SEMANTIC WEB
(CREATING AND QUERYING)

Prof. Vidya S. Dandagi[1] and Dr. Nandini Sidnal[2]

[1]Department of Master of Computer Applicaton
KLE DR. M.S.Sheshagiri College of Engg.& Technology,Belgaum , India
[2]Department of Computer Science and Engineering
KLE DR. M.S.Sheshagiri College of Engg.& Technology,Belgaum , India

ABSTRACT

Semantic Web is a system that allows machines to understand complex human requests. Depending on the meaning semantic web replies. Semantics is the learning of the meanings of linguistic appearance. It is the main branch of contemporary linguistics. Semantics is meaning of words, text or a phrase and relations between them. RDF provides essential support to the Semantic Web. To represent distributed information RDF is created. Applications can use RDF created and process it in an adaptive manner. Knowledge representation is done using RDF standards and it is machine understandable. This paper describes the creation of a semantic web using RDF, and retrieval of accurate results using SparQL query language.

KEYWORDS

Resource Description Framework (RDF), SparQL, Ontology, Semantic Web ,URI metadata.

1. INTRODUCTION

Humans can read, understand and analyze the web pages easily unlike the computers that can read and interpret but not analyze on their own. Representing web pages in World Wide Web is in HTML (Hyper Text Markup Language). It is a language that is useful for visual presentation. It contains a set of "markup" symbols to display on the web that are generally designed for human utilization. Current search engines are very robust but still return irrelevant results. Mining the relevant information from the search results is a major challenge as it does not happen in real time and fetching relevant information is the necessity of the day. The possible solution for the aforesaid challenge is the semantic web.

Semantics is the learning to understand the meaning of data. It emphasizes on the relationship between the words, symbols, phrases and signs. Semantic web which is a mesh of data represents meanings through connectivity, expressing multiple viewpoints and uses logic in making models to share information across applications. Thus semantic web structures the contents of unstructured data meaningfully and creates knowledge modeled management systems [1].

Semantic web indicates to be a result for presently faced web problems. Semantic web is an upgraded version of existing web. Semantic web is a network of data which represents meanings, and prepares models using business rules logic to distribute information across applications. Meaningful contents are structured by the semantic web. The main creating component is the ontology. Information retrieval is more effective as data is stored in machine processable format [2].

The final objective of the semantic web is to index the web contents semantically so that it can be understood by machines [3].

2. WHAT IS ONTOLOGY?

Direct description of concepts is Ontology. It describes different features and attributes of the concept. A domain consists of classes. Properties of a class describe the various features and attributes of the concept. An Ontology which composes of a Knowledge base consists of group of individual instances of classes. Usage of the term "ontology" is referred by different ways. Pidcock [4] writes that "People use the word to mean different things, e.g.: glossaries and data dictionaries, thesauri and taxonomies, schema and data models, and formal ontologies and inference", and according to Uschold [5] "Ontology may take a variety of forms, but necessarily it will include a *vocabulary of terms*, and some *specification of their meaning*. This includes definitions and an indication of how concepts are inter-related which collectively impose a structure on the domain and constrain the possible interpretations of terms" [6]. The main function of the ontology is to facilitate the knowledge base [7].

