ACADEMICIA

ISSN: 2249-7137 Vol. 11, Issue 10, October 2021 Impact Factor: SJIF 2021 = 7.492

ACADEMICIA
An International Multidisciplinary Research Journal
(Double Blind Refereed & Peer Reviewed Journal)

DOI: 10.5958/2249-7137.2021.02093.0

A REVIEW ON SEMANTIC WEB MINING

Dr. Ajay Rana*; Vijay Maheshwari**

*Shobhit Institute of Engineering and Technology,
(Deemed to be University), Meerut, INDIA
Email id: ajay.rana@shobhituniversity.ac.in,

**School of Computer Science and Engineering,
INDIA
Email id: vijay@shobhituniversity.ac.in

ABSTRACT

Due to the enormous quantity of information in various forms, retrieving the most relevant papers from the web is challenging. The enormous quantity of data is tough for computers to comprehend, but it is simple for people to comprehend. The semantic web, often known as Network 3.0, is an online of data that allows competent computers to comprehend data on web pages. The Semantic Web is a technique for improving the precision of information retrieval systems. Online mining is the use of data mining methods to extract information from web data, such as web pages, connections between documents, and web site use records, among other things. Semantic web mining aims to bring the semantic web and web mining together. The primary goal is to use semantic web tools to convert unstructured data into machine comprehensible data so that machines can reply to human inquiries in less time and save tiresome effort, and to use web mining tools to automatically extract information buried in huge quantities of online data. The different Semantic-web methods and difficulties are the subject of this paper.

KEYWORDS: *Semantic Web Mining, Semantic Web Approaches, Semantic Web, Web 3.0, Web Mining.*

1. INTRODUCTION

Both the Semantic Web and Web Mining study fields build on the success of the World Wide Web. They work effectively together because they each handle a different aspect of a new problem presented by the current World Wide Web's (WWW) enormous success[1], [2]. The majority of data accessible on websites is unstructured, making it difficult for computers to comprehend, yet the volume of data is so large that it can only be handled effectively by machines. With Web information input, the World Wide Web has evolved into a new communication medium. This is reinforced by informational, cultural, social, and evidentiary values, to name a few. With the availability of different Search Engines such as Google, Yahoo, and others, people are increasingly turning to them to get information from Web sites. Existing search engines, on the other hand, are unable to differentiate between specific user requests. The semantic web tackles the first half of the problem by attempting to make data machine-understandable, while web mining tackles the second by automatically extracting valuable information from this data. Web mining necessitates the creative use of data mining and/or text mining methods, as well as unique approaches[3], [4]. Web data mining is one of the most difficult jobs in data mining. The goal of semantic web mining is to bring together the two fields of semantic web and web mining. Fig. 1 illustrates the layered structure of semantic web.

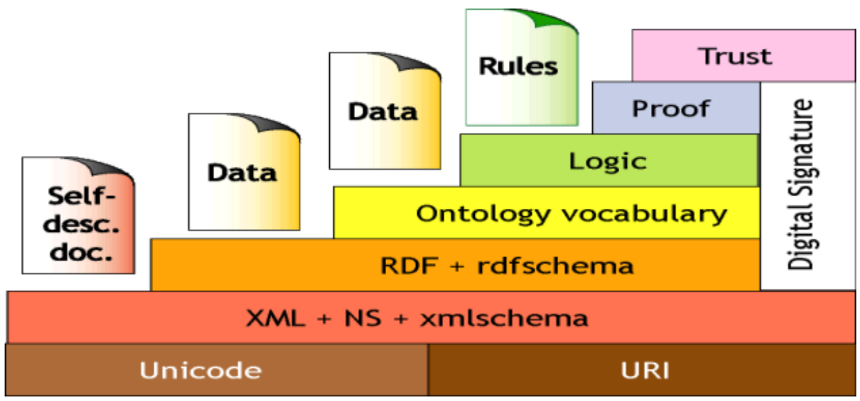

Fig. 1: Layer structure of semantic web[5]**.**

1.1. Semantic Web:

Tim Berners-Lee, the creator of the World Wide Web, has a vision for the Semantic Web[6]. Semantics, in general, is concerned with the structure of sentences and what they really imply. When a user alters a sentence's pattern, the pattern changes, but the semantics stay the same. When a computer comprehends the semantics of a document, it does more than simply analyze the sequence of letters that make up that text; semantic web aids the machine in comprehending the meanings behind the web page. The present WWW contains a huge quantity of material that is frequently unorganized and comprehensible only by humans. The human brain, for example, employs logic:

a. Ramesh is a father.

b. A father is a parent.

