ISSN: 2321-7782 (Online)
Volume 2, Issue 1, January 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Paper
Available online at: www.ijarcsms.com

Overview of Xml based Knowledge Representation using Scripts

Pranita P. Deshmukh[1]
Department of Computer Science and Engineering
Prof Ram Meghe College of Engineering and Management,
Badnera, Amravati University
India

Dr. M. S. Ali[2]
Prof Ram Meghe College of Engineering and Management,
Badnera, Amravati University
India

Abstract: ***A knowledge representation (KR) is an idea to enable an individual to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than taking action in it. The knowledge acquired from experts or induced from a set of data must be represented in a format that is both understandable by humans and executable on computers. Knowledge representation research involves analysis of how to reason accurately and effectively and how best to use a set of symbols to represent a set of fact within a knowledge domain. A symbol vocabulary and a system of logic are combined to enable inferences about elements in the knowledge representation to create new knowledge representation sentences by using various techniques.***

Keywords: Knowledge Representation, Semantic net, Frames, logic, Scripts, Pattern Matching, Fuzzy Rules.

I. INTRODUCTION

Knowledge representation research involves analysis of how to reason accurately and effectively and how best to use a set of symbols to represent a set of fact within a knowledge domain. A symbol vocabulary and a system of logic are combined to enable inferences about elements in the knowledge representation to create new knowledge representation sentences.

Logic is used to supply formal semantics of how reasoning functions should be applied to the symbols in the knowledge representation system. Logic is also used to define how operators can process and reshape the knowledge. Examples of operators and operations include negation, conjunction, adverbs, adjectives, quantifiers and modal operators. The logic is interpretation theory. These elements – symbols, operators, and interpretation theory – are what give sequences of symbols meaning within a knowledge representation.

In applying knowledge representation systems to practical problems, the complexity of the problem may exceed the resource constraints or the capabilities of the knowledge representation system. Recent developments in knowledge representation include the concept of the Semantic Web, and development of XML-based knowledge representation languages and standards, including Resource Description Framework (RDF), RDF Schema, Topic Maps, DARPA Agent Markup Language (DAML), Ontology Inference Layer (OIL), and Web Ontology Language (OWL).

A script is a structured representation describing a stereotyped sequence of events in a particular context. For Example, when we go to a restaurant, we usually 'enter the restaurant', 'wait', 'sit down', 'get the menu and decide what to eat', 'order the dish', 'wait until the dish has come', and so on. This sequence can be said to be script knowledge in the situation of 'eating at a restaurant'.

Commercial applications of script-like structured objects: work on the basis that a conversation between two people on a pre-defined subject will follow a predictable course. Certain items of information need to be exchanged. Others can be left

unsaid (because both people know what the usual answer would be, or can deduce it from what's been said already), unless (on this occasion) it's an unusual answer.

A knowledge representation (KR) is an idea to enable an individual to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than taking action in it [2]. The knowledge acquired from experts or induced from a set of data must be represented in a format that is both understandable by humans and executable on computers. For Good Knowledge Representation Languages, there should some qualities:

1. Expressive

2. Concise

3. Unambiguous

4. Independent of context

a) What you say today will still be interpretable tomorrow.

5. Efficient

a) The knowledge can be represented in a format that is suitable for computers.

b) Practical inference procedures exist for the chosen format.

6. Effective

a) There is an inference procedure which can act on it to make new sentences.

II. LITERATURE REVIEW

Ref 7]:- KR is the study of how to know at the same time be represented as comprehensibly as possible and reasoned with as effectively as possibly. The simplest analysis shows difference between procedural and declarative knowledge. KR is very important for knowledge based systems. A selected KR scheme should have appropriate inference methods to allow for reasoning. Popular KR schemes are Rules, Semantic Nets, Schemata (Frames and Scripts) and Logic. Balance must be found between effective representations, efficiency and understandability for effectiveness. Effective KR should be used to represent the most important aspects of the real world, such as action, space, time, mental events.

