Semantic Annotation Framework For Intelligent Information Retrieval Using KIM Architecture

Sanjay Kumar Malik¹ , Nupur Prakash² and SAM Rizvi³

¹ University School of Information Technology, GGS Indraprastha University, New Delhi ² University School of Information Technology, GGS Indraprastha University, New Delhi ³ Deptt. of Computer Science, Jamia Millia Islamia, New Delhi <u>samsam_rizvi@yahoo.com</u>

ABSTRACT

Due to the explosion of information/knowledge on the web and wide use of search engines for desired information, the role of knowledge management(KM) is becoming more significant in an organization. Knowledge Management in an Organization is used to create , capture, store, share, retrieve and manage information efficiently. The semantic web, an intelligent and meaningful web, tend to provide a promising platform for knowledge management systems and vice versa, since they have the potential to give each other the real substance for machine-understandable web resources which in turn will lead to an intelligent, meaningful and efficient information retrieval on web. Today, the challenge for web community is to integrate the distributed heterogeneous resources on web with an objective of an intelligent web environment focusing on data semantics and user requirements. Semantic Annotation(SA) is being widely used which is about assigning to the entities in the text and links to their semantic descriptions. Various tools like KIM, Amaya etc may be used for semantic Annotation.

In this paper, we introduce semantic annotation as one of the key technology in an intelligent web environment, then revisit and review, discuss and explore about Knowledge Management and Semantic Annotation. A Knowledge Management Framework and a Framework for Semantic Annotation and Semantic Search with Knowledge Base(GATE) and Ontology have been presented. Then KIM Annotation platform architecture including KIM Ontology(KIMO), KIM Knowledge Base and KIM front ends have been highlighted. Finally, intelligent pattern search and concerned GATE framework with a KIM Annotation Example have been illiustrated towards an intelligent information retrieval

KEYWORDS

Semantic Annotation, Knowledge Management, Semantic Search, KIM, KIMO, Knowledge Base, GATE, NLP, Intelligent pattern search.

1. Introduction

Semantic web is an extension of the current web to a more intelligent, meaningful and efficient web where information is given formal semantics. It will bring meaning to the content of web pages, where software agents roaming from page to page can carry out sophisticated tasks. Various semantic web technologies are being used to realise the vision of semantic web as shown

DOI: 10.5121/ijwest.2010.1402

in figure 1.1 like XML(Extensible Markup Language), RDF(Resource Description Framework),URI(Uniform Resource Identifier),OWL(Web ontology Language), Semantic Annotation etc. Semantic Annotation is one of the key technology and has various aspects like KIM Annotation platform, KIM Ontology, KIM Knowledge Base and front ends, GATE, Intelligent and entity pattern search etc which have been illustrated in the figure 1.1 below.



Inference Engines

Fig 1.1 : Semantic Annotation as a key technology in an Intelligent Web Environment.

A body of formal knowledge about entities is called Knowledge base where it should be possible to identify, describe and interconnect the entities in a general, flexible and standard fashion. The goal of knowledge discovery is to find useful piece of knowledge within the data with little or no human involvement. It is a process which aims at the extraction of intersting (non trivial, implicit, previously unknown and potentially useful) information from data in large databases(Fayad et al,1996)[8] ie,to discover novel and interesting from raw data. Knowledge Management is responsible for discovery or manipulation of knowledge contained in heterogenous information resources.Semantic Search refers to meaningful search as the search engines knows the meaning of several words. After populating the knowledge base, the IE(Information Extraction) fuctionality may be used for paern-matching grammars which is based on GATE platform. Most important for Ontology are XML, XML Schema, RDF, RDF Schema, and OWL where Data can be defined and linked using RDF and OWL so that there is more effective discovery, automation, integration and reuse across different applications. and XML Schema is used for data exchange. Inference Engines process the knowledge available in the semantic web by deducing new knowledge from already specified knowledge[23]. Semantic Annotation for knowledge management is one of such significant concern and semantic web may be a promising platform for developing knowledge management systems[1]. Managing heterogenous and distributed knowledge sources available in different forms has become a challenge today. Inorder to facilitate access and reuse of knowledge, knowledge management is used for managing knowledge

resources[6]. The objective of knowledge management is to locate the organization knowledge, access and use it and better manage it. In knowledge management, semantic annotation provides a more precise description of the knowledge contained in a document and it's semantics in the domain. Semantic annotation is about assigning entities in the tesxt links to their semantic descriptions. Automatic semantic annotation enables : highlighting, indexing, information extraction and retrieval, entity pattern and intelligent pattern search for an intelligent web platform.

