Learning and Analytics in Intelligent Systems 4

Suresh Chandra Satapathy · K. Srujan Raju · K. Shyamala · D. Rama Krishna · Margarita N. Favorskaya *Editors*

Advances in Decision Sciences, Image Processing, Security and Computer Vision

International Conference on Emerging Trends in Engineering (ICETE), Vol. 2

Springer

Learning and Analytics in Intelligent Systems

Volume 4

The main aim of the series is to make available a publication of books in hard copy form and soft copy form on all aspects of learning, analytics and advanced intelligent systems and related technologies. The mentioned disciplines are strongly related and complement one another significantly. Thus, the series encourages cross-fertilization highlighting research and knowledge of common interest. The series allows a unified/integrated approach to themes and topics in these scientific disciplines which will result in significant cross-fertilization and research dissemination. To maximize dissemination of research results and knowledge in these disciplines, the series publishes edited books, monographs, handbooks, textbooks and conference proceedings.

More information about this series at http://www.springer.com/series/16172

Suresh Chandra Satapathy ·
K. Srujan Raju · K. Shyamala ·
D. Rama Krishna · Margarita N. Favorskaya
Editors

Advances in Decision Sciences, Image Processing, Security and Computer Vision

International Conference on Emerging Trends in Engineering (ICETE), Vol. 2

Springer

Editors
Suresh Chandra Satapathy
School of Computer Engineering
Kalinga Institute of Industrial Technology
(KIIT) Deemed to be University
Bhubaneswar, Odisha, India

K. Shyamala
Department of CSE
Osmania University,
University College of Engineering
Hyderabad, Telangana, India

Margarita N. Favorskaya
Institute of Informatics
and Telecommunications
Reshetnev Siberian State University
of Science and Technology
Krasnoyarsk, Russia

K. Srujan Raju
Department of CSE
CMR Technical Campus
Hyderabad, Telangana, India

D. Rama Krishna
Department of ECE
Osmania University,
University College of Engineering
Hyderabad, Telangana, India

ISSN 2662-3447 ISSN 2662-3455 (electronic)
Learning and Analytics in Intelligent Systems
ISBN 978-3-030-24317-3 ISBN 978-3-030-24318-0 (eBook)
https://doi.org/10.1007/978-3-030-24318-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Representation Techniques that Best Followed for Semantic Web - Web Mining

K. Vaishali[1,2] and Sriramula Nagaprasad[1,2(✉)]

[1] Jyothismathi Institute of Technology and Sciences, Karimnagar, India
nagkanna80@gmail.com
[2] Tara Government Degree College, Sangareddy, India

Abstract. Web Mining is a Data Mining Technique used widely in mining billions of information from the World Wide Web (WWW) as faster as possible with the exact match of data. The huge information available in WWW with various formats, like: text format files, images, documents and other forms of data like structured, semi structured and unstructured forms. The amount of this information is increasing day by day. Data mining is the technique used to extract the data available in the internet. Web mining technique is used to determine and mine information from data sources related to web which are documents in web, contents in web, server logs and hyperlinks. The Semantic Web is used to provide information in a defined meaning that enhanced the interoperability between human and machines, which created the space for the machines to handle most of the decisions and tasks.

This paper gives a brief idea regarding representation techniques that are best used in semantic web.

Keywords: Web mining · Web mining techniques · Semantic web · Data mining technique · Representation techniques

1 Introduction

One of the major applications of data mining techniques which is used in discovering models or patterns for the content, structure and usage within the web pages from the WWW - World Wide Web is popularly known to be web mining. Web mining as the word describes, the data or information will be gathered by mining the web. This can be applied for both structured and unstructured information in the form of browser activities, page content, website, server logs, link structure and different sources. Semantic web and web mining are the quick rising technologies in the study areas. Web content, web usage and web structure mining are three important web mining types used to satisfy the whole process of data mining in web mining [1]. Web content mining methods are used for Semantic Annotation creation from web page content; on the other part it also profits the content that is structured already in RDF, XML or Ontology. To understand and serve better requirements of Web-based applications and to discover usage patterns, Web usage mining methods are used effectively [1].

S. C. Satapathy et al. (Eds.): ICETE 2019, LAIS 4, pp. 693–699, 2020.
https://doi.org/10.1007/978-3-030-24318-0_79

2 Web Mining

Discovering patterns from the large databases with knowledge discovery has been processed with the help of a field in Computer Science known as Data Mining. Moreover the information are extracted and transformed to an understandable format that can be used further for prediction in the future or for any other purpose. To extract knowledge data from web such as logs used for websites, hyperlinks used between documents, documents of web etc, Data Mining technique use one of its applications – web mining [9]. One of the major differences between web mining and data mining is, web mining find patterns which are very useful from the web data such as logs, hyperlinks and documents, whereas data mining works with data from the database and find out the patterns that are useful. The raw data from data mining will be always in structured form which can be further used in mere prediction, but in web mining the raw data that is available will be either semi-structured or unstructured which then converted to structured format for knowledge extraction. Web mining techniques are divided in three types: (1) web content mining, (2) web structure mining and (3) web usage mining.

