of the Art and Future Directions", Gerd Stumme, Andreas Hotho, Bettina Berendt, Volume: 4, Issue: 2, Knowledge and Data Engineering Group, University of Kassel, 2006.

[5] "Using Semantic Web Mining Technologies for Personalied E-Learning Experiences", P. Markellou, I. Mousourouli, S. Spiros, and A. Tsakalidis (Greece),Proceeding (461) Web-based Education, 2005.

[6] "Web Usage Mining for Semantic Web Personalization" Baoyao Zhou1, Siu Cheung Hui, and Alvis C. M. Fong, 2006.

[7] "Formal Concept Analysis: Mathematical Foundations", Ganter B., and Wille R., Springer, Heidelberg, (1999).

[8] "Fuzzy Sets and Fuzzy Logic: Theory and Applications", Klir G.J., and Yuan B, Published by Prentice Hall, (1995).

[9] "Web Mining for Web Personalization", Eirinaki M., and Vazirgiannis M, ACM Transactions on Internet Technology, 3(1) (2003) 1-27.

[10] "Web Mining Research: A Survey", Kosala R., Blockeel H, ACM SIGKDD Explorations, 2 (2000) 1-15.

[11] "The Semantic Web", Berners-Lee T., Hendler J., and Lassila O, Scientific American, (2001).

[12] "Using Description Logics for Ontology Extraction", Todirascu A., Beuvron F., Galea D. and Rousselot F, Workshop on Ontology Learning, at the 14th European Conference on Artificial Intelligence, Berlin, (2000).

[13] "Ontology Learning for the Semantic Web", Maedche A., and Staab S, IEEE Intelligent Systems, 16(2) (2001) 72-79.

[14] "Ontology Discovery for the Semantic Web using Hierarchical Clustering", Clerkin P., Cunningham P., and Hayes C, Workshop on Semantic Web Mining, at ECML/PKDD, (2001) 27-38.

[15] "Ontology Merging for Federated Ontologies on the Semantic Web", Stumme G., and Maedche A, Workshop on Ontologies and Information Sharing, at IJCAI, Seattle, USA , (2001).

[16] "The Personal Reader: Personalizing and Enriching Learning Resources Using Semantic Web Technologies", Peter D., Nicola H., Wolfgang N., and Michael S, In Proc. of AH 2004, LNCS 3137, pp. 85–94, (2004).

[17] "Personalization in Distributed e-Learning Environments", Peter D., Nicola H., Wolfgang N., and Michael S, In Proc. of WWW2004, New York, USA, pp. 170-179, (2004).

[18] "A Personal Agent Application for the Semantic Web", Kumar S., Kunjithapatham A., Sheshagiri M., Finin T., Joshi A., Peng Y., and Cost R.S, AAAI Fall Symposium on Personalized Agents, North Falmouth, (2002).

[19] "Semantic Web Mining: State of the art and future directions", Stumme, G., Hotho, A., Berendt, B, Web Semantics: Science, Services and Agents on the World Wide Web 4(2) (2006).

[20] "Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions" Sankar K. Pal, Varun Talwar, and Pabitra Mitra, IEEE Transactions on Neural Networks, Volume: 13 Issue:5, Sep 2002.

[21] "Implementation of Web Housing and Integration of Web Traffic with Traditional Business - Web Mining", AIEPIC Project 2009-2010.

[22] "The Semantic Web as the apotheosis of annotation, but what are its semantics?" Yorick Wilks, IEEE Computer Society, 2008.

[23] "Intelligent Information-1. Sharing Systems", T.Malone et al., Comm. ACM, vol. 30, no. 5, 1987, pp. 390–402.

[24] "Representation Formalisms and Methods", Kalyanpur et al., Tech. report, Univ. Maryland, 2001.

[25] "Semantic Email: Adding Lightweight Data Manipulation Capabilities to the Email Habitat", O.Etzioni et al. Proc. 6th Int'l Workshop the Web and Databases, ACM Press, 2003, pp. 12–13.

[26] "Semantic Email: Theory and Applications", L.Mcdowell, O. Etzioni, and A. Halevy, Web Semantics: Science, Services and Agents on the World Wide Web, vol. 2, no. 2, 2004, pp. 153–183.

[27] "Semantic Email", L.McDowell et al. Proc. 13th Int'l Conf. World Wide Web (WWW 04), ACM Press, 2004, pp. 244–254.