In this paper, first section introduces semantic annotation as one of the key technology in an intelligent web environment and in the second section, we revisit and review, analyse and discuss about knowledge management, semantic search, semantic annotation and its key aspects and a framework for Semantic Annotation and Semantic Search with Knowledge Base(GATE) and Ontology is presented. Third section, discusses about KIM platform, it's architecture, KIM Ontology(KIMO), KIM Knowledge Base, KIM Front Ends, and GATE framework. Fourth section illustrates intelligent pattern search and fifth section illustrates an example of KIM Annotation and it's entity extraction and finally conclusions and future scope.

2. Knowledge Management and Semantic Annotation

2.1 Knowledge Management(KM) Framework

It is responsible for the discovery or manipulation or management of knowledge contained in heterogenous information documents. For this, knowledge has to be structured, modeled and interlinked for the user. The challenge is the metadata so that machines can correctly interpret and quickly process the data as managed knowledge. The main process in a knowledge management system is the possibility to find knowledge sources, which are relevant for the problem at hand, as well as the process of providing knowledge sources which can be used in resolving some problems[1]. Knowledge Management introduces the challenge of making semantics of knowledge as explicit as possible[3] which may be done by encoding knowledge in XML based dialects[4][5]. Knowledge Management approach encompasses the following processes[2][1] as shown in figure 2.1.



Figure 2.1: Knowledge Management Framework

The Knowledge Management framework consists of the following key steps:[2][1]

- Knowledge Capturing: from various sources in different formats.
- Knowledge Representation: in some common representation formalism.
- Knowledge Processing: to compute interdependencies between knowledge items or to resolve conflicts
- Knowledge Sharing: between knowledge items for searching
- Knowledge Usage: for problem solving.

All the above are related with Domain Ontology which contains domain axioms to derive new information using inference engine.

2.2 Framework for Semantic Annotation and Semantic Search with Knowledge Base(GATE) and Ontology

In semantic search, a word is no longer a non informational string, but rather an informational concept and thus, the search engine knows the meaning of many words and 'knows' what a person is looking for. The meaning of the words may be stored in the brain of the search engine which is ontology. Ontology consists of many concepts which are interlinked with each other by relations such as 'subclass of', 'instance of' and many more. Semantic interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged[7]. The main obstacle of semantic interoperability is semantic heterogeneity of the information to be exchanged.



Fig 2.2: A Framework for Semantic Annotation and Semantic Search with Knowledge Base(GATE)

The above figure 2.2 illustrates a Framework for Semantic Annotation and Semantic Search with Knowledge Base(GATE) and Ontology for intelligent information retrieval. Knowledge Base consists of GATE:NLP and IE functionalities of GATE can be reused and adapted to fulfill the requirements of intelligent reasoning. Sesame[21] repository is another crucial architecture of knowledge base which provides functionalities for storage and querying of ontology. Semantic Annotation process is associated with Ontology which is the brain of the search engine and stores words meanings. Semantic Annotation provides the functionality to perform indexing and

retrieval. In semantic search, search engine knows what a person is looking for because it knows the meaning of many words and thus associated with Ontology reasoning.

2.3 Semantic Annotation

It is a technique which makes it possible to add semantics to unstructured and semi-structured documents on the web. The process of tying semantic models and natural language together is referred to as semantic annotation which may be characterised as the dynamic creation of inter relationships between ontologies ontologies(shared conceptualisation of domains) and documents[8]. Present search engines acquire metadata by extracting keywords from the web without any clue about the actual meaning of the keywords. Semantic annotation is a type of information extraction(IE) which may be achieved by different ways like information extraction using grammar rules or by recognizing concepts and instances from the ontology in unstructured texts. Semantic Annotation is about assigning to entities and relations in the text links to their semantic descriptions in an ontology like KIMO(KIM Ontology). This sort of metadata provides both class and instance information about the entities/relations. Automatic semantic annotation enables many applications like : highlighting, categorisation, semantic search, generation of more advanced metadata, smooth traversal between unstructured text and formal knowledge[8]. The KIM platform provides infrastructure and services for automatic semantic annotation, indexing, and retrieval of documents and allows scaleable and customizable ontology-based information extraction(IE) as well as annotation and document management[9]. Benefits of semantic annotation are searching and extracting desired information when the search engines lack the commonsense to extract desired information from textual representation spread over different sources. There are possibilities of misleading the search engines inorder to get a higher ranking. Semantic Annotation may be viewed as enriching a document by creating a connection between passages in a text and their semantic descriptions. Presently, a computer machine has no idea about the meaning of a set of characters. When this set of characters refers to a concept in an ontology, it becomes possible for the computer machine to understand the meaning of the text without any human intervention. Furthermore, various software systems like softbots etc are able to apply intelligent reasoning techniques on the information on the web. The figure 2.3 below illustrates the semantic annotation of text[20].



