International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 4, Issue 6, June 2015 **ISSN 2319 - 4847**

Research Paper on Xml Based Knowledge Representation using Scripts

[1]Ms. Pranita P. Deshmukh , Ms. Nupoor M. Yawale[2], Ms. Rupali A. Meshram[3], Ms. Komal R. Hole[4] , Ms. Ashwini H. Bhuskat[5]

PRMIT&R, Badnera (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.

1. 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.

2. RELATED WORK

To develop an approach for building knowledge-acquisition tools that provide strong support for a wide range of medications and knowledge-based system types [1][2]. 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[3].

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.

Knowledge-based systems for subjects using ETM vs. subjects using EXPECT. The experiments howed 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[3].

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[4].

A script is a structured representation describing a stereotyped sequence of events in a particular context [5]. 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[6].

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[7].

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 WebKB site[8].

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[9].

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. Sepsis 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[10].
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[11].

3. 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[12].

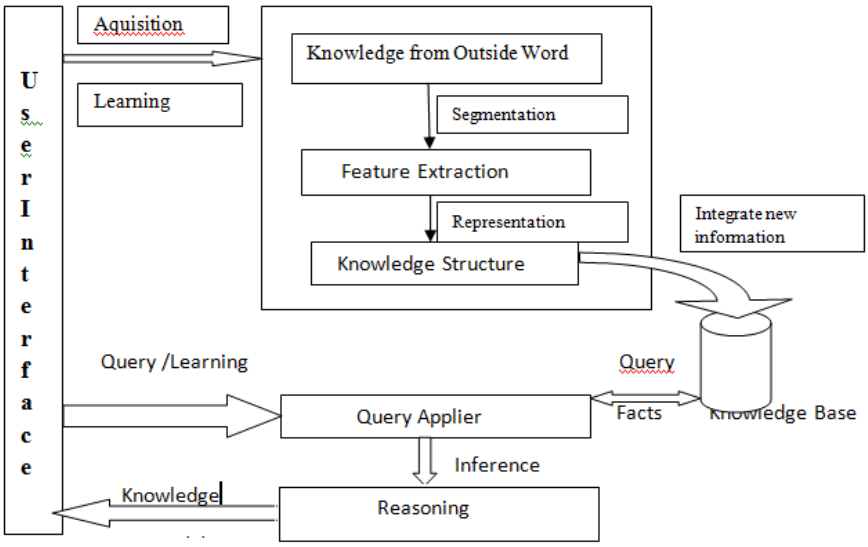


Figure 1. Knowledge base system model/Architecture

Table 1. 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.

4. KNOWLEDGE REPRESENTATION TECHNIQUES

4.1 Knowledge Representation Using Script

A script is a term proposed by Schank, and it refers to a form of knowledge representation. 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'. 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[12].

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[13]. 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[14][15][16]. 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 table.

Table 2: Example of 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

4.2 Information Retrieval

Over the years, the volume of information available has been increasing continuously, and never has so much information been so readily available and shared among so many people. The role of searching applications has therefore changed radically from systems designed for special purposes with a well defined target group to general systems for almost everyone. Unfortunately the unstructured nature and huge volume of information accessible over networks have made it increasingly difficult for users to shift through and find relevant information. Numerous information retrieval techniques have been developed to help to deal with this problem[17].

4.3 Ontology's and Knowledge Representation

Exactly what the knowledge representation is? This question can be answered as:

1) A Knowledge representation is a surrogate. Most of the things that we want to represent cannot be stored in a computer e.g., bicycle, birthdays, motherhood, etc., so instead, symbols are used as a surrogate for the actual objects or concepts.
2) A knowledge representation is a set of ontological commitments. Representations are imperfect approximations of the world, each attending to some things and ignoring others. A selection representation is therefore also a decision about how and what to see in the world. This selection is called the ontological commitment.
3) A knowledge representation is a fragmentary theory of intelligent reasoning. To be able to reason about the things represented, the representation should also describe their behavior and intentions. While the ontological commitment defines how to see, the recommended inferences suggest how to reason.
4) A knowledge representation is a medium for efficient computing. Besides, guidelines on how to view the world and how to reason, some remarks on useful ways to organize information are given.

5) A knowledge representation is a medium for human expression. The knowledge representation language should facilitate communication.

Over the past few years, many ontology development and query languages have been developed and this is still a continuing effort. As a branch of symbolic Artificial Intelligence, knowledge representation and reasoning aims at designing computer systems that reason about a machine-interpretable representation of the world, similar to human reasoning. Knowledge-based systemshave a computational model of some domain of interest in which symbols serve as surrogates for real world domain are facts, such as physical objects, events, relationships, etc[19][20].

4.4 Means of Ontological Knowledge Representation

If the knowledge in the KB is incomplete the semantic ranking algorithm performs very poorly: RDQL queries will return fewer results than expected, and the relevant documents will not be retrieved, or will get a much lower similarity value than they should. As limited as that might be, keyword-based search will likely perform better in these cases. To cope with this, our ranking model combines the semantic similarity measure with the similarity measure of a keyword-based algorithm[21].

