Online Social Network Message Filtering Using Ontology

Thilagavathi . N
Assistant professor
Department of IT
Sri ManakulaVinayagar
Engineering College
Pudhucherry, India

Indhumathi. S
UG student
Department of IT
Sri ManakulaVinayagar
Engineering College
Pudhucherry, India

Sowmiya. V
UG student
Department of IT
Sri ManakulaVinayagar
Engineering College
Pudhucherry, India

Vinothini. J
UG student
Department of IT
Sri ManakulaVinayagar
Engineering College
Pudhucherry, India

***Abstract*— Online Social Network (OSN) is a vast environment where people communicate with each other and share their ideas. The ideas that are shared by the person are posted on the user profile. So that all the views of the person are shared among the people. The major drawback in OSN is the irrelevant messages are shared in the user's wall without the knowledge of the user. The user doesn't require these irrelevant messages on their wall, so it needs to be filtered. This problem can be overcome by creating an Ontology that includes set of rules based on the user's interest to avoid irrelevant messages. Ontology is a generic knowledge that represents agreed domain semantics that can be reused by different kinds of applications or tasks. The content that is posted in the wall is analyzed using content analysis method. This paper proposes a method for the online social network using ontology to clear unwanted messages posted in the wall.**

Keywords*— Ontology, Content analysis, Rule based system, Keyword Extraction.

I. INTRODUCTION

Semantic web is not a separate web; it is the extension of already existing one, which provides information in a well-defined form. Semantic web is used to convert the unstructured or semi structured data into a web of data. Semantic web content are easily accessed by the humans, but it is not easily accessed by the computers. Semantic web are build using the similar meaning, patterns, structures, relations from the existing web. Semantic web are used in data mining with the help of ontology. Ontology is the backbone of semantic web, by using ontology the semantic web are transformed into understandable form. Some of the SW technologies are XML and RDF.

Ontology is study of nature and categories of being and their relationships. Ontology acts as backbone to semantic web. The web content are not understand by the user, to transform the web content into user understandable form ontology is used in data mining. Ontology is a technique that is used in the semantic web to do the transformation. Ontology is not only used to semantic web transformation but also for defining the classes, relations and properties of each individual.

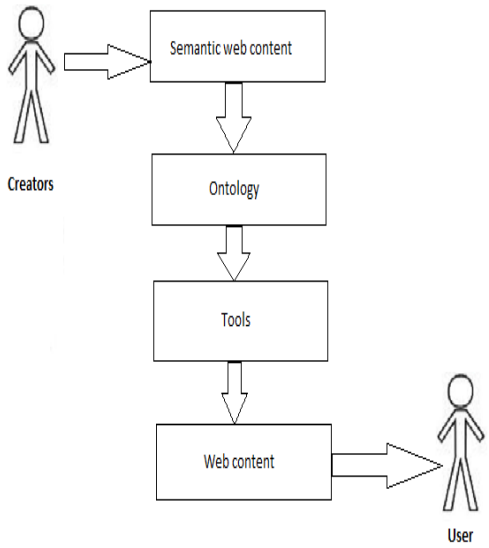

Figure 1: Ontology

Online social network (OSN) plays an important role in today human life. It is the place where the individuals share their idea and communicate with each other in
day today life. The major drawback in OSN is irrelevant messages are posted on user wall by some persons without the knowledge of the user. Those messages are not read or the user does not give any importance to those messages, sometime it gives user a bad impression about the person who made the post on the user wall. Thus there is a need for an automated system to eliminate that irrelevant post on the user wall which is not based on user interest.

The basic unit of OSN is post made by the person on the user wall. User interest is obtain from the likes and dislikes made by the user, most visited pages by the user and based on the user interest that are specified by the user in their profile creation. This task is very challenging since it needs to consider the user behavior and semantic information of the post in user wall. Using the information collected from the user behavior Ontology is created. Then the post or messages posted on the user wall are collected and segmented using the Ontology created. The content collected and segmented are stored in the database. Based on the data stored in the database and the rules framed by the user the messages are restricted on the user wall. By

retrieving the semantics of the word posted based on the wall, the prescribed domain of the word is identified using Ontology.

