CHAPTER 1

INTRODUCTION

1.1 KNOWLEDGE SYSTEM

During recent decades, knowledge is the aspiring elementary resource mandatorily required by all intelligent information processing systems. It is difficult to define knowledge precisely. It can be considered as broader, deeper or detailed understanding of any concept coming from a source (Irfan & Shaikh 2010). The term knowledge refers to the fluid mix of framed experience, values, contextual information, in addition to expert insight that facilitates a framework to evaluate and incorporate innovative experiences and information (Whittaker et al 2003). Therefore, knowledge is not just an explicit tangible “thing”, like information, but information combined with experience, context, interpretation and reflection (Denning 2002). The acquisition and effective deployment of domain knowledge is essential for the success of any domain. Knowledge Engineering (KE) is an engineering discipline that involves the integration of knowledge into computer systems for solving complex problems typically necessitating a high level of human expertise (Whittaker et al 2003). KE provides two views namely, transfer view and modeling view. The traditionally used KE is the transfer view that works under the assumption of applying the knowledge engineering techniques to transfer human knowledge into intelligent system, whereas in the modeling view knowledge engineers attempt to model the knowledge of domain experts into intelligent systems (Newman & Conrad 2000).

Knowledge engineers use knowledge engineering techniques to build intelligent systems, expert systems, knowledge based systems, knowledge based
decision support systems, expert database systems and more. More specifically, knowledge engineering addresses uncertain process requirements by emphasizing the acquisition of knowledge about a process and representing this knowledge in a Knowledge-Based System (KBS) (Wilson 1993). A KBS maintains a knowledge base that contains domain knowledge, knowledge about knowledge, business heuristics, procedural rules, factual data, and more (Grimm et al 2007), (Abel et al 2004). (Studer et al 1998), outlined the relationship of Knowledge Engineering to Software Engineering, Information Integration and Knowledge Management.

The mode in which we, as humans, process knowledge is by reasoning, i.e. the process of arriving at conclusions. Analogously, the knowledge available in a knowledge base is processed by a computer and thereby drawing conclusions from it, i.e. by deriving novel statements that proceeds from the given ones (Mahalakshmi & Geetha 2008). The web-based construction knowledge system is developed to facilitate knowledge sharing and reusing for the project teams and construction corporations. The construction knowledge system can be used to capture and share both explicit and tacit knowledge under multiple projects for contractors (Intan & Mukaidono 2002).

Another significant constituent of KBS is the inference engine, which derives new knowledge and utilizes the existing knowledge for decision-making and problem solving (Sajja 2008). Knowledge-based systems generally apply powerful abstract constructs to model a domain and operate robustly with a small number of instances of constructs (Abel et al 2004). In brief, the KBSs are expected to extract fresh constructive knowledge from existing knowledge that is already represented and stored (Mattos 1991).

Knowledge management (KM) basically gains insight into the subject of intellectual capital management to form a corporation with competitive
advantage. Though Knowledge Management needs knowledge technology, it cannot be merely inherited through a whole computer code package (Akhavan et al 2010). KM helps the capture, storage, and dissemination of knowledge in a useful way (Lai 2007). (Wang et al 2011) forwarded the theoretical framework of Knowledge Management (KM) and explored the method of knowledge discovery method based on KM, mainly elaborating the definition and normal process of KM, giving the concept, overall structure and main function modules. There are numerous forms of IT tools for sharing and organizing knowledge. During the past few years, Knowledge management (KM) has become a very important subject of study among practitioners within the field of management, applied science, and knowledge technology (Akhavan et al 2010). Knowledge Management (KM) manifests as set of organising principles and heuristics which shape management routines, structures, technologies and cultures within organisations (Moayer & Gardner 2012).

The scope of knowledge is broad and knowledge could exist in numerous forms like declarative, explicit and specific. Strategies for managing knowledge has become a very important issue in the past few decades, and therefore the KM community has developed a large variety of technologies and applications for each educational analysis and sensible applications (Lai 2007). Knowledge management (KM) is one among the rising topics in the several fields of knowledge (Kebede 2010). E-learning system among KM is historically analyzed as a knowledge resource repository, wherever the KM strategies are often enforced to extend the effectiveness of knowledge dissemination (Qwaider 2011). Another way of knowledge management done in academic libraries was proposed by Ralph and Ellis(2009). They investigated the use of the knowledge base of QuestionPoint as a knowledge management tool, which is capable of improving reference services in academic libraries. It also explored the librarians’ perceptions and problems of using the knowledge
base of QuestionPoint and its impact on reducing response time and duplication. The emergence of a knowledge-based economy (k-economy) has spawned a “new” notion of workplace literacy, changing the relationship between employers and employees((Kefela 2010).