Fig.2.3 :Semantic annotation of text [20]

2.4 Annotation tools

For semantic annotation, there are various manual, semiautomatic and fully automatic tools developed by various organizations like mindswap.org and ontotect.org etc. These tools are applicable for an application domain in knowledge management for extracting the relevant and desired information from huge knowledge base of an organization. Various annotation tools are being used like : Amaya[10], Mangrove, SMORE , KIM etc. Amaya is an example of manual annotation tool which allow users to manually create annotation. Lixto is an example of automatic annotation which includes automatic annotation components but require intervention by knowledge workers. It allows users to create wrappers interactively for converting unstructured resources into structured ones[11].

3. KIM(Knowledge and Information Management) for Semantic Annotation

The Kim platform (http://www.ontotext.com/KIM) provides services and infrastructure for semantic annotation, indexing and retrieval[12]. In other words, it presents a plaform for indexing documents, annotating these documents to an ontology or knowledgebase and perform semantic queries on the indexed documents. In contrast to the most of the existing IE(Information Extraction) SYSTEMS, KIM provides for each entity references in the text (i) a link(URI) to the most specific class in the ontology (ii) a link to the specific instance in the knowledge base. KIM, a product of OntoText Lab, is an extensible platform for semantics–based knowledge management which offers IE(Information Extraction) facilities for metadata creation, storage and conceptual search[8]. It has a server which stores results in a central knowledge base and performs ontology based IE and for querying and annotating documents and minimal descriptions are added to the semantic repository and finally an API for their management is provided[8].

3.1 KIM Architecture

It consists of : KIMO(KIM Ontology), Knowledge Base, Information Extraction, indexing and retrieval, front ends(KIM API) including KIM Server functionality. KIM Ontology and knowledge bases are kept in the Sesame RDF(S) repository and the ontology middleware module[13]. KIM provides a matrix infrastructure for IE(Information Extracton),annotation and document management based on GATE[14]. KIM, as a software platform, is domain and task independent as are GATE, Sesame, and Lucene. Lucene information retrieval engine has been adopted to index documents by entity types and measure relevance according to entities, along with token and stems.

3.1.1 Ontology and KIMO

The task of annotating webpages to an ontology is very complex and time consuming. Various problems are encountered like a large number of different languages, ambiguity of words and billions of documents on the web, and difficulty in performing analysis on webpages with less relevant contents. Ontology is the key concept for describing semantics of information exchange. It is defined as "specification of a shared conceptualization of a particular domain" as they provide a shared and common understanding of a domain that can be communicated across people and application systems, and thus facilitate knowledge sharing and reuse[15][16]. Ontology may be viewed in different layers viz, upper, middle, and lower ontology layer. The upper ontology is domain independent, middle ontology contains domain specific concepts and

lower ontology is the biggest part of the ontology which holds the instances[17]. Ontology based semantic annotation allows to resolve anomalies in searches. For example, if a document collection was annotated using geographical ontology, it would become easy to distinguish 'Niger', the country from 'Niger', the river in searches, because they would be annotated with references to different concepts in the ontology. Secondly, annotation using Ontology can support for the integration of information from heterogenous sources. KIMO is the foundation Ontology of KIM which can be further populated as Manual, Automatic and Semi-automatic population as follows:[20]



Figure 3.1.1 : KIMO Ontology Populated

- Manual Population: filling instances manually but time consuming
- Automatic Population: takes less time but the quality is not guaranteed ie, results may not be perfect.
- Semi-automatic population: Instances are harvested automatically but after suggestions, feedback, and manual verification.

KIMO is a light weight ontology sufficient for simple definition of concept classes, their appropriate attributes and relations and includes various real world instances like organizations, locations etc containing about thousands of explicit statements. There is a need of adding concepts and instances to make the ontology cover a broader domain and be more useful to satisfy enough requirements to serve as good foundation. KIMO is encoded in RDF(S) and consists of about 250 concepts and 40 properties and is being extended continuously. It consists of many concepts , which are connected to each other by relations such as 'subclass of', 'instance of', and many more[20].

