Implementation of FOAF, AIISO and DOAP ontologies for creating an academic community network using semantic frameworks

Himanshu Ahuja, Dr. Sivakumar R.
Department of Computer Science, CHRIST (Deemed to be University), India

Article Info

ABSTRACT

Web 2.0 delivers the information which is then displayed in human readable content, omitting the crucial information which can be drawn from the data by the applications. Web 3.0 or semantic web is an extension to the current web, with an ambition to determine the drawbacks of the current web. The semantic web has already proven its influence in several communities around the globe, such as social media, music industry, healthcare domain, online blogs or articles, etc.; Among the several tools and technologies, ontologies or vocabularies are the foundation pillar for the semantic web. In this paper, the developed system aims at improving the collaboration and academic relations among staff which is directly related to our education community by providing a better networking platform which lets the agents discuss their achievements, titles, domain interests, and various other activities. Results have been analyzed to show how new facts, information can be implied from the presented knowledge of several agents and help generate a relationship graph by utilizing various semantic tools. The system discussed in this paper processes all the information in a format which can be understood by both humans and the machines, to interpret the underlying meaning about it and provide effective results.

Corresponding Author:

Himanshu Ahuja,
Department of Computer Science,
CHRIST (Deemed to be University),
Bengaluru, Karnataka -- 560 029, India.
Email: ahuja.himanshu@mca.christuniversity.in

1. INTRODUCTION

In 1989, Tim Berners-Lee invented the web. He assisted in modeling the Internet as we know it today [1]. World Wide Web and the Internet are two distinguished topics, the Web is something more to than just the HTML stack. We have that trivial ideology that Web is just the consolidation of various technologies which are adhered with fundamental philosophical principles. To become the part of the World Wide Web, a page or an application should have two necessities: it should be link-able, and it should permit any client to access it. Web 1.0 "Static Web", the web before the year 1999 is referred to as *Read-Only* web [2]. The only functionality provided to the user was reading the static content which was available on the website. The paucity of interaction between users leads to the next stage in the development of the Web 2.0. It is referred to as *Participative Web*, the transformation of static web pages displaying just information to the ability to collaborate at a global scale and share the content via various web-based forums.

The Semantic Web or Web 3.0 or Linked Data is the next evolution in Web Development. Basically, it is a bridge to fill the gap between human interaction and web applications. Web 3.0 aims at sharing and connecting services by understanding the meaning behind the underlying data.

The Semantic Web associates the context of data and facts presented, rather than just linking a web page. Now, this can have a huge impact on how the data can be processed and presented to the user. Working on the semantics of the data than on presentation of it will help to build the web of Linked Data than the web of billion documents. In other words, the web as we know today should transform into a global graph of the boundless structure. To attain this goal, proper tools and language must be provided to machines so that they can read and then process the data semantically.

To make the Semantic Web a reality, the data available on the web should be presented in the common format which can be understood and accessed by the Semantic Tools. Data Accessibility will not provide much help we have to set the relationships among data. This collection of co-related datasets on the web is referred to as Linked Data. RDF (Resource Description Framework) is a standard model for data exchange on the World Wide Web. To identify resources on the web uniquely RDF uses URIs and describes with attributes and values.

In RDF, the resource is anything which contains a URI, such as *https://example.org/cicero* and property is a name of the resource like 'given-name' or 'birth-date' and property is the value assigned such as *11-11-2011* or *Cicero*, in some cases property can be just another resource. Resource, Property, and Value form a statement which in RDF terms is Subject, Predicate, and Object of that statement. And it is written as P (S; O). Subject S has a predicate (property) P with value O. Think of this relationship is as a labeled edge between two nodes: [S] –P–>[O]. Figure 1 describes a statement; *Esbern is an Elder Nord of Tamriel* in RDF.

```
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:foaf="http://xmlns.com/foaf/0.1/">
<foaf:Person rdf:ID="me">
<foaf:name>Esbern  Nord</foaf:name>
<foaf:title>Mr.</foaf:title>
<foaf:givenname>Esbern</foaf:givenname>
<foaf:family_name> Nord</foaf:family_name>
</foaf:Person>
</rdf:RDF>
```

Figure 1. Esbern profile as displayed in RDF document

2.   RELATED WORK

A trivial term which is used by various organizations internally need not necessarily infer the exact semantics. This yields to a problem when companies must work collaboratively. In such cases, both the companies must come to an agreement of the mutual social terminologies contextually. The authors [3] proposed Upper Ontologies define conclude vocabularies which are discovered common across all the domains.