Languages represented semantically are Web Services Description Language (WSDL), and OWL-S which is a OWL-based web service ontology. They describe the type of Services and the Service Provider information. WSDL is used for representing the underlying functionality of a Web service. OWL-S is an upper ontology for explaining the property and capability of web services in Web Ontology Language [8].
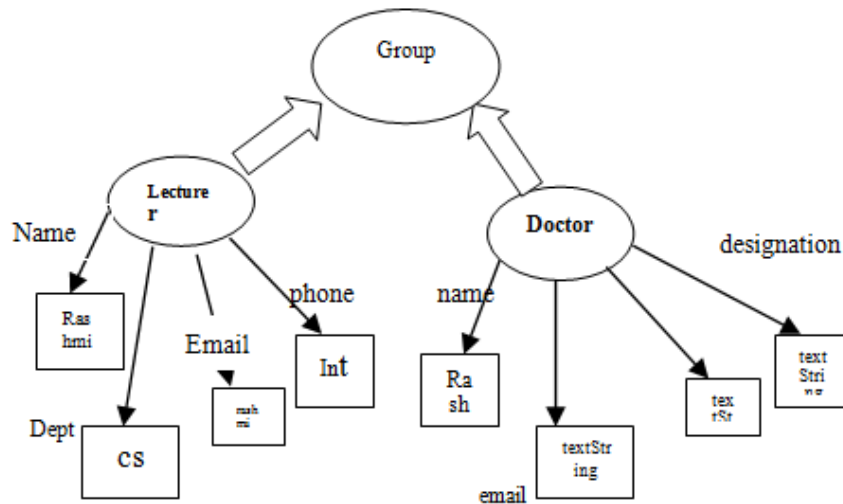

Figure 1. Example of Ontology

The above figure 1 describes the example of ontology. A class Group is created and has subclass lecturer and doctor and this subclass have properties like name, Dept, phone and email. Lecturer is a subject, has-a is a predicate, email is the object which is of type string. Doctor is a subject, is-a predicate and name is an object which is a String.

2.1. RDF

RDF is a basis for handling data about data. To exchange machine understandable information on the web, RDF provides affinity between applications. "RDF documents are made up of three entities resource, properties and statements" [9] .Resources are groups of web pages that are addressed by URIs (Uniform Resource Identifies). Properties are particular attributes and

they describe relationship between resources. Statements are object-attribute-value triples. A simple modeling language which is there on top of RDF is the RDF Schema. It includes classes and properties. The Standard syntax for RDF and RDF schema is given by XML [9].
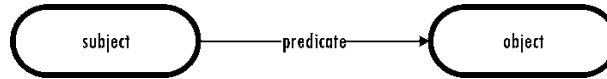

Figure 2. A RDF statement

The subject and object names are the two things in the world, with the predicate being the name of a relation that connects these two things. Semantic web uses RDF to extract the relevant information by identifying subject predicate and object.

The resource http://xmlns.com/foaf/0.1/Group has a property called mbox whose value is Rashmimahendrakar@gmail.com

Here http://xmlns.com/foaf/0.1/Group is the subject mbox is the predicate and Rashmimahendrakar@gmail.com is the object.

RDF created by Single person as shown below:

```
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" >
<rdf:Description rdf:nodeID="A0">
<rdf:predicate rdf:resource="http://xmlns.com/foaf/0.1/Group"/>
<foaf:mbox>Rashmimahendrakar@gmail.com</foaf:mbox>
<foaf:phone>9448130159</foaf:phone>
<foaf:topic>10Am-1pm</foaf:topic>
<foaf:workday>mon-sat</foaf:workday>
<foaf:member>CSE</foaf:member>
<foaf:name> Rashmi G</foaf:name>
<foaf:title>Assprof.</foaf:title>
</rdf:Description>
```

3. ARCHITECTURE OF SEMANTIC WEB

Berners-Lee's vision [10], the semantic network consists of seven levels. It is a layered architecture (Figure 3).

The first layer is the Uniform Resource Identifier (URI) and Unicode which is the starting point of the structure for the entire system. Unicode is a unique number for every character. It is independent of the underline language. This layer is accountable for encoding of the resources. URI is incharge for resource identification and allows exact retrieval of information.

The Second layer of XML + NS (Namespace) + XML Schema is used to pin point common syntax in semantic web. XML namespace identifies various markup vocabularies.

The third layer of RDF + RDF Schema imparts a semantic model which reports the information on the Web in the graph form.

The fourth layer of ontology vocabulary layer takes into account the definition of shared knowledge. It relates the semantic relationships between the various kinds of information.

The fifth layer of a logic layer is accountable for providing axioms and inference principles. It gives the origin for intelligent services.

The sixth layer of Proof is validation of statements and to infer a conclusion.