3.1.2 KIM Knowledge Base

A Knowledge base is a body of formal knowledge about entities where it is possible to identify, describe and interconnect the entities in a general, flexible and standard fashion. An ontology is a set of schema for the knowledge base and both should be kept into a semantic store which provides storage and retrieval according to the syntax and semantics of the selected formalism. Knowledge base can have entity knowledge which may be populated (acquired from other trusted sources) or automatically extracted by semantic annotation or any knowledge discovery method such as data mining etc. KIM Ontology contains about 250 concepts and 40 properties grouped into three upper level concepts. KIM Knowledge base has entity descriptions which are being stored in RDF(S) repository like in KIMO where each entity has information about it's type, attributes and relations. It is prepopulated with various entities of general importance (80,000

entities)[18] and we can enrich it using data of various organizations. For intelligent information extraction, one needs tp populate the knowledge base with information domain data which may be done by GATE framework. To annotate a document, one needs to have a newly created one or loaded from persistence and also need a semantic Annotation API component. To annotate a text, just pass some string to the method.

3.1.3 KIM Front Ends

They provide full access to KIM functionality, including semantic indexing, semantic repositories, metadata annotation services and metadata management. For example, the KIM plug-in for internet explorer available at http://www.ontotext.com.kim, provides some light weight semantic annotations. On it's first tab, the plug-in displays the Ontology and some classes are selected by the user ehere each class has a color for metadata highlighting. User requests semantic annotation of some viewed web page by pressing annotation button. KIM server returns automatically created metadata with its class and instance identifiers. Results are highlighted in browser window and are hyperlinked[8].

3.1.4 GATE Framework

In KIM, IE is based on the GATE frameworkwhere GATE'S document management functionalityas well as the tokenizer, part-of-speech-tagger and the sentence splitter are reused in the platform along with pattern-matching grammars and named entity (NE) coreference functionality[20]. In the context GATE version, GATE grammar components have been added to handle ontology concepts and math rules where the semantic Gazetteer component is responsible for spotting ontology concepts or instances in the text[20]. Therefore, KIM is acting as a powerful resource of semantic annotation and information extraction and it provides extensive NLP(Natural Language Processing) functionality as well as described API to access the functionality.

4. Intelligent and Meaningful Search



Figure 4.1 : Intelligent/Meaningful Search

The more concepts and instances from the ontology are recognised in a text, the better is the text understandable by a computer and to achieve this, a broad and extensive ontology is the first requirement. In the above figure 4.1, for an intelligent or meaningful search, the following may be considered: [20]

- **Proper Noun Coreference**:to detect proper nouns that refer to the same semantic descriptions. The GATE framework[22] provides excellent functionality to perform this task in the orthomatcher module.
- Analysis of subjectives: A lot of ambiguous words have several different meanings/properties and the need is to use the difference in properties to analyse the actual meaning of a word. Eg, word 'plane' may refer to aeroplane or a flat level surface. The Ontology should be broad enough to contain such words/information.
- **Document Subject Analysis** : to describe the subject/domain of the document being analysed to find which semantic description, the ambiguous word is actually referring. eg, it may refer to 'sunil' who is boxer in the context and not the cricketer having the same name. Either, we create an Ontology or may use KIMO as the foundation of our domain ontology. Then, populate the instances of application into it. In database, populate the organizational information using KIM document store population tool as in fig 4.2 below:

* KIM Document Store Population Tool	
Kill Server: Exit Host: localhost Port: 1099 Help Source: Document(s): Documents and Settings/New/Desktop/igit.txt Image: Content in the set of the se	Process Status Directory files: 1 Start time: Sun May 09 11:49:48 GMT+05:30 2010 End time: Sun May 09 11:50:00 GMT+05:30 2010 Total time (ms): 11236 -=[STATISTICS]=- Document count: 1 Document size (kb): 5 Create time (ms): 771 Parse features time (ms): 40 Annotation time (ms): 8372 Store time (ms): 250 Index sync time (ms): 0 Index opt time (ms): 1422 End Time: Sun May 09 11:50:00 GMT+05:30 2010 Finished.
Optimize Index Start Pause Stop	

Fig 4.2 :KIM document store Population Tool.

After populating the database with ontology and documents using KIM document store population tool, one can search for queries. We may reuse the information extraction functionality regarding pattern-matching grammars. This system is based on GATE platform which has proved its quality and extensibility in several IE and NLP applications. The extension of the GATE components which is integrated in KIM make it possible to relate the grammar-rules

to Ontology concepts. Refer figures 4.3 and 4.4, KIM has developed a web user interface that allows traditional access methods(keyword search) and semantic one(entity search/pattern search). The latter returns either a set of entities that satisfy the query or a set of documents that refer to these entities. One could see the content with the associated metedata on the document level(title,author etc) [19]. KIM Web UI is another front end of KIM which offers a powerful search interface[8]. Following figures, fig 4.3 and fig 4.4 illustrating intelligent pattern search.

