ISSN:0975-9646

A Review on Semantic Web Mining

Amruta Arun Joshi[1], Vrishali A. Chakkarwar[2]

[1]*Research Scholar, Government Engineering College,*
[2]*Asst. Professor, Government Engineering, Dr. B.A.M. University, Maharashtra, India*

***Abstract*—Web mining is being a huge and rich data source. Web mining is the process of extracting useful information from server. Some users might be looking at only textual data, whereas some others might be interested in multimedia data. Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. We just put some keywords relevant to aim of extraction as a request and we get number of pages indexed as per the information. Semantic Web Mining aims at combining the two fast-developing research areas Semantic Web and Web Mining. The amount of ontologies and semantic annotations available on the Web is constantly growing. This new type of complex and heterogeneous graph structured data raises new challenges for the data mining community. In this paper we are going to develop the web mining technology which will be based on ontology and decision tree.**

***Keywords*—Semantic web mining, ontology learning, association rule mining.**

I. INTRODUCTION

Semantic web mining aims at combining semantic web and web mining. Most data on the web are unstructured and they can only be understood by humans. But the data on the web is huge which cannot be processed by humans, so this huge amount of data can be processed only by machine. The semantic web manages the first part of the challenge by making the data machine-understandable while the web mining takes care of the second challenge by automatically extracting the useful knowledge from the available data. Many researchers work on improving the result of web mining by introducing the semantic structure into the web, and make use the web mining technique to build the semantic web.

Semantic web mining is all about the machine understandable web page which makes the web more intelligent and also provides a good service to the user. This means that the content or the information available should be mined so that the machine can understand it easily.

A. Semantic Web

Semantic web is based on a vision of Tim Berners Lee [3]. Semantic based web mining is a combination of two fast growing areas semantic web and web mining. These two fields tell the current challenges of World Wide Web. The current version of WWW is 2.0 which are having some drawbacks due to overhead of information which is often unstructured. Web is rich in information and to retrieve this information data is not well organized and structured. To obtain human readable information in structured format the technology named semantic web can be used where efficiently and effectively data can be organized in machine understandable way.

Semantic web is made of different layers. The bottom most layer is the World Wide Web and the next layer is the XML, RDF, Ontology, logic, proof and trust [3].

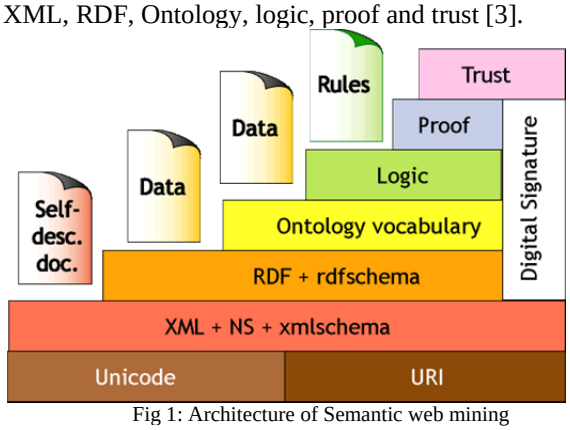

Fig 1: Architecture of Semantic web mining

i. Unicode: an international encoding standard by which each letter, digit, or symbol is assigned a unique numeric value that applies across different platforms and programs.
ii. URI: **Uniform Resource Identifier** (**URI**) is a string of characters used to identify a name of a resource. It is named as locator.
iii. XML: **Extensible Markup Language** (**XML**) is a markup language that defines a set of rules for encoding documents in a format which is both human-readable and machine-readable. XML only carries information and it contains user defined tags. XML layer only provides structure of a data on web.
iv. RDF: **Resource Description Framework (RDF)** framework for describing resources on web. It is visible to machine not to people. RDF is defined in XML. RDF as three parts: subject, predicate, and object.
v. Ontology: Ontology is an agreed vocabulary that provides a set of well-founded constructs to build meaningful higher level knowledge for specifying the semantics of terminology systems in a well-defined and unambiguous manner.
vi. Logic and Proof: An (automatic) reasoning system provided on top of the ontology structure to make new inferences. Thus, using such a system, a software agent can make deductions as to whether a particular resource satisfies its requirements or not (and vice versa).
vii. Trust: The purpose of final layer of the layered architecture is to know trustworthiness of the information by asking questions in Semantic Web. This assures the quality of that information. The Semantic Web uses in reasoning while searching

towards the exactness on web data for the search query hence we can say that Semantic Web helps the Web machine process better. In this paper we are interested in finding out the reasoning from the grammar as used by Semantic Web.

