Using Semantic Web Technology and Data Mining for Personalized Recommender System to Online Shopping

Rana Alaa El-Deen
Business Information Systems Department
Arab Academy for Science,
Technology & Maritime Transport
Cairo, Egypt
Rana-aast@hotmail.com

Shereen Morsi
Business Information Systems Department
Arab Academy for Science,
Technology & Maritime Transport
Cairo, Egypt
shereen_morsi@aast.edu

Nermeein Magdi
Business Information Systems Department
Arab Academy for Science,
Technology & Maritime Transport
Cairo, Egypt
nermeen-magdi@gmail.com

***Abstract*— The growth of the World Wide Web provides online users a large amount of information, which has led to the emergence of web recommendation systems to recommend items, services or pages to users. Although the recommendation systems assist in retrieving and accessing interesting items automatically by mainly rely on acquiring users' historical data and matching items with the preferences of users, but still have limitations due to many issues. The issues can be identified as content overspecialization, cold start problem, and sparsity of data structures. In addition, the current recommender systems fail to make full use of the semantic information about items and the relations among them. The aim of this paper is to improve traditional recommender systems by incorporating users' information from social networks and developing users' information ontology to make personalized recommendations. Therefore, the paper proposes a framework for semantic recommender system employing user profile ontology, and products/items ontology. The proposed model includes the using of data mining techniques for knowing which classification algorithm fit well user's data analysis, which will be next using in developing the user profile ontology. Then, building a semantic application and integrate the system for checking the accuracy of the recommendation model after the validation for each techniques built. This paper is presenting the experimental results of the first step by using data mining technique of proposed model. The results show that by analyzing users' data by applying different types of classification algorithms and based on the TPR results, the Decision table algorithm gives the highest TPR by (0.871). This means that the best classifier algorithm that can be used for building user profile recommender model.**

Keywords*— *Semantic Web, Ontology, Recommender system, Data Mining, e-commerce, Online Shopping.

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, sparsity, 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].

978-1-5386-4371-6/18/$31.00 ©2018 IEEE

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 are 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. Such constructs can be leveraged to build essential higher level knowledge in a particular domain such as e-commerce applications. Therefore, ontologies will be used in our proposed recommender system model. The paper is structured as: Section II presents related works. In Section III, a brief overview of Semantic Web is given. Section IV, introduces the Semantic Web proposed model. In section V, describes the implementation of the proposed model, while section VI contains a conclusion and future work.

II. RELATED WORKS

In the last decade, the recommender systems researchers have explored hybrid recommender systems as an approach for developing effective recommender systems. Hybrid filtering combines two or more recommendation techniques for improving the performance of recommender system. And in recent studies have shown the effectiveness of this type such as - Wei et al. [8] proposed a hybrid recommendation technique that driven from collaborative filtering and deep learning to overcome the cold start problem for new items. The result of their recommendation approach showed significant improvement in the cold-start problem. While Colombo-Mendoza et al. [9] proposed a hybrid recommender system based on context-awareness, semantic web technologies, and ontology for movies recommendation. Their hybrid recommender systems showed improvement in the recommendation performance.

In addition to Romadhony et al. [10] proposed a personal recommendation technique that combines collaborative filtering and content-based to avoid cold start problem for new items. The evaluation of recommender technique has been conducted by a survey on users who dealt with the recommender system, and the result showed significant improvement in avoiding the cold-start problem. Similarly, Ting et al. [11] propose a personalized recommender system based on weblog mining and weighted bipartite graph. Their experimental results indicate that combining weblog mining and the weighted bipartite graph is feasible and improves the recommendation results. W. Carrer-Neto et al. [12] proposed a hybrid recommender system combines knowledge and social networks that applied on movie graphic domain. The users of the system had an option of choosing the product to buy while their social network was influencing on the recommendations directed to them. The proposed system presents a social based contents recommender system that makes use of the semantic web principles to build a mechanism to help users in finding content that is relevant to their preferences. The model rely on collaborative and knowledge filtering that provide suggestion based on inherit recommendations from users' social network and semantic relations between contents of users preferences. In addition, the proposed model introduces the concept of "social aperture," by mean of which users can decide the impact of social networks that will have on the recommendations.

While, H.Liao [13] proposed a hybrid personalized recommender model rely on the combination of collaborative filtering and content-based. The model designed to effectively recommend information to users to users' interest that already knew, while recommender system could learn users' new interest in users' similar action. The researcher designed an online book system to carry out the research, and the data for analysis of personalized recommender system was getting from movie lens application. The model provided significant results. While Cheng et al. [14] proposed a hybrid recommendation model based on content-based and knowledge-based that aims to overcome the problems of recommender system. They proposed an ontology-based, personalized recommender system to solve these problems. The proposed model is applied on movie application by building ontololgy-based based on the internet movie database (IMDB by using protégé –OWL API from the protégé project. The recommendation model is based on an approach of multi-agent personalization. It is consists of three elements a)random selection agent which is responsible for avoiding the over specialization problem b)semantic discovery agent which aims to solve the problems of overspecialization and new item c)data mining agent which is responsible for resolving the new user problem.