🖉 KIM WEB UI - Windows Internet Explorer	
🕞 🕤 👻 http://localhost:8080/KIM/screen/Documer	ntsResult.jsp?m=1 💽 🐓 🗙 Google
File Edit View Favorites Tools Help	
😪 🕸 🗄 🖌 🏉 Jnu.com 🔀	KIM WEB UI 🗙 🖌 🕅 👻 🖄
	ATTERII SEARCH $ angle$ A predefined patternis $ angle$ A keyword search $ angle$ A browse ontology
Document Query Result >	Title
07/05/2010 22:33	igit o held the post of Principal, Indira Gandhi Institute of Technology, for four years (2002-06)
1-1 of 1 Documents Available Query Options NEW QUERY EDIT QUERY	Change the showing documents per page: 15

Fig.4.3: Intelligent pattern search, screen 1

KIM WEB UI - Windows Internet Explorer		
- Intp://localhost:8080/KIM/screer	/DocumentDetail.jsp?UniqueID=28selectedDoc=0	▼ 49 X Google
ile Edit View Favorites Tools Help		
🕸 🔡 🕶 🏀 Jnu.com	KIM WEB UI X	🙆 • 🔊 - 🖶
9		
	TITY PATTERN SEARCH $ angle$ A predefined patterns $ angle$ A keyword seaf	CH $ angle$ A browse ontology $ angle$ A core search $ angle$ A semantic search $ angle$ A timelines
ocument Detail ›		
		back to Document Query Result Show All Docum
Feature Name 🔺	Feature Value	
Jate	07/05/2010 22:33	
KeyEntities	Prakash Designation Singh Indraprastha University Nupur Prakash Ministry of Cor Mauritus Board of Examiners Institutes of India Advisory Board Governing Counc Roorkee Board of Studies National University Engineering College Institute of Ch Electronics & Communication Engineering Computer Science & Technology Depart	m Society of India Institute of Electronics Asian Development C-DAC School of Advanced Computing University I Basic Bucation Department of Computer Science Annjab Liniversity Computer Engineering & Technology University ritered Accountants University School of Information Indira Gandhi Institute of Technology CGS Indraprasha U unet of Training and Technica Education GSSP University Delh' College of Engineering Tab'u SGS Transgarah D
CeyPhrases	principal, workshop, constituent college, seminar, student	
ATTLE	igit	
UBL	E:\Documents and Settings\New\Desktop\/ait.txt	
Document Content Index Gradel Institute of Technology was ee Brojneering education to inome of the court beginning by opening three 3. Fach Engineering 1. Compute Stocker and Engineering (60) 2. Bectonics and Communication Engineering 3. Nechanical and Automator Engineering (60) The stockers at IGIT are selected through a se also equipped with good manageal and co in follow and mitigation to begin at good and Engineering to begin at good additional and inflationation to begin at good additionation to begin at good additional and inflationation to begin at good Additionationation to begin at good additionation to begin at good additionation to begin at good additionation to begin at good additionationation to begin at good additionation to begin at good additio	tablished by Department of Training and Technical Education, Govt. of Dehi, , keeping in view the global technological developments and to meet the require the first institute became the constant college of the University. Denat technological development of the state of the state of the state (sol) 	In the year 1998 as the first woman Engineering College in India. The institution was specifically established it ments of the Indian's in I I II Century. Out of a number of institutions toolay affiliated with the prestige rig from the prestigours of clamps of Dehi College of Engineering at Kadhmere Gate, Dehi, 1GT has made a india level. These students not only go through the rigons of academics during the four years engineering progra is dynamic progress at the Good I prestide of Instituted of Instituted of Instituted activities alongwith imprive departments.
Contact Phone No. FAX Principal 011-23669497, 23900351 011-23669 Administration 011-23669857, 011-2369255 Dept. of CEC 011-23696740 Dept. of ACE 011-2369742 Dept. of APAIe Science 011-23669040 Dept. of Apale Science 011-23669040	872	

Fig.4.4:Intelligent pattern search, screen 2

5. Annotation Example using KIM

Following are shown the illustrations of KIM annotation of Prof Nupur Prakash's web page, IGIT(Indra Gandhi Institute of Technology), GGS Indraprastha University and entity extraction of the same. We attempt to annotate the information of an Organization which can be in the form of web pages, using KIM platform which semantically annotates the web pages stored in the HTML format using KIMO.

Following are the screen shots of KIM's Semantic Annotation of Prof Nupur Prakash's web page, IGIT(Indira Gandhi Institute of Information Technology, Delhi), GGSIPU (Guru Gobind Singh Indraprastha University, Delhi) and Entity Extraction of the same.