FOAF Ontology describes people, social relations, various activities, and their acquaintances but it is not close to the academic communities where professional achievements matter in the formation of relationships. Development of new ontology FOAF – Academic is described in this paper [4] by the authors. It is built on top of FOAF vocabulary by restricting the target audience as only academic communities.

In [5] Ontology UniGrad which extends FOAF has been presented. Academicians and researchers find it ponderous to query and get relative results. This issue can be resolved by applying Semantic Web technology and Ontologies such as FOAF. A new model is proposed to develop an end-user centric social network which will represent relationships and entities to improve retrieval performance of the semantic network.

In [6] the authors discuss the outcome of the study of describing mutual social communications with aspect to their sentimental depth based on their social relations and individual relationships from an online form of communication. The issue with acquiring a distinct depiction of relationships must be processed in a stepwise design approach by constructing an ontology or a vocabulary consisting of various social relations.

Social networking using FOAF vocabulary makes a compelling part of all the data available on the Semantic Web. The authors of [7] performed an experimental survey of several FOAF profiles which are found on various social networking websites is analyzed and studied to understand the intersection of FOAF data and how profiles can be merged from multiple sources and create a larger unified social network.

Relationships involving social and structural changes are not part of the primary focus for semantic web analysts. In [8] many FOAF documents were collected which were in RDF format to study the usage and scalability of these documents. A heuristic model is proposed which will help in the discovery and identification of FOAF documents available on the Web and deriving information about a person from their FOAF document.

A large portion of data available is in the form of unstructured data and social networking websites contribute the most in the generation of unstructured and unmanaged user-related data every day. Semantic technology is developed by the authors of [9] which can be used to access the world's largest social network, Facebook will be able to generate interesting facts and relationships of targeted subjects, like friends with mutual connections or similar interests of a group.

In this paper [10] the authors have presented how ontology coordinating plays a vital rule in all the domains of Semantic Web. Such as merging of different ontologies, alignment, mapping, and transformation. Ultimately how this will help in solving the major drawbacks of several database application domains. The authors [11] have proposed a solution to answer the classic why-question which is useful in keyword-based approaches. An experimental study was conducted to match against various ontology methods with the proposed system, sentence extraction from the Monge-Elkan similarity function.

A large portion of data available is in the form of unstructured data and social networking websites contribute the most in the generation of unstructured and unmanaged user-related data every day. In this paper [12] Semantic technology is developed which can be used to access the world's largest social network, Facebook will be able to generate interesting facts and relationships of targeted subjects.

Software Project Management [SPM] older versions were only focused on the metadata which was in the format of XML. This paper [13] focuses on addressing the several issues faced while developing Project Document Ontology (PDO) which is based on RDFS and it takes inspiration from various popular vocabularies.

The web application is developed in this paper [14] which is based on FOAF Ontology which will contribute to academic institutions. This system is developed using the PHP scripting language. When a social graph is developed and presented in FOAF format, it enables machines to provide reasoning power and mining of useful information without hitting the database and increasing the load with every request.

This paper [15] reviews how the imminent metadata has the potential to expose the information available on the Semantic Web and how it can empower the academic libraries. This paper highlights how academic libraries can adopt the Semantic Web metadata power which will enable richer bibliographic records and retrieval.

Ontology is proposed by the authors of [16] for scientific research approach to gathering the data from different websites which contain information about scientific publishers and store it in the knowledge base with certain parameters and keep updating it frequently. It only requires constructing a knowledge base by adding semantic based annotations.

3.   RESEARCH METHOD

The development of the proposed system is carried out based on several parameters which are described further in this section. The generated RDF file will be stored in TripleStore namely AllegroGraph for further analysis of the triples and ultimately building an academic community knowledge base. Since institutions do not publicly disclose information about its members and academic data particularly in RDF format is not publicly available. We have aimed to create our own dataset comprising of institution members and their details in RDF format.