II. RELATED WORK

Many researches are done for the filtering of messages, providing privacy to the users in Online Social Network. The exact solution to filter the messages is not found. K. Strater et al proposed a system where the possibility of the threads and the vulnerability level is to be identified so that better privacy can be provided [1].But few messages cannot be restricted by using that system. Apart from the vulnerability level and other parameter. Sriram et al proposed a system which includes short text classification where different types of categories like News, Events, Opinions, Deals, Private Messages [2] are used to classify the text in Twitter but it doesn't find all the words because the keywords used to find the categories are minimum so that new words are posted on the wall of same categories cannot be found. This system includes a symbol representation to find the categories of message. The @ symbol is used to denote that the message is private. With this symbol the content classification is done. This paper proposes an approach to classify the content posted in user wall from OSN, ontology is created based on the interest and likes made by the user in OSN and assigning rules to eliminate the content which are not needed by the user.

III. SYSTEM ARCHITECTURE

Online Social Networks (OSN) is one of the important communication media to share several types of content, including free text, image, audio, and video data. There is a possibility of posting or commenting other posts on particular public/private areas, called in general as walls. Information filtering can therefore be used to give users the ability to automatically control the messages posted on their own walls, by filtering out unwanted messages. Rules are specified using Ontology to filter the messages that are blocked by the users. The users may have interest on specific domain that will be added into the account of the users. Message posted by other users will be of any domain, but the system will post the message to the user's wall that belongs to the user's interest. Messages that are not related to user's interest will be blocked and will not be posted in the wall.

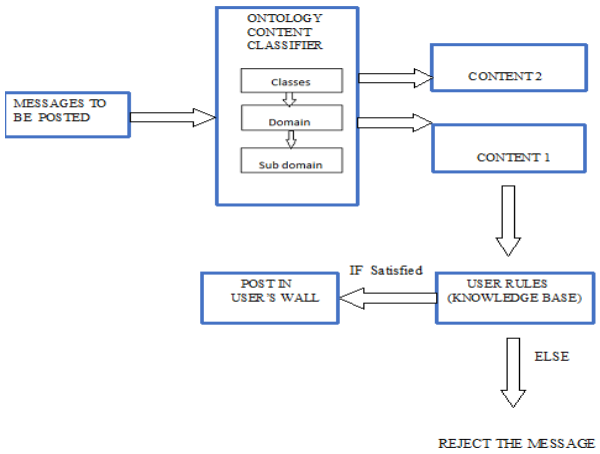

Figure 2: Architecture diagram on online social network message filtering using ontology

IV. RESEARCH PROPOSAL

STEP 1: USER PROFILE CREATION

To be a member of Social Network the user need to create a profile. All the information about the user is given during registering into the site. The user's information is one of the major input that to be considered. Based on the interest specified by the user the messages are to be decided whether to be post it on the wall or not. Then the user can login using the user ID and password that is created by the user. By logging on it directs to the user profile where the user can see the posts that are made by the friends of the user and also by the other users.

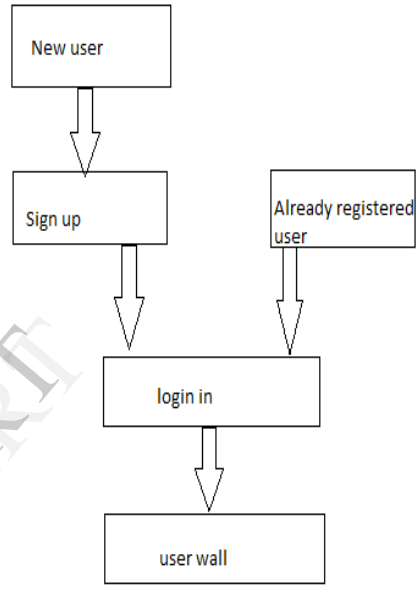

Figure 3: User Profile Creation

STEP 2: ONTOLOGY CREATION

Ontology is also known as content classifier which consists of class, subclasses and relationship between the classes are created in the ontology using the tool protégé. Ontology is created to classify the messages posted on the user wall. The ontology is automatically generated based on the user interest and likes made by the user on their profile. The classes are generated based on the interest that is specified by the user during the registration of the user profile and based on the likes that are done by the user and also based on the most visited pages by the user. By using the keywords specified in the object value the content is classified. The classification is used to identify the word which domain it belongs to.

