International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 03 , March 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

ROLE OF ONTOLOGY IN SEMANTIC WEB MINING

Mamta, Research scholar

ABSTRACT

The Semantic Web (SW) has been the vision for the next generation of the web, where information is desired to be useful not only for the people but also for the computers. Semantic Web Mining aims at combining the two areas Semantic Web and Web Mining. Web Mining is contributing towards the development of knowledge. Semantic web is an extension of current web and web mining technique is an extension of data mining technique. Semantic interpretability arises whenever two queries do not share the similar kind of information and numerous intended meanings are related with the same word. A solution to this problem is provided by the third basic component of the Semantic Web, collections of information called ontology.

Keywords: Semantic interpretability, Semantic web architecture, ontology

I. INTRODUCTION

With the rich development of information in online, the World Wide Web (WWW) is a productive area of data mining. The demand of World Wide Web (WWW) has made it a fertile ground for exhaust information. Due to the attributes of huge, heterogeneous and dynamic and unstructured nature of web data, web data research has faced a lot of challenges for data mining principles, or web mining. The web mining include a wide array of issues, aimed at retrieving actionable knowledge form the web, and comprise researchers from information retrieval, database technologies, and artificial intelligence [1]. Internet, Web and distributed computing infrastructures continue to gain in popularity as a means of communication for organizations, groups and individuals alike. In such an environment, characterized by large distributed, autonomous, diverse, and dynamic information sources, access to relevant and accurate information is becoming increasingly complex. This complexity is magnified by the evolving system, semantic and structural heterogeneity of these potentially global, cross-disciplinary, multicultural and rich-media technologies. Clearly, solutions to these challenges require addressing directly a variety of interoperability issues [2]. To accomplish this interoperability between ambiguous information systems is extremely monotonous, complex and error-prone task.

Therefore, the potential for research perspective in web management & enhancements by developing a standard, adjustable but perspective, adaptive and distributed framework for the support of dynamic and heterogeneous infrastructure is apparent.

Over the last decade, there is an explosive growth in the information available on the World Wide Web (WWW). Today, web browsers provide easy access to large 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. This heap of resources has prompted the need for developing automatic mining techniques on the WWW, thereby giving rise to the term "Web Mining". The "Semantic Web" aims to address this problem by providing machine interpretable semantics to provide greater machine support for the user. These two areas pave way for the extraction of relevant and meaningful information from the web, thereby giving rise to the term "Semantic Web

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Mining" [3]. Semantic Web Mining aims at combining the two fast-developing research areas Semantic Web and Web Mining. This survey analyzes the convergence of trends from both areas: Growing numbers of researchers work on improving the results of Web Mining by exploiting semantic structures in the Web, and they use Web Mining techniques for building the Semantic Web. Last but not least, these techniques can be used for mining the Semantic Web itself. [4].

Semantic Web Mining aims at combining the two areas Semantic Web and Web Mining by using semantics to improve mining and using mining to create semantics. Web Mining aims at discovering insights about the meaning of Web resources and their usage In Semantic Web, the semantics information is presented by the relation with others and is recorded by technology.

A user-oriented semantic search engine is the need of today. These fields if explored in a right manner will provide unlimited opportunities to extract knowledge from the data available across the globe. **[5].** Most of the data on the Web is still in the form of unstructured text. Knowledge extraction from unstructured text is highly desirable but extremely challenging due to the inherent ambiguity of natural language. But from different domains to find a proper domain that can exploit the maximum benefits of this approach, as well as to learn the typical patterns for the rewriting and reconstruction rules is the future work. [8] This idea is to make World Wide Web (WWW) intelligent, efficient and machine readable by providing tools to find, exchange and interpret information to a limited extent by adding metadata.

This paper is related to the field of Semantic Web Mining. In particular, we analyze and compare the techniques under which data stored in web search engines' logs to discover usage patterns, and the aim is to enhance efficiency of search tools as well as to help users to find information on the web.

II. LITERATURE SURVEY

The foreground of semantic web mining is from artificial intelligence. The Semantic Web vision given by **Tim Berners-Lee et. al. [1]**, which is currently supported by the World Wide Web consortium, is quite determined and has to be gradually realised (and in particular outreached to industry) in the long term. Thus, this grand vision both represent and stand on an ongoing research framework, which has early roots in computer science, more precisely in formal logics, knowledge representation and reasoning, and databases. The vision of the Knowledge Web network experts on the evolution of some topics related to the Semantic Web is presented. Finally, the current research directions, which aim at supporting the scaling up of semantic technologies from closed intranets to the open internet, are discussed.

