SOCIAL NETWORK DATA RETRIEVAL USING SEMANTIC TECHNOLOGY

ABHILASH SRIVASTAV, ALOK CHAUHAN*

School of Computing Science and Engineering, VIT University (Chennai Campus), Chennai, Tamil Nadu - 600 127, India.
Email: alok.chauhan@vit.ac.in

ABSTRACT

Social network data analysis is an important problem due to proliferation of social network applications, amount of data these applications generate, and potential of insight based on this big data. The objective of the present work is to propose architecture for a semantic web application to facilitate meaningful social network data analytics as well as answer query about concerned ontology. Proposed technique, on the one hand, links tools based on semantic technology provided by social network applications with data analytics tools and, on the other hand, extends this link to ontology authoring tools for further inference. The results obtained from data analytics tool, results of query on generated ontology, and benchmarking of the performance of data analytics tool are shown. It has been observed that a semantic web application utilizing above-mentioned tools and technologies is more versatile and flexible and further improvements are possible by applying generic data mining algorithms to the above scenario.

Keywords: Social network, Semantic web, Ontology, R-Studio, Graph API, Token access.

INTRODUCTION

With the increasing social networking demand, it has become more popular for connecting the people together through which people can interact with each other by sharing information, interest, and ideas together; due to this, the user is continuously increasing social network day-by-day. Directly, search information from raw data of any social network data will be devastating. Moreover, the other unsolved problem is that there is no mechanism to define relationships among people who are connected with social network. This type of problem plays an important act in the social networking.

Through semantic web, the information is stored in the semantic web in a structured way of information by that a computer can easily identify and analyze that how many users are connected with social network, so it can be implemented by SNS; it means that social network service provides all the services related to social networking sites such as Facebook and Twitter [2]. Hence, semantic web and social network service can combine with each other to build up the semantic social network service. Using this service, a computer system can be more helpful in activities such as locating the data in different formats. Semantic social network service can improve the efficiency of information sharing and integration. In this paper, the social network will be retrieved by performing the semantic search operation and by using an ontology which will integrate with each other, and one more thing is that in this research, the social network data will be analyzed with the help of Rtool which shows the relationship among all the people who are connected with the social network. The format of semantic web is in resource description framework (RDF) format, that is, RDF that contains the owl file.

LITERATURE SURVEY

By focusing on different surveys, there is a problem to retrieve data from semantic web because in that there is no technique to retrieve data easily, semantic web is vast and more information is stored about social networking sites. Accessing the data may be critical through web which is connected with the ontology; the ontology is the backbone of semantic web [6]. Hence, the social network data can be retrieved, but it cannot be analyzed in a specific manner due to the absence of any tool which can analyze the social sites such as Facebook and Twitter.

The ontology uses the owl file which generates RDF files of particular domain which has to be created because the social network uses the primary data source [7]. The ontology which contains the SPARQL Query, RDF, and OWL file provides only some related information through which the user cannot understand and analyze the social network data. Social network using semantic technology does not provide full information through semantic web to the user because it does not give any representation of the social network, i.e., how people are connected to each other and how they communicate.

Overview of ideal semantic web application

The ideal semantic web application simply works on some files, that is, RDF, web pages, and some database files [2]. The main question is that how this application will perform. It works by following procedure: First of all, the data is fetched from database in the form of web pages; then, the converter helps convert the database files into RDF files that are stored in the separate database and it can be accessed through RDF query engine [4]. The engine helps retrieve the data in the form of RDF; basically, RDF uses the XML file which is also known as owl file. The RDF query engine uses the query in structured form. In Fig. 1, we discussed the application of RDF query engine which converts the file into RDF format; this engine manages and controls the RDF database where the RDF is stored (the data is collected from web pages for a specific application). The application contains RDF parser which helps convert the web pages data into RDF format [5]. According to the requirements, the data is updated very fast for the sharing of data from one side to another. This process is the part of abstraction; it means that all operations are already done at the back end. The admin and other
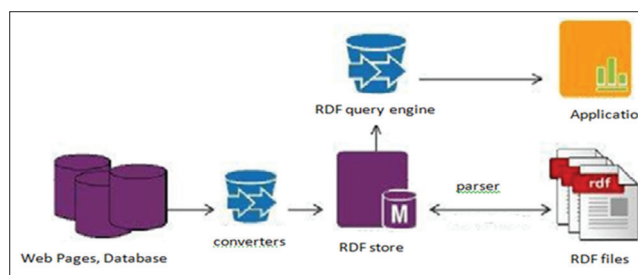