Ref 6]:- The comparison between five representation schemes and the objective is to analyses the power and expressiveness of a system. Each knowledge representation schemes has advantages and disadvantages. Combination of two or more representation scheme may be used to for making the system more efficient and improving the knowledge representation.

Ref 21]:- RDF/XML can be (extended to be) used in various knowledge representation cases and It proposed intuitive notations (FE, FCG and FO) covering at least all the presented cases. Although these high-level notations are unlikely to be widely adopted, they show some ways to improve other notations in readability, expressivity and "knowledge normalizing effect" – for example, Notation-3, Tim Berners-Lee's "academic exercise", which does not (yet) have a special syntax for extended quantifiers, collections, functions and definitions. This article complements the lexical and ontological conventions proposed in to permit knowledge sharing. We are now working on the import and export of FE, FCG, KIF and RDF/XML in WebKB-2, along the lines presented in this article. More information can be found, and testing can be done, on the Web KB site.

Ref 3]:- To develop an approach for building knowledge-acquisition tools that provide strong support for a wide range of medications and knowledge-based system types. To achieve this goal, we equipped a knowledge-acquisition tool with a library of knowledge-acquisition scripts (or KA Scripts) which represent prototypical procedures for modifying knowledge-based systems. The advantage of

KA Scripts is that they provide a context for relating individual changes to different parts of the knowledge-based system enabling the tool to analyze each change from the perspective of the overall modification. This kind of analysis complements previous approaches for interpreting changes to a knowledge-based system and enables a knowledge- acquisition tool to provide a more precise guidance. Because KA Scripts are problem-solving method independent, they can be used to support modifications of any kind of knowledge-based system. Furthermore, because KA Scripts represent varied procedures for modifying different aspects of a knowledge-based system, they can support a wide range modification. It has implemented a script-based knowledge-acquisition tool called ETM that supports modification to knowledge-based systems developed within the EXPECT framework. In implementing ETM we addressed several research issues that concerned the development of a KA Script library, the coordination of KA Scripts, and the model of interaction with the user. To evaluate ETM, we carried out an experiment that compared the performance in modifying

Knowledge-based systems for subjects using ETM vs. subjects using EXPECT. The experiment showed that subjects using ETM outperformed the ones using EXPECT, especially in the more complex modification tasks. In this first experiment they are chose subjects that were already familiar with EXPECT but not with ETM. Author has expect that the difference in performance will be more significant in their future experiments involving subjects not familiar with EXPECT. One important extension to our approach is to give advice on how to start a modification, not just how to complete it. In fact, three of our four subjects made the comment that they would like help in where to start the modification. One way to achieve this goal is by integrating KA Scripts of different level of abstraction. The more abstract KA Scripts would plan the overall modification while the more specific ones would take care of the details.

Ref 4]:- There are various knowledge representation schemes in AI. All KR techniques have their own semantics, structure as well as different control mechanism and power. Combination of two or more representation scheme may be used for making the system more efficient and improving the knowledge representation. They are trying to build the intelligent system that can learn itself by the query and have a power full mechanism for representation and inference. The semantic net and script are very powerful techniques in some respects so the aim is to take the advantage of these techniques under one umbrella. The comparison between various hybrid KR techniques is shown in table with the proposed one.

Ref 9]:- In this paper author has presented a first approach toward the creation of intelligent brows able documents, applied to cutaway diagrams. Author identifies the tags in the drawing and correlates them with the occurrences of the same tags in the legend adjoining the graphics. Moreover, we identified the need for real component algebra for document analysis applications and research and defined the major requirements for such an environment: edibility and interchangeability, which we obtain through scripting basic image treatment components and interoperability, to obtain by defining a sound xml-based description format for image and document analysis results.

The very first implementation made of this component algebra, based on the scripting language and own graphic analysis library, developed an ad hoc that is anything but generic and extendible. It is clear that, in order to achieve a realistic framework, the final representation should encompass most graphical entities and attributes. Sags is a very important initiative in that direction, and we are currently investigating its possibilities. However, at the printing of this document, it is still unclear whether this W3C recommendation remains suited for free, unlimited use, and other challengers, like Mpeg7_Vrml also need to be considered.