McIlraith et.al [2] proposed the markup of Web services in the DAML family of Semantic Web markup languages. This markup enables a wide variety of agent technologies for automated Web service discovery, execution, composition, and interoperation. The authors present one such technology for automated Web service composition.

Abraham and Romos [3,4] proposed the study of ant colonies behavior and their self-organizing capabilities is of interest to knowledge retrieval/ management and decision support systems sciences, because it provides models of distributed adaptive organization, which are useful to solve difficult optimization, classification, and distributed control problems, among others [16][17][18].

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

G kick [5] proposed the potential of the Semantic Web to solve real-world problems in inter-device communication, finding, sorting and classifying information, is tremendous. To achieve this it is necessary to understand that its power is more applicable to certain types of information than it is to others.

Gracia and Mena [6] proposed the Semantic Web must be able to make explicit the semantics of Web resources via ontology, which software agents use to automatically process these resources.

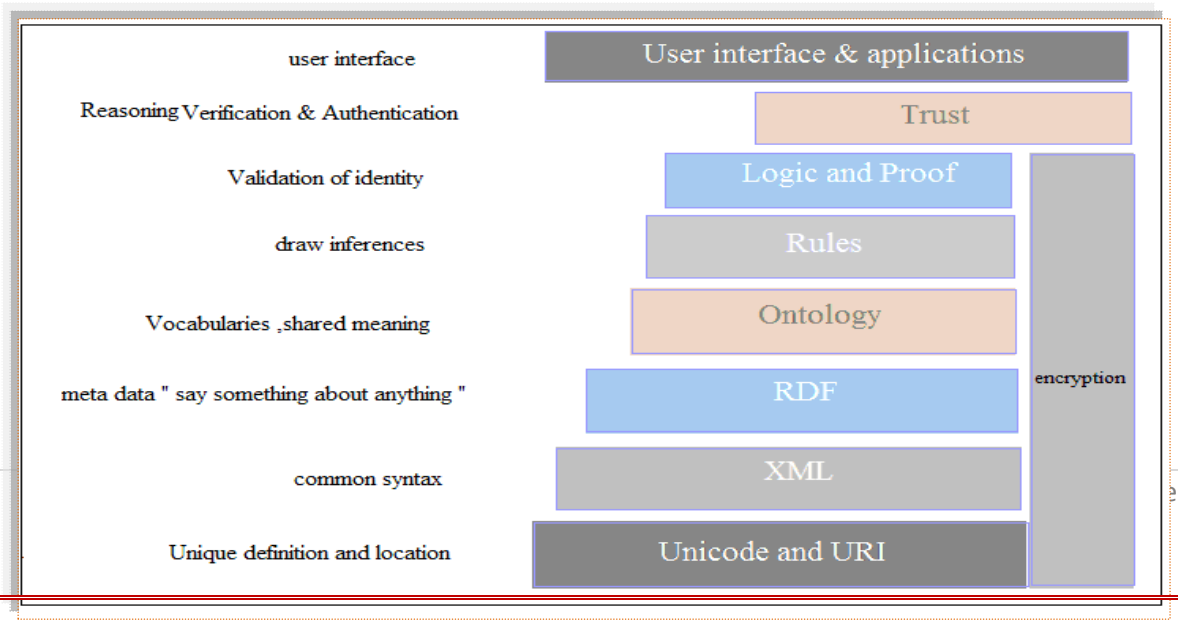
Shi et. al. [7] proposed survey of heterogeneous information network analysis. They have introduced basic concepts of heterogeneous information network analysis; examine its developments on different data mining tasks.

Rana and Singh [9] have proposed a semantic web mining interface, which is competent to handle heterogeneity issue and provide meaningful information in non-redundant way. Their work focused on finding the most ambiguous words and finding the relatedness measure with other important keywords in the query.

III. SEMANTIC WEB

The basic idea of Semantic Web [12] 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 ontology 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 [12].