Fig. 1: Structure of ideal semantic web application

authorized users know all the information about these processes. The overall view is that the process of data fetching from the web that is not in a meaningful way and does not facilitate about analytics procedure needs to be taken care of.

PROPOSED WORK

In proposed system, it overcomes all problems which are present in the existing system. The solution is that how to understand and take some insight of social network data by not only using semantic web but also using ontology; we have to use Rtool for analyzing the social network data, i.e., Facebook data, within that, we have to install Rfacebook package in Rtool; hence, from that, the data will be gathered and it will be visualized using different plots, which gives some knowledge for performing this operation, accesses the token from Facebook, and also accesses the API Graph Explorer on getting access token through which extraction can be done easily.

The data is directly fetched in the Rtool; the data is in the form of comma-separated values (CSV) format which can be converted to owl format; owl means web ontology language which is used in semantic web [7]. Graph API Explorer uses the semantic web services that will be achieved by accessing the token from social network, so it can analyze the friends' location and relationship status and shows the graph of connected friends which can be shown through connected graph script that is used in Rtool. The ontology worked as a semantic web through which it identifies the various domains which exist in social network, by retrieving the data in the form of RDF files that is extracted by using query engine [2]. To perform implementation of social network, first of all, access the token from Facebook, then install the Rfacebook and dependency package, and as Graph API Explorer gets clicked on get access token, it goes for analyzing the social network data by plotting graph such as pie chart and bar chart. In Fig. 2, we discuss the overall structure of an application of social network.

### Detailed description of flowchart

Graph API explorer
The Graph API is one of the high-level platforms of Facebook components which are available in the Facebook developer options [1]. In Fig. 3, this API is generally used to access the data from Facebook, that is, friends' data; it includes friends' list, friends' location, friends' gender, and friends' relationship status. This Graph API is to be copied in the R-Studio as access token; the API uses simple and consistent view of social graph that usually represents the object of social network. The Graph API uses the RDF files that are used to apply semantic technology.

#### *Analytics on friends' location and relationship status*
In Fig. 4, inference of friends' location that determines how many friends or people are from the same location (for that operation, view the location data of friends) is shown. In this figure, there are total four locations that are shown in the form of pie chart. Pie chart is used because it scatters the data in different locations; this chart is implemented on console window of R-Studio, so it is best way to analyze friends' location data. The data is shown in the form of percentage as well as color form. According to cities, it is scattered to get the inference of friends' location. In Fig. 5, inference of friends' relationship status, that is, single, complicated, and open relationship, is determined that shows how many friends come under which category of status. The percentage describes total amount of status in a particular category that is scattered in the form of pie chart.