Fig 5.1 :Example of KIM Annotation of Prof Nupur Prakash's web page, IGIT, GGSIPU



Fig 5.2: Entity Extraction of annotation performed of Example of KIM Annotation of Prof Nupur Prakash's web page, IGIT, GGS IP University

6. Related Work

Some work has already been done on knowledge management, semantic web knowledge base sytems and semantic annotation methods and frameworks. Some efforts have been carried out onn semantically enhanced information extraction systems providing automatic annotation with refernce to classes in the ontology and instances in the knowledge base has been already done based on which indexing or retrieval are performed and is extended using the ontology. Tools like KIM and others have been used for the purpose.Various semantic web based approaches to knowledge management for different applications have already been provided. Our research work is inspired by the above work and the challenge of task of annotationg web pages to an ontology which is complex and time consuming task and to provide necessary framework and illustration for semantic annotation.

7. Conclusions and Future Work

In this paper, we revisit knowledge management and Semantic Annotation as a key technology in an Intelligent Web Environment and present a Knowledge Management Framework and a Framework for Semantic Annotation and Semantic Search with Knowledge Base(GATE) and Ontology.Finally, an intelligent pattern search and concerned GATE framework with a KIM Annotation Example have been illiustrated towards an intelligent information retrieval.

The above provides the researchers an idea for indepth research to realise the vision of intelligent or efficient retrieval of information on web. The researchers may use tools like KIM or it's Ontology(KIMO) or similar platforms for finding better results towards an intelligent or efficient retrieval of information on web. In future, the researchers may devise new methods of using Ontology as a semantic annotation technique and new semantic search techniques may be developed taking into consideration about factors like accuracy, speed, strengths and weaknesses.New algorithms may be devised to evaluate performance of annotation. Since Natural Language Processing(NLP) has the capacity to give better results by Linguistic Document Annotation and therefore Linguistic Annotation Languages may be developed. Ranking or disambiguity algorithms may be written to avoid disambiguity problems .

8. References

[1] Nenad Stajonovic, Siegfried Hardschuh, "A framework for knowledge management on the semantic-web", pp-1, pp-2, http://kaon.semanticweb.org, http://ontobroker.semanticweb.org

[2] S Staab, HP Schnurr, R Studer, Y Sure, "Knowledge Processes and Ontologies ", IEEE Intelligent systems

[3] T. Berner's LEE, W3C/MIT, R. Fielding Day Software and L Masinter Adobe Systems, "Uniform Resource Identifier: Generic Syntax, "Network Working Group, http://www.ietf.org/rfc/rfc3986.txt, Jan 2005

[4] Paul V Biron, Ashok Malhotra, "xml Schema part 2: Data types second edition, "W3C recommendation, http://www.w3.org/TR/XMLSchema-2/, Oct 2004

[5] Alpesh Gajbe, Rajiv Nair, "Node Oriented Knowledge Management", CDAC, NCOSS, 2009

[6]Yuabo, Guo, "A requirement driven framework for benchmarkingsemantic web knowledge base system.

[7] Brian(Bo) Newman, "A framework for characterizing Knowledge Management methods, practices, and technologies", January1999, in support of "The introduction to knowledge management", George Washington University Course EMGT 298.T1, Spring 1999

[8] John Davies, Rudi Studer, Paul Warren, "SemanticWeb technologies, trends and research in ontology based systems", Wiley(pp-29,36,42,43,44,45)

[9] Atnas Kiryakaov, Borislav Popor, Dimitar Manav, Damyan Ognyanoff, Rosen Marinov, Ivan Terziev, "Automatic Semantic Annotation with KIM"

[10] V.Quint, I.Valton, "An introduction to Amaya", W3C 1997(http://www.w3.org/TR/NOTE-amaya-970220.html)

[11] R Baumgartner, R Flesca, Gottlob G, "Visual web information extraction with Lixto", in proceedings of the international conference on very large databases, 2001.

[12] Atanas Kiryakov, Borislav Popov, Damyan Ognyanoff, Dimitar Manaov, Angel Kirilov, Miroslav Goranov, "Semantic Annotation, indexing and retrieval", to appear in Elsevier's Journal of web semantics, vol 1, ISWC2003 special issue(2), 2004, http://www.websemanticsjournal.org/

[13] Kiryakov A, Simon K, Ognyanov D, "Ontology Middleware: Analysis and Design", March 2002

[14] Cunningham H, Maynard D, Bontcheva K, , Tablan V, "GATE: a framework and graphival development environment for robust NLP tools and applications", in proceedings of 40th Anniversery meeting of the association for computational linguistics, 2002.