According to Tim Burners-lee's vision, The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new functionality as machines become much better able to process and "understand" the data that they merely display at present. The essential property of the World Wide Web is its universality [13]. The semantic network Constituted by seven levels is constituted of a layered architecture [13]. As shown in figure:

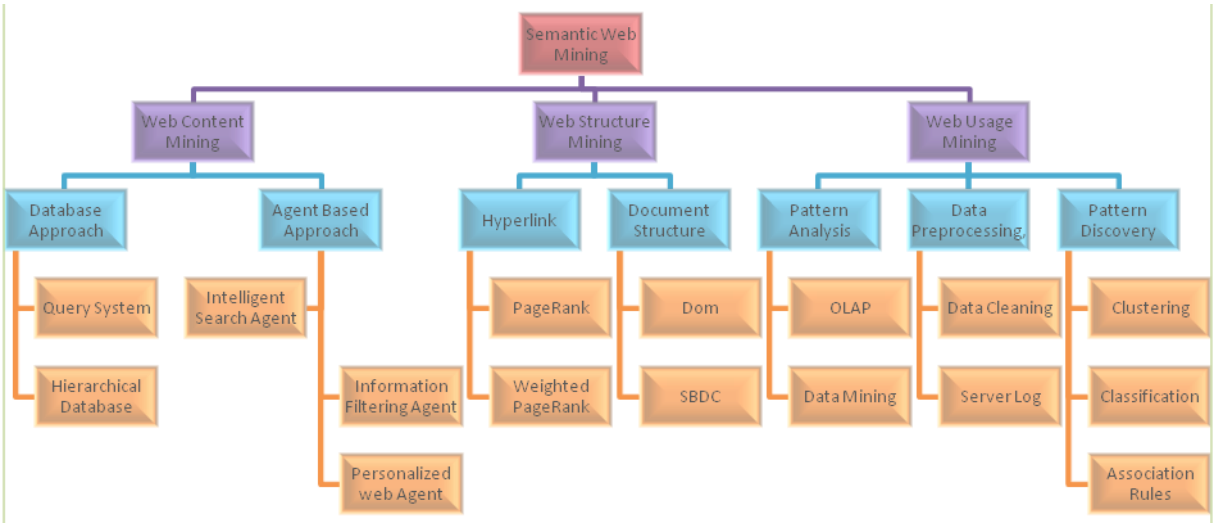


International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 03 , March 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

IV. WEB MINING

Web mining is the application of data mining techniques to the content, structure, and usage of Web resources. Three areas of Web mining are commonly distinguished: content mining, structure mining, and usage mining [27]. In all three areas, a wide range of general data mining techniques, such as association rule discovery, clustering, classification, and sequence mining, are employed and developed further to reflect the specific structures of Web data and the Web related application questions [21]. Web mining can be generally defined as[l]: Extract interested, useful patterns and implicit information from the WWW resources and behavior.

Figure shows the classification of Web mining:



1.1 Web Content Mining (WCM)

WCM is a major technique to find valuable contents and documents from the web. The web contents are describing in two forms: text and multimedia contents. Text content further consists of semi-structure content likes HTML data and unstructured text [15]. On other side multimedia content contains of picture, sound, tape and structured articles that provide semantic description. The current development of WCM technique have encouraged developers to make more intelligent approaches for knowledge accessing, such as Information Retrieval (IR) [7] and Database approach (DB) [14]. IR uses intelligent agent approach [17] to enhance the information searching and extracting the information from the users inferred or solicited profiles. DB uses database approaches to determine the data on the web and integrate them so that more difficult queries could be

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

searched. The rapid growths in WCM techniques have allowed system to increase knowledge deliverance of information through combining of several approaches such as agent based approach and database approach [17].

1.2 Web Structure Mining

Web structure mining is the way of discovering valuable information from the interconnected hypertext document on the web. This technique is work with the topology of hyperlinks consisting of web pages as nodes and hyperlinks as edges. It is appropriate technique to calculate the relatedness of each web page [19]. Web structure mining executes on two phases: hyperlink and document structure.

1.3 Web Usage Mining

Web usage mining is an innovative approach to automatically identify the user interaction patterns from web services and measures user behaviour, when the user works on the web. It helps to identify type of contents in which user are more interested. Today various business firms and e-commerce societies are follows these rules for evaluating life time value of client and gives better link according their browsing behaviours. Web usage mining retrieves desire knowledge from server log, proxy log, browser log and managed databases. Web server log contains the history of page log and proxy server executes between customer browser and web server. It works on three forms such as data pre-processing, pattern discovery and pattern evolution.

V. THE RELATIONSHIP BETWEEN SEMANTIC WEB AND WEB MINING

The Semantic web is an extension of current web and web mining technique is an extension of data mining technique. Semantics can improve the results of Web Mining by taking advantage of structures in the Web. Web Mining can improve the Semantic Web by finding new semantic structures to enrich the semantics. The central idea of this work is to proposed new innovative model that is called semantic web mining (SWM).