Network graph
In Fig. 6, the network graph shows how the friends are connected to each other and how they can share the information, interest, and ideas. The network graph shows the internal structure of social network. This type of graph connects the friends by nodes and edges. In the figure of network graph, five friends are interconnected to each other in that there is one common friend who is connected to all friends. According
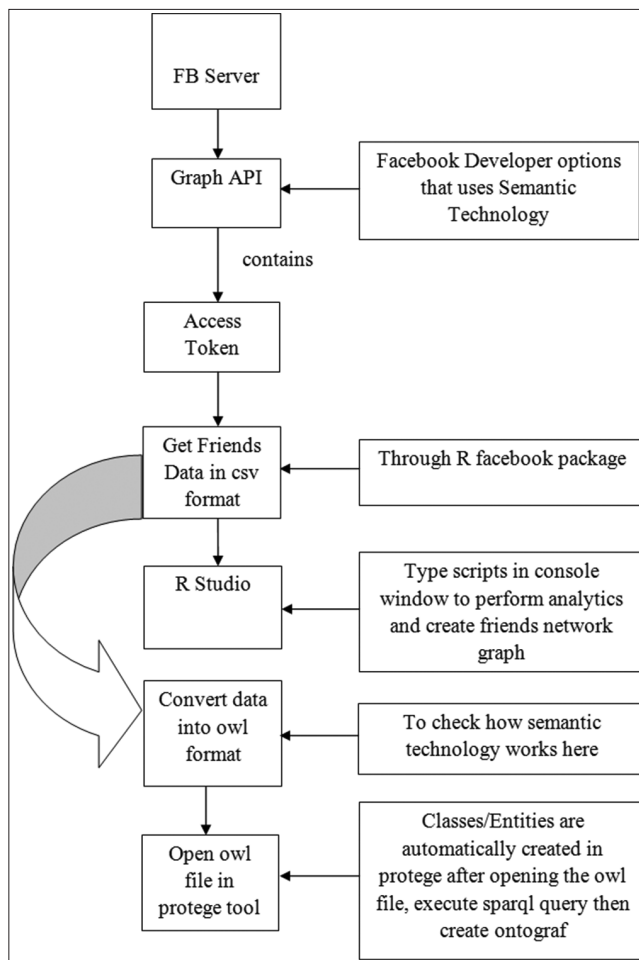
Fig. 2: Flowchart of application
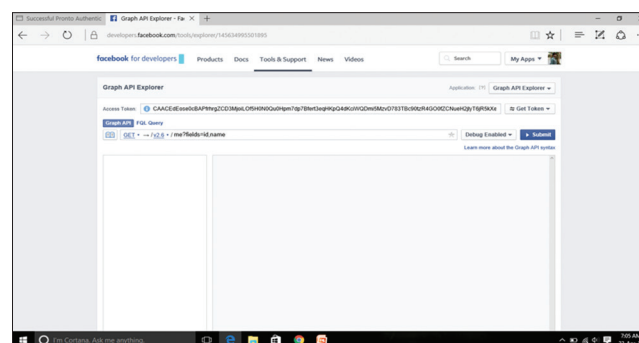
Fig. 3: Graph API

to the figure, the nodes are connected to each other, some nodes are scattered in different areas in network graph, this type of nodes shows that the friends are not mutual or not interconnected to each other, but it connected with one person and not with all. Basically, the network graph consists of nodes and edges, edges connect the nodes, and nodes consist of entities that are friends' name. When the friends are connected to each other, then it shows as a mutual friend from that they can share information easily with each other. Analytics on network graph tells full information of social network, so the user can easily get some inference of social network. The network graph shows the structure of friends' network. The total numbers of friends are already shown in network graph. The network graph contributes maximum number of social network analytics when the data is retrieved using semantic services.

Conversion of CSV file

The CSV file gets converted through online processing; first of all, convert the CSV file to JSON file, then convert into RDF file, finally RDF file is converted into owl file that is web ontology language file as shown in Fig. 7. Basically, the owl file consists of several XML and RDF data [6]. During conversion of CSV to owl, one thing always is to be remembered that the conversion is not done directly, but step-by-step, we can convert the file to analyze the behavior of semantic web through ontology. The owl file will be treated as ontology file which is used for analyzing that how semantic web works here. The owl file consists of classes or entities. Ontology file consists of RDF, XML, JSON,

etc. [6]. Online conversion of data from CSV to owl is also available; otherwise, we can access it from graph API web ontology language file; it represents the classes or individual entity. We can also implement the ontology using RDF file that is RDF It uses the syntax and semantics or behavior of social network. The RDF stores the data in the triple format at the back-end side of semantic web. The human can also directly read the owl file after conversion from CSV to web ontology language.