Q.Wang, X.Yuan & M.Sun[15] proposed a hybrid recommender system based on item features and demographics information, and it focused on searching for set neighboring users shared with same interests which help to improve system scalability. For improving the accuracy of recommendation, each feature in the model is given a different weight of similarity among users in calculation step. A comparison experiment was used and performed on the dataset of movie lens. The experiment showed that the proposed algorithm could provide recommendations with higher accuracy compared to methods of traditional collaborative and content-based filtering. Yu [16] used the ontology to enhance collaborative filtering recommendation based on community-based. Their results show that collaborative filtering based on community achieves better performance than the traditional method.

III. SEMANTIC WEB OVERVIEW

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."*[17,18]. 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 to

annotated, discovered, published, and composed automatically [19, 20], 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)[21].

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 aim of the semantic web is providing an infrastructure from machine-understandable web resources that can be sharing and processing information with automated tools such as search engines. The W3C initiatives for the semantic web are XML, RDF, OWL, and ontologies. [22, 23]. 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 [24]. 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 is required for describing and interpreting the same concepts within agents.

IV. ONTOLOGY

Ontology is defined as specification of shared conceptualization that presents formally explicit [24]. This means that the machines can understand the ontology after validating the concepts through a group or a community. It is some kind of knowledge representation. Conceptualization means that the domain of interest is presented as a hierarchy of classes, subclasses and relationships [24].

The traditional search engine ranks billion of a webpage to discover a particular term or keyword, which appears in web documents [25]. The traditional search engine does not understand the meaning of the term and its relationships with other terms, so the search engine restricts the type of question the user asks for. Therefore, the semantic web has appeared to make web search more efficient. It enables the web search engine to understand the question that the user asks for. The 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 [26]. Ontology building [27]. Ontology matching is the process of finding correspondences between two ontologies in order to integrate the required data between autonomous agents [28].

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 depend 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; and the description of the data in the semantic web is preserved in ontologies for being ready for inference.

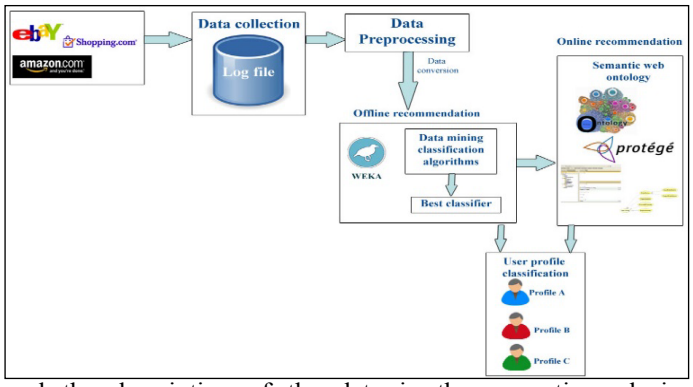

Fig. 1.Proposed framework for the personalized recommender system

V. THE PROPOSED FRAMEWORK

A. Process designing

The process of designing and implementing a recommender system as shown in figure 1 involves the following steps:

- 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.
- 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.

- 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.
- 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.

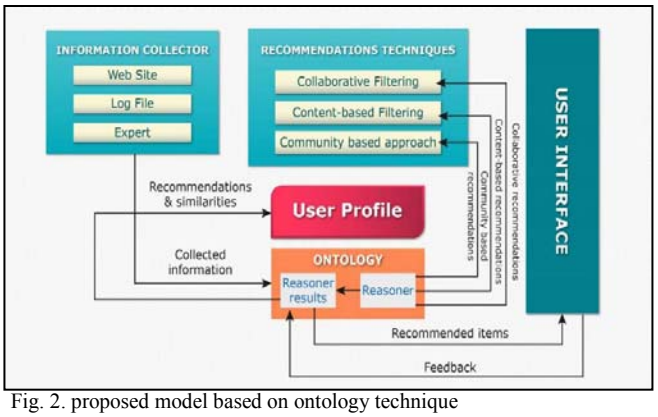

Fig. 2. proposed model based on ontology technique

From figure 2, the proposed 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 these information is to ensure users' satisfaction when shopping online. that will contain item
- Recommendation Techniques (analysis layer), this step consists of two modules; capture user's preferences and user's friends preferences. User's preferences is collected from the profile of user and friend's prefrences is collected from social network data (facebook social network). 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 genertate 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.