On-to-Knowledge uses the general-purpose Resource Description Framework to represent metadata. The tool suite uses RDF Schema, a simple, Web-based RDF Vocabulary Description Language (http://www.w3.org/TR/rdf-schema/) to describe application-specific attributes and their corresponding semantics—for example, class hierarchies and domains and ranges of properties. [22] OIL is particularly effective as a development, delivery, and exchange language for ontology such as the Tao, which is complex and evolves with the current understanding of biology. Specifically, OIL inherits the best of both the frame and the description logic worlds. The frame-based modeling style and the range of epistemological constructs offered by OIL's syntax is comfortable and intuitive for most ontologies[23].

The RDFS produced by Protégé conforms to the W3C standard, except for the range definition of properties. RDFS only allows a single type for a range constraint; this is too limited for Protégé. This inconsistency can be handled by simply allowing multiple range constraints. The RDFS specification document indicates that we should specify a super class for multiple range classes, but this syntactic solution is not desirable from an ontological-engineering perspective because ranges can be disjunctive[24]. GeoShare lets users integrate heterogeneous information by resolving structural, syntactical, and Semantic heterogeneities. In the GeoShare spatial data infrastructure, we use the BUSTER search module as the core component of an Open GIS Consortium- compatible catalog service. The knowledge is represented in ontologies formalized in SHIQ description logic which can be migrated to OWL in the near future[25]. IHO S-100 Marine Data Model of the international Hydrographic Organization recently introduced new paradigm for integration and usage of various data on marine safety and the marine environment also includes a metadata model defined in Uniformed Modeling Language(UML).[26] This model describes main concepts and conceptual structures which provides syntactic interoperability between systems. Orymold is desktop software for the contextual management of gene expression data. The technical aspects of the implementation and the features of Orymold can be described by building a semantic model of the model organism *Oryza sativa*. This model can be used to include macro and microscopic pictures of organs, tissues, and cell types and to integrate gene expression data [27].

The ontology will be used to facilitate salivaomics data retrieval and integration across multiple fields of research together with data analysis and data mining. The ontology will be tested through its ability to serve the annotation ('tagging') of a representative corpus of salivaomics research literature that is to be incorporated into the SKB[10]. User profiles are semantically enriched using association rules, Bayesian networks and ontological knowledge in order to improve an agent's performance. Herein, the user profile is enhanced by enriching it with the semantics of ontology[11]. FAQ system on the Personal Computer (PC) domain, which employs ontology as the key technique to pre-process FAQs and process user query. It is also equipped with an enhanced ranking technique to present retrieved, query-relevant results. Specifically, the system uses the wrapper technique to help clean, retrieve, and transform FAQ information collected from a heterogeneous environment, such as the Web, and stores it in an ontological database. During retrieval of FAQs, the system trims irrelevant query keywords, employs either full keywords match or partial keywords match to retrieve FAQs, and removes conflicting FAQs before turning the final results to the user[12]. Search activities are performed by analyzing the subtopic correlation degrees with the user profile using Kendall, WebJaccard and cosine measures along the training subtopic sequence, measuring the session boundary recognition accuracy to identify the best threshold value. [13].

5. PROPOSED WORK

This work aims at the XML based knowledge representation using scripts. As a test of the proposed knowledge representation scheme, it shall be applied to specific domain. Then all the documents will be converted into xml document, and save in document base. These XML document convert it into ASCII format and apply pattern matching algorithm on them, Such that parse document using XML tags, apply fuzzy pattern rules on it. Then generate the

knowledge which later converted into scripts for representing knowledge. And also whenever query is made display respected results.