As a result, Ramesh must be a parent, but machines are incapable of applying this reasoning. The Semantic Web attempts to solve this issue by giving machine-interpretable semantics, allowing machines to better assist users. If a computer can recognize the meaning behind a web page, it may readily assist people in accessing knowledge rather than unstructured content, enabling knowledge to be handled automatically.A basic search engine cannot comprehend the connections between keywords, phrases, or parts of speech inside a search phrase, but a semantic search engine can, thanks to ontologies, understand the underlying meaning of the whole phrase. A detailed explanation of words and reasoning in a topic area is ontology. The usage of ontologies allows computers to provide meaningful results from semantic data. By making the message as plain as possible, a machine can comprehend or at the very least use it.A semantic search engine, for example, would be able to tell the difference between the following sentences, which are made up of the same keywords but have distinct meanings:

a. By using java language semantic web mining.

b. Semantic web mining by using java language.

The sentences in the example above are made up of the same keywords, but the subject connections have been inverted. Because the connections between the keywords or phrases are unknown in conventional online search, which is dependent on ranking algorithms, the engines would return similar or almost identical results, even if two totally distinct queries were asked. Additional issues with conventional online search occur when the terms are either too precise, resulting in few or no results, or too broad, resulting in irrelevant results.

The current web is about documents, while the semantic web is about objects on web pages, such as people, locations, organizations, and ideas.

2. DISCUSSION

2.1. Semantic Web Architecture:

For the Semantic Web, Berners-Lee proposed a layer structure. The Semantic Web is built in layers, one on top of the other, allowing for a more uniform development process[7].

2.1.1. Uniform Resource Identifier (URI):

The web's cornerstone is the URI, which binds the rest of the web together. The URI's goal is to clearly provide an identity that will be used to represent a resource in a consistent manner, identifying information representation structures such as classes, attributes, and people. It becomes easy to combine all data that relates to a particular resource since there is no ambiguity. A URI is just a description, not a location; it simply identifies something. For example, one individual's web page might be abcd.com, which can be used to identify that person. As a result, the object abcd is associated. A person may utilize his page, but it is only visible to other users. Users and software can understand precisely what they are being referred to since URIs are globally unique and each occurrence of the same identifier signifies the same thing.

2.1.2. Unicode:

Unicode is a character encoding system that allows all user languages to read and write on the web in a uniform format.

2.1.3. Extensible Markup Language (XML):

XML is a web-based data transmission and storage language. The purpose of XML is to transport and describe data, not to display it. User-defined tags may be found in XML documents. The structure of an XML document is described using an XML schema. In the semantic web, the XML namespace is used to prevent conflicting data or names. The purpose of the XML layer is to provide the fundamental syntax and structure of online data.

2.1.4. Resource Description Framework (RDF):

RDF is a format for storing and organizing data. RDF addresses the issue of data linkage. A resource is anything that has an identity, and URIs(Uniform Resource Identifiers) are used to identify it. The term "description" refers to a container that holds multiple assertions that describe the resource. To create and comprehend claims, you'll need a framework. The machine can analyze the query and come up with a response using RDF.

2.1.5. Resource Description Framework Schema (RDFS):

RDFs may be used to organize classes and attributes into generalization/specialization hierarchies. The purpose of RDF and RDFS is to give metadata to upper-layer technologies so that this information may be shared and reused across these technologies or between these technologies and other applications.