One of the key advantages of introducing KM practices in organizations is its positive impact on organizational performance. KM totally mediates the impact of organizational culture on organizational effectiveness and part mediates the impact of organizational structure and strategy on organizational effectiveness. Finally, the results of diverse research show that KM affects organizational performance in an exceedingly positive manner; however, this relationship is incredibly troublesome to prove (Rasul & Vuksic 2012). KM strives to bring a successful market business by having a proper, structured initiative. It can also enhance the employment of knowledge in wider areas where a good structure memory knowledge system plays a vital role. This suggests that the knowledge of suppliers are often maintained, categorized, retrieved and managed effectively (Mollahosseini & Barkhordar 2010). The knowledge engineering-related KM covers the following tasks, such as

i) Knowledge acquisition

ii) Knowledge representation

iii) Knowledge discovery and retrieval

iv) Decision making

1.1.1 Knowledge Acquisition

The utilization and development of the knowledgeable systems in varied subject areas is found in agriculture, chemistry, technology, engineering, geology, medicine, house technology etc. Knowledge acquisition is a dynamic process. Cognitive structure and cognitive process have great influence on knowledge acquisition and knowledge discovery (Hongxia et al 2010). Knowledge
acquisition is an essential part in the development of a knowledgeable system. It consists of human experts’ knowledge that is extracted by the knowledge engineers through the method referred to as knowledge Acquisition. Knowledge Acquisition (KA) is the process of acquiring knowledge, and its formalized structure, that will allow some particular task to be performed by a computer system (Potter 2003). The knowledge acquisition is the accumulation, transfer, and transformation of problem resolution experience from human consultants to bug. Exploiting knowledge from consultants is a tough task and has been known as a bottleneck in knowledgeable system development. It is among the foremost common cited problem within the construction of knowledgeable system (Amin & Khan 2009).

Secondly, evaluation of knowledge acquisition tools and methodologies remains problematic. The essential problem is that any analysis needs those involved to truly build a KBS. However, completely different Problem Solving Methods (PSMs) can be applied to constant problems. There is an elementary problem with solely evaluating the first stage of developing a KBS. KA analysis transpires to deal with the “knowledge-engineering bottleneck”. Network fault knowledge acquisition is a necessary part of intelligent network management (Wu et al 2005). Process planning knowledge discovery technology based on knowledge model driving is researched and is applied (Jia et al 2007). Knowledgeable systems has been simple to demonstrate the technology on little issues, however there are many problems with the ad-hoc approach in addressing giant scale knowledge-acquisition (Cao & Compton 2005). Knowledge acquisition technique based on structural analysis is formed by applying structural modeling, which is on basis of knowledge structural model (Zhang et al 2004).
1.1.2 Knowledge Representation

A Knowledge representation is a concept to change to thinking strategy instead of acting. Knowledge representation plays a primary role in agricultural intelligent information system as a knowledge-based system(Yuanyuan et al 2010). There are two basic elements of knowledge illustration, i.e. reasoning and illation. There are three wide views of knowledge illustration. a) KR as applied epistemology: All intelligent system presupposes knowledge that is diagrammatical in an exceedingly mental object that consists of knowledge structures and programs, b) KR as tell-ask module: KR system ought to offer a minimum of two operations: (i) For a given mental object K, with the facts f. It should be leading to a brand new mental object, K'. (ii) The mental object K is being queried by a few reality f. Outcome depends upon KR paradigm used, may be yes, no, unknown, affirmative with a confidence issue, c) KR as the embodiment of AI systems: There are identical interconnected units that are put together for representing numerous ideas. An inspiration is diagrammatical distributed sense and is indicated by an evolving pattern of activity over a set of units.