(1) Web Content Mining
Through this technique, information is extracted from the data available from the web in the form of web documents. The data available will be in various forms like images, audio, text, video, table etc. Most research in web content mining has been processed using knowledge extraction from text data. NLP - Natural Language Processing and IR - Information Retrieval technologies are also widely utilized in web content mining. In recent years image processing is also getting influenced for extracting data from images.

(2) Web Structure Mining
Web structure mining technique is mainly designed to focus on web structured data. For instance when the data is considered to be a graph, then the web pages are set of nodes belongs to that graph and edges are hyperlinks that connects different nodes in the web pages. This always deals with information that is structural from the web. Document structure and hyperlinks are the two different classifications of web structure mining. Hyperlinks are used mainly for structural units which connect web page location to different web page location, either in the same page or in different pages in the web. Then the content is organized in format known to tree-structured which is based on XML and HTML tags within the specific page [9] (Fig. 1).

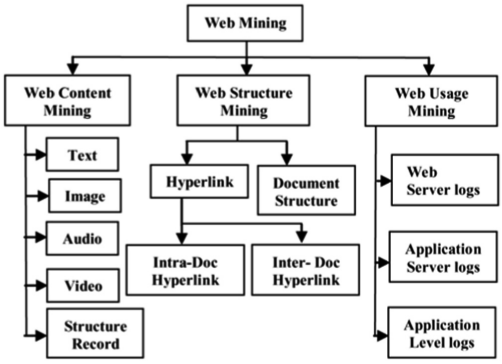

Fig. 1. Techniques used in web mining

(3) Web Usage Mining

Mining technique that process with user's web usage patterns from the web logs in web sites is said to be web usage mining. To realize the user's patterns in browsing and for better service of web-based applications at present and modification in future web usage mining technique always provides its best. The data that are used for this mining technique are identity, location, browsing patterns of the user, etc which are obtained from website usage log. Web server data, application server data and application level data are the three types of web usage mining. User log information like IP address, time of access and reference, etc are collected from the web server and logged in web server data [9]. Various business events are tracked and logged using application server data. Events defined within the same logged application and sourced by creator are tracked and logged by application level data.

3 Semantic Web

The transformation of information oriented web to knowledge oriented web is carried by semantic web a joined progress lead with the standards followed by international body - W3C. Semantic web is a powerful extension of (WWW) World Wide Web. Semantic web provide a standard for expressing web page relationships by allowing the machines to understand and accelerate the exact meaning of information that are hyperlinked [10]. Tim Berners-Lee who is the inventor of WWW and Director of W3C has coined the term "Semantic Web" – for the data from web that can be processed by the machines [10]. Computer basically does not understand the textual data that are unstructured, hence semantic web help computers to interpret the read data by adding meta-data to the pages in the web. Moreover this would never add any AI to machines, nor will construct self-awareness to the system, but will definitely provide machine tools for finding information, exchanging data and interpreting for a few levels [11]. The ultimate aim of semantic web is converting the web which includes semi-structured or unstructured documents to a web of data by adding semantic content in the web pages.

4 Semantic Web Representation Techniques

To communicate and to express semantic of information several models and accessing strategies are utilized. W3C has suggested standard systems such as XML (Extensible Mark-up Language), RDF (Resource Description Framework) and OWL (Web Ontology Language) [3].

1 Extensible Mark-Up Language (XML)

XML (Extensible Mark-up Language) strategies have the power of recouping data from the web. In engaging customers, to make their own specific marks, it licenses them to portray the content adequately. Along these lines, the set of information and the semantic connection streams of that information are able to be addressed [5, 6].

2 Resource Description Framework (RDF)

By utilizing their own domain vocabularies, RDF (Resource Description Framework) has been entitled with the capability of storing data which are retrieved and used by resources on accessing the WWW [3, 4]. The three categories of content elements available with RDF are,

(a) Resources (entities are recognized by using URIs)
(b) Literals (atomic series such as numbers, strings, etc)
(c) Properties (binary associations recognized using URIs - Uniform Resource Identifiers) [2].

An extremely efficient method for representing several type of information which is defined in web is RDF [3].

3 Web Ontology Language (OWL)

When compared with RDF, OWL is said to be more complicated language with enhanced ability for interpreting. Nature of the resource and their relationships are accurately identified by OWL. For representing the information of semantic web, the OWL utilizes ontology which is a demonstration of proper clear clarification of common procedure and basic input [4, 6]. OWL accurately identifies the sources' character and association. Developers of Ontology have expressed the attention on domains that is class based and properties such as representing rules and atomic distinct concepts in some further semantic languages too. Sir Berners-Lee has examined the architecture of Semantic Web in seven layers [7] (Fig. 2),

(1) URI
(2) XML, NS, & XML schema
(3) RDF & RDF schema
(4) The Ontology Vocabulary
(5) Logic
(6) Proof and
(7) Trust