Classes and individuals in protege tool

After opening the owl file, the classes or entities are automatically generated in Protege tool; otherwise, we can also create classes itself; using classes, we can also create separate individuals of subclass. Basically, the classes or entities define the properties of specific domain as mentioned in Fig. 8. The SPARQL query is also generated after creation of classes; on the basis of classes, the OntoGraf will also be generated. The social network data is stored in the form of classes of each individual that shows the class hierarchy as shown in Fig. 8.

SPARQL query

SPARQL query is an RDF query language that is known as semantic query language for data store, from which the user is able to retrieve data. The SPARQL query is generated after creating classes or entities in an ontology using Protege tool; if any user wants to access the data, then first of all execute the query from a specific domain. SPARQL query contains prefix of RDF and prefix of owl, subject, and object of entities [8]. In Fig. 9, there is a separate figure that tells about the execution of SPARQL query; in the left-hand side of Fig. 9, SPARQL syntax is created; and in the right side, SPARQL query is executed in the form of subject and object of classes; both subject and classes are combined together during execution of SPARQL query because classes or entities can also be in the form of subject and object, and for that purpose, SPARQL query is used. SPARQL query is fully defined for a particular class of specific domain. The social network data is also present in the form of subject and object during SPARQL query execution.
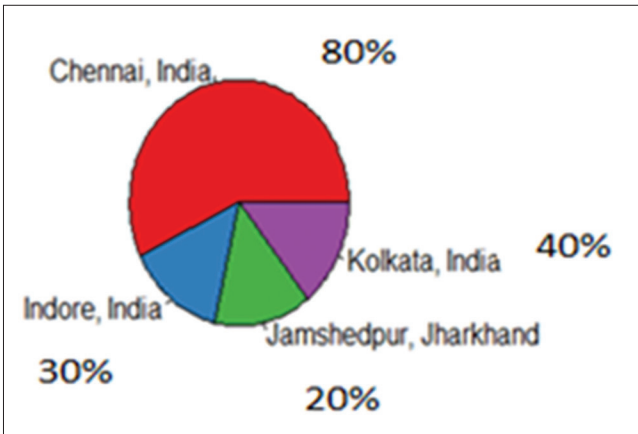

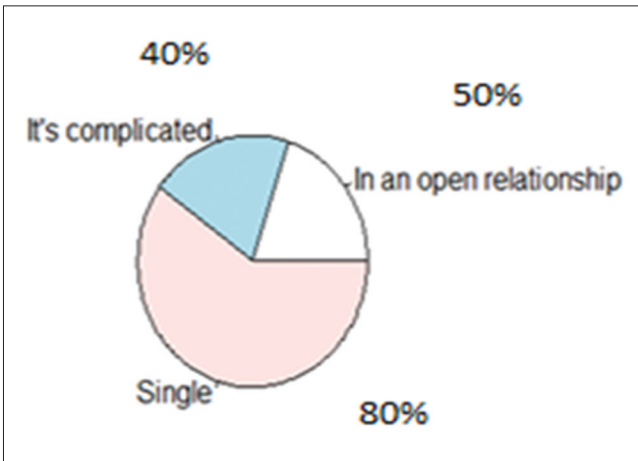
Fig. 4: Inference of friends' location

Fig. 5: Inference of relationship status
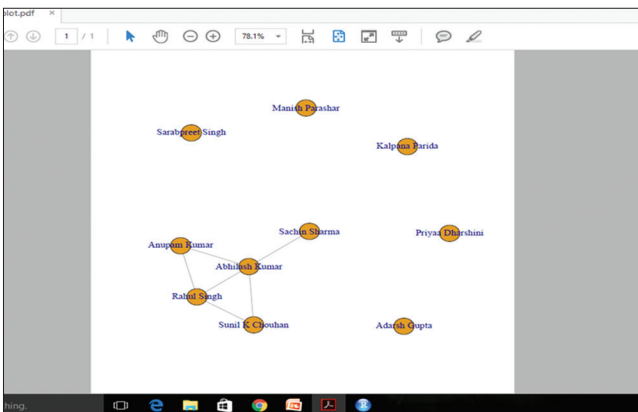