Ref 11]:- Author has presented a scheme for management of multilingual OCR system using XML as the universal medium for data description and exchange. They have shown that our XML based representation can handle different requirements of multilingual document image management systems. Using XML based unified representation framework we have provided solutions for ground truth management, modular multiparty OCR development and semantics or concept based search of OCR documents. Out approach provides a generic solution which can be emulated for other multilingual multiscript OCR environments.

III. ARCHITECTURE

The KR system must be able to represent any type of knowledge, "Syntactic, Semantic, logical, Presupposition, Understanding ill formed input, Ellipsis, Case Constraints, Vagueness". In our previous paper we have proposed the model for effective knowledge representation technique that consist five different parts the K Box, Knowledge Base, Query applier, reasoning and user interface as shown in fig 8. This time the total emphasis is on knowledge representation. This section used to describe the new hybrid knowledge representation technique which is the integration of script and semantic net KR technique. The semantic net & script KR technique are explained in next subsection.

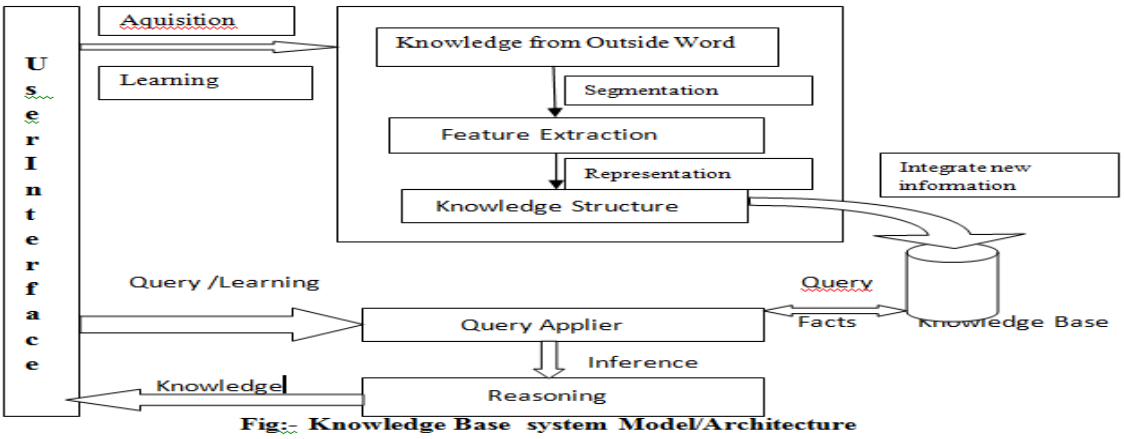


Fig:- Knowledge Base system Model/Architecture

IV. COMPARISON BETWEEN PROCEDURAL AND DECLARATIVE KNOWLEDGE

Sr. No.	Procedural Knowledge	Declarative Knowledge
1.	Hard to Debug	Easy to Validate
2.	Black Box	White Box
3.	Obscure	Explicit
4.	Process Oriented	Data Oriented
5.	Extension may affect stability	Extension is easy
6.	Fast, direct execution	Slow (requires interpretation)
7.	Simple data can be used	May require high level data type
8.	Representation in the form of sets of rules, organized into routines and subroutines.	Representation in the form of production system, the entire set of rules for executing the task.
9.	Focuses on tasks that must be performed to reach a particular.	Refers to representations of objects and events, knowledge about facts.

V. KNOWLEDGE REPRESENTATION TECHNIQUES

KNOWLEDGE REPRESENTATION USING LOGIC :

Logic is a formal language for representing facts in a precise and unambiguous way. There two basic forms of computational logic are propositional logic and predicate logic. A proposition is nothing more than a statement that is either true or false. In propositional logic, we use symbols, such as letters of the alphabet, to represent propositions, premises, or conclusions. For example, the mail carrier comes Monday through Friday.