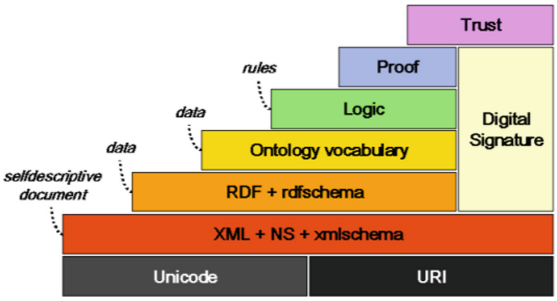

Fig. 2. Architecture of semantic web - seven layers

- URI – Responsible for encoding process of resource and their identification.
- XML, NS, and XML schema layer – Responsible for
  (1) Division made towards content information, structural information and design performance by following linguistic
  (2) Providing Standard Language Format.
- RDF and RDF schema – By using Semantic model, this layer defines the information on WWW and its types.
- Ontology Vocabulary layer – This layer is mainly focused on disclosed semantics between data in the way of characterizing the shared knowledge and the relations of semantic inside various types of data.
- Logic layer – The foundation of intelligence services like logical reasoning by providing inference principles and axioms are taken care by this layer.
- Proof and Trust layers - Mechanisms based on digital signature and encryption are used for recognizing alteration made with the papers for the purpose of enhancing the web security.

5 Ontology and Web Ontology Language (Owl) – A Best Representation Technique for Semantic Web

The backbone of semantic web is Ontology - a representation technique. Ontology has been defined by different literatures in different ways; some of them have been mentioned here,

(1) It is a official demonstration which contains the group of ideas and associations [8].
(2) It is an explicit specification of conceptualization [12].
(3) It is a term in philosophy and its meaning is “theory of existence” [13].
(4) It is a body of knowledge describing some domain, typically common sense knowledge domain [13].

The best technique followed in semantic web which is understandable by both humans and machines is strongly said to be Ontology. Semantic web - meaning

assigned to the web, is followed by Ontology. The creation of Ontology is a semi-automatic procedure. All the data of information that is extracted from a semi-structured or unstructured data forms a structured format and then inserted into the knowledge base is known as Ontology creation. To improve the results of user's query, the information which is available from knowledge base is utilized in process of web mining. For authorizing Ontologies or knowledge bases a unit of knowledge representation languages or languages of ontology like web ontology language (OWL) is used. This language is categorized as formal semantics, and RDF/XML oriented serialization for semantic web [14].

6 Conclusion

After analyzing various categories of web representation techniques in the process of extracting knowledge source from WWW information for semantic web, it is very well may be reasoned that the information which are unstructured, present in the web pages can also be verified and checked to make ontologies for colonizing knowledge base in the search of web. Data embedded with knowledge base are given in organized way so that the machine will recognize perfectly. Data that are retrieved from knowledge base are then utilized with computer system to give better enhanced results for requested web user queries. In this manner semantics can be appended to the present web through knowledge extraction method for making ontologies towards the formation of semantic web.

References

1. Sitha Ramulu V, Santhosh Kumar ChN, Sudheer Reddy K (2012) A study of semantic web mining: integrating domain knowledge into web mining. Int J Soft Comput Eng (IJSCE) 2 (3). ISSN 2231-2307
2. Stumme G, Hotho A, Berendt B (2006) Semantic web mining: state of the art and future directions. Web Semant: Sci Serv Agents World Wide Web 4(2):124–143 Semantic Grid – The Convergence of Technologies
3. Jeon D, Kim W (2011) Development of semantic decision tree. In: Proceedings of the 3rd international conference on data mining and intelligent information technology applications, Macau, 24–26 October 2011, pp 28–34
4. Sugumaran V, Gulla JA (2012) Applied semantic web technologies. Taylor & Francis Group, Boca Raton
5. Domingue J, Fensel D, Hendler JA (2011) Handbook of semantic web technologies. Springer, Heidelberg
6. Jain A, Khan I, Verma B (2011) Secure and intelligent decision making in semantic web mining. Int J Comput Appl 15(7):14–18. https://doi.org/10.5120/1962-2625
7. Yong-Gui W, Zhen J (2010) Research on semantic web mining. In: Proceedings of the international conference on computer design and applications, Qinhuangdao, 25–27 June 2010, pp 67–70. https://doi.org/10.1109/iccda.2010.5541057u

8. Jayatilaka ADS, Wimalarathne GDSP (2011) Knowledge extraction for semantic web using web mining. In: The international conference on advances in ICT for emerging regions, ICTer2011
9. Srivastava J, Desikan P, Kumar V. Web mining – concepts, applications and research directions
10. Berners-Lee T, Hendler J, Lassila O. The Semantic Web. http://semanticweb.org/wiki/Semantic_Web
11. Wilson TV. How Semantic Web Works. HowStuffWorks.com: http://www.howstuffworks.com/semantic-web.htm
12. Gruber T (1993) A translation approach to portable ontology specifications
13. Obitko M. What is Ontology. http://www.obitko.com/tutorials/ontologies–semantic–web/what-is-ontology.html
14. Web Ontology Language. http://en.Wikipedia.org/wiki/Web_Ontology_Language