[15] D Fensel, "Ontologies: Silver Bullet for knowledge management and electronic commerce", springerverlag, Berlin D, to appear (2000)

[16] P Borst, JM Akkermans, JL Top, "Engineering Ontologies", International Journal of Human-Computer Studies, 46(1997) 356-406

[17] Jochem Prins, "Semanticc search@ILSE Media" towards a Dutch semantic web infrastructure

[18] Atanas Kiryakov, Borislav Popov, Damyan Ognyanoff, Dimitar Manov, Angel Kirilov, Miroslav Goranov. Semantic Annotation, Indexing, and Retrieval. To appear in Elsevier's Journal of Web Semantics, Vol. 1, ISWC2003 special issue (2), 2004. http://www.websemanticsjournal.org/

[19] Atanas Kiryakov, Borislav Popov, Dimitar Manov, Damyan, Rosen Marinov, Ivan Terziev, "Automatic semantic annotation with KIM", Ontotext Lab, Sirma AI EAD,135, TSarigradoko Shose, Sofia 1784, Bulgaria

[20] Jochem Prins Pim, Van den Broek, "Semantic Search @ Ilse Media, Towards a Dutch Semantic Web infrastructure", Master Thesis Artificial Intelligence, 2004

[21] J Broekstrn, A Kampman, F Van Harmelan, "Sesame: A generic architecture for storing and querying RDF and RDF Schema", in the proceedings of first international semantic web conference

[22] Cunningham H, Maynard D, Bontcheva K, and Tablan V(2002), "GATE:A framework and Graphical development environment for robust NLP tools and applications", in the proceedings of the 40th Anniversary meeting of association for computational linguistics(ACL02).