The seventh layer of trust is accountable for providing authentication and trust mechanisms. To intensify web security, digital signatures and encryption technology is used.

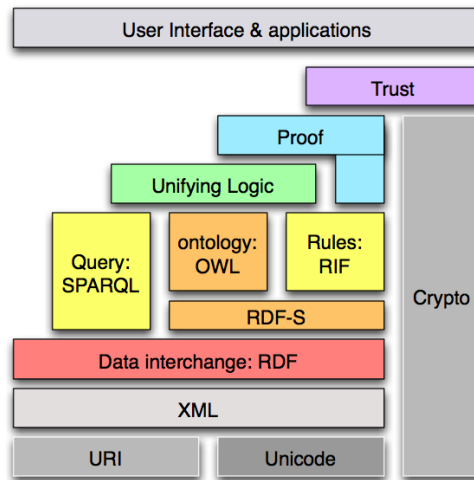XML, RDF (S) and the Ontology are the prime components in the Semantic web architecture.

Figure 3. Architecture of Semantic Web. [11]

Semantic web lets the user get answers to queries such as "Which type of music is played in UK radios stations?"[12].
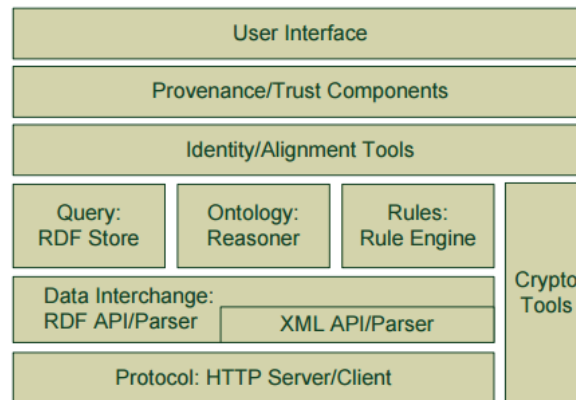

Figure 4. [12]Example of Semantic web

4. RELATED WORKS

Tarik Alafif claims that to know the information about objects, Semantic web search engine takes the help of ontologies. It uses semantic web document for retrieving search results. To manage the queries Semantic search engine takes the help of machine learning, Natural Language Processing, web mining and information retrieval techniques. To retrieve search results which are in a single range, from a domain of the user query. Tarik Alafif proposes a "Domain and Range Identifier that is incorporated into an existing Semantic web search engine". This module uses

ontologies; these ontologies include the keywords, range, link and content, which are represented as semantic web documents to validate the domain. This helps to retrieve more focused information, by identify the domain of the user query and then to identify and classify the range in that domain [13].

Duygu Tumer studied different keyword search engines (Google, yahoo, msn) and the Semantic search engine (Hakia). Results of precision and normalized recall ratio for keyword–based search engines were noted down for various cutoff points. Hakia recovered more applicable results in comparison to Google. Google recovered at least one applicable document for all queries, but Hakia failed to recover. Semantic search performance for keyword based search engine was less than the semantic search engines. So, semantic search engines help in giving more precise information with less strain [14].

Junaidah Mohamed Kassim claims that semantic search engine resolves complex queries and holds semantic information   about web resources. Semantic search combines both semantic web and search engine to recover appropriate results. The processing of semantic search engine is to recover appropriate concepts from the user sentence. Concepts are set up to create a query against the ontology. Semantic search systems can retrieve semantically related items since the information is encoded in ontology [15].

Jun Zhai states that the   semantic information retrieval is by ontology which supports logic reasoning. Semantic retrieval has better representation in recall and precision since it emphasizes matching based on knowledge and semantics. Ontology-based information retrieval is by creating domain ontology. A collection of data from the sources and annotate them to create an ontology. By ontology reasoning, the search engine completes semantic matching for the users search request, to find out the dataset [16].

Machines can understand text documents through tags which improves the semantic web. The tags represent the meaning of the text. Information retrieval is excellent when the algorithm is implemented based on TFIDF (Term Frequency –Inverse Frequency) [17].