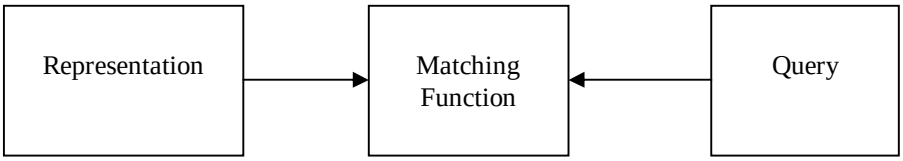


Figure 2: Simple model for Information Retrieval

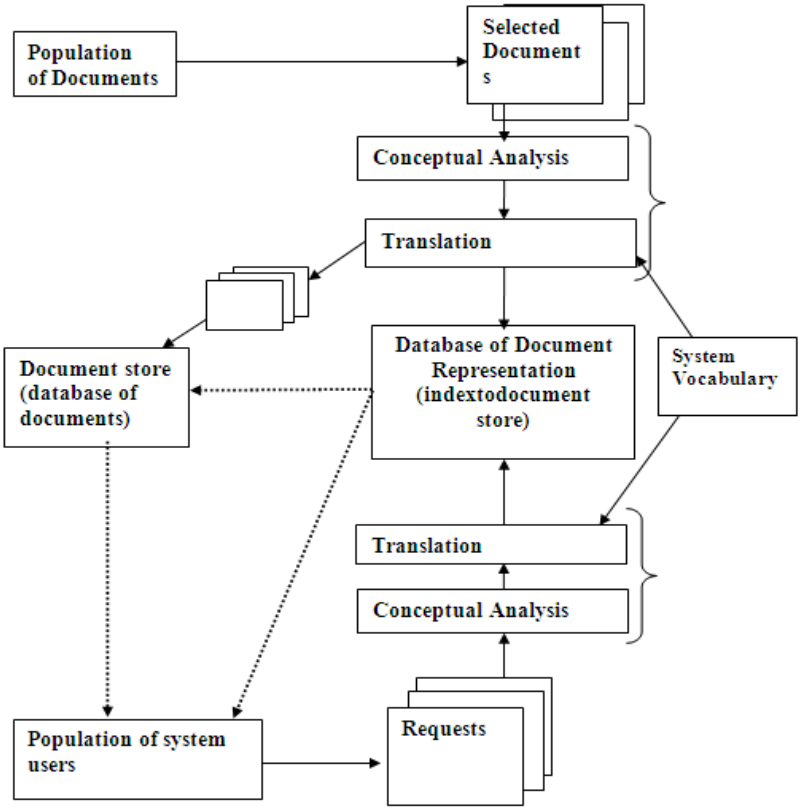


Figure 3: System Flow Diagram

In Fig.3, a more detailed view of the major functions in most information retrieval systems is shown. At the top of the figure, the input side of the system is designed as a set of selected documents. These documents are organized and controlled by the indexing process, which is divided into two parts, a conceptual analysis and a transformation. The conceptual analysis, and a content analysis recognizes the contents of the document and the transformations transforms this information into the internal representation. This representation is stored and organized in a database for later (fast) retrieval. The bottom of the figure defines the output side, which is very similar to the input side. The conceptual analysis and transformation of requests recognizes the semantics and transform this knowledge into a representation similar to the representation used in the indexing process.

Then, some type of search strategy (matching function) is used to retrieve documents by comparing the description of the request with the description of the documents. If the request process is iterative, as shown in the figure by broken lines, it uses either document or document description in the process of obtaining the information needed. As the following shows that the system architecture as per the flow of data, and the storage of database, also checking the documents as per need, and send request some of them, finding out the some of the document in it. In this way, the portioning the two ways as per the given in flow diagram are mentioned below.

5.1 Clustering

A clustering is essentially a set of such clusters, usually containing all objects in the data set. Additionally, it may specify the relationship of the clusters to each other, for example a hierarchy of clusters embedded in each other. Clustering can be roughly distinguished as:-

- Hard clustering: Each object belongs to a cluster or not.
- Soft clustering (also: fuzzy clustering): Each object belongs to each cluster to a certain degree (e.g. a likelihood of belonging to the cluster).

In the presented work, fuzzy approach is provided to the clustering which gives efficient way to define relationships. In comparison with hard clustering methods, in which a pattern belongs to a single cluster, fuzzy clustering algorithms allow patterns to belong to all clusters with differing degrees of membership.
Clustering is this ……..

Algorithm:

Input:a→Length of word
Si…n→ character
Pi…n→ Character Array
C → No. of clusters

Output: Cluster membership values
{ Pim| i=1…….N, m=1…….c}

Steps:

- Initialization [i ,p=0, c=0]
- Initialize and normalize membership values.
- Convert input string to character array.
- While (p<=a.length)
- For (i=p; i<a.length; i++)
- S.Add (a[i].ToString());
- C= C +S[i];
- Add C;
- End For
- C= null;
- P++;
- End While

Detailed Design

This project proposed two flow charts as they are following

- Xml Clustering
- Data mining & Knowledge Representation

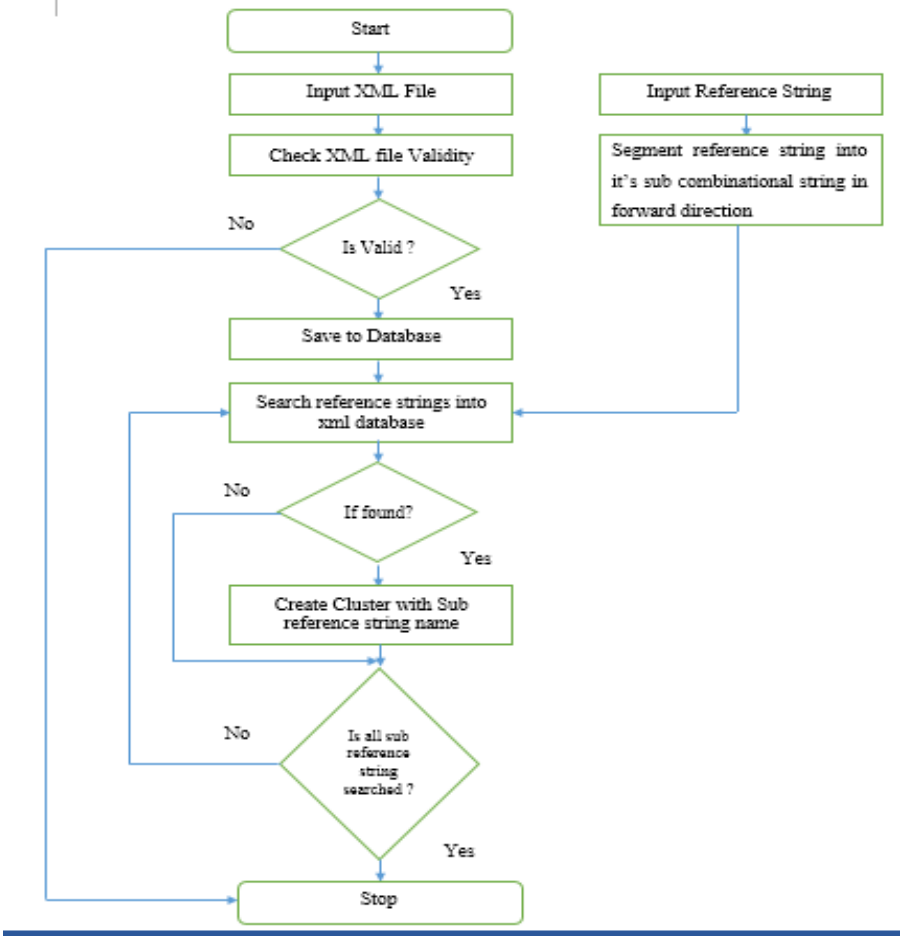


Figure 4. Data Flow Diagram(XML Clustering)