Fig. 7: Web ontology file

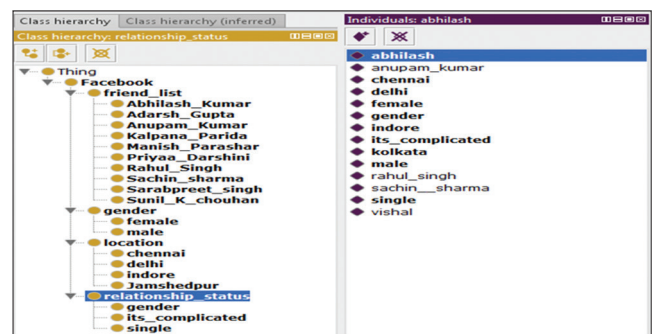**Fig. 6: Friends network graph**

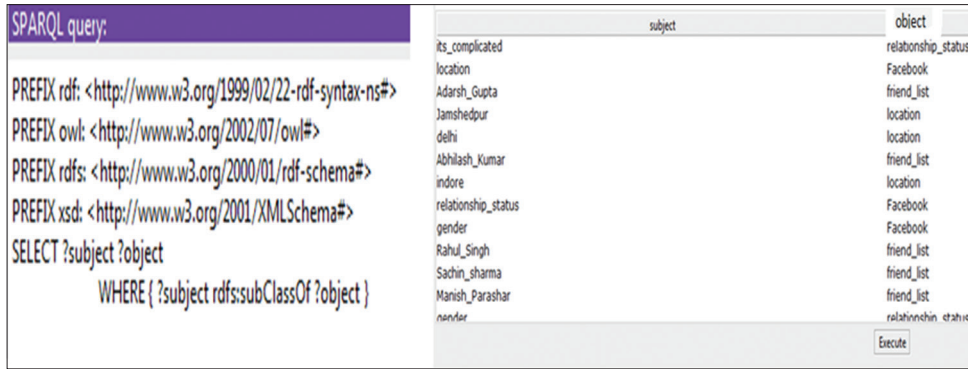**Fig. 8: Classes and individuals**
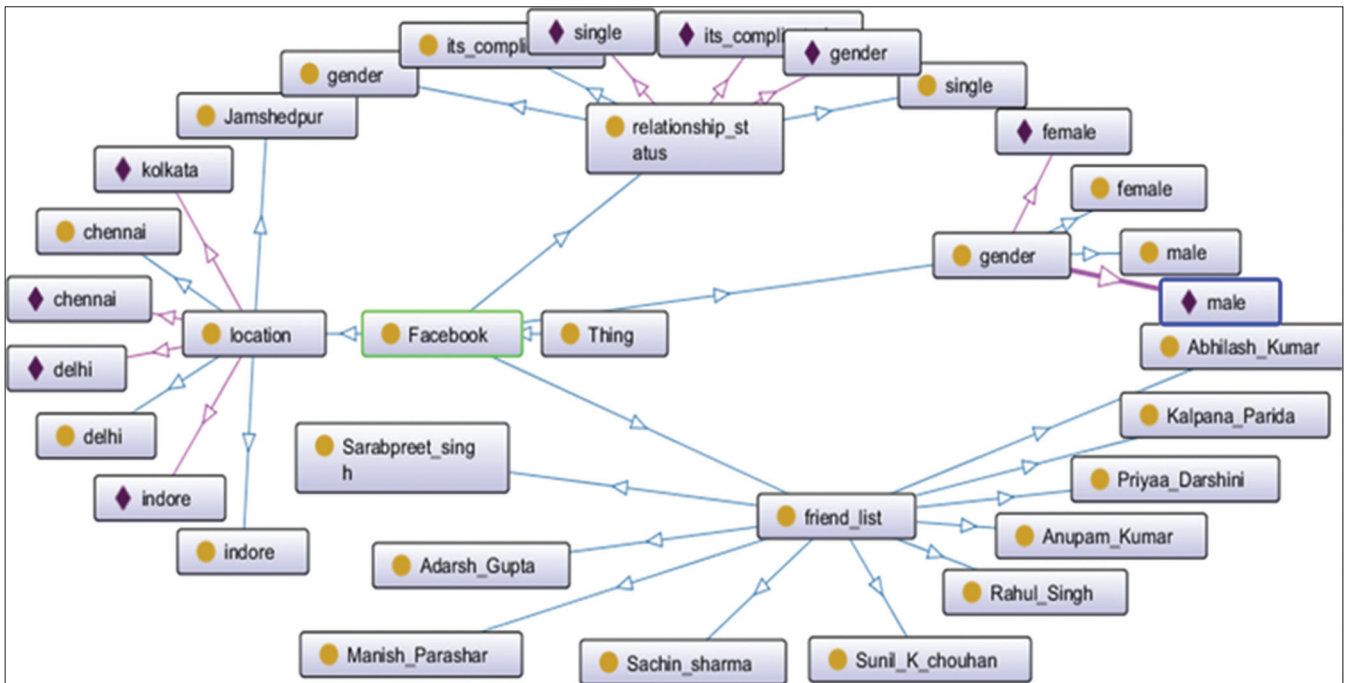
Fig. 9: (a and b) SPARQL query execution