3.1. Ontologies and vocabularies used in proposed system

FOAF (friend of a friend) is the most successful examples of Semantic Web. It describes people and their social network. FOAF is an open project and anyone can use it to describe their personal information such as name, gender, interests and social contacts without the need of a centralized database. AIISO (Academic Institution Internal Structure Ontology) is RDF/XML based Ontology which describes the classes and properties of an academic institution. DC (Dublin Core) ontology is used to describe digital and physical resources. It can be used to demonstrate basic resource description or a full-fledged combination of metadata vocabularies of various other metadata standards to achieve interoperability. SIOC (Semantically-Interconnected Online Communities) Ontology describes terms which can be used to provide information and describe various online communities on the Web. It is used in conjunction with FOAF Ontology for asserting profile web page or social online accounts.

Bibliographic Ontology is RDF based ontology which describes bibliographic things on the web. Citations, document classification or just describing document using RDF and using SPARQL to query the same information. DISCO (DDI-RDF Discovery) vocabulary defines RDF schema for the disclosure of research and survey data available on the web. It is extended from DDI Lifecycle and Codebook. DOAP (Description of a Project) Ontology is used to describe software projects, notably open source ongoing or completed projects by a group of people or individuals. It also aims to provide support for popular programming languages which are used in the development of the project.

4. PROPOSED APPROACH

4.1. System architecture

The architecture of the developed semantic web-based academic community network provides services to several agents of academic institutions. Figure 2 describes system architecture which consists of six primary components which are explained below:

- User/Agent input analysis: It comprises of details which will be entered by an agent who wishes to join the academic network. The entire process of registration is further divided into subsections which are explained in the later section.
- Filtration of the vocabulary classes: The developed system extensively uses a wide number of vocabularies for agents to describe their academic details and achievements.
- Identifying the related agents: Based on the agent role, the registration page is customized and required vocabularies to support the role are made available.
- Construction of RDF document: This comprises of the creation of RDF document in RDF/XML serialization format. User input is mapped to vocabularies class and then written to a file.
- Agent profiling and networking: Once the agent has registered to the network. RDF file can be downloaded in supported serialization formats and it is displayed on the homepage.
- Social relationships and graphs: The Graphical model used to represent the relationships between the agents of an institution. It is visualized using and queried using SPARQL language.
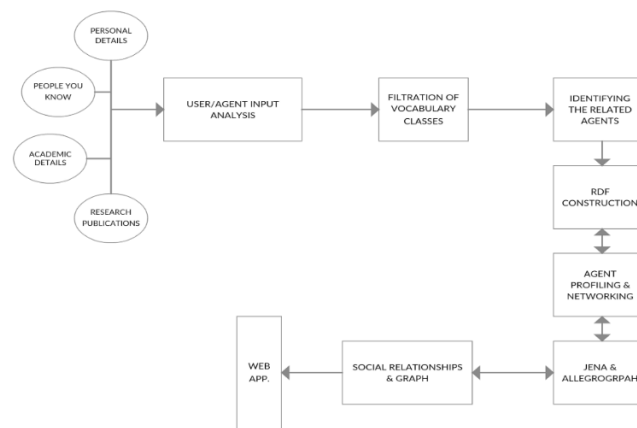

Figure 2. System architecture of the proposed system

4.2. Modules description

Proposed system comprises of 4 primary modules: Agent registration, RDF Document Construction, Agent Profiling and lastly Triples storage. Figure 3 shows the FOAF Profile Generation form.

a. Registration process: Agents who are willing to join the academic network of their institution must register themselves and provide their academic details. Register form is divided into 4 sub-modules:
  - Personal Details: In this sub-module, there are 14 parameters values which will be provided by the agent. Fields such as an 'email-address' is a unique identifier which is required for every RDF document for uniquely identifying the agent in the academic network.
  - People You May Know: This sub-module is crucial to expose the most prominent feature of FOAF vocabulary, 'people you may know'. This ultimately helps crawlers to create a network of connections which the agent may know and develop a vast social graph by linking known contacts in a group and providing suggestions whenever a new member has registered.
  - Academic Details: This sub-module comprises several vocabularies such as AIISO, AIISO-roles, AIISO-participation, Bibliographic, and Description-of-a-project, etc. The agent will be prompted to fill in required academic details which will help to create a network in the institution and link other agents based on similar attributes. This will help to create groups of agents based on their own choices selected and provide a better collaboration service between them.
  - Research Publications: Based on the agent role in an academic institution, this sub-module will be activated. This sub-module will create an agent's research publications details in the RDF model and store them in a TripleStore.