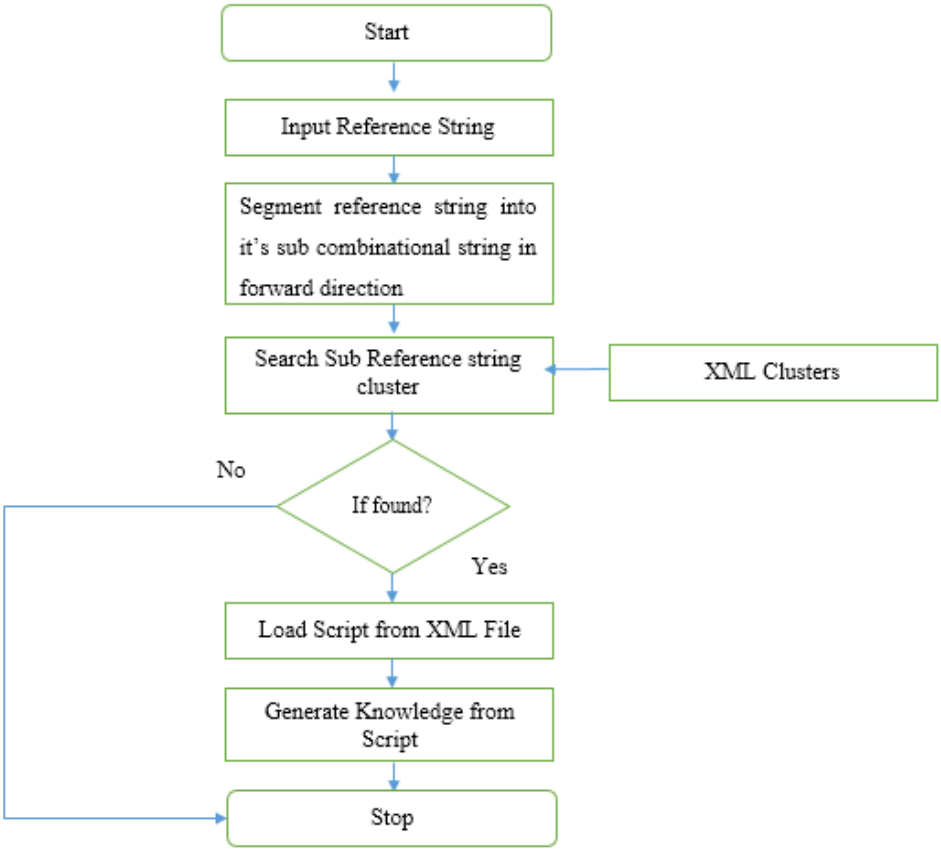
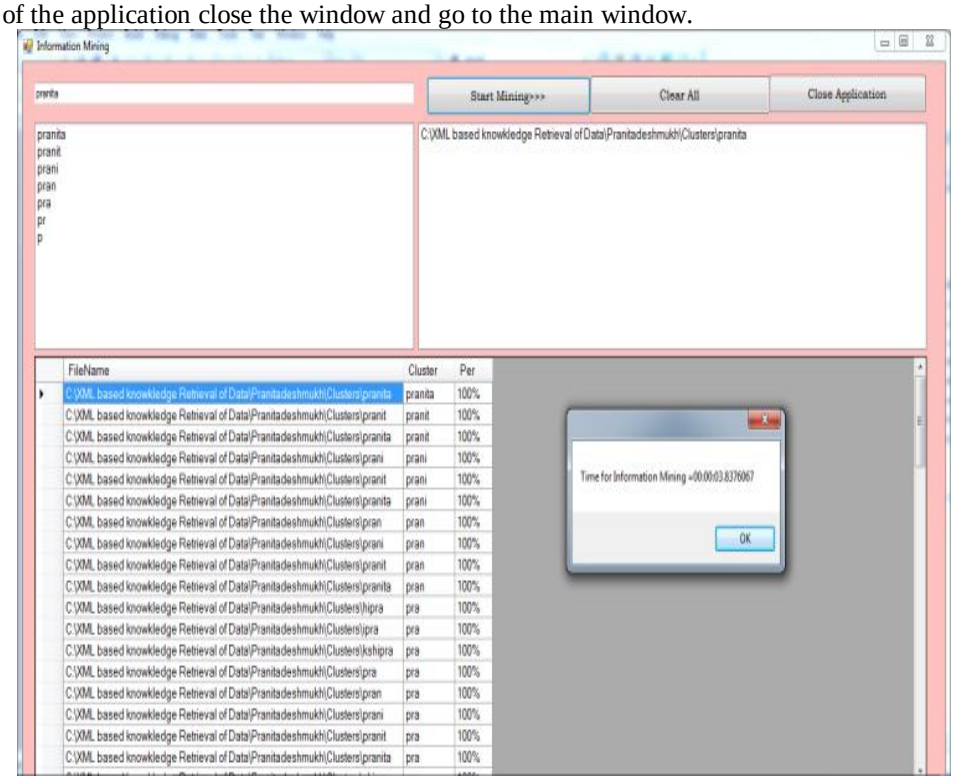


