International Journal of Progressive Research in Science and Engineering
Volume-1, Issue-5, August-2020
www.ijprse.com

Semantic Web Technology and Data Mining for Personalized System to Online E-Commerce

S N Kakarwal[1], Pradip M Paithane[2], Sushant S Khedgikar[3]

[1]*Professor, CSE Department, PES College, Aurangabad, Maharashtra, India.*
[2]*Assistant Professor, Computer Engineering Department, VPKBIET, Baramati, Maharashtra, India.*
[3]*Assistant Professor, Electronics Department, PES College, Aurangabad, Maharashtra, India.*

Corresponding Author: pradip.paithane@vpkbiet.org

Abstract: News-papers, blogs, and web-pages are a rich and diverse source of textual information. However, the information contained in these sources cannot be physically extracted, verified, and indexed, mainly because it comes in a massive size. Moreover, the extraction of some information sometimes requires specific knowledge or technical background. This is the case in the news domain where we need to extract the relevant news from a lot of available information. In order to scale knowledge extraction to the large size of available textual information, and build extractors specific to a certain field various techniques are applied over the unstructured data so that it can be made available to the users. This could help the researchers and the news readers or users to find applicable information in less time and with great ease. This study aims to review all the approaches and techniques done so far, for the information retrieval, search capability and its analysis and it also proposed an idea for better searching that reduces the time complexity to extract the data and also reduces human intervention. This is a better idea to put forward which also helps in filtering of irrelevant data and thus integrates only the relevant data to create a better space for the news data.

Key Words: - Information retrieval, Semantics, Natural Language Processing, Ontology, Semantic Web.

I. INTRODUCTION

Today, the World Wide Web is considering the main source of extensive information. The current search engines are generally unable to customize the results according to the users' preferences. Recently, the recommender system was proposed as an alternative approach for information retrieval. Recommender Systems are an intermediary program that intelligently generates a list of information, which matches the users' preferences and being used among different online business such as entertainment sites, social networks, medical applications, financial investments, and e-commerce applications. Recommender systems in e-commerce rely on the historical data and users' behavior to recommend items or services. Recommender systems are built on top of web mining usage [1], which is concerned with finding user navigational patterns by extracting the required information about access time and view pages from weblogs. Web mining usage has become popular in e-commerce, which becomes related to the development of the website. The extraction of navigation paths in recommendation applications in web mining usage is through web page addresses from web server, visit logs and patterns [2], [3]. Although the recommender systems filtering information and recommend items based on users' preferences, but still face many problems such as cold start and content overspecialization. In the last decade, the technology of the semantic web is used for enhancing the machines' understanding of the web by building an appropriate infrastructure of intelligent agents that have the ability to perform complex actions for users [4]. Semantic Web is about intelligently integrating information, providing semantic-based access to the Internet, and extracting information from texts to explicitly declaring the knowledge embedded in many web-based applications [5]. Finally, the semantic web is for making machines interpretable by implementing large-scale interoperation, and reliable web services. This means that the creation of web machine understandable and interoperable services, which intelligent agents can execute, compose, and discover automatically [6]. Therefore, there is an increasing effort to define web pages and objects regarding semantic information by using ontologies [7]. Several standards such as Resource Description Framework (RDF) and Web Ontology Language (OWL) have been developed to establish the layer structure of the semantic web [7]. As the Semantic Web becomes more usable with specific standards, more online businesses are starting to include ontologies domain in their online applications such as-Amazon.com, YouTube, Flickr, delicious, eBay, due to the continuous development and use of Semantic Web and Web 2.0 technologies. On the internet, the underlying of several ontology applications is relying on semantic web. Tagging and semantic annotation of web pages are also spreading widely on the World Wide Web, towards realizing the semantic web. Ontology provides a set of well-founded constructs that define significant concept and their semantic relationships.

II. SEMATIC WEB

Tim Berners Lee, the inventor of World Wide Web, conceived semantic web. According to Lee semantic web is defined as "extended Web of machine-readable information and automated services that amplify the Web far beyond current capabilities."[8,9]. The purpose of the semantic web is distinguished by more meaningful information presentation for humans and machines, which assist in providing a description of contents and services by form of machine-readable. Semantic web enabled services annotated, discovered, published, and composed automatically [10,11], by making the machines capable of understanding and representing the information by semantics way as human being. This means mediating between the need of users and the available resources of information by providing intelligent access to heterogeneous, distributed information, enabling software (intelligent agents) [12]. Semantic web can be viewed as an extension towards the meaning of the current web, which is supposed to provide machine accessible meaning for any applications that construct on web. This means that the indication of semantic web, by the ability of machines for solving well-defined problem by using well-defined data and executing well defined operations. Instead of asking machines to assume people's language, it works by asking people making extra effort so that the machines are able to process the data by specific way. The W3C initiatives for the semantic web are XML, RDF, OWL, and ontologies. [13, 14]. Semantic-based web data mining is a combination of semantic web and web mining. The result of web mining is helping in building the semantic web. The semantic web' knowledge can improve the effectiveness of web mining and makes it easier to achieve. According to Berners-Lee's vision, the semantic network is a seven-layered [15]. The first step is putting data on the Web in a form that machines can naturally understand. In order to determine the meaning of the collection of the documents, it is requiring the using the model theory of Resource Description Framework (RDF) and OWL model theory. The consumers of Web resources are referred to as agents. This sharing of information among different agents requires semantic markup, such as web page annotation with information on its content that is understood by searching the agents on the Web. This kind of annotation will be given in some standardized, expressive language and will make use of certain classes. For checking that different agents have a common understanding of these classes, ontologies are required for describing and interpreting the same concepts within agents.