In standard computing, the knowledge is put on hold in knowledge base, whereas in AI the mental object is employed to store the knowledge needed for resolving the actual task (Tanwar et al 2010). (Wang & Meng 2008) discussed the knowledge representation about the part description, and manufacturing resources knowledge which are manufacturing operations, cutting tools, machine tools. A knowledge representation - "GCS ternary" method for the model of capacitated vehicle routing problems (CVRP) was introduced on the basis of knowledge representation theories in artificial intelligence (AI) and knowledge engineering (KE) in order to decrease the deficiencies of model representation in decision support systems (DSS) for vehicle routing problems (VRP) in logistics distribution systems of e-commerce(Sun & Hu 2005).
Currently, there are several techniques for representing knowledge, i.e. List and tree, which are employed to represent the hierarchal knowledge. Concept Knowledge Tree (CKT) is an ideal method to represent structural and relational knowledge. (Liu et al 2013) presented a knowledge acquisition and representation approach using the fuzzy evidential reasoning. Behavior Tree is an elegant and understandable solution to procedural knowledge representation for creating reusable logic(Geng et al 2011). (Lei et al 2010) proposed a agent-based knowledge representation, focusing basic framework and the key points of the agent-based knowledge representation. Linguistics networks are those where nodes and links are not stored the propositions. Rule-based knowledge illustration primarily utilized in problem-solving contexts that involve production rules containing if-then or situation-action pairs. Rules generally contain:

- Initial state.
- Goal state.
- Legal operators, which are the things allowed doing.
- Operator restrictions.

Logic-based representations might use deductive or colligation that contain:

- Facts and premises.
- Rules of propositional calculus and rules of predicate
- Calculus that permits use of extra info regarding objects within the proposition, use of variables and functions of variables.
- Measures of certainty involve certainty factors (Tanwar et al 2010).

Different types of knowledge representation usually utilized in knowledge and data engineering is tree, quad tree, and XML structure.
**Tree:** Tree representation is the most ordinarily used, attributable to its simple, low cost to grasp and implement. It provides a modeling technique that is simple for humans to grasp and simplifies the extraction method. Most of the tree algorithms are often enforced in each serial and parallel type, whereas others will solely be enforced in either serial or parallel type. Parallel implementation of tree algorithms is fascinating in-order to confirm quick generation of results particularly with the classification or prediction of enormous knowledge sets. It conjointly exploits the underlying design. However, serial implementation of tree is simple to implement and fascinating once small-medium knowledge sets are not concerned (Matthew & Anyanwu 2009).

**Quad tree:** A quad tree may be a hierarchical system applied to represent geometric knowledge. It is particularly helpful for raster-based image processing and vital for several applications, e.g., image processing, pattern recognition, computer vision, process of spatial information, etc. The hierarchical system of quad tree illustration for regions or objects on a picture makes storage a lot more economical. The initial approach to represent geographical knowledge is that the quad tree system with pointers (Chang et al 1997). The quad tree is portrayed by a tree whose node is either a leaf or an enclosed node with four children. The foundation represents a block and its four sons are represented by NW, NE, SW, and SE quadrants. The quad tree is obtained by recursively subdividing the image into quadrants, sub-quadrant, till every block consists entirely of constant points. Every node within the tree desires different fields to keep up the connection with different nodes. There is one pointer for its father’s node and there are four tips to the nodes of 4 sons. The latter field holds the info price of this node. A routine quad tree desires abundant less storage than the initial image. However, the employment of pointers is inevitable; a routine quad tree still consumes a substantial quantity of
space for storing. The memory area will be saved more if a pointer-based quad tree is replaced by a linear writing technique. A linear quad tree may be a linear array of structure parts and it is specified implicitly. The linear structure stores location code and attributes price of any node in a much sorted array. As a result, it desires no tips to represent the nodes and is more room saving than a routine quad tree. Linear quad tree may be an efficient system for raster-based spatial knowledge to scale back the quantity of knowledge storage and also the demand of memory and time throughout processing. However the way to improve the linear quad tree manufacturing potency remains a drag to be studied. It is a basic operation for geographic system to overlay 2 maps. The quad tree system has been used because the underlying illustration for maps in many experimental GIS with success (Elkaffas et al 2011, Yehua et al 2001).