Figure 5: Data Flow Diagram(Data Mining & Knowledge Representation)

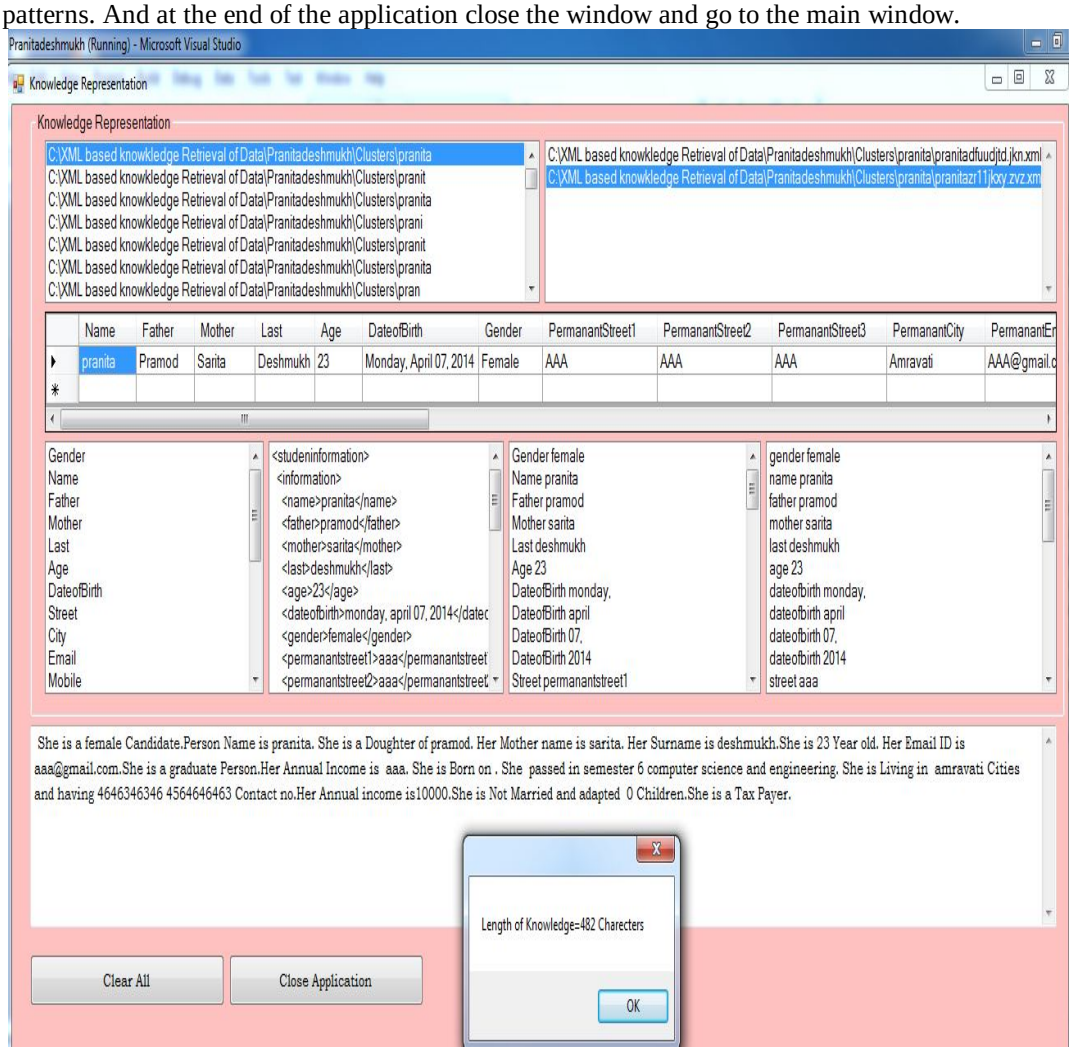
5.2 Implementation:- System Execution Results

This snapshot shows that the keyword to be mined is entered. Their respective files are displayed. All clusters of respective keyword are displayed. The probabilities of presence of particular cluster in particular files are displayed. And at the end of the application close the window and go to the main window.



This is knowledge representation window shows that the all xml files of respective mined keyword are displayed. Select the particular file from it. The information in tabular format from particular file whichever we selected are displayed. Also the attributes which are given in file are displayed. Xml tag file are displayed. The attributes and their respective

information are displayed. Then finally represent the knowledge using scripts from given data by applying fuzzy rules and fuzzy patterns. And at the end of the application close the window and go to the main window.