1. A (premise) Today is Sunday.

2. B (premise) The mail carrier will not come today.

3. C (conclusion)

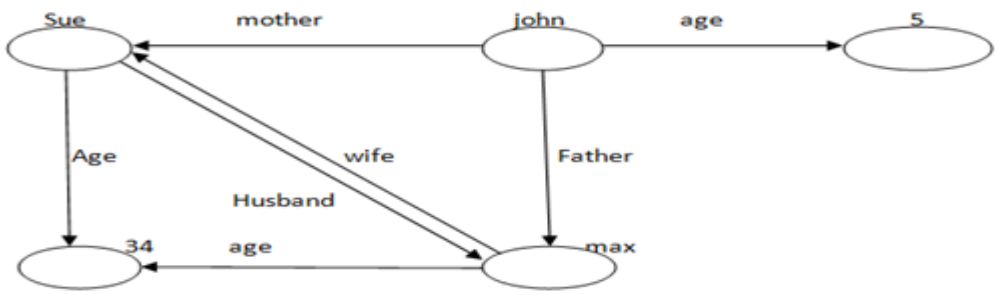
To form more complex premises, two or more propositions can be combined, using logical connectives such as AND, OR, NOT, IMPLIES, and EQUIVALENT. It can be manipulated using the rules of propositional logic to infer new conclusions.

Because propositional logic deals primarily with complete statements and whether they are true or false; its ability to represent real-world knowledge is limited. Predicate logic is useful for showing logic relationships and their reasoning. In predicate logic statements can be split into words. Predicate logic employs the notions of constant, variable, function, predicate, logical connectives and quantifiers to represent facts. A predicate with one argument expresses a property of an object for e.g. Student (smith). A predicate with two or more arguments expresses a relation between objects for e.g. likes (smith, Mary). Predicate with no arguments is just simple proposition logic.

KNOWLEDGE REPRESENTATION USING SEMANTIC NET:

Semantic networks are an alternative to predicate logic as a form of knowledge representation. Semantic networks focus on the relationships between different concepts. They are graphical depictions of knowledge composed of nodes and links (arcs). Nodes are represented by objects, and arcs representing relationships between those objects. An object can be any physical item, such as a book, a car, a desk, or even a person. Nodes can also be concepts, events, or actions.

Some of the most common arcs are of the IS-A or HAS-A type. IS-A is used to show a class relationship (i.e., that an object belongs to a larger class or category of objects). HAS-A links are used to identify characteristics or attributes of object nodes. But the representation of semantics networks in a computer is very complex. As only objects and there associations are stored in International Journal of Scientific and Research computer, therefore retrieving the actual and correct information from semantics networks is a very difficult task.



KNOWLEDGE REPRESENTATION USING FRAME:

A frame is a data structure that includes all the knowledge about a particular object. This knowledge is organized in a special hierarchical structure that permits a diagnosis of knowledge independence. Frames are basically an application of object-oriented programming for artificial intelligence and Expert System. Frames provide a concise structural representation of knowledge in a natural manner. The knowledge in a frame is partitioned into slots. A slot can describe declarative knowledge (e.g., the color of a car) or procedural knowledge.

A frame includes two basic elements: slots and facets. A slot is a set of attributes that describe the object represented by the frame. Each slot contains one or more facets. The facets (sub slots) describe some knowledge or procedural information about the attribute in the slot. Most artificial intelligence systems use a collection of frames linked together in a certain manner to show their relationship. This is called a *hierarchy of frames*. The hierarchical arrangement of frames permits inheritance frames.

KNOWLEDGE REPRESENTATION USING PRODUCTION RULES:

Production rules are the most popular form of knowledge representation for expert system and automated decision support (ADS) systems. Knowledge is represented in the form of condition/action pairs: IF this condition (premise) occurs, THEN some action (result or conclusion) will occur. Consider examples, "If the stop light is red AND you have stopped, THEN a right turn is okay." Two types of rules are common in artificial intelligence: knowledge and inference rules. Knowledge rules (declarative rules) state all the facts and relationships about a problem. Inference rules (procedural rules) advice on how to solve a problem, given that certain facts are known.