**XML:** XML (eXtensible Mark-up Language) has recently gained unequalled importance as an elementary standard for economical knowledge management and exchange. Knowledge destined to be broadcasted over the web is henceforth pictured using XML, to warranty its ability. Due to the unprecedented web exploitation of XML, XML-based comparison, particularly for heterogeneous documents, becomes a central issue within the knowledge retrieval and info communities. Its applications move over version management, amendment management and knowledge deposition, XML question systems, similarly because the classification/clustering of XML documents gathered from the web against a group of DTDs are declared in an XML knowledgebase (Tekli et al 2012). The eXtensible Mark-up Language (XML) is seeing multiplied use, and guarantees to fuel even a lot of applications in near future. However several of those XML documents, particularly those getting down seemingly on the web, are not Document Type Descriptors (DTDs) (Nierman & Jagadish 2012).
Gathering information from limited number of organization was discussed here because the knowledge management was a brand new concept of Iranian organization. The methodology of relational data mining was discussed which can take into account several coexistence measures at the same time. The result provided greatest accuracy rate and robustness even if the data set contained missing values Akhavan et al (2010), Debeljak et al (2012), and Kumar & Rathee (2011). In most cases KM was a natural and long awaited development in Information System (IS). However, a review was conducted and defined the common characteristics of knowledge management and e-learning system and enhanced the organizational performance by creating, accumulating, organizing and utilizing knowledge. And found that knowledge model presented here was a useful starting point to gain a deeper insight into KM elements and their influence to the organizational performance and defined knowledge management and supplier knowledge management. Although the progression was best described as evolutionary and a number of factors have contributed at different times accelerated the progression to take place at this point in time (Kebede 2010), (Qwaider 2011) and Rasul & Vuksic (2012).

The integration of e-Learning systems and Knowledge management technology was investigated here (Qwaider 2011). Hence, knowledge management supports business management to find, choose and keep reasonable supplier network in the best manner in which the firm maintained its competitive advantages. The specific aspect of inferring semantics automatically from raw video data was addressed here. In AI various intelligence scheme was represented and each have advantages and disadvantages and determined several important weakness factors which affected the implication of knowledge management. Here they were presented a model for integration of e-learning system and knowledge management. Empirical study was conducted to investigate the relation between Knowledge Management strategies and organizational performance and understood the
importance of knowledge management. Firms to cost switching and raise barriers to entry were created by the supplier knowledge management and applied it to cut production costs, innovating new products and to attract customers and suppliers Mollahosseini & Barkhordar (2010), ThangaRamya& Rangarajan (2011).

A positive effect of knowledge management practice on organizational performance was confirmed through the result investigated and provided a conceptual frame work to enhance the supplier knowledge management. A knowledge engineering approach was proposed here to manage knowledge and the approach was called Knowledge management through knowledge Engineering (KMKE). Here, the challenging factors were identified for the implementation of knowledge management portals for Iranian organizations. Knowledge engineering with knowledge management was combined here to solve the major problems in most KM systems encounter Rasul & Vuksic (2012), Lai (2007), and Akhavan et al (2010). It was for adventitious presence of authorized genetically modified GM material in conventional non-GM maize crops. Networks of protein knowledge were computed in which interconnected metabolic, disease, enzyme and gene function data to motivate possible application. Information technology, organisation and knowledge were the three component presented here for knowledge management. Knowledge management was a process that transformed individual knowledge into organizational knowledge Debeljak et al (2012), Lange et al (2007) and Rasul & Vuksic (2012). And illustrated that how knowledge management was enhanced in an organization and provided the benefits of both. The discovered knowledge may become obsolete in a short period of time due to frequent change in the visitor’s interests. Knowledge Base System(KBS) consists of a database-type repository for maintaining the patterns, and rules to improve the relation between the website and its visitors by consulting the knowledge based system. (Qwaider 2011), Velasquez & Palade (2007) and Kumar & Rathee
(2011). The developed method was to extract knowledge from distributed networks from heterogeneous life science databases. Quantitative approach based on Web 2.0 technologies were presented here. And Knowledge Management was a logical progression within the knowledge hierarchy framework (Qwaider 2011).