B. *Implementation of data mining step*

In this step, the dataset was collected from one of the companies that conduct e-commerce in Egypt. The dataset contained data on user access sequences, which are extracted from the web server log records to be preprocessing by using data mining. The dataset is consist of the online order log file for three months. It contains about 300 records and 26 attributes (personal information, educational level, brand, product name, type, product description, product category, product quantity sold, product price, payment method, rating, and satisfaction level). Data cleaned and formatted to data mining format for analysis in order to extract and construct the user access sequences. As the selection and implementation of the suitable data mining technique is the one of the main step in analysis layer, the different type of classification algorithms were applied over the dataset. In addition, the ten-fold cross validation method is used for accuracy testing for all classification types that selected for analysis. In ten-fold cross-validation, the dataset is equally divided into ten folds (partitions) with the same distribution. In each test, nine folds are used for training, and one fold (unseen dataset) is used for testing. The procedure of the test is repeated ten times. The final algorithm's accuracy is the average of the ten trials [29].

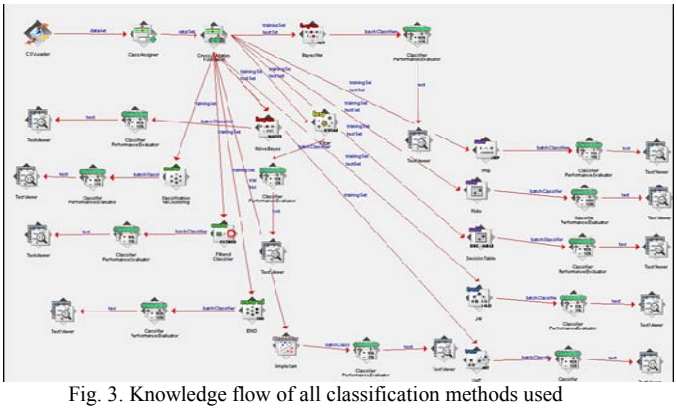

Fig. 3. Knowledge flow of all classification methods used

Figure 3 shows all classifiers algorithms used in the experiment. All the Knowledge flows start with three main nodes; the first node is loading csv file, which is responsible for displaying the dataset file; the second node is the class assigner that sets the new class attribute; and the third node is for the class validation fold maker, which responsible for separating the testing set from the training set. Finally, all sets are connected.

C. *Compared classifiers parameters*

The below table no. 1 represents a comparative analysis between eleven classification algorithms. The classification algorithms were chosen from each major classifier groups by choosing two classifiers from each family (Decision Table, SimpleCart, Bayes Net, K Star, Naïve Bayes, Classification via Clustering, Filtered Classifiers, JRIP, END, Ridor, J48).

The comparison among classification types was based on true positive rate (TP), false positive rate (FP), precision that is the probability of randomly selected retrieved relevant document [30], and recall which is the probability of randomly selected relevant document that retrieved in search [31]. F-measure parameter that combines precision and recall is the harmonic mean of precision and recall [31]. Receiver Operating Characteristic (ROC), which explains the binary classifier system performance as its discrimination threshold, is varied by comparing the two operating characteristics (TPR and FPR) as the criterion changes.

According to the comparison of results through performance measured parameter for each classifier and showed that the TP Rate of the decision table is the highest rate by 0.871 and classification via clustering is the lowest one that recorded 0.363. The FP Rate decision table is the lowest which record 0.078 and the simple cart was recorded the highest FP rate with 0.482. The precision of decision table is the highest by 0.883 and simple cart is the lowest by 0.232. The Recall of decision table is the highest by 0.871and classification via clustering is the lowest by 0.363. The F-measure of simple cart is 0.313 which is the lowest results and the decision table was recorded the highest which record 0.872. And the Roc Area of END is the highest and simple cart is the lowest.

Table 1 Classifiers Parameters

Classifier Name	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Decision Table	0.871	0.078	0.883	0.871	0.872	0.955
Filtered Classifier	0.868	0.079	0.878	0.868	0.869	0.952
END	0.865	0.083	0.872	0.865	0.864	0.957
J48	0.861	0.086	0.862	0.861	0.859	0.88
JRIP	0.858	0.089	0.865	0.858	0.857	0.894
Ridor	0.842	0.083	0.839	0.842	0.839	0.879
Bayes Net	0.716	0.093	0.764	0.716	0.735	0.886
Naïve Bayes	0.541	0.189	0.564	0.541	0.549	0.76
K Star	0.508	0.263	0.49	0.508	0.496	0.696
Simple Cart	0.482	0.482	0.232	0.482	0.313	0.483
Classification via Clustering	0.363	0.37	0.321	0.363	0.337	0.497