The knowledge engineer separates the two types of rules: Knowledge rules go to the knowledge base, whereas inference rules become part of the inference engine. Example of the knowledge rules:

Rule 1: IF an international conflict begins, THEN the price of gold goes up.

Rule 2: IF the inflation rate declines, THEN the price of gold goes down.

Rule 3: IF the international conflict lasts more than seven days and IF it is in the Middle East, THEN buy gold.

Inferences rules contain rules about rules and thus are also called met rules. They pertain to other rules (or even to themselves). Inference (procedural) rules may look like this: Rule 1: IF the data needed are not in the system, AND THEN request them from the user.

Rule 2: IF more than one rule applies, and THEN deactivate any rules that add no new data.

KNOWLEDGE REPRESENTATION USING SCRIPT:

A script is a term proposed by Schank, and it refers to a form of knowledge representation [8]. A script is a structured representation describing a stereotyped sequence of events in a particular context [5]. For Example, when we go to a restaurant, we usually 'enter the restaurant', 'wait', 'sit down', 'get the menu and decide what to eat', 'order the dish', 'wait until the dish has come', and so on. This sequence can be said to be script knowledge in the situation of 'eating at a restaurant'.

A script is a remembered precedent, consisting of tightly coupled, expecting-suggesting primitive action and state-change frames. A script is a structured representation describing a stereotyped sequence of events in a particular context. Scripts predict unobserved events. Scripts can form a coherent account from disjoint conversations. As compared to scripts, a frame is a relatively large chunk of knowledge about a particular object, event, location, situation or other element. The frame describes the object in great detail. Script, on the other hand, is a knowledge representation scheme that instead of describing an object, describes a sequence of events.

It is active type information which contains class of events in terms of contexts, participants and sub-events represented in the form of collection of slots or series of frames which uses inheritance and slots. Scripts predict unobserved events and can build coherent account from disjointed observations. Scripts basically describes the stereotypical knowledge i.e. if the system in not given the information dynamically then it assumes the default information to be true Scripts are beneficial because real world events do follow stereotyped patterns as human beings use previous experiences to understand verbal accounts.

A script is used for organizing the knowledge as it directs the attention and recalls the inference. They provide knowledge and expectations about specific events or experiences and can be applied to new situations. For example: "Rohan went to the restaurant and had some pastries". It was good now meaning derived from the above text one gets to know he got the pastries from the restaurant and that for eating and that was good. Script defines an episode with the known behavior and describes the sequence of events. The script consist the following.

• Current plans (Entry condition, Result)

• Social link(Track)

• Played roles,

• Scene.

• Probs.

• Anything indicating the behavior of the script in a given situation.

An example of script for class room is shown in Fig: Script Structure for class Room

Script Lecture Room	
Track : Class room Props: table, chair, chock board, chock box, lecture stand , projector Roles: T= Teacher S= Student	Entry condition : T has prepare lecture T has prepare notes The class is open T has attendance register. Result: T has imparted knowledge S: Acquired Knowledge

Script Lecturer Room Contd.	
Scene 1 Entering T: enter the classroom T: moves the lecture stand T: Switched on the projector T: Look the student	Scene 2 Lecture T: lecture notes on lecture stand T: select the lecture no. T: Explain the lecture S: listen the lecture S: ask the question T: Use the board T: go to the scene 4 at the "No student class" T: Explain T: ask the question

Script Lecturer Room Contd.	
Scene 3 question solving T: gave question S: discussion S: Solve the question T: Solve the question	Scene 4 existing T:took the attendance T: Collect the sheet T: Leave the class room