2.1.6. Web Ontology Language (OWL):

A precise explanation of words and reasoning in a topic area is known as ontology[8]. We may bring the semantic stuff using ontology. The web ontology language was created using description logics and web languages and may be used to define ontologies. OWL meets the semantic web's criteria for requiring little human input while still meeting software needs for a language with correct meaning. Ontologies are useful for clearly representing things and their relationships, which may be either direct or inverse.

2.1.7. Rules, Proof & Trust layer:

- *Proof Layer:*

The proof layer is used to validate that the findings generated by the agents can be trusted or to authenticate the agent's activity.

- *Rules Layer:*

The Rules Layer is intended to be used as a framework for developing new conclusions and expressing them for the deployment of the semantic web.

- *Trust Layer:*

The purpose of the trust layer is to offer a method for information providers and users to trust and be confident in each other.

2.2. Web Mining:

Online mining is the use of data mining methods to extract information from web data, such as web documents, hyperlinks between papers, and web site use records, among other things[9]. Thus, finding legitimate, previously unknown, and possibly valuable patterns in the enormous quantity of online data, patterns that represent them in succinct form and manageable orders of magnitude, is a difficult task. Web mining, like other data mining applications, may benefit from data structure, but it can also be used on semi structured or unstructured data. Web mining aids

ISSN: 2249-7137 Vol. 11, Issue 10, October 2021 Impact Factor: SJIF 2021 = 7.492

the transition from human-understandable material to machine-understandable semantics in this way. Fig. 2 illustrates the categorization of web mining methods.

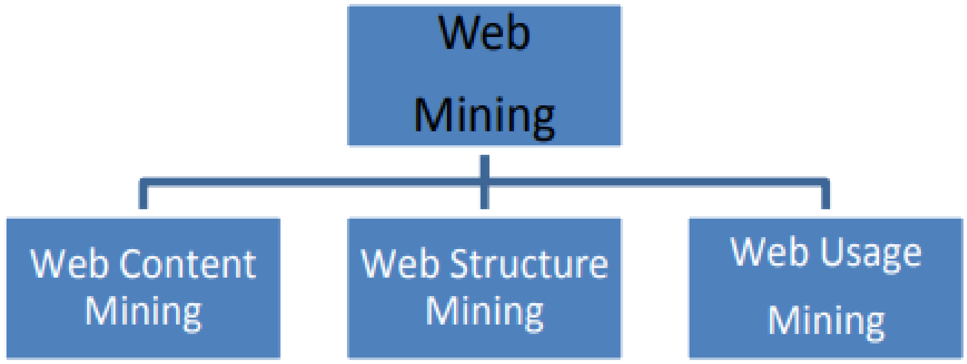

Fig. 2: Taxonomy of web mining[10]**.**

2.2.1. Web Content Mining:

Web content mining is the technique of collecting information from the contents of Web pages by analyzing the content of web resources. It examines web page content as well as web searches. The study is mostly focused on text mining methods, although expansions to multimedia material are starting to appear. Textual, picture, audio, video, metadata, and hyperlinks are all examples of data kinds found in online content. Web content may be unstructured (plain text is often associated with unstructured data), semi-structured (HTML documents), or structured (XML documents) (extracted from databases into dynamic Web pages). Unstructured data is strongly linked with big data. Large datasets that are challenging to examine using conventional methods. As a result, NLP methods are used to retrieve information.

2.2.2. Web Structure Mining:

Online pages serve as nodes, while hyperlinks serve as edges linking related sites in a typical web graph. The technique of finding structural information from the web is known as web structure mining. We combine structure mining with the page rank algorithm to ensure that the material is relevant and of sufficient quality. Based on the kind of structural information utilized, this may be further split into two types.

Furthermore, a web page's content may be arranged in a tree-structured manner using the different HTML and XML elements on the page.

- *Hyperlink:*

A hyperlink is a structural element in web pages that connects one place to another, either inside the same page or on a separate page. An intra-document hyperlink leads to a different section of the same page, whereas an inter-document hyperlink connects two distinct pages.