5.3 Experiment Result

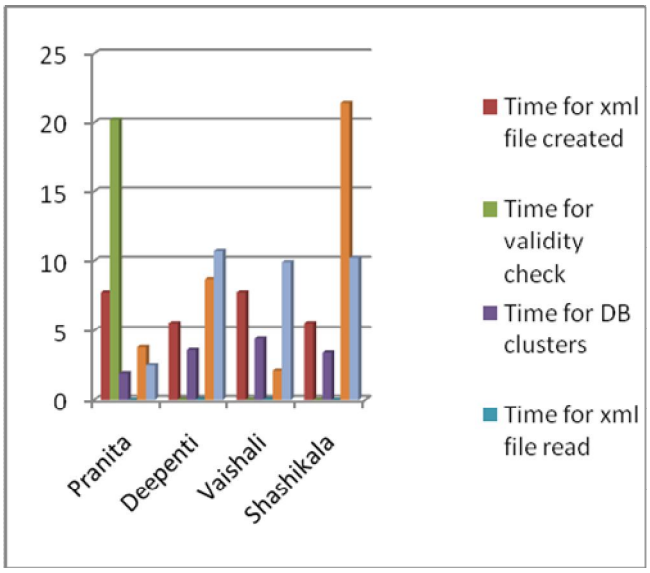
Table 3: Combinational sub reference string

Sub-String	***Weight(%)***	***Time for segmentation***
Pranita	100	0.0056 Sec
Pranit	85.71	
Prani	71.42	
Pran	51.14	
Pra	42.85	
Pr	28.57	
P	14.28	

Table 4. Mining & Knowledge Generation Result

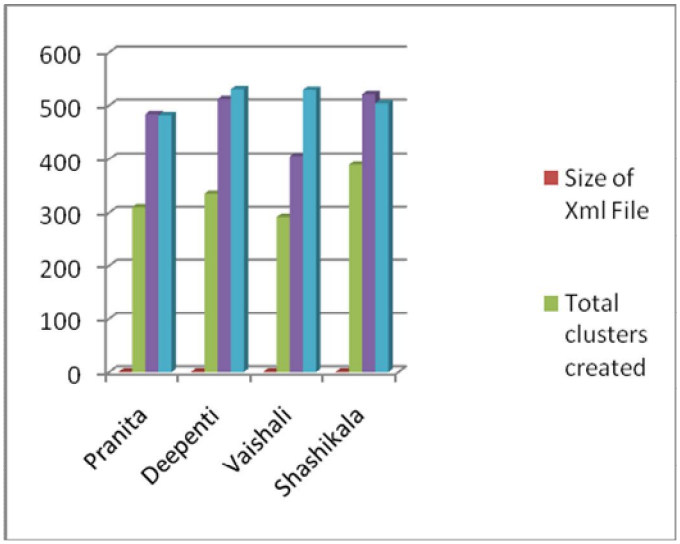
Sr.No	***Cluster name***	***No of XML files***	***No of script Matched***	***Time for knowledge Generation***
1	Pranita	2	Name,mother name	0.0234 Sec
2	Pranit	4	Name,mothername,father name	0.01435 Sec
3	prani	7	Name,mothername,sister name	0.2341 sec
4	p	16	Name,city,mothername,fathername,address….	12.345 sec

5.4 Simulation Results:



Graph 1: Graphical Results on input strings

The above graph shows that the time of information mining depends on the no. of clusters. And the no of clusters depends on length of keyword.



Graph .2: Graphical Results on input string

The above graph shows that the total no of clusters and length of knowledge for particular keyword.

5.5 Comparative Results and Discussion

The experimental results shown above derive the need of efficient clustering. To define the efficiency the proposed algorithm is compared with some other clustering algorithms and the comparative results are discussed. To deal with the comparisons, here the previous results given on the different clustering algorithm are considered. The proposed clustering algorithm is compared with hierarchical clustering algorithm, Self-organization map algorithm (SOM) and Expectation maximization (EM) clustering algorithm. The brief description algorithm given as follows.

- **Proposed algorithm:**
  It is a partitioning method. Objects are classified as belonging to one of the group. This method minimizes the overall within-cluster dispersion by iterative reallocation of cluster members.
- **Hierarchical algorithm (HCA):**
  This algorithm combines or divide existing groups, creating hierarchical structure that reflects the order in which groups are merged or divided.
- **Self- organization Map algorithm (SOM):**
  To use this algorithm is to regard the objects in the input space represented by the same node as grouped into a cluster. Each objective the input is presented to the map and the best matching node is identified.
- **Expectation Maximization algorithm (EM):**
  EM is the distance based algorithm that assumes the dataset can be modeled as a linear combination of multivariate normal distributions and the algorithm finds the distribution parameters that maximize a model quality measure, called log likelihood.

The reasons behind selecting these algorithms are:

- Popularity
- Flexibility
- Applicability

All these algorithms are compared according to the following factors:

- Number of clusters
- Size of dataset

According to above discussion and results of previous algorithm, it is found that

- As value of clusters become increases, the performance of SOM algorithm becomes lower. However the performance of proposed algorithm and EM algorithm become better than the hierarchical clustering algorithm.
- The Proposed Algorithm and EM algorithm has better for huge database and SOM & HCA algorithm better for small database.