B. Web Mining

Web mining is the use of data mining to discover and extract information from web pages. In simple web mining is collecting interesting information from World Wide Web. The classification of web mining techniques represented in below Figure [3].

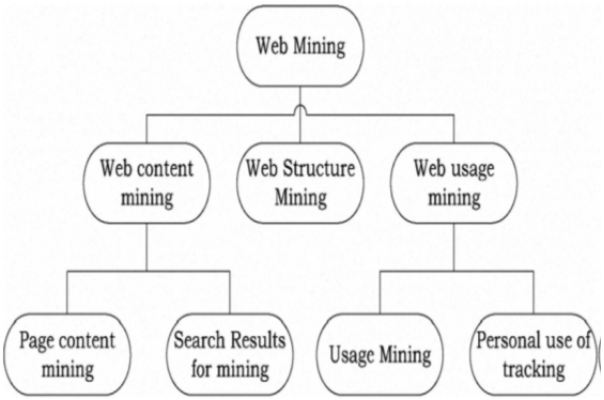

Fig 2: Classification of web mining

i. Web Content mining: Web Content Mining is the process of extracting information from the contents of Web documents. It examines content of the web pages as well and web searching. Content data corresponds to the collection of facts a Web page was designed to convey to the users. Web content may be unstructured (plain text), semi- structured (HTML documents), or structured (extracted from databases into dynamic Web pages). Such dynamic data cannot be indexed and consist what is called "the hidden Web". A research area closely related to content mining is text mining.

ii. Web-Structure mining: It mainly operates on hyperlink structure. Web-structure mining focus on mining set of pages ranging from a single website to a web as whole. Usually web content mining and web structure mining are used together.

iii. Web-Usage mining: It mainly focus on the request made by the visitor or user, this information are mostly collected in web server log. Web content mining and web structure mining focus the primary data on the web page while the web usage mining uses the secondary data derived from the user's interaction with the web page. This includes cookies, registration data, bookmark, user profiles.

C. Semantic web mining

Semantic web mining is a combination of two fast growing technologies semantic web and web mining. The Semantic Web is a Web of data. There is a lot of data we all use every day, and it's not part of the Web. The vision of the Semantic Web is to extend principles of the Web from documents to data. Data should be accessed using the general Web architecture using, e.g., URI-s; data should be related to one another just as documents (or portions of documents) are already. This also means creation of a common framework that allows data to be shared and reused across application, enterprise, and community boundaries, to be processed automatically by tools as well as manually, including revealing possible new relationships among pieces of data. Different aspects used in semantic web mining are shown in figure [4].

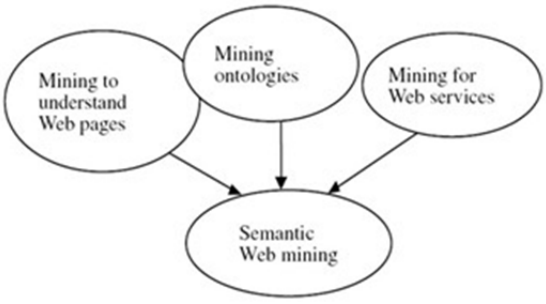

Fig 3: Aspects of semantic web mining

A. Mining to understand the web page

One way to acquire required knowledge for semantic web is to enhance the content of the web. In the past few years the World Wide Web is grown in large and consists of huge distributed database which consist of text, video, audio, image and documents. For semantic web one must consider this huge distributed database. Semantic knowledge is not efficient for semantic web. From the three web mining techniques web content mining plays an important role in gathering knowledge from developing semantic web.

B. Ontology Mining

The relationship between ontologies and web mining techniques is mutual relationship. Mining web content can develop ontologies e.g. mining data in a particular field can be used to find related concept in that field and add them to the ontology. Ontologies can help to improve the intelligence of the current web. The content that is for ontologies can be used for better understanding of the web page.

C. Web service and Web mining

Web mining and data mining can be used to develop a web service. Data mining can be used to extract resource from a web page. Data mining can be used as an intermediate between client and the server. Whereby web service and web mining are two different components, a combination of this two will bring a potential web service technology.

II. ONTOLOGY LEARNING

The term ontology is to mean a *specification of a conceptualization*. Ontologies are (meta) data schemas, providing a controlled vocabulary of concepts, each with an explicitly defined and machine processable semantics. By defining shared and common domain theories, ontologies help both people and machines to communicate concisely, supporting the exchange of semantics and not only syntax [1].