- *Document Structure:*

According to the types of data to be mined, web mining may be classified into three categories. A broad variety of conventional data mining methods, including association rule discovery, clustering, classification, and sequence mining, are used and expanded in all three domains to reflect the unique structures of online resources and the unique issues presented in Web Mining.

2.2.3. Web Usage Mining:

Web usage mining is the use of data mining methods to find interesting usage patterns in web usage data or server logs in order to better understand and fulfill the requirements of web-based applications. It also analyzes user clicks from web server logs. The identity or origin of online users, as well as their browsing activity on a website, is captured by use data.

Directly addressing the issues connected with web log mining is a key aspect of the Web use mining method. It's all about detecting user surfing habits on the internet using information gleaned from web logs.

2.3. Semantic Web Approaches:

2.3.1. Ontology Approach:

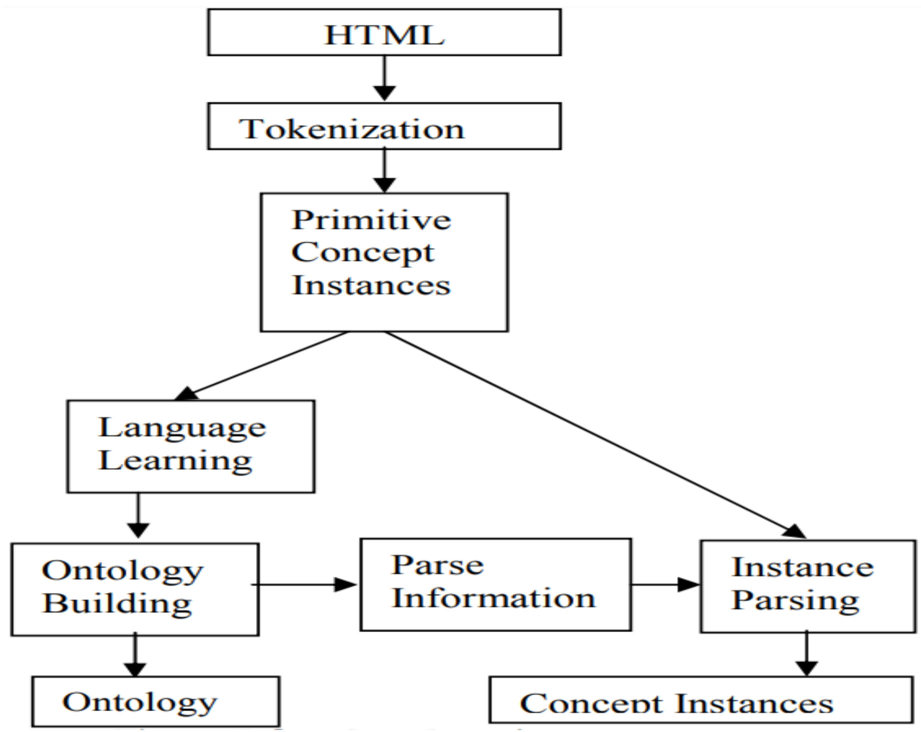

Fig. 3: Ontology learning process[11]**.**

The semantic web's backbone is ontology. An ontology is a set of URIs having a meaning that is typically only obliquely explained. A formal ontology language is used to represent ontologies. Ontology is a collection of ideas and their interrelationships that are linked to a certain knowledge area. Ontology expertise is very helpful in defining the structure and scope of online content mining. Ontology is a collection of things, ideas, and other entities that exist in a given region, as well as the connections that exist between them. Getting an ontology from the internet is a difficult job. It is based on Web content mining techniques and blends machine learning techniques with approaches from areas such as information retrieval, using them to find and make explicit the semantics in data. The methods generate interim findings that must be combined into a machine-readable format. Ontology learning process is illustrated in Fig. 3.