5. CHALLENGES

Some of the possible solutions and major challenges in creating a semantic web that may reduce not only the storage space and search time but also returns focused and relevant search results [18] are:

5.1.1 Availability of content

Creating a Semantic web content for Semantic web is one of the biggest challenges. Annotations are being generated for the web content, but the tools annotate only static pages and focus only on creating new content. Dynamically generated content is not considered and existing content is being excluded from the Semantic web.

Possible Solution: Creating a set of annotation services for static and dynamic web documents which   includes multimedia and web services. Annotations of static content can be implemented by using wrapping technology. Annotations of dynamic content can be done by annotating the query which retrieves the concerned content.

Annotations of multimedia content can be implemented by   two alternatives image and video processing. Information extraction from the textual explanation of images, transcripts of

automatic speech recognition and the reports of events in video sequences   can be done by using NLP techniques.

5.1.2 Ontology Availability, Development and Evolution

Construction of the kernel ontologies can be used by all domains. Ontology development process may provide methodological and technological support which includes knowledge acquisition, conceptual modeling. Evolution of ontologies and its relationship to the annotated data can be implemented by configuration management tools.

Possible Solution: IEEE Standard Upper Ontology Group aims to create kernel ontologies. Machine Learning techniques are extremely beneficial for knowledge acquisition and conceptual modeling. Semantic web languages RDFS, OIL are used for ontology coding.

5.1.3 Organizing, Storing and mining the information

Organizing and grouping of Semantic Web content can be implemented by semantic indexes. Mining of the content can be explored by Peer to peer architecture. Semantic routing of topics can be obtained by using agents and negotiation techniques, such as 'semantic distance used in WordNet. In this way we can make the semantic content more scalable so that we can add continuous content to the Semantic web.

5.1.4 Multilinguality

Accessing information in several languages can be implemented by Semantic web. Multilinguality plays a important role at the level of ontologies and annotations.

Possible Solution: Internationalization and localization techniques can be used to personalize information access based on the native language of the user. Annotations of content can be performed in various languages which allow providers to annotate content in their native languages. To support for Multilingual existing resources such as WordNet, EuroWordNet may be explored.

5.1.5 Visualization

Users are demanding easy and relevant content, so intuitional visualization of content is gaining importance.

Intuitive visualization will help in easy reorganization of relevant content for the user. The use of semantic indexes and routers for the storage, organizing and finding information will help in visualization. Using Java3D, Shockwave3D, 3D real-time graphic representation can be explored using semantic relationship and three-dimensional interface can be automatically created. This method utilizes less space and helps the users to interact with the site realistically.

5.1.6 Semantic Web Languages Standardization

Possible Solutions: To construct the semantic web tools are essential. Semantic web language is relying on the tools. The tool to standardize the Semantic web language is the W3C Semantic Web Activity.

6. PROBLEM DEFINITION

An attempt to understand the concept and operations on semantic web has been made. A frame work is designed and developed to identify the context of search text, effectively parse and interpret it. Further an algorithm to parse the search text semantically is to be developed. Thus an automated process that interprets the search text and generates the search results efficiently and effectively is to be developed along with a semantic search algorithm.

7. IMPLEMENTATION

Data is collected from two colleges in Belgaum, retrieving of relevant data when there is a name ambiguity. The data is stored in RDF format and SPARQL is used for retrieving data. The software used for the implementation is Jena, which is the Semantic Web Framework for Java. Joseki is a Web-based SPARQL endpoint.