Figure 3. Agent FOAF profile registration process

b. Creation of RDF document: Once the agent registers successfully, the next step is the generation of RDF file which will contain all the agent-provided details and map it to its respective vocabularies. APIs provide functionalities to create or manipulate RDF documents. By default, it will create an RDF document in the form of RDF/XML which can be later changed to any of the following formats: JSON-LD, N-Triples or Turtle. In this module, the RDF file is created and stored in the file system as presented in Figure 4.

Figure 4. RDF file of an agent

c. Agent Profiling: In this module, an agent is presented with an RDF file along with various facts and details. Agent's related triples are queried and any new inferred facts from the TripleStore are displayed on the homepage. The query is performed with the help of SPARQL APIs. Suppose, to fetch all the triples related to one agent, the following query is used as shown in Figure 5.

```
{
    String subjectString = "<http://xmlns.com/foaf/0.1/Person/" + email + ">";
    AGQuery sparql = AGQueryFactory.create("SELECT ?p ?o  WHERE {" + subjectString + " ?p ?o.}");
    QueryExecution qE = AGQueryExecutionFactory.create(sparql, agModel);
        try {
            ResultSet results = qE.execSelect();
            if(results.next() ==null ){
                System.err.println("Account Does Not Exist. Please Register");
                return;
            }
            while (results.hasNext()) {
                QuerySolution result = results.next();
                RDFNode s = result.get("s");
                RDFNode p = result.get("p");
                RDFNode o = result.get("o");
                jsonObject.put(p.toString(), o.toString());
            }
        }
}
```

Figure 5. Java SPARQL query execution

d. Storing Triples in AllegroGraph: RDF File can be stored in the agent's local disk and can be viewed or downloaded from the website homepage. Agent homepage shows their information from the TripleStore by performing SPARQL queries and connected or mutual colleagues' information when OWL Reasoning is enabled in the TripleStore.

5. RESULTS AND DISCUSSIONS

For the development of an academic community network, several agent's RDF profiles were created which consists of nearly 1800 triples. Developed system fetches triples and performs analysis by using SPARQL queries. As a result, the social graph is established by examining the relationship between related triples and finally presented to the agent. Figure 6 shows the Triples visualized using Gruff Tool. As shown in Figure 6, all the triples which are available in AllegroGraph are being mapped to a graph view. RDFS++ reasoning is enabled which helps to link social relationships in the academic community and infer new or additional facts.
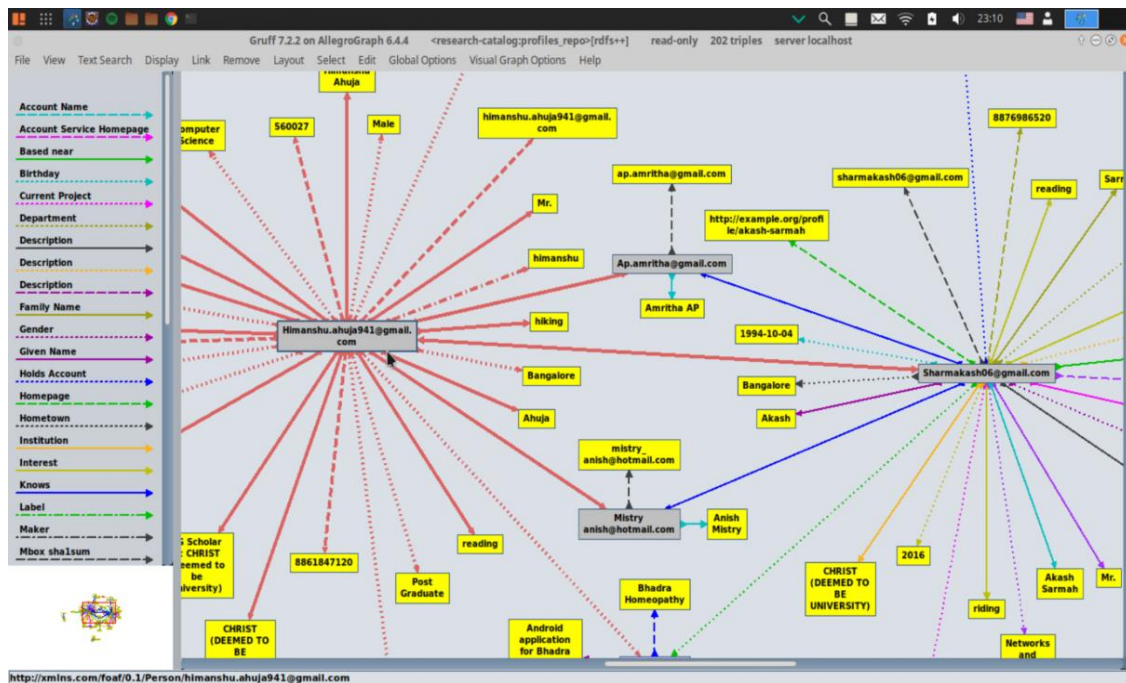