Figure 3: High Level View of semantic web mining

The objective of SWM is to get a higher understanding of user behavior at the time of Web surfing in order to better support for the users on the Web. The Semantic Web mining offers to add structure to the Web, while Web Mining can learn implicit structures. This is an interesting way for Semantic Web Mining to create itself as the dependence between the Semantic Web and Web Mining increases. The resulting research benefits many areas of industry such as “e-activities”, health care, privacy and security, and knowledge management and information retrieval.

VI. ONTOLOGY

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

A program that wants to compare or combine information across the two databases has to know that these two terms are being used to mean the same thing. Ideally, the program must have a way to discover such common meanings for whatever databases it encounters. A solution to this problem is provided by the third basic component of the Semantic Web, collections of information called ontology. In philosophy, ontology is a theory about the nature of existence, of what types of things exists; ontology as a discipline studies such theories.

Artificial-intelligence and Web researchers have co-opted the term ontology is a document or file that formally defines the relations among terms. The most typical kind of ontology for the Web has taxonomy and a set of inference rules. The taxonomy defines classes of objects and relations among them. For example, an address may be defined as a type of location, and city codes may be defined to apply only to locations, and so on. Classes, subclasses and relations among entities are a very powerful tool for Web use. We can express a large number of relations among entities by assigning properties to classes and allowing subclasses to inherit such properties. If city codes must be of type city and cities generally have Web sites, we can discuss the Web site associated with a city code even if no database links a city code directly to a Web site. Inference rules in ontology supply further power. Ontology may express the rule "If a city code is associated with a state code, and an address uses that city code, then that address has the associated state code." A program could then readily deduce, for instance, that a Cornell University [11].

VII. CONCLUSION & FUTURE WORK

The semantic web is a highly dynamic and distributed system which contains incomplete and uncertain knowledge; therefore an autonomous and distributed software system was desired to optimally exploit the knowledge and information available on the web. Ontology helps to discover information across the databases. With the use of sophisticated techniques to syntactic and structural interpretability have been solved, the problem of semantic interpretability is still remain and only partially solved.

REFEERENCES

[1] A.J. Gerber, A. Barnard, A.J. Van der Merwe (2008), **'**A Functional Semantic Web Architecture'. Published in Proceedings of 5th European Semantic Web Conference, ESWC08,Tenerife Spain,1-5 June 2008, ISBN: 3-540-68233-3.

[2] A. Karageorgos and N. Mehandjiev (2004), 'A Design Complexity Evaluation Framework for Agent-Based System Engineering Methodologies'. Published in 'Engineering Societies in The Agents World', LNCS, Vol. 3071/2004, pp. 258-274, 2004.

[3] A. Maedche and S. Staab (2001), 'Ontology Learning for the Semantic Web', Published in IEEE Intelligent Systems, Vol. 16, Issue 2, pp. 72-79, March 2001.

[4] Andrea Moro and Roberto Navigli, (2015), SemEval-2015 Task 13: Multilingual All-Words Sense Disambiguation and Entity Linking, SemEval 2015, Association for Computational Linguistics, pp 288-297.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

[5] Anatoly Gladun, Julia Rogushina, Francisco Garcia-Sanchez,Rodrigo Martinez-Bejar, Jesualdo Tomas Fernandez-Breis ,'An application of intelligent techniques and semantic web technologies in E-learning environments' published in 'Expert Systems with Applications' Vol 36,issue 2, part 1,March 2009 , pg 1922-1931.

[6] Andreas Montoyo, Armando Suarez, German Rigau and Manuel Palomar, (2005), Combining Knowledge- and Corpus-based Word-Sense-Disambiguation Methods, Journal of Artificial Intelligence Research 23, pp 299-330.

[7] Aris M.oksal and Amit seth, (1999), A Semantic Interoperability in global information systems: A brief introduction to the research area and the special section, portal.acm.org,vol. 28, No. 1, pp 5-12.

[8] Basim Alhadidi, Singh A, (2013), Knowledge Oriented Personalized Search Engine: A Step towards Wisdom Web, International Journal of Computer Applications, Vol.76, No.8, pp. 1-9.

[9] Gerd Stumme, Andreas Hotho and Bettina Berend, (2006), Semantic Web Mining State of the art and future directions, Journal of Web Semantics – Elsevier, pp 124-143.

[10] Hai Leong Chieu and Hwee Tou Ng, (2002), A Maximum Entropy Approach to Information Extraction from Semi-Structured and Free Text, AAAI, 2002, pp 786-791.