2.3.2. Semantic Based Web Mining:

Semantic-based web mining combines the semantic web with web mining, two rapidly evolving areas. It's also known as Semantic-Web Mining or Semantic Web-Mining. The semantic web attempts to solve this problem by making data intelligible to both machines and humans. Online mining, on the other hand, is concerned with automatically collecting valuable knowledge or information from vast amounts of data on web sites. Web pages are mined by the machine in semantic based web mining so that the machine can better comprehend the information on the web pages. It works by retrieving XML and RDF documents, as well as ontologies and metadata. Semantic web mining entails sifting through data sources and information pertaining to web-based information management systems. Web mining will lead to semantic web mining. The aim of semantic web mining is to make the web more accessible. Ontology, semantic web content, and web services are the three main categories of requirements for the semantic web.

2.4. Semantic Web Challenges:

2.4.1.Huge Quantityof Available Data:

Existing technology has not yet been able to predict all semantically duplicated words on the web, which includes vast amounts of data on billions of web pages.

2.4.2. Unclearness:

These are hypothetical terms like "young" or "tall." This is due to the ambiguity of user questions, the difficulty of matching queries to provider contents, and the difficulty of combining several knowledge bases with overlapping but carefully distinct ideas.

2.4.3. Term's Adaptability:

This is a set of exact ideas with varying values. A teacher, for example, could offer a set of test criteria that correlate to a number of different unique student abilities, each with a different likelihood.

2.4.4. Ontology Inconsistency:

These are logical inconsistencies that will inevitably emerge with the creation of big ontologies and the merging of ontologies from different sources.

3. CONCLUSION

We looked at two rapidly emerging research topics in this paper: web mining and semantic web. Semantic Web Mining, as a unified field, provides new methods to enhance both fields. By using the new semantic structures on the Web, semantic-based web mining may enhance the outcomes of web mining. Because of the availability of background information, the Semantic Web may make Web mining more simpler, and Web mining can also create new semantic structures in the Web. Many industries benefit from the study, including e-commerce, health care, privacy and security, search engines, knowledge management, and information retrieval.

REFERENCES

1. J. Sivakumar and K. S. Ravichandran, "A review on semantic-based web mining and its applications," *International Journal of Engineering and Technology*. 2013.

2. V. A. and A. A., "Semantic Web Mining using RDF Data," *Int. J. Comput. Appl.*, 2016, doi: 10.5120/ijca2016908022.

3. T. A. Al-asadi, A. J. Obaid, R. Hidayat, and A. A. Ramli, "A survey on web mining: Techniques and applications," *Int. J. Adv. Sci. Eng. Inf. Technol.*, 2017, doi: 10.18517/ijaseit.7.4.2803.

4. A. Gök, A. Waterworth, and P. Shapira, "Use of web mining in studying innovation," *Scientometrics*, 2015, doi: 10.1007/s11192-014-1434-0.

5. [5] M.-R. Koivunen and E. Miller, "W3C Semantic Web Activity." https://www.w3.org/2001/12/semweb-fin/w3csw (accessed Sep. 13, 2018).

6. A. Gangemi, V. Presutti, D. Reforgiato Recupero, A. G. Nuzzolese, F. Draicchio, and M. Mongiovì, "Semantic Web Machine Reading with FRED," *Semant. Web*, 2017, doi: 10.3233/SW-160240.

7. A. Harth, M. Janik, and S. Staab, "Semantic Web Architecture," in *Handbook of Semantic Web Technologies*, 2011.

8. L. Yu, "OWL: Web Ontology Language," in *A Developer's Guide to the Semantic Web*, 2014.

9. M. Spiliopoulou, B. Mobasher, O. Nasraoui, and O. Zaiane, "Guest editorial: Special issue on a decade of mining the Web," *Data Min. Knowl. Discov.*, 2012, doi: 10.1007/s10618-012-0257-y.

10. K. Griazev and S. Ramanauskaite, "Web mining taxonomy," 2018, doi: 10.1109/eStream.2018.8394124.

11. M. El Ghosh, H. Naja, H. Abdulrab, and M. Khalil, "Ontology learning process as a bottom-up strategy for building domain-specific ontology from legal texts," 2017, doi: 10.5220/0006188004730480.