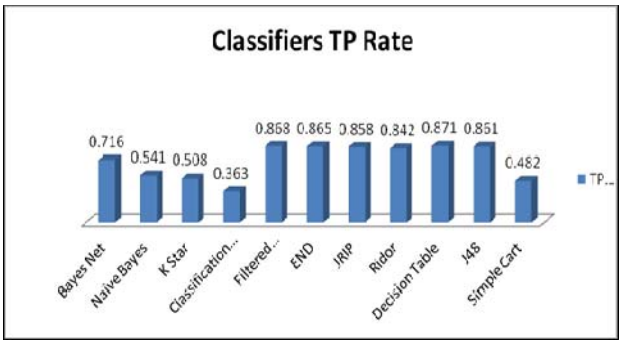

Fig. 4. TP rate of selected classification types

Based on the analysis result from WEKA software, figure 4 displays the results of TPR for eleven classification algorithms which were selected and representing on the X-axis. The results show that Decision table algorithm gives the highest TPR by (0.871) and the Classification via Clustering gives the lowest TPR by (0.363).

VI. CONCLUSION AND FUTURE WORK

In this paper, we presented the proposed framework of a personalized recommender system that is composed of two phases using data mining technique for finding the best classifier to classify users based on personal data, preferences, and characteristics for giving the accurate recommendation to them according to ontology base knowledge. The data-mining phase was implemented 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] Jiawei Han and Michelin Kamber, "Data Mining Concepts and techniques "Morgan Kaufmann, 2001. Publishers 2nd ed. ISBN: 1558609016. ISBN 13:9781558609013.

[2] 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.

[3] 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.

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

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

[6] 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.

[7] 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.

[8] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, “Collaborative filtering and deep learning based recommendation system for cold start items,” Expert Syst. Appl., vol. 69, pp. 29– 39, 2017.

[9] L. O. Colombo-Mendoza, R. Valencia-García, A. Rodríguez-González, G.Alor- Hernández, and J. J. Samper-Zapater, “RecomMetz: A contextaware knowledge-based mobile recommender system for movie showtimes,” Expert Syst. Appl., vol. 42, no. 3, pp. 1202–1222, 2015.

[10] A.Romadhony, S.Al Faraby, B.Pudjoatmodjo, 'Online shopping recommender system using hybrid method,' Information and Communication Technology (ICoICT), 2013 International Conference of , vol., no., pp.166,169, 20-22 March 2013.

[11] Y. Ting, C. Yan, and M. Xiang-wei, “Personalized Recommendation System Based on Web Log Mining and Weighted Bipartite Graph,” 2013 Int. Conf. Comput. Inf. Sci., pp. 587–590, 2013.

[12] W. Carrer-Neto, M. Hernández-Alcaraz, R. Valencia-García and F. García-Sánchez, 'Social knowledge-based recommender system. Application to the movies domain', Expert Systems with Applications, vol. 39, no. 12, pp. 10990-11000, 2012.

[13] H.Liao, 'A new web service model of hybrid personalized recommendation,' *Natural Computation (ICNC), 2013 Ninth International Conference on* , vol., no., pp.868,872, 23-25 July 2013.

[14] S. T. Cheng, C. L. Chou, and G. J. Horng, “The adaptive ontology-based personalized recommender system”, Wireless Personal Communications, vol. 72, no. 4, pp. 1801–1826, 2013.

[15] W. Chen, Z. Niu, X. Zhao, and Y. Li, “A hybrid recommendation algorithm adapted in e-learning environments,” World Wide Web, vol. 17, no. 2, pp. 271–284, 2014.

[16] X. L. Zheng, C. C. Chen, J. L. Hung, W. He, F. X. Hong, and Z. Lin, “A Hybrid Trust- Based Recommender System for Online Communities of Practice,” IEEE Trans. Learn. Technol., vol. 8, no. 4, pp. 345–356, 2015.

[17] 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.

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

[19] 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.

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

[21] 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.

[22] 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.

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

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

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

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

[27] K.Laudon,J.Laudon,Management Information Systems, Ninth Edition, Chapter 8: Telecommunications Networks,and the Internet.Page 285.

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

[29] T. Gruber. 'Towards principles for the design of ontologies used for knowledge sharing', Int. J. of Human and Computer Studies, 43:907– 928, 1994.

[30] A. H, Q. A., M. N. and E. M., 'A Comparison Study between Data Mining Tools over some Classification Methods', *SpecialIssue*, vol. 1, no. 3, 2011.

[31] A. M. Ragab, A. Y. Noaman, A. S. AL-Ghamdi and A. I. Madbouly, ,“A Comparative Analysis of Classification Algorithms for Students College Enrollment Approval Using Data Mining, ”in Interaction Design in Educational nvironment, 2014.