Figure 6. Triples visualized using Gruff Tool

Triples which are stored can be used to perform reasoning of any order, from a simple SPARQL query to a complex SPARQL sub-queries. For instance, suppose we want to list all the team members who have worked on a project entitled "Bhadra Homeopathy" in our academic network without writing any query. It is possible to select a node in a graph and list all the linked or connecting nodes from it. In this case, we have selected a node with label 'Bhadra Homeopathy' and made it as a primary node, and the result is visualized in Figure 7. While in Figure 8 shown Tree format representation of Triples.

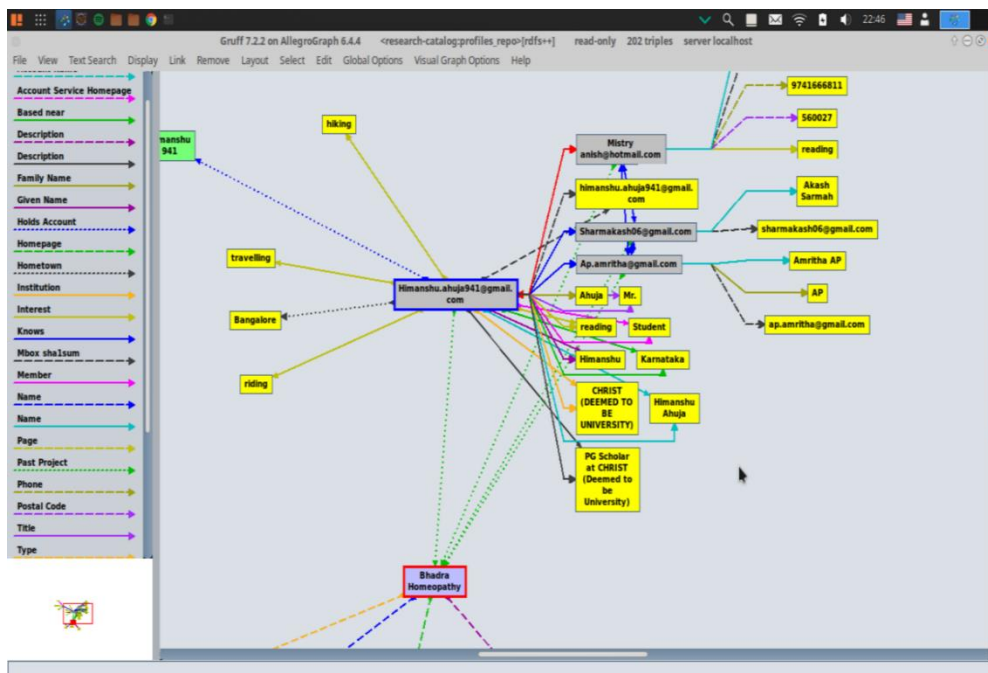

Figure 7. Model node related details exposed

Figure 8. Tree format representation of triples

We can create SPARQL endpoints to query a set of triples faster for future references as visualized in Figure 8. A SPARQL endpoint allows us to perform a query on the available knowledge base and results are displayed in machine-readable formats. As a part of an academic network, it is possible for an agent to have an RDF document stored in a cloud or a remote machine. Rather than providing the same information and generating another RDF document, the agent can upload the RDF document and use it normally like it is part of the current system.

Development of the proposed system with the help of available Semantic technologies will result in expanding the academic network and benefit the academic community. This approach will help us to build an institutional based network where researchers, scholars, students, faculties or any related agents based on their mutual interests, domains, subjects, and projects can work collaboratively, yield positive results, gain mutually and result in a better academic connection.