A. Ontology:

The Ontology is defined as specification of shared conceptualization that presents formally explicit [16]. It is some kind of knowledge representation. Conceptualization means that the domain of interest is presented as a hierarchy of classes, subclasses and relationships [16]. The traditional search engine ranks billion of a webpage to discover a particular term or keyword, which appears in web documents [17]. Semantic web search engine is implemented through the agents, which deal with the different information sources to find the suitable answer. Moreover, the ontology enables the use of description logic, which is used to define some facts to infer new knowledge. In addition, ontology is considered the semantics building technique. Ontology is classified into ontology semantic annotation, ontology building, and ontology matching "mapping". Ontology semantic annotation is defined as the process of metadata' adding to web pages which makes it more meaningful and accessible by the search engines and the autonomous agents [16]. Ontology matching is the process of finding correspondences between two ontologies in order to integrate the required data between autonomous agents [18]. As the ontology is a collection of concepts' definitions and shared understanding that comes from the fact of all agents interpret the concepts of the same ontology and the same using of concepts standards will enable the reuse of the defined information. This means that, the using of the information is not annotated for a specific system; however, the annotation depends on shared standards, which make it promising to be accepted by different computer systems. By using ontologies in the Semantic Web, users can influence the advantages of the following two features: (i) data is displayed by using of the common vocabulary and rules of grammar; (ii) the description of the data in the semantic web is preserved in ontologies for being ready for inference.

The Semantic Web is an extension of the World Wide Web through standards by the World Wide Web Consortium (W3C). The standards allow simple data formats and allow exchange protocols on the Web, most commonly the Resource Description Framework (RDF). According to the W3C, The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. The Semantic Web is thus considered as an integrator for various contents, information applications and systems. In this paper you will know about the work done so far in second section of the paper, and in the

third section you will get to know about the new proposed methodology for a better and improved system to gather information from all the possible sources and present it to the user on a single platform. The fourth section gives the data set used for this system. In the fifth section we will get to know how the limitations of the work done so far can be overcome using the proposed methodology and it also gives the future

scope of the work done. The sixth section gives the conclusion of the entire work done.

III. PROCESSING DESIGN

The process of designing and implementing a recommender system as involves the following steps:

A. Data Collection:

The initial step involves the collection of the dataset for executing the data mining algorithms. This step also includes the collection of the ontology and semantic information for representing and organizing the information content.

B. Data Preprocessing:

The requiring of the data preprocessing is to clean and transform the collected dataset into formats that are suitable for data algorithms. In addition, the step includes the ontology construction for representing the information content.

C. Data Mining Algorithms:

It's considered the core process of the recommender system model, where the dataset is analyzed, and mining algorithms are applied to generate and discover best-recommended results that will be used in ontology method.

D. Implementation of semantic web:

The ontology will be implemented in protégé, and the reasoning technique was used to detect the user profile similarities and provide each user profile recommendations according to personal preferences, actions, and similarity with another user profile.

E. Ontology model includes four steps:

Information collector is the interaction layer between the system and user and provides the recommended items to users. In this layer, the system picks the user's file when answers some questions and connect to his/her account, gathering all data and displaying recommended information through the website.

Ontologies layer Personalized Recommender techniques based on two ontologies which consist of user information ontology and products/items ontology. Products/items ontology contains information about all items that offered online. The main purpose of this information is to ensure users' satisfaction when shopping online.

Recommendation Techniques (analysis layer), this step consists of two modules; capture user's preferences and user's friend's preferences. User's preference is collected from the profile of user and friend's references is collected from social network data. The analysis of entire user profile is required firstly for analyzing all data that will be used secondly in building user information ontology. Then recommender model generates. Recommendations by looking for similar items for user information ontology through products/items ontology.

User Profile will contain the user preferences. The user profile will be constructed by considering direct feedback from the users.