[23] Berners LEE, Godel, Turing, "Thinking on the Web", Wiley, pp 39

Authors

	Saniav Kumar Malik has more than 13 years of experience in academics and
	industry in India and abroad and is presently working as Asst Professor in
	University School of Information Technology of Technology GGS Indraprastha
100	University School of Michael He is MCA from BMS College of Engineering
1 22	Dengelore University and MTeab(IT) from CCS Indroprosthe University. Delhi and
	bangalore University and Mitech(11) from GGS indraprastina University, Denn and
	presently pursuing PhD from GGS IP University. He has been teaching
	B.Tech/MCA/MBA/MTech/MSc programmes. He has been heading the University
	Server Room for internet and web activities of University since several years and
	created various websites indoa and abroad. He is Lifetime Member of CSI and his
	research areas of interest include Semantic Web and Web Engineering and has
	presented/published nearly 50 research papers in national and international
	Journals/Conferences of repute and attended conferences in USA. Prior to joining
	academics, he worked as Software Engineer in Dubai and USA.
	Prof(Mrs)Nupur Prakash is working as a Principal Indira Gandhi Institute of
-	Technology a constituent college of GGS Indraprastha University. She has worked
	as Additional Director (IT) at Institute of Chartered Accountents of India Dalhi and
	as Auditional Director (11) at institute of Charlefee Accountants of India, Denni and
	was holding the charge of Dean in the University School of Information Technology
A AND	at GGS Indraprastha University, Delhi (2006-2008) prior to this assignment. She
	received her B.E. (Electronics & Communication Engineering) and M.E. (Computer
	Science & Technology) degree from University of Roorkee (now I.I.T. Roorkee)
	and completed her PhD (Computer Engineering & Technology) from Punjab
	University, Chandigarh in 1998 in the area of Natural language Processing using
	Artificial Neural Networks. She possesses a work experience of around 25 years in
	research and academics. Her major areas of interest are Mobile Communication,
	Natural Socurity Countergraphy Secure Wineless Communication and Natural
	Network Security. Cryptography, Secure wireless Communication and Natural
	language Processing. She has authored around 50 research papers and articles for
	language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences
	language Processing . She has authored around 50 research papers and articles for various national and international journals/conferences.
	 Instantiation and reaction and
	 Indexervery, Cryptography, Secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an international processing of the processin
	 Indexervery, Cryptography, Secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia
00.	 Indexerverse in the section of the sec
	 Indexerverse in the second seco
	 Indexervity, Cryptography, secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his
	 Indexervity, Cryptography, secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian
	 Indexion R Security, Cryptography, Secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has
	 Indexion R Security, Cryptography, Secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, MIS, Mathematical
	 Indexion R Security, Cryptography, Secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software
	 Indexion R Security, Cryptography, Secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software Migration. Re-Engineering, Bioinformatics and Web based applications. He has
	 Indexional Section V, Cryptography, Sectine Writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software Migration, Re-Engineering, Bioinformatics and Web based applications. He has been instrumental in designing courses as Chairman and Member of BOS and
	 Indexion R Security, Cryptography, Secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software Migration, Re-Engineering, Bioinformatics and Web based applications. He has been instrumental in designing courses as Chairman and Member of BOS and Academic Bodies of Various Universities and Engineering Colleges. He has also
	 Indexion R Security, Cryptography, Secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software Migration, Re-Engineering, Bioinformatics and Web based applications. He has been instrumental in designing courses as Chairman and Member of BOS and Academic Bodies of Various Universities and Engineering Colleges. He has also worked as Chief Manager (EDP/IT) in Goa Shipward under the Ministry of Defence
	 Interview Security, Cryptography, secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software Migration, Re-Engineering, Bioinformatics and Web based applications. He has been instrumental in designing courses as Chairman and Member of BOS and Academic Bodies of Various Universities and Engineering Colleges. He has also worked as Chief Manager (EDP/IT) in Goa Shipyard under the Ministry of Defence, Government of India.
	 Interview Security, Cryptography, secure writeless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software Migration, Re-Engineering, Bioinformatics and Web based applications. He has been instrumental in designing courses as Chairman and Member of BOS and Academic Bodies of Various Universities and Engineering Colleges. He has also worked as Chief Manager (EDP/IT) in Goa Shipyard under the Ministry of Defence, Government of India. He has effectively trained Industrial Computer Professional in the area of acfuture and hands on avariance through under the ministry.
	 Interview Security, Cryptography, Secure wireless Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software Migration, Re-Engineering, Bioinformatics and Web based applications. He has been instrumental in designing courses as Chairman and Member of BOS and Academic Bodies of Various Universities and Engineering Colleges. He has also worked as Chief Manager (EDP/IT) in Goa Shipyard under the Ministry of Defence, Government of India. He has effectively trained Industrial Computer Professional in the area of software development and hands-on experience through various Training of Evertime MDA
	 Intervioux Security, Cryptography, Secure writeress communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software Migration, Re-Engineering, Bioinformatics and Web based applications. He has been instrumental in designing courses as Chairman and Member of BOS and Academic Bodies of Various Universities and Engineering Colleges. He has also worked as Chief Manager (EDP/IT) in Goa Shipyard under the Ministry of Defence, Government of India. He has effectively trained Industrial Computer Professional in the area of software development and hands-on experience through various Training of Executive MBA programmes. He was a consultant and acted as a Programme
	 Intervolit Security, Cryptography, Secure writeless communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software Migration, Re-Engineering, Bioinformatics and Web based applications. He has been instrumental in designing courses as Chairman and Member of BOS and Academic Bodies of Various Universities and Engineering Colleges. He has also worked as Chief Manager (EDP/IT) in Goa Shipyard under the Ministry of Defence, Government of India. He has effectively trained Industrial Computer Professional in the area of software development and hands-on experience through various Training of Executive MBA programmes. He was a consultant and acted as a Programme Director (MIS) for Abu Dhabi University in UAE and for Jawaharlal Institute of
	 Intervolk Security, Cryptography, Secure whereas Communication and Natural language Processing. She has authored around 50 research papers and articles for various national and international journals/conferences. Dr. S.A.M. Rizvi is a Ph.D. in Computer Science & Engineering, and an internationally recognized Computer Scientist of repute, presently working as an Associate Professor at the Department of Computer Science, Jamia Millia Islamia (Central Government University), New Delhi, INDIA. Dr Rizvi has more than 26 years of experience out of which he has served seven years as Professor of Computer Science and Academic Administrator in India and Abroad prior to his joining Jamia Millia Islamia. He has taught in USA, Australian, UAE and Indian Educational (Credit Hour) Systems. An Expert in Software Engineering, who has published numerous papers in the field of Software Engineering, MIS, Mathematical Modeling and Bioinformatics. His current research interests include Software Migration, Re-Engineering, Bioinformatics and Web based applications. He has been instrumental in designing courses as Chairman and Member of BOS and Academic Bodies of Various Universities and Engineering Colleges. He has also worked as Chief Manager (EDP/IT) in Goa Shipyard under the Ministry of Defence, Government of India. He has effectively trained Industrial Computer Professional in the area of software development and hands-on experience through various Training of Executive MBA programmes. He was a consultant and acted as a Programme Director (MIS) for Abu Dhabi University in UAE and for Jawaharlal Institute of Technology (JIT), in India. He has authored several research papers and articles for