6. CONCLUSION

This paper presents a semantic web-based application for the academic community to express academic achievements, projects, publications, and personal information in RDF format. Presenting data in W3C recommended format provides several benefits. RDF document is machine readable and can be processed and interpreted by both humans and machines. An extensible number of ontologies and vocabularies were used in the development of this system which helped to present the academic details in the form of triples. Later, they were stored in a TripleStore named AllegroGraph and were further analyzed and processed to infer new facts, information and generate a relationship graph among several agents. Ontologies and vocabularies enable us to reuse and share available knowledge among several agents. This research targeted specifically to the academic community to showcase how Semantic Technologies can be used to improve the academic experience among several institutions. The proposed system will help to build a network between several agents of the institution based on their research domains, interests and projects they have worked on, etc.; which will result in better collaboration between agents and gain mutual benefits.

REFERENCES

[1] "Tim Berners-Lee," W3.org, 2018. [Online]. Available: https://www.w3.org/People/Berners-Lee/. [Accessed: 23-Dec- 2018].
[2] "Web 1.0 vs Web 2.0 vs Web 3.0 vs Web 4.0 vs Web 5.0 – A bird's eye on the evolution and definition," Flatworldbusiness.wordpress.com, 2018. [Online]. Available: https://flatworldbusiness.wordpress.com/flat-education/previously/web-1-0-vs-web-2-0-vs-web-3-0-a-bird-eye-on-the-definition/. [Accessed: 23- Dec- 2018].
[3] J. Grabarske and D. Heutelbeck, "An Upper Ontology for the Social Web," in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2012.
[4] E. Kalemi and E. Martiri, "FOAF-Academic Ontology: A Vocabulary for the Academic Community," in *2011 Third International Conference on Intelligent Networking and Collaborative Systems*, 2011.
[5] M. Challenger, "The Ontology and Architecture for an Academic Social Network," *IJCSI International Journal of Computer Science Issues*, vol. 9, no. 2, 2012.
[6] J. Hauffa, *et al.,* "Beyond FOAF: Challenges in Characterizing Social Relations," in *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, 2011.
[7] J. Golbeck and M. Rothstein, "Linking Social Networks on the Web with FOAF: A Semantic Web Case Study," *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, 2008.
[8] Li Ding, *et al.*, "How the Semantic Web is Being Used: An Analysis of FOAF Documents," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences,* 2005.
[9] R.-C. Wang, *et al.*, "Social Network Data Retrieving Using Semantic Technology," in *2013 IEEE 37th Annual Computer Software and Applications Conference Workshops*, Japan, 2013, pp. 322–327.
[10] Fakhar F., "Semantic Constraints Satisfaction Based Improved Quality of Ontology Alignment," *Bulletin of Electrical Engineering and Informatics*, vol. 2, no. 3, pp. 182-189, 2013.
[11] A. Karyawati, *et al.*, "Ontology-Based Sentence Extraction for Answering Why-Question," *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, IEEE, 2017.
[12] R.-C. Wang, *et al.*, "Social Network Data Retrieving Using Semantic Technology," in *2013 IEEE 37th Annual Computer Software and Applications Conference Workshops*, Japan, 2013, pp. 322–327.
[13] I.-C. Hsu, *et al.*, "Using Linked Data for intelligent information retrieval," in *the 6th International Conference on Soft Comp. and Intelligent Systems, and the 13th International Symposium on Advanced Intelligence Systems*, 2012.
[14] Al-Mukhtar, *et al.*, "The Implementation of FOAF Ontology for an Academic Social Network," *International Journal of Computer Science Engineering & Technology*, vol. 4, no. 1, pp. 10-14, 5p, Jan 2014.
[15] D. G. Campbell, *et al.*, "Academic Libraries and the Semantic Web: What the Future May Hold for Research-Supporting Library Catalogues," *The Journal of Academic Librarianship*, vol. 30, no. 5, pp. 382–390, 2004.
[16] M. Ghanem, *et al.*, "Towards a Scientific Research Based on Semantic Web," *Procedia Computer Science*, vol. 73, pp. 328-335, 2015.

BIOGRAPHIES OF AUTHORS

Himanshu Ahuja is a research scholar pursuing his Master of Computer Applications (MCA) in CHRIST (Deemed to be University) Bangalore. In 2015, he received his bachelor's in computer applications (BCA) from Presidency College Bangalore. His research interests are in Social Networking, Semantic Web, and Ontologies.
ahuja.himanshu@mca.christuniversity.in

Dr. Sivakumar R. is an Associate Professor in the Department of Computer Science, CHRIST (Deemed to be University). He has a teaching and research experience of 17 years. His research interests are in Computer Networks and Security, Internet of Things (IoT) and Web Engineering.
sivakumar.r@christuniversity.in