Advantages of using scripts:-

1. Details for a particular object remain open and

2. Reduces the search space.

Disadvantages:-

1. Less general than Frames

2. It may not be suitable for all kind of Knowledge

VI. COMPARISON BETWEEN KNOWLEDGE REPRESENTATION TECHNIQUES

Sr. No.	Knowledge Structure	Logic	Semantic Net	Frame	Script	Production Rule
1	Merits	1. Facts asserted independently of use. 2. Assurance that only valid consequences that are aseerted. 3. Its providing a better way to do reasoning by providing a way	1. Easy to follow hierarchy. 2. Easy to trace associations. 3. Easy to visualize and understand. 4. It is flexible	1. Expressive power 2. Easy to set up slots for new properties and relations. 3. Easy to create specialized procedures. 4. Easy to include default	1. It provide a high degree of flexibility 2. Ability to predict events 3. A single coherent interpretation may be build up from a collection of	1. Simple syntax 2. Easy to understand 3. Simple interpreter 4. Highly modular 5. It is flexible(to add or

		of deducting new form old one. 4. It can be use for providing the statements. 5. Quantified and extential statements are easily represented.		information and easy detect missing values 5. Very flexible	observations	modify) 5. Naturalness of expression .
2	Demerits	1. Not possible to store and represent infinite sentences. 2. Processing is not efficient and is time consuming. 3. Less expressive 4. Used for representing statics facts only. 5. Inefficient with large data sets. 6. Very slow with large knowledge bases	1. Meaning attached to nodes might be ambiguous. 2. Exception handling is difficult 3. Difficult to program	1. Difficult to program 2. Difficult for inference 3. Lack of inexpensive software	1. Complex concepts, influences and dependencies are difficult to represent. 2. The precise structure is hard to select to achieve an optimal performance. 3. Less general than frames. 4. May not be suitable to represent all kinds of knowledge.	1. Hard to follow hierarchies. 2. Inefficient for large systems 3. Not all knowledge can be expressed as rules 4. Poor at representing structured descriptive knowledge .
3	Nearest data structure	Rule based system	Graph	Class in object oriented programming	Class of events	1. Tree-structure 2. Divide and conquer,
4	Searching algorithm	1. Breadth First 2. Depth first 3. Forward Chaining 4. Backward Chaining	1.Intersection Search 2. Inheritance 3. Breadth First 4. Depth First 5. Heuristic Search	1.Inheritance 2. frame matching	1. Inheritance 2. Tree-Structure	1. pattern matching algorithm(Rete algorithm) 2. *forward chaining* and *backward chaining*

VII. CONCLUSION

There are various knowledge representation schemes in Artificial intelligence. All KR techniques have their own semantics, structure as well as different control mechanism and power. In this paper, trying to build the intelligent system that can learn itself by the query and have a power full mechanism for representation and inference. The semantic net and script are very powerful techniques in some respects so the aim is to take the advantage of these techniques under one umbrella.

KR is the study of how what we know can at the same time be represented as comprehensibly as possible and reasoned with as effectively as possibly. The simplest analysis shows difference between procedural and declarative knowledge. KR is very important for knowledge based systems. A selected KR scheme should have appropriate inference methods to allow for reasoning. Popular KR schemes are Rules, Semantic Nets, Schemata (Frames and Scripts) and Logic. Balance must be found between effective representations, efficiency and understandability for effectiveness. Effective KR should be used to represent the most important aspects of the real world, such as action, space, time, mental events. Future work is to improve the parameters considered.