1.1.3 Knowledge Discovery and Retrieval

Knowledge Discovery in Databases (KDD) is the nontrivial extraction of implicit, antecedently unknown and probably helpful knowledge from knowledge in Databases. Knowledge processing is the extraction of hidden prophetic knowledge from massive knowledgebases. It is a strong technology with nice potential to assist organizations concentrate on the foremost necessary knowledge in their knowledge warehouses (Neelamadhab et al 2012). Matthias et al (2007) have developed a method for extraction of knowledge networks from distributed, heterogeneous life science databases. Knowledge mining offer an efficient approach for knowledge retrieval at semantic-level. Basic issues of knowledge retrieval systems are examined and a conceptual framework of such systems is proposed in (Yao et al 2007). Knowledge mining is the process of effort hidden, unknown, and probably helpful information and knowledge from an outsized range of knowledge (Tiple, & Dhande 2012).

Knowledge processing poses several challenges to the analysis community. The most challenges in knowledge processing are given below:

- Knowledge may be placed at totally different sites and the quantity of knowledge might exceed the T limit.
- Knowledge processing computationally intensive processes involving very massive knowledge sets. Usually, it is necessary to partition and distribute the information for multiprocessoring and to attain acceptable time and house performance.
In several applications domain knowledge that is to be deep-mined is either created with high rate or is available in streams. In those cases, knowledge must be deep-mined quickly and with efficiency to be usable and updated.

- Security can be a major concern for firms or alternative organizations as unfair knowledge processing can result in supplying no useful knowledge (Rao 2010).

**Knowledge mining:** Knowledge mining is the core step, which ends with the discovery of hidden and helpful knowledge from large databases (Srinivas et al 2010). The vital reason that attracted a good deal of attention in knowledge technology is the creation of helpful knowledge from massive collections of knowledge. With the big quantity of knowledge kept in files, databases, and alternative repositories, it's more and more vital, to develop powerful tool for analysis and interpretation of such knowledge and for the extraction of attention-grabbing knowledge that would facilitate in decision-making (Neelamadhab et al 2012). A proper definition of knowledge discovery in databases is given as follows: “Knowledge mining is that the non trivial extraction of implicit antecedently unknown and probably helpful knowledge regarding knowledge” (Srinivas et al 2010).

Conventional systems for knowledge discovery (KD) and data mining (DM) have the ability to extract valid rules from huge data sets(Horeis &sick 2007). To take complete advantage of knowledge; the information retrieval is solely not enough, it needs a tool for automatic account of knowledge, extraction of the essence of knowledge keep, and also the discovery of patterns in knowledge. There is a large volume of knowledge and the aim is to derive helpful knowledge so as to help in improving business (Neelamadhab et al 2012). Web knowledge mining framework attempts to determine useful knowledge from derived data, complex format, and high dimensional data
obtained from the interactions of the users through the Web (Madasamy 2012).

Knowledge mining is the method of automatic classification of cases supported knowledge patterns obtained from a dataset. Variety of algorithms are developed and enforced to extract knowledge and find out knowledge patterns, which will be helpful for call support. Knowledge processing, popularly referred to as Knowledge Discovery in Databases (KDD), refers to the nontrivial extraction of implicit, antecedently unknown and probably helpful knowledge from knowledge in databases. Many knowledge processing techniques are square measure pattern recognition, clustering, association, classification and clump. The projected work can specialize in challenges associated with integration of clump and classification techniques. Classification has been known as a very important drawback within the rising field of knowledge mining. Given our goal of classifying massive knowledge sets, we have a tendency to focus principally on call tree classifiers. Call tree classifiers square measure quicker as compared to alternative classification strategies. Choice trees often regenerate into easy and straightforward manner to know classification rules. Finally, tree classifiers obtain similar and typically higher accuracy when put next with alternative classification strategies. Clump is that the unattended classification of patterns into clusters. The community of users has competed heap stress on developing quick algorithms for clump massive datasets. It teams similar objects along in a very cluster and dissimilar objects in alternative cluster (Kumar & Rathee 2011).

**Rule mining**: Knowledge mining is a locality of current analysis and development in engineering science, that is attracting a lot of attention from a good variety of individuals. It aims to extract numerous sorts (models) of hidden, attention-grabbing, antecedently unknown and doubtless helpful knowledge (i.e. rules, patterns, regularities, customs, trends, etc.) from databases, wherever the degrees of collected knowledge are often measured in
gigabytes. Association Rule Mining (ARM) may be a well-known knowledge processing analysis (Yanbo & Wang 2008).