Ontology Learning aims at the integration of a multitude of disciplines in order to facilitate the construction of ontologies, in particular machine learning. Because the fully automatic acquisition of knowledge by machines remains in the distant future, we consider the process of

ontology learning as semi-automatic with human intervention, adopting the paradigm of balanced cooperative modeling for the construction of ontologies for the Semantic Web. This objective in mind, we have built an architecture that combines knowledge acquisition with machine learning, feeding on the resources that we nowadays find on the syntactic Web, viz. free text, semi-structured text, schema definitions (DTDs), etc.

Ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where Ontology is a systematic account of Existence. For AI systems, what "exists" is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, in the context of AI, the ontology of a program can be defined by a set of representational terms. In such ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms. Formally, ontology is the statement of a logical theory. Common ontologies are used to describe ontological commitments for a set of agents so that they can communicate about a domain of discourse without necessarily operating on a globally shared theory. An agent commits to ontology if its observable actions are consistent with the definitions in the ontology.

A. The Relationship between Semantic Web and Ontology

Ontology and the Semantic Web strive to express and enable semantic relations among represented entities. Semantic relations are meaningful associations between two or more concepts, entities, or sets of entities (Khoo and Na, 2006). As a new information representation system, ontology aims to substantiate the rich variety of semantic relations among the concepts it represents – a characteristic that distinguishes it from other representation and organization systems. Hodge grouped typical information representation systems into three general categories: term lists, classifications and categories, and relationship lists. Term lists emphasize lists of terms usually presented with definitions. Classifications and categories emphasize the creation of subject sets. Relationship lists emphasize the connection between terms and concepts.

III. APPLICATIONS OF SEMANTIC WEB MINING

Semantic Web technologies can be used in a variety of application areas; for example: in data integration: whereby data in various locations and various formats can be integrated in one, seamless application.

Resource discovery and classification: to provide better, domain specific search engine capabilities; in cataloging for describing the content and content relationships available at a particular Web site, page, or digital library.

Intelligent software agents: to facilitate knowledge sharing and exchange.

Describing collections of pages: represent a single logical "document"; for describing intellectual property rights of Web pages.

At present, the Semantic Web is increasingly used by small and large business. Oracle, IBM, Adobe, Software AG, or Yahoo! are only some of the large corporations that have picked up this technology already and are selling tools as well as complete business solutions. Large application areas, like the Health Care and Life Sciences, look at the data integration possibilities of the Semantic Web as one of the technologies that might offer significant help in solving their R&D problems [4].

IV. FUTURE SCOPE

In semantic web mining lot of work has been done. But there is no any perfect search engine is developed which retrieves semantics from huge information. In our project some of the drawbacks can be removed using decision tree and association rule mining. Due to which data is divided by making decision and then appropriate data is mined by association rule.

V. CONCLUSION

Semantic web mining is a new area in web mining. The combination of these two areas will bring a great success to World Wide Web. But due to the lack of global standards and lack of rugged database management system to manage semantic web mining opens up new avenues for the researchers to develop KIMS (Knowledge extraction management system) for unstructured data available on the web this area is slowly developing. If these fields explored in a right manner it will provide unlimited opportunities to extract knowledge from the goldmine of unstructured data available across the globe.

REFERENCES

[1] C.S.Bhatia 1 Research Scholar and Dr. Suresh Jain2 Supervisor, Department of Computer Engineering, Mewar University, Chittorgarh "Semantic Web Mining: Using Ontology Learning and Grammatical Rule Inference Technique"2011 IEEE.

[2] Nitesh R Pathak Assistant Professor, Thadomal Shahani Engineering College, Bandra (West) "Semantic Web Mining: Using Ontology Learning" Volume: 3 | Issue: 9 | September 2014.

[3] Sivakumar J#1, Ravichandran K.S*2 "A Review on Semantic-Based Web Mining and its Applications" International Journal of Engineering and Technology (IJET).

[4] Ramkumar.R (MCA) Pope John Paul II College of Education "AN INTEGRATION OF SEMANTIC WEB IN WEB MINING"

[5] Mahendra Thakur1, Geetika S. Pandey2 "Performance Based Novel Techniques for Semantic Web Mining" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012, ISSN (Online): 1694-0814.

[6] Victoria Nebot ⇑, Rafael Berlanga "Finding association rules in semantic web data" Knowledge-Based Systems 25 (2012) 51–62.

[7] Aarti Singh, Associate professor, MMICT&BM, M.M. University, Mullana, Hariyana. "Agent Based Framework for Semantic Web Content Mining". International Journal of Advancements in Technology ISSN 0976-4860.

[8] Sumaiya Kabira,b, Shamim Ripona*, Mamunur Rahmanb and Tanjim Rahmanb "Knowledge-Based Data Mining Using Semantic Web" IERI Procedia 7 (2014) 113 – 119.