6. CONCLUSION

The aim of this dissertation, as the indicated in the beginning was to investigate and discuss introduction of ontologies and fuzzy rules to knowledge representation and thereby the promotion of semantics in the document representation propose. The main research question was how to recognize concepts in information objects and queries, represent these in the information mining system, and use the knowledge about relations between concepts captured by ontologies in the querying process. In the dissertation, we have focused on the following three main aspects related to the research question.

1. Recognition and representation of information in the documents and queries and the mapping of this knowledge into the ontologies.
2. Improvement of mining process by use of similarity measures derived from knowledge about relations between concepts in ontologies, and
3. How to weld the ideas of such ontological and fuzzy rules and ontological similarity into realistic scalable information mining scenario, And the conclusion first discusses and concludes on these three elements separately, after which an overall conclusion is given and perspectives that can be used as the subject of further work is indicated.

7. FUTURE SCOPE

The ontology's can be adapted to the document space within multi-disciplinary domains where different terminology is used. The objective is to enhance the user-experience by improvement of search result quality for large-scale search systems. In the same domain, also we will increases The Xml tags, Attributes and Fuzzy rules on it. Also we can use Indexing Concept in it, for which we will reduces the searching timing and calculate the no of turns for particular search word. Also improve parameters as per consideration.

This project was intended to compare between some data clustering algorithm. Through my extensive search I was unable to find any study that attempts to compare between the four clustering algorithm under investigation. As a future work, comparisons between these four algorithm (or may other algorithms) can be attempted according to different factors other than those considered. One important factor is normalization. Comparing between results of algorithms using normalized data or non- normalized data will give different results. Of course normalization will affect the performance of the algorithm and the quality of the results. Another approach may consider using data clustering algorithms on applications such as object and character recognition or information retrieval which is concerned with automatic storage and retrieval of documents.

REFERENCES

[1]. Marcelo tallis, “A script based approach to modifying knowledge based system”2011.
[2]. PushpaKesarwani, Prof. A.K. Misra, “Selected integrated approach for knowledge representation by comparative study of knowledge scheme”, International Journal of Scientific and Research Publications, Volume 3, Issue 2, February 2013.
[3]. Tilotma Sharma1, Deepali Kelkar2,“A Tour Towards Knowledge Representation Techniques”, International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 2, Issue 2,2012.
[4]. DovDori, “The Visual Semantic Web: Unifying Human and Machine Semantic Web Representations with Object-ProcessMethodology”.Technion, Israel Institute of Technology, Haifa, 2010.
[5]. PoonamTanvar, Dr. T.V. Prasad, Dr. KamleshDatta “Hybrid Technique for Effective Knowledge Representation & A Comparative Study”2010.
[6]. Philippe Martin, “Knowledge representation in RDF/XML, KIF, Frame-CG and Formalized-English”. Distributed System Technology Centre Griffith University - PMB 50 Gold Coast MC, QLD 9726, 2004.