The gathered information is related to staff members of Department of Computer Science and Engineering, KLE Dr.M.S.Sheshgiri College of Engineering and Technology, Belgaum, and the information of the doctors from the KLE'S Dr.Prabhakar Kore Medical college . The Figure 1 shows the flowchart of generating an ontology using RDF.
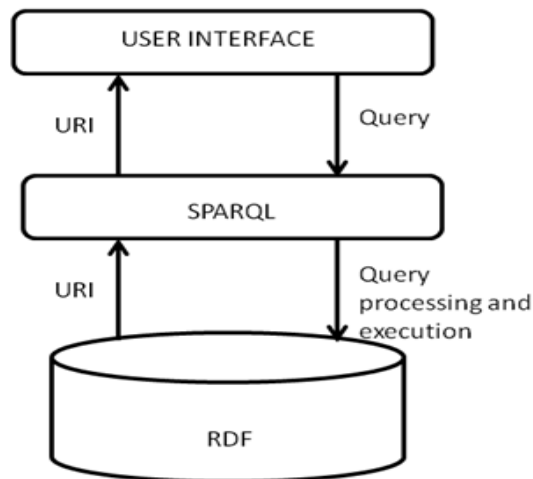

Figure 5. Generation of Ontology using RDF

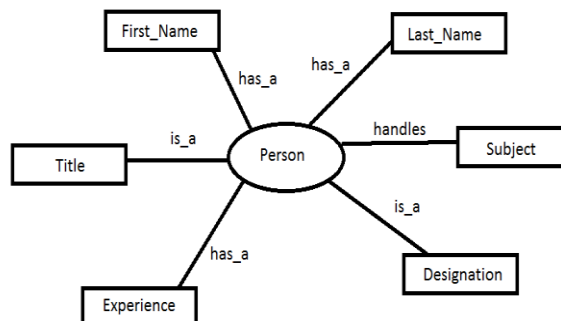The information is stored graphically as shown below in Figure 6.

Figure 6 Information Storage.

Person has a Last name. In Figure 6, Person is the subject, has-a is a predicate, Last Name is the object.

Creating the RDF for a single person:

> Step-1: Start
> Step-2: Create a model.
> Model model= ModelFactory.*createDefaultModel*();
> model.setNsPrefix("rdfs",RDFS.*getURI*());
> model.setNsPrefix("foaf",FOAF.*getURI*());
> Step-3: Create the resource for the model.
> Resource subject=model.createResource rdf:predicate
> rdf:resource="http://xmlns.com/foaf/0.1/Group");
> Step-4: Add the properties to the created resource.
> subject.addProperty(FOAF.*name*,"Rashmi G");
> subject.addProperty(FOAF.*title*,"Assprof.");
> .subject.addProperty(RDF.*type*,FOAF.*Person*);
> Step-5: Direct the output to the required file.
> For Example:d://f6.rdf
> Step-6: Stop.

7.1. SPARQL

SPARQL is a query language. RDF data is queried by SPARQL [19]. FOAF is called as Friend of a Friend. The aim of the FOAF is to create a social network using the Semantic Web technology. The usage of FOAF is to build application which cannot be implemented in traditional web [20].The main component of FOAF project is the FOAF ontology. It has a collection of terms that can be used to describe a person attributes such as name, home page, e-mail address, interest, and people oneself knows, etc.[21].

An RDF document can be created to describe oneself by using the FOAF ontology, and an individual can join the friend's network as well.

The RDF data is represented as a Triple pattern. A triple is a data representation in the form of "subject-predicate-object" [22]. Domain specific database is a triplestore which is used for the storing and retrieval of triples. RDF validator is the tool for creating the RDF data model [22]. Sample Sparql Query for Querying the RDF:

> prefix foaf: <http://xmlns.com/foaf/0.1/>
> select *
> from <c://f6.rdf>
> where
> {
> ?Person foaf:givenname ?Givenname.
> ?Person foaf:title ?Title.
>   }

8. RESULTS AND DISCUSSIONS

The paper deals with creating a semantic web using RDF to ease the process of human-computer interaction. The proposed system uses real-time data collected from two colleges in Belgaum.