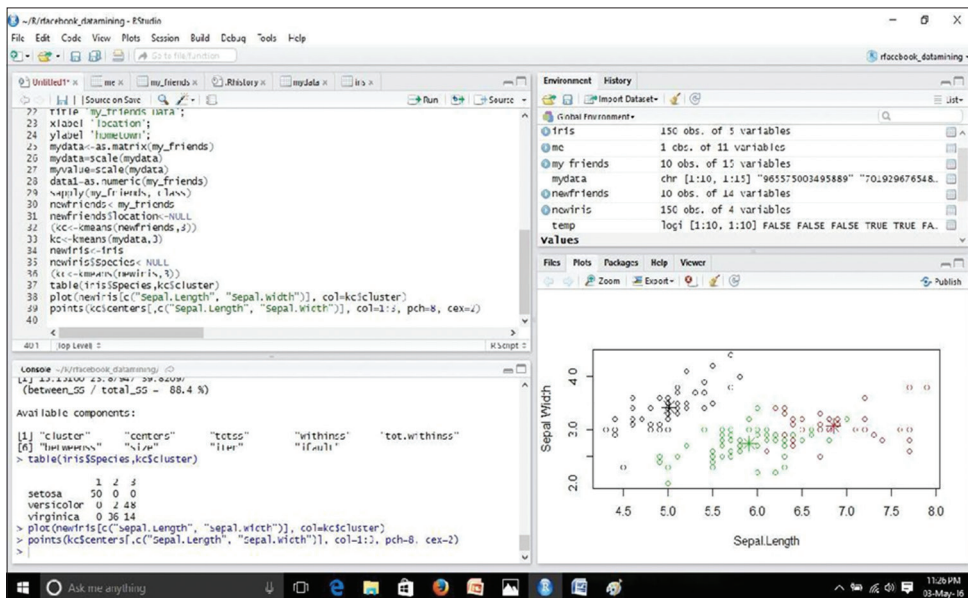**Fig. 10: OntoGraf of social network**

Fig. 11: Performance analysis

OntoGraf

This is the last and final step of analyzing that how semantic web works in social network. In Fig. 10, the OntoGraf derives the relationship between all entities. OntoGraf is the part of ontology that is implemented in Protege tool that shows semantic relationships after creating the classes and SPARQL queries. Hence, in this ontoGraf, the relationship between all entities is shown, which can be characterized after the creation of classes that is called an entity. All modules are interconnected in OntoGraf which is related to social network. The relationship tells the behavior of subclass, domain, range, and individual class. The relations about the friend list, gender, relationship status, and location are mentioned in the diagram. Generally, the ontology design is used for knowledge sharing by following certain principles [6].

PERFORMANCE ANALYSIS

By analyzing all the tools for analytics such as Tableau and Google charts, these tools produce some problems such as being time-consuming, some complexity is there, and they do not give clarifications about a particular project for analytics. To overcome all these complexities, we can use R-Studio which is user-friendly and provides some specific property for analyzing and it is fast to implement for any type of analytics project. To check the processing of R-Studio, first of all implement the clustering method that is k-means. In Fig. 11, we analyze the k-means algorithm to measure the performance of R-Studio. It can be measured by analyzing the modules that are grouped together after getting k-means output; basically, this k-means is already defined in the tool that includes the iris and species data; just we have to know about how k-means works to measure the performance of the particular software. R-Studio overcomes all complexities which are in previous tools. The Rtool provides some functionality, in which we have to install some packages for proceeding to the analytics and data retrieving part from the social network which uses semantic technology by just installing Rfacebook package inside it. The Rtool is the latest tool for analytics and visualization in a particular domain; sometimes, Rtool may be slow during implementation due to some complexities. The micro benchmarking is one type of measuring performance of a very small code, so it takes only milliseconds for this type of code. The overall performance of Rtool is availability, ease of learning, data handling capabilities, graphical capabilities, and customer service support.