[7]. Ying Zhang, Alex Golovinsky, Mark Yim, Craig Eldershaw, "AnXML-based Scripting Language for Chain-type Modular Robotic Systems", Palo Alto Research Center, 3333 Coyote Hill, Palo Alto, CA, USAyzhang, yim, celdersh @parc.com and alexg@mit.edu,2010.
[8]. Thomas R. Gruber , "Towards Principles for the Design of Ontology Used for Knowledge Sharing", International Journal of Human Computer Studies, November 1995, Vol. 43, Issues 5-6,1993.
[9]. Dieter Fensel, "Ontology Based Knowledge Management", IEEE Computer, Nov 2002, Vol 35 No 11. pp56-59, 2002.
[10]. Pablo Castells,Mirium Fernandez, David Vallet, "AnAdaptation of the Vector Space Model for Ontology Based Information Retrieval" . IEEE Transactions on Knowledge and Data Engineering, Feb 2007, Vol.19,No.2.
[11]. M. Kerber, "Knowledge Representation I", Introduction to AI 06-08775, 2004/05.
[12]. B. Lamiroy1 and L. Najman2,3, "Scan-to-xml: Using Software Component Algebra for Intelligent Document Generation"2010,
[13]. 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 lpp7-18Volume 14, Number 1, February 2009.
[14]. 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.
[15]. Gaurav Harit1, K. J. Jinesh2, Ritu Garg3 C.V Jawahar2, and Santanu Chaudhury3," Managing Multiligual OCR Project using XML"2008.
[16]. Alexandre Delteil1, Catherine Faron2, and Rose Dieng1, "Building Concept Lattices by Learning Concepts from RDF Graphs Annotating Web Documents",1990.
[17]. Jean-François Baget 1,2, Olivier Corby1, Rose Dieng-Kuntz1, "Griwes: Generic Model and Preliminary Specifications for a Graph-Based Knowledge Representation Toolkit",2010.
[18]. H. Cherfi1, O. Corby1, C. Faron Zucker1,2, K. Khelif1 and M.T. Nguyen1," Semantic Annotation of Texts with RDF Graph Contexts"2009.
[19]. 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.
[20]. Victoria Uren, Bertrand Sereno, Simon Buckingham Shum &Gangmin Li, September, 2003,"Interfaces for Capturing Interpretations of Research Literature".
[21]. Kossmann D. and Stocker K., "Iterative Dynamic Programming: A New Class of Query Optimization Algorithm," ACM TODS, March 2000.
[22]. Ibraraki T. and Kameda T. "Optimal Nesting for Computing N-Relational Joins," ACM Transactions on Database Systems, vol. 9, no. 3, pp. 482-502, 1984.
[23]. Oszu M. T. and Valduriez P., "Principles of Distributed Database Systems", Prentice Hall International, NJ, 1999.
[24]. H. Balakrishnan, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica, "Looking up data in P2P systems", Communications of ACM , Vol.46, No.2, 2003.
[25]. N. Daswani, H. Garcia-Molina, and B. Yang, "Open problems in data-sharing peer-to-peer systems", Proc. of the 9th International Conference on Database Theory (ICDT'03), 2003.
[26]. D. S. Milojicic, V. Kalogeraki, R. Lukose, K. Nagaraja, J. Pruyne, B. Richard, S. Rolins, and Z. Xu, "Peer-to-peer computing", HP Lab technical report, HPL-2002-57, 2002.
[27]. D. Barkai, "Technologies for sharing and collaborating on the net", Proc. of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02), 2002.
[28]. S. Chaudhuri, G. Das, and V. Narasayya, "A Robust Optimization- Based Approach for Approximate Answering of Aggregate Queries," Proc. 20th ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '01), 2001.
[29]. B. Babcock, S. Chaudhuri, and G. Das, "Dynamic Sample Selection for Approximate Query Processing," Proc. 22nd ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '03), pp. 539-550, 2003.
[30]. S. Chaudhuri, G. Das, M. Datar, R. Motwani, and V. Narasayya, "Overcoming Limitations of Sampling for Aggregation Queries," Proc. 17th IEEE Int'l Conf. Data Eng. (ICDE '01), pp. 534-542, 2001.
[31]. S. Chaudhuri, G. Das, and V. Narasayya, "A Robust Optimization- Based Approach for Approximate Answering of Aggregate Queries," Proc. 20th ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '01), 2001.
[32]. S. Acharya, P.B. Gibbons, and V. Poosala, "Aqua: A Fast Decision Support System Using Approximate Query Answers," Proc. 25th Int'l Conf. Very Large Data Bases (VLDB '99), 1999.
[33]. C. Gkantsidis, M. Mihail, and A. Saberi, "Random Walks in Peerto- Peer Networks," Proc. IEEE INFOCOM '04, 2004.
[34]. M. Charikar, S. Chaudhuri, R. Motwani, and V. Narasayya, "Towards Estimation Error Guarantees for Distinct Values," Proc. 19th ACM Symp. Principles of Database Systems (PODS '00), 2000.
[35]. S. Chaudhuri, R. Motwani, and V. Narasayya, "Random Sampling for Histogram Construction: How Much Is Enough," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '98), pp. 436- 447, 1998.

[36]. X. Li, Y.J. Kim, R. Govindan, and W. Hong, "Multi-Dimensional Range Queries in Sensor Networks," Proc. First ACM Int'l Conf. Embedded Networked Sensor Systems (SENSYS '03), 2003.
[37]. KING, J.C. "Centralized versus decentralized computing: Organizational considerations and management options". ACM Comput. Suru. 15, 4, Dec, 319-349 1983.
[38]. TANENBAUM, A. "Computer Networks". Prentice-Hall, Englewood Cliffs, N.J., 1981.
[39]. CERI, S., AND PELAGATTI, G. "Distributed Databases: Principles and Systems". McGraw-Hill, New York, 1984.
[40]. Osama Abu Abbas ," Comparisons Between Data Clustering Algorithms" Computer science department, Yarmouk University, Jordan, The international Arab Journal of Information Technology, Vol.5,no.3,July 2008.
[41]. Dan W. Patterson, "Introduction to artificial Intelligence and Expert System".
[42]. Elaine Rich, Kelvin Knight, Shivshankar Nair", Artificial Intelligence"(Third Edition).

Author Profile

P.P. Deshmukh has received her ME. Degree in Computer Science & Engineering from P.R.M.C.E.A.M, Badnera, Amravati, India. Currently she is working as an Assistant Professor at Prof. Ram Meghe Institute of Technology & Research, Badnera, Maharashtra, India

N.M. Yawale has received her ME. Degree in Computer Science & Engineering from P.R. Patil College of Engineering, Amravati, India. Currently she is working as an Assistant Professor at Prof. Ram Meghe Institute of Technology & Research, Badnera, Maharashtra, India.

R. A. Meshram has received her M.Tech. degree in Computer Science & Engineering from Govt. College of Engineering, Amravati, India. She has published one paper in national level and three papers in international journals. Her research interests include data mining, pattern classification. Currently she is working as an Assistant Professor at Prof. Ram Meghe Institute of Technology & Research, Badnera, Maharashtra, India.

K. R. Hole has received her M.E. degree in Computer Science & Engineering from Sipna College of Engineering & Technology, Amravati, India. She has published five papers in international journals. Her research interests include Artificial Intelligence, Algorithm. Currently she is working as an Assistant Professor at Prof. Ram Meghe Institute of Technology & Research, Badnera, Maharashtra, India.

A.H. Bhuskat has received her ME. Degree in Computer Science & Engineering from P.R.M.C.E.A.M, Badnera, Amravati, India. Currently she is working as an Assistant Professor at Prof. Ram Meghe Institute of Technology & Research, Badnera, Maharashtra, India.