IV. CONCLUSION

In this paper, the proposed outline of a personalized recommender system that is composed of two phases using data mining technique for finding the best classifier to categorize users based on personal data, preferences, and characteristics for giving the accurate recommendation to them according to ontology base knowledge. The data-mining stage was applied by applying a different type of classification algorithms over the dataset in order to extract and construct the user access sequences. The comparative study was developed to show the best classifier algorithm used for the dataset by measuring its performance parameters to get the best classifier that fit online shoppers' behavior and attitudes based on the obtained dataset. The analysis showed that the decision table classifier gives the highest accuracy. The lowest accuracy is given by clustering and simple cart. The analysis of best classifier (decision table) will be assist and used when building the ontology model that will helping online users for finding items.

REFERENCES

[1]. Robin Wolz, Chengwen Chu, Kazunari Misawa, Michitaka Fujiwara, Kensaku Mori,Daniel Rueckert," Automated Abdominal Multi-Organ Segmentation With Subject-Specific Atlas Generation" , IEEE Transactions On Medical Imaging, Vol. 32, No. 9, Pp. 1723-1730,September 2013

[2]. M.Harsh vardhan,S.Visweshwara Rao," GLCM Architecture For Image Extraction", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 3, Issue 1, January 2014

[3]. Pradip M. Paithane, S.N. Kakarwal, "Automatic Determination Number of Cluster for Multi Kernel NMKFCM algorithm on image Segmentation", Intelligent Systems Design and Applications, Springer Cham, pp 870-879,2018.

[4]. Pradip M. Paithane, S.A. Kinariwal, "Automatic determination number of clusterfor NMKFC-means algorithms on Image Segmentation", IOSR-JCE,Volume 17,Issue1, Ver.2,pp12-19, Feb 2015.

[5]. Leonid K Antanovskii, “Implantation of the scale invariant feature transform algorithm in MATLAB”, Weapons and combat systems divisions defense science and technology group Australia, DST-Group–TR–3347,2017. [1] Jiawei Han and Michelin Kamber, “Data Mining Concepts and techniques “Morgan Kaufmann, 2001. Publishers 2nd ed. ISBN:1558609016. ISBN 13:9781558609013.

[6]. Jaideep Srivastava, Robert Cooley, Mukund Desh- Pande, Pang-Ning Tan "Web Usage Mining: Discovery and Applications of Usage patterns from web data," ACM SIGKDD, Jan 2000.

[7]. Kosala and Blockeel, "Web Mining Research Survey," SIGKDD: SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery and Data mining, ACM, 2(1):1- 15,2000.

[8]. Heflin, J., & Hendler, J. A Portrait of the Semantic Web in Action. IEEE Intelligent Systems 16(2), 54-59, 2001.

[9]. T. Gruber. Towards principles for the design of ontologies used for knowledge sharing. Int. J. of Human and Computer Studies, 43:907–928, 1994.

[10]. T. Ruotsalo, “Methods and Applications for Ontology-Based Recommender Systems,”, Ph.D. thesis, 2010. Collaborative filtering,” Expert Syst. Appl., vol. 53, pp. 204–218, 2016.

[11]. Suleyman Salin, Pinar Senkul, “Using Semantic Information for Web Usage Mining Based Recommen-dation” “24th International Symposium on Computer and Information Sciences, 2009. ISCIS 2009.

[12]. L. Yu, “Using ontology to enhance collaborative recommendation based on community,” Proc. - 9th Int. Conf. Web-Age Inf. Manag. WAIM 2008, pp. 45–49, 2008.

[13]. Berners-Lee, T., Hendler, J., & Lassila, O. The Semantic Web. Scientific American 284, 34–43, May 2001.

[14]. Berners-Lee, T., Fischetti, M., & Dertouzos, T. M. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor. San Francisco: Harper, 1999.

[15]. D. Fensel, "Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce", Springer, 2001.

[16]. M. Paolucci, T. Kawamura, T.R. Payne, and K.P. Sycara, "Semantic Matching of Web Services Capabilities", In Proceedings of International Semantic Web Conference, 2002, pp.333-347.

[17]. D. Tidwell, "Web Services-The Web’s Next Revolution", IBM Web Service Tutorial, 29 Nov. 2000. http://www106.ibm.com/developerworks/edu/ws-dwwsbasics-i.html.

[18]. Bettina Berendt, Andreas Hotho, Dunja Mlad enic, "A Roadmap for Web Mining: From Web to Semantic Web.".

[19]. WANG Yong-gui, JIA Zhen, “Research on Semantic Web Mining” 2010 International Conference on Computer Design and Applications (ICCDA 2010).

[20]. T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web", Scientific Am, May 2001, pp. 34– 43.

[21]. T. Gruber. Towards principles for the design of ontologies used for knowledge sharing. Int. J. of Human and Computer Studies, 43:907–928, 1994.

[22]. Z.ahmed,' Domain specific Information Extraction for Semantic Annotation', DIPLOMA THESIS, 2009.