References

1. Arnheim, R. 1969 Visual Thinking. University of California Press, Berkeley, California.
2. Berners-Lee, T. 1998. What the Semantic Web can represent, http://www.w3.org/DesignIssues/RDFnot.html
3. Berners-Lee, T. and Hendler, 2001. J. Scientific publishing on the semantic web. http://www.nature.com/nature/debates/eaccess/Articles/bernerslee.htm
4. Berners-Lee, T. Hendler J. and Lassila, O. The Semantic Web. Scientific American, May 2001.
5. Marcelo tallis, A script based approch to modifying knowledge based system.
6. Pushpa Kesarwani*, Prof. A.K. Misra, February 2013, Selected integrated approch for knowledge representation by comparitive study of knowledge scheme, International Journal of Scientific and Research Publications, Volume 3, Issue 2.
7. Tilotma Sharma1, Deepali Kelkar2, A Tour Towards Knowledge Representation Techniques, International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 2, Issue 2
8. M. Kerber, 2004/05 "Knowledge Representation I", Introduction to AI 06-08775.
9. B. Lamiroy1 and L. Najman2,3, "Scan-to-xml: Using Software Component Algebra for Intelligent Document Generation",
10. Ying Zhang Alex Golovinsky Mark Yim Craig Eldershaw, "An XML-based Scripting Language for Chain-type Modular Robotic Systems" Palo Alto Research Center, 3333 Coyote Hill, Palo Alto, CA, USA.
11. Gaurav Harit1, K. J. Jinesh2, Ritu Garg3 C.V Jawahar2, and Santanu Chaudhury3," Managing Multiligual OCR Project using XML".
12. Alexandre Delteil1, Catherine Faron2, and Rose Dieng1, "Building Concept Lattices by Learning Concepts from RDF Graphs Annotating Web Documents".
13. Jean-François Baget 1,2, Olivier Corby1, Rose Dieng-Kuntz1, " Griwes: Generic Model and Preliminary Specifications for a Graph-Based Knowledge Representation Toolkit"
14. Victoria Uren, Bertrand Sereno, Simon Buckingham Shum & Gangmin Li, September, 2003 "Interfaces for Capturing Interpretations of Research Literature".
15. H. Cherfi1, O. Corby1, C. Faron Zucker1,2, K. Khelif1 and M.T. Nguyen1," Semantic Annotation of Texts with RDF Graph Contexts".
16. Philippe Fournier-Viger, Evaluating Spatial Representations and Skills in a Simulator-Based Tutoring System, IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 1, NO. 1, JANUARY-MARCH 2008.
17. LI Guoliang, FENG Jianhua, ZHOU Lizhu, Keyword Searches in Data-Centric XML Documents Using Tree Partitioning*, TSINGHUA SCIENCE AND TECHNOLOGY, ISSNl l1007-0214l l02/ 21l lpp7-18 Volume 14, Number 1, February 2009.
18. LI Yingchun , A Systems-Science-Based Knowledge Explanation Method, TSINGHUA SCIENCE AND TECHNOLOGY ISSN 1007-0214 11/20 pp49 – 56 Volume 6, Number 1, March 2001.
19. Randall Davis, Howard Shrobe, and Peter Szolovits, What Is a Knowledge Representation?, AI Magazine Volume 14 Number 1 (1993) (© AAAI).
20. Dov Dori, The Visual Semantic Web: Unifying Human and Machine Semantic Web Representations with Object-Process Methodology,
21. Philippe Martin, Knowledge representation in RDF/XML, KIF, Frame-CG and Formalized-English.

AUTHOR(S) PROFILE

Miss. P. P. Deshmukh has pursuing of Master of Engineering in Computer Science and Engineering at Sant Gadge Baba Amravati University, Amravati, Maharashtra, India. She had pursued her Bachelor of Engineering in Computer Science and Engineering from Sipna College of Engineering, Sant Gadge Baba Amravati University.

Dr. M. S. Ali completed his B.E. (Electronics & Power) from Govt. College of Engineering, Amravati, and Nagpur University in 1981 with First Division. He completed his Masters' degree M.Tech. in Power Electronics from I.I.T. Powai, Mumbai in the year 1984. He obtained his Ph.D. in Electronics Engineering from Sant Gadge Baba Amravati University, Amravati in the year 2006. His main areas of interest are Operating Systems, Intelligent Systems and Java Technology. He is a recognized Ph.D. guide in Computer Science & Engineering in the Faculty of Engineering & Technology at SGB Amravati University. He is the founder Principal at Prof Ram Meghe College of Engineering & Management, Badnera-Amravati since 2009.