Information of the staff members of KLE Dr.M.S.Sheshgiri such as name, designation, email etc, was collected and the information of the Doctors of KLEs Dr.Prabhakar Kore Medical college such as name, specialization and working hours etc, was taken and RDF is generated for all staff members. SPARQL is used to retrieve the query results. Taking into consideration name ambiguity the search is more effective. For example Dr.Rashmi Medicine from Dr.Prabhakar Kore medical college and Dr.Rashmi from Computer Science Engineering department of KLE Dr.M.S.Sheshgiri.

Figure 7 illustrates the main menu.

Figure 7. The main Menu

Query if given in SPARQL as Rashmi is a Medicine, the details of Dr.Rashmi will be retrieved.The below figure 8 shows the query entered.
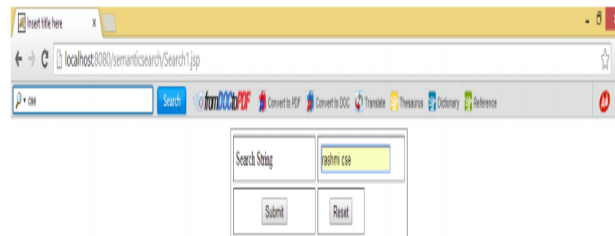

Figure 8. The Query Entered

The details of the employee Rashmi is a Medicine are displayed..The below figure 9 illustrates it.
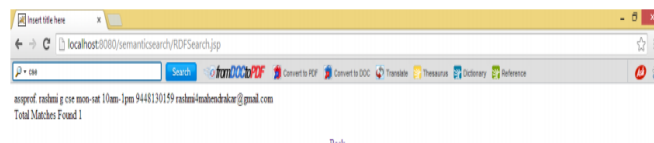

Figure 9. The query results

Semantic web is the future web where retrieval is faster and accurate. The study can be further extended for efficient building of semantic web for large organization so that retrieval of information is faster and efficient. For example, If the Query is to retrieve the details of "Dr. Rashmi" such as email, contact number, website etc." In a conventional search engine it returns multiple results of all Rashmi's details which are available, but with the proposed semantic representation the search engine will return only the concerned Dr Rashmi's details.

9. CONCLUSION

Semantic Search Engine is, no doubt, the future of World Wide Web, but the challenges such as availability of content, Multilinguality has to be addressed.
This work gives an insight of the creation of a semantic web. This web is created by RDF which expresses the semantics of concepts. Querying, using SPARQL to retrieve the relevant data. Searching the web is improved by ontology.

10. FUTURE SCOPE

In future we will be representing the information of all the institutions and create a semantic search algorithm which will give accurate results.

11. ACKNOWLEDGEMENTS

The authors are immensely thankful for the valuable guidance provided by the department of Computer Science KLE DR.M.S. Sheshgiri College of Engineering and Technology, Belgaum, Karnataka, India and Dr. KLES Prabhakar Kore Medical College, Belgaum, Karnataka, India.

REFERENCES

[1]. Omelayenko, Borys. "Learning of ontologies for the web: the analysis of existent approaches." *First International Workshop on Web Dynamics in Conjunction with the Eighth International Conference on Database Theory London, UK*. 2001.

[2]. Ahmed, Zeeshan, and Detlef Gerhard. "Web to Semantic Web & Role of Ontology." *arXiv preprint arXiv:1008.1331* (2010).

[3]. Ram Kumar Rana , Nidhi Tyagi A Novel Architecture of Ontology-based Semantic Web Crawler *International Journal of Computer Applications (0975 – 8887) Volume 44– No18, April 2012*

[4]. Pidcock, Woody, "What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?", 2003. http://www.metamodel.com/article.php?story=20030115211 223271

[5]. Uschold, Mike, "Building Ontologies: Towards a Unified Methodology", AIAI-TR-197, 1999. http://www.aiai.ed.ac.uk/project/pub/documents/1996/96- es96-unified-method.ps

[6]. Grobe, Michael. "Rdf, jena, sparql and the'semantic web'." *Proceedings of the 37th annual ACM SIGUCCS fall conference*. ACM, 2009.

[7]. Sharma, A. K. "Accessing the Deep Web Using Ontology." *Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference on*. IEEE, 2010.