Association rule mining is one in every of the foremost necessary and well researched techniques of knowledge mining. It aims to extract attention-grabbing correlations, frequent patterns, associations or casual structures among sets of things within the dealings knowledge bases or different knowledge repositories. Association rules are wide employed in numerous areas like telecommunication networks, market and risk management, internal control etc. numerous association mining techniques and algorithms are going to be in short introduced and compared later. Association rule mining is to search out association rules that satisfy the predefined minimum support and confidence from given knowledge. The matter is conventionally rotten into 2 sub issues. One is to search out those item sets, whose occurrences exceed a predefined threshold within the knowledgebase; those item sets are referred to as frequent or massive item sets. The second problem is to come up with association rules from those massive item sets with the constraints of lowest confidence. By checking the boldness this rule is often determined as attention-grabbing or not. Then different rules are generated by deleting the last things within the antecedent and inserting it to the resultant. Additionally, the confidences of the new rules are checked to see their interest. Those processes iterated till the antecedent becomes empty. Since the second sub problem is sort of easy, most of the researches concentrate on the primary sub problem.

The primary sub-problem is often additionally divided into 2 sub-problems: candidate massive item sets generation method and frequent item sets generation method. To decide those item sets whose support exceed the support threshold as massive or frequent item GESTS. Those items set that are expected or have the hope to be massive or frequent are referred to as candidate item sets. In several cases, the algorithms generate a very sizable amount of association
rules, usually in thousands or maybe millions. Further, the association rules are typically terribly massive. It is nearly not possible for the top users to understand or validate such sizable amount of complicated association rules, thereby limiting the quality of the knowledge mining results. many methods are planned to scale back the quantity of association rules, like generating solely “interesting” rules, generating solely “no redundant” rules, or generating solely those rules satisfying sure different criteria like coverage, leverage, raise or strength (Kotsiantis & Kanellopoulos 2006).

Now-a-days, holding previous customers is mostly used attracting new customers. Business organizations are adopting completely different methods to facilitate their customers in a variety of ways that, in order to make these customers stick to it. Association Rule Mining (ARM) and knowledge agglomeration may be a explicit reasonably knowledge processing problem for giant set of three-dimensional knowledge points. ARM has a tendency to hunt for relationship among different things within the dataset whereas the knowledge spared is typically not uniformly occupied so it can turn out into different clusters. Knowledge agglomeration identifies the distributed and also the jammed places, and hence discovers the general distribution patterns of the dataset. Association Rule Mining (ARM) (Neelamadhab et al 2012) is one in many of the methods that have two- fold benefits to the business after applying the market analysis. These are: 1) It helps customers to urge all the connected things from one place which saves their time from visiting all parts of the shop. 2) It helps in having better organization by having items sold along in the near place (Khattak et al 2010).

In data mining based knowledge extraction, the two techniques such as neural networks and decision tree are considered as good choice for non-experts. Decision Tree classification algorithm can be implemented in a serial
or parallel fashion based on the volume of data, memory space available on the computer resource and scalability of the algorithm (Anyanwu & Shiva 2009). Importantly a systematic methodology using data mining and knowledge management techniques are proposed in Shaw et al (2001). Baqueiro et al (2009) introduced an integration study which combines Data Mining (DM) and Agent Based Modeling and Simulation (ABMS). However, knowledge retrieval has been realized by various approaches and the knowledge retrieval level and efficiency has improved significantly Tiple & Dhande (2012). More specifically the application of data mining techniques in knowledge management has been reviewed in Silwattananusarn & Tuamsuk (2012). And discussed about the application of data mining techniques in Knowledge management based on the four types which were, (i) knowledge resource, (ii) knowledge types, (iii) data mining tasks and (iv) data mining techniques.