CONCLUSION

In this paper, the data retrieving operation from social network is done using semantic technology and it is analyzed by using Rtool to get some insight and share some knowledge. This operation is tested for gathering the data from social network such as Facebook. From this experiment, we can create ontology from social network data that is about Facebook data in the CSV format; it can be converted as owl that is web ontology language. It is already planned to create a separate ontology for social network data; from that, we can get some information about how semantic web is working in a real time and also for social network. In future, the demand of analytics is more in the field of social network, weather forecasting, and for statistical purpose. And finally, it provides the result of different analysis after retrieving the data. The process uses various services for information sharing and knowledge sharing.

FUTURE WORK

Here, we focused only on some specific tools for data retrieval operation from social network; using specific or limited tools is not sufficient to analyze the whole social network or internal structure of social network because we cannot apply some specific algorithm for data retrieval operation. Hence, to access the whole social network data, we have to apply data mining algorithm for data retrieval operation through semantic technology. In future, we can apply some specific data mining algorithm to gather the data using semantic technology.

REFERENCES

1. Available from: http://www.facebook.com.
2. Wang RC, Su TH. Social Network Data Retrieving Using Semantic Technology, IEEE Conference Workshop Taiwan; 2013.
3. Berners T, Lassila O. The semantic web. Scientific American.com. 2001.
4. Klyne Y, Carroll J, editors. Resource Description Framework (RDF): Concept and Syntax. Available from: https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/: W3C Recommendations; 2009.
5. Brickley D, Guhal RS, editors. RDF Vocabulary Description 1.0: RDF. Available from: https://www.w3.org/TR/2004/REC-rdf-schema-20040210/: W3C Recommendation; 2010.
6. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing. Int J Hum Comput Stud 1995;43:4-5, 907-28.
7. W3C Web Ontology Language Working Group, editor. OWL 2 Web Ontology Language Document Overview. W3C Recommendation; 2009.
8. Prud'hommeaux E, Seaborne A, editors. SPARQL Query Language for RDF. W3C Recommendation; 2011.