[8]. Jenice Aroma R, MathewKurian *A Semantic web: Intelligence in InformationRetrieval*2013 IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN 2013)

[9]. Berendt, Bettina, et al. *A roadmap for web mining: From web to semantic web*. Springer Berlin Heidelberg, 2004.

[10]. Research on Semantic Web Mining. WANG Yong-gui1. Dept of Software.Liaoning Technical University. Huludao, Liaoning, China. yghI2000@163.net.

[11]. http://www.obitko.com/tutorials/ontologies-semantic-web/semantic-web-architecture.html

[12]. Semantic web architectureA Harth, M Janik, S Staab - Handbook of Semantic Web Technologies, 2011 - Springer

[13]. Alafif, Tarik, and Sreela Sasi. "Domain and range identifier module for semantic web search engines." *Data Science & Engineering (ICDSE), 2012 International Conference on*. IEEE, 2012.

[14]. Tumer, Duygu, Mohammad Ahmed Shah, and Yiltan Bitirim. "An empirical evaluation on semantic search performance of keyword-based and semantic search engines: Google, Yahoo, MSN and Hakia." *Internet Monitoring and Protection, 2009. ICIMP'09. Fourth International Conference on*. IEEE, 2009.

[15]. Kassim, Junaidah Mohamed, and Mahathir Rahmany. "Introduction to semantic search engine." *Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on*. Vol. 2. IEEE, 2009.

[16]. Zhai, Jun, and Kaitao Zhou. "Semantic retrieval for sports information based on ontology and SPARQL." *Information Science and Management Engineering (ISME), 2010 International Conference of*. Vol. 1. IEEE, 2010.

[17]. Jin, Yi, Zhuying Lin, and Hongwei Lin. "The research of search engine based on semantic web." *Intelligent Information Technology Application Workshops, 2008. IITAW'08. International Symposium on*. IEEE, 2008.

[18]. Benjamins, Richard, et al. "The six challenges of the Semantic Web." (2002).

[19]. Zhai, Jun, and Kaitao Zhou. "Semantic Retrieval for Sports Information Based on Ontology and SPARQL." *Information Science and Management Engineering (ISME), 2010 International Conference of*. Vol. 1. IEEE, 2010.

[20]. Paolillo, John C., and Elijah Wright. "The challenges of FOAF characterization."*Proceedings of the 1st Workshop on Friend of a Friend, Social Networking and the (Semantic) Web*. 2004.

[21]. Ding, Li, et al. "How the semantic web is being used: An analysis of foaf documents." *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*. IEEE, 2005.

[22]. Kumari, Sneha, et al. "SPARQL: SEMANTIC INFORMATION RETRIEVAL BY EMBEDDING PREPOSITIONS." *International Journal of Network Security & Its Applications* 6.1 (2014).

Authors

Vidya S Dandagi
Asst . Prof .
Department of Master Of Computer Applications
K.L. E M.S.Sheshgiri College of Engg. & Tech.
Udyamhag , Belgaum
Karnataka

Dr. Nandini Sidnal
Professor
Department of Computer Science
K.L. E M.S.Sheshgiri College of Engg. & Tech.
Udyamhag , Belgaum

KarnatakaRecognitions & Awards:

- Best paper award at International conference on soft Computing System, Tamilnadu, India, April 2015.

- Best Paper award at Asia Pacific conference on Parallel and Distributed Computing Technologies VIT,

  Vellore, India, 13-15 Dec 2004 .

- Best Project award at REC Coimbatore National Conference on software agents and - Embedded systems,

  NCSAES, Sri Ramakrishna Engineering College , Coimbatore, Tamilnadu, India, 11-12 July 2003.

- First prize in "Prathiba" National level paper presentation at Goa College of Engg, Faramagudi, Goa, India,

  2003.

- Book Publications: Cognitive Computing for Bidding in E-Auctions, Scholar's Press. C Programming

  Laboratory Hand Book for Beginners by Wiley publications