From Tiple & Dhande (2012), SivaKumar et al (2011) and Shaw et al (2001) the deeper need for the use of data mining and knowledge management for marketing decision support was understood. It has provided methods which generated unexpected patterns with respect to managerial intuition by evoking managers. The use of data mining for decision support highlighted the unique and interesting issues in customer relationship management for real-time interactive marketing, customer profiling and cross-organizational management of knowledge. To discuss the application of data mining techniques in knowledge management a query based summarizer has been proposed in SivaKumar et al (2011). More Specifically the journals from 2007 to 2012 has analyzed and classified to discuss the applications of data mining techniques in knowledge management Silwattananusam & Tuamsuk (2012). All the knowledge discovery process are referred and applied in many data mining techniques like outlier analysis, prediction, classification, association mining and clustering. In most cases data mining technology has integrated and
suggested knowledge retrieval as a new research field. Therefore most new-generation database applications demand intelligent information management to enhance efficient interactions between database and the users Kohail & El-Halees (2011), Nihalani et al (2009). A research has been presented in Ngai et al (2009) on the application of data mining and knowledge retrieval model has been proposed here to combine knowledge search with data mining. Meteorological data mining was a form of data mining concerned with finding hidden patterns inside largely available meteorological data in which the information retrieved could be transformed into usable knowledge, Yima et al (2004). Generally data mining was the main part of knowledge management. Here the data mining technology was integrated for the purpose of knowledge retrieval Tiple & Dhande (2012), Silwattananusarn & Tuamsuk (2012).

**Clustering:** Clustering is an unsupervised learning process, for grouping data into subgroups (Kavitha & Punithavalli 2010). Clustering text data streams is an important issue in data mining community and has a number of applications such as news group filtering, text crawling, document organization and topic detection and tracing etc (Liu et al 2008).

**Neural network:** A neural network model is a learning system made up of simple units configured in a highly interconnected network. Neural networks are normally classified as one of soft computing techniques (Yao 2003). Artificial neural networks (ANNs) have proven to be a powerful and general technique for machine learning. As machine learning has been applied to complex real-world problems, many researchers have found themselves turning to systems that combine knowledge from multiple sources. A standard approach is to incorporate existing knowledge about a domain into an empirical learning system, so as to produce a more accurate solution. Artificial neural networks (ANNs) have been shown to be a powerful technique for empirical learning, but until recently ANNs were largely unable to take advantage of existing problem-
specific knowledge (Maclin et al 1993).

**Self Organizing Map** (SOM): The Self-Organizing Map is a very popular artificial neural network (ANN) algorithm based on unsupervised learning. The SOM is used in various knowledge mining tasks. It provides several very beneficial properties, like vector quantization and projection (Polzlbauer 2004)

**Knowledge Retrieval:** Knowledge Retrieval (KR) is the science of finding out documents, information among documents, knowledge regarding documents, relative databases (Sivakumar et al 2011). Knowledge Retrieval Systems (KRS) is that the next generation retrieval systems for supporting knowledge discovery, organization, storage, and retrieval. Such systems are utilized by advanced and knowledgeable users to tackle the difficult problem of knowledge seeking (Tiple & Dhande 2012). Knowledge retrieval (KR) focuses on the knowledge level and the way to extract. Knowledge retrieval systems offer knowledge to users in an exceedingly structured means. They are completely different from knowledge retrieval systems and knowledge retrieval systems in illation models, retrieval ways, result organization, etc. The core of knowledge retrieval is unit retrieval subsystems. Knowledge retrieval gets results through Boolean match. Info retrieval uses partial match and best match.

Knowledge retrieval is additionally supported by partial match and best match. Considering illation perspective, knowledge retrieval uses deductive illation, and knowledge retrieval uses inductive illation. Considering the restrictions from the assumptions of various logics, ancient logic systems cannot build economical reasoning in an exceedingly affordable time. Associative reasoning, figurative reasoning and therefore the plan of unifying reasoning and search could also be effective ways of reasoning at the net scale. From retrieval model perspective, knowledge retrieval systems specialize in the linguistics and knowledge organization. Knowledge retrieval organizes the
information and documents by assortment, whereas knowledge retrieval organizes knowledge by connections among knowledge units and therefore the knowledge structures (Tiple & Dhande 2012).