[11] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar, (2005), Web Mining - Concepts, Applications & Research Directions, Foundations and Advances in Data Mining Volume 180 of the series Studies in Fuzziness and Soft Computing, pp 275-307.

[12] Jaideep Srivastava, Robert Cooley, (2000), Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, ACM SIGKDD, pp 1-12.

[13] Jeffrey M. Bradshaw (1997), 'An Introduction to Software Agents'. In Book 'Software Agents', J.M. Bradshaw (Ed.), AAAI Press, 1997, pages 3-46.

[14] Jin Zhao, Min-Yen Kan and Yin Leng Theng, (2008), Math Information Retrieval: User Requirements and Prototype Implementation, ACM, 2008, pp 16-20.

[15] Jorge Gracia and Eduardo Mena, (2008), Web-Based Measure of Semantic Relatedness, WISE 2008, LNCS, Vol.5175, pp. 136–150.

[16] Karim Heidari, (2009), The Impact of Semantic Web on E-Commerce, World Academy of Science, Engineering and Technology",Vol 25, pp 303-306.

[17] Manoj Manuja & Deepak Garg, (2011) Semantic Web Mining of Un-Structured Data: Challenges And Opportunities, International Journal of Engineering (IJE), Volume (5): Issue (3), pp 269-276.

[18] Mohammed Maree and Mohammed Belkhatir, (2015), Addressing semantic heterogeneity through multiple knowledge base assisted merging of domain-specific ontologies, Knowledge-Based Systems, Vol 73, No. 3, pp199-211.

[19] Manoj Manuja & Deepak Garg, (2011) Semantic Web Mining of Un-Structured Data: Challenges And Opportunities, International Journal of Engineering (IJE), Volume (5): Issue (3), pp 269-276.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

[20] N. Shadbolt, W. Hall and T. B. Lee (2006), 'The Semantic Web Revisited'. In IEEE Intelligent Systems, Vol. 21, Issue 3, pp 96-101, 2006.

[21] P. Ravi Kumar and Ashutosh Kumar Singh, (2010), "Web Structure Mining: Exploring Hyperlinks and Algorithms for Information Retrieval", American Journal of Applied Sciences, pp 840-845.

[22] P.Saravana kumar/ R.Iswarya, (2014), Predictive Analysis of Users Behaviour in Web Browsing and Pattern Discovery Networks, International Journal of Latest Trends in Engineering and Technology (IJLTET), pp 239-245.

[23] Paul Warren, (2006), Knowledge Management and the Semantic Web: From Scenario to Technology, IEEE Computer Society, pp 53-59.

[24] Pedersen T, Lexical Semantic Ambiguity Resolution with Bigram-Based Decision Trees, Lecture Notes in Computer Science Volume 2004, 2001, pp 157-168.

[25] Peter Mika, 'Ontologies are us: A unified model of social networks and semantics' , Published in journal of Web Semantics ,Vol 5,Pg. 5-15,2007.

[26] Prasanna Desikan, Jaideep Srivastava, Vipin Kumar, and Pang-Ning Tan, (2002), "Hyperlink Analysis: Techniques and Applications", (IJCSE) International Journal on *Computer Science* and Engineering pp 1-42.

[27] R. Cooley, B. Mobasher, and J. Srivastava, (1997), Web Mining: Information and Pattern Discovery on the World Wide Web, ICRI-1997, LNCS-1836, pp 1-10.

[28] Rana V, (2012), Blueprint of an Ant-Based Control of Semantic Web, Vol. 2, No. 4, IJoAT, pp. 603-612.

[29] Roberto Navigli, (2009), Word Sense Disambiguation: A Survey, ACM Computing Surveys, Vol. 41, No. 2, 1-69.

[30] Siddharth Patwardhan, Satanjeev Banerjee and Ted Pedersen, (2005), SenseRelate::TargetWord - A Generalized Framework forWord Sense Disambiguation, Association fo Computational Linguistics-ACM, pp 73-76.

[31] Singh A, Basim Alhadidi, (2013), Knowledge Oriented Personalized Search Engine: A Step towards Wisdom Web, International Journal of Computer Applications, Vol.76, No.8, pp. 1-9.

[32] T. Berners Lee, J. Hendler, O. Lassila (2001),'The Semantic Web'. Published in 'The Scientific Americans', Vol. 5, Issue 1, pp 36, 2001.

[33] Valentina Tamma and T.R. Payne (2008), 'Is a Semantic Web Agent a Knowledge-Savy Agent?'. Published in 'IEEE Intelligent Systems', pp 82-85, July 2008.