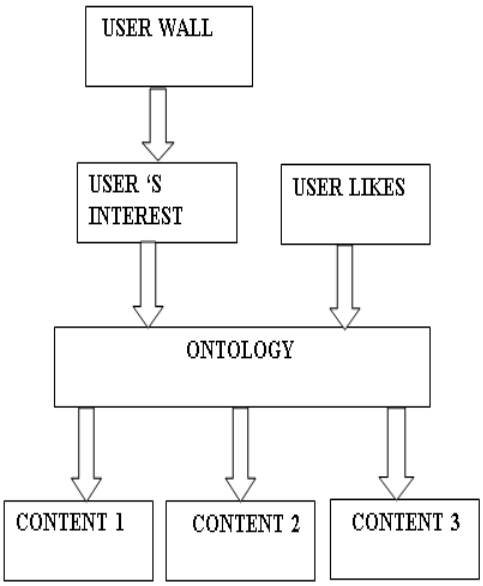

Figure 4: Ontology Creation

STEP 3: ANALYSING THE CONTENT

Content analysis is a summarizing, quantitative analysis of messages that relies on the scientific method (including attention to objectivity, intersubjectivity, a priori design, reliability, validity, generalizability, replicability, and hypothesis testing) and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented. The content is analyzed by using the Ontology created system by analyzing the content the meaning of the word is to be found. The word with similar meaning of the word is also being considered it is also added to the ontology dynamically, Which helps the system to block the messages. A single word has more than one meaning or different word with same meaning those words are collected by the content analysis which is done by the tool WordNet, which gives the noun, adjectives and so on. On collecting those words is added to ontology which generates the dynamic ontology. So that the ontology is generated dynamically and is used for the process.

STEP 4: BLOCKING THE CONTENT

The content is blocked based on the rules that are formed using the classes and subclasses created in ontology dynamically. Following the rules specified by the user the content which is to be posted on the wall and what are the posts that are to be blocked is identified. Documents processed in content-based filtering are mostly textual in nature and this makes content-based filtering close to text classification. The content which is irrelevant to the user's interest and the likes, those messages are blocked to post from the user wall. The rules are formed based on the ontology created and based on the relationship created between the classes and subclasses. The post that are posted on the user wall are given to the ontology in the protégé tool, it is blocked based on the rules that formed. The blocked messages are not posted on the user wall. Other messages that are not based on the rule are posted on the user wall.

V. CONCLUSION

Ontology is created based on the domain knowledge.An ontologies a catalog of the types of thingsthat are assumed to exist in a domain of interest*D*from the perspective of a personwho uses a language*L* for the purpose of talking about domain.Ontologies are used not only to represent a domain of interest, but also define concepts, describe relations among them and insert individuals. The existing system uses a method for filtering unwanted messages in a social networking using machine learning. To filter the unwanted messages posted on the user wall in a social networking content based filtering method is used. The proposed system uses ontology concept to filter the messages. It is done by creating the rules in order to block the unauthorized user and unwanted messages posted by the people on the user wall. The messages posted on the wall is analyzed by the content analysis method. Then the content is filtered based on the rules that are specified by the user. This provides a better result than the machine learning technique. Unwanted messages, unauthorized people and also the advertisements can also be blocked in the future.

REFERENCES

[1] K. Strater and H. Richter, "Examining Privacy and Disclosure in a Social Networking Community," Proc. Third Symp. Usable Privacy and Security (SOUPS '07), pp. 157-158, 2007.

[2] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short Text Classification in Twitter to Improve Information Filtering," Proc. 33rd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '10), pp. 841-842, 2010.

[3] Liping Jing, Lixin Zhou, Michael K. Ng, Joshua Zhexue Huang, "Ontology-based Distance Measure for Text Clustering" Comm. ACM, vol. 90, no. 12, pp. 29-38, 1998

[4] Belkin, N.J., Oddy, R.N. and Brooks, H.M. ASK for information retrieval: Part I. Background and theory. J. Doc. 38, 2 (June 1982), 61-71.

[5] Sankaranarayanan, J., Samet, H., Teitler, B. E., Lieberman, and M. D., Sperling, J. TwitterStand: news in tweets. In Proc. ACM GIS'09 (Seattle, Washington, Nov. 2009), 42-51.

[6] http://protege.stanford.edu/publications/ontology_ development/ontology101-noy-mcguinness.html.

[7] http://en.wikipedia.org/wiki/Data_mining.

[8] http://searchcrm.techtarget.com/tip/Major-issues-in-data-mining

[9] http://en.wikipedia.org/wiki/Semantic_Web