An analysis was developed for the types of errors generated by human expert in building a Knowledge Based System (KBS). (Shih & Tseng 2009) constructed a knowledge-based system to expand query keywords based on the derived strategies and then selected relevant keywords according to geographical distance between entities of concept and learners. A structured method was selected using a questionnaire as a knowledge acquisition tool by the presenting a step by step knowledge acquisition process. A process based on decision tree approaches was presented for automated building of knowledge based on soil resources mapping. It showed that the process was moderately successful in extracting knowledge from the existing soil maps and other GIS data related to soil development. Automated Soil mapping using Thematic-Mappers, bi-temporal images and GIS data was easily built based on decision tree algorithm. Cao & Compton (2005), Amin & Khan (2009) and Zhou et al (2004). A novel algorithm was suggested to change customers from an undesired status to a desired one while maximizing objective function. Intelligent and efficient Customer Retention Management (CRM) was created because of the distributed information provided by the most data mining technique on customer profile Yang et al (2003). Complete knowledge acquisition processes have been adapted through structured method by the use of questionnaire as a knowledge extraction tool. Using two data mining algorithm churn prediction model was built by the credit card data collected from a real Chinese bank. This phenomenon was called customer “churning” or “attrition,” which was a major problem for these companies and makes it hard for them to stay profitable Amin & Khan (2009), Swati & Kulkarni (2011). The usefulness of this Soil Ecological Knowledge (SEK) for restoration was best
considered in the context of the severity of the original perturbation. Importantly a specific method for sampling pixels based on modes of environmental histograms have proved to be effective in terms of reducing noise and constructing representative sample sets. And found a greedy heuristic algorithm to solve both problems efficiently and presented an ensemble-based decision-tree algorithm that use a collection of decision trees, rather than a single tree, to generate the actions. Because of these methods it was easy to build knowledge bases for automated soil mapping than using the conventional knowledge acquisition approach and evaluated the performance of the resultant knowledge bases Heneghan et al (2008) Swati & Kulkarni (2011).

1.1.4 Knowledge-Based Decision Making

Decision support system is created to help people make decisions by providing access to information and analysis tools. It is a way to model data and make quality decisions based upon it (Martinsons & Davison 2007). Decision support systems (DSS) constitute a class of computer-based information systems including, knowledge-based systems that support decision making activities. Making the right decision is usually based on the knowledge and the ability to sift through and analyze the data to find trends that can create solutions and strategies.

DSS are interactive computer based systems and subsystems intended to help decision makers use communications technologies, data, documents, knowledge and models to complete decision process tasks. Development of innovating algorithms for access and maintenance of the data and procedures; techniques and tools to carry out data warehouse and data mining, offering a good fulfillment as for time of answer and extensive capacity of backup to requirements such as operational as strategic Decision Support Systems(Ruiz et al 2003).
A decision support system may present information graphically and may include an expert system or artificial intelligence (Ben-Zvi 2010). An expert system is a knowledge-based system to be used instead of or together with a human operator to make decisions in the framework of a professional task with explanations for users (Yang et al 2003). DSS is considered as one of the most popular areas in information systems during the time period. Diverse DSS are developed to support decision makers at all levels in an organization including systems that can support problem structuring, operations, financial management and strategic decision making, even extending to support for optimization and simulation (Vohra & Das 2011).

The intelligent management of parts and tools in flexible manufacturing systems for knowledge based decision support system was proposed here. Much knowledge management focused on identifying, storing, and disseminating process related knowledge in an organized manner have illustrated the activities regarding admissions in the higher institutes where decision support systems were required for taking the admissions Yima et al (2004), Ozbayrak & Bell (2003). The students profile can be taken by the specific data mining techniques like clustering and identified the promising groups for target. It has been widely demonstrated that business simulation games were an effective way to engage students in Decision Support Systems (DSS). Intelligent Decision Support Systems (IDSS) technologies were suited to provide decision support in the higher education environments because of the generated and presented relevant information and knowledge. Importantly it created a need to know how managers in different parts of the world make decisions, and how computer based Information Systems (IS) could support decision making Vohra & Das (2011), Ben-Zvi (2010). Information system (IS) issues were examined to found out how computer based decision support was working in different parts of the world. For flexible manufacturing system, Knowledge Based Decision Support

1.2 KNOWLEDGE SYSTEM TO SOIL DATABASE

Edaphology is a domain that is bothered with the influence of soils on living things, notably plants. The term is additionally applied to the study of soil influences. An agricultural soil science explores soil’s physical and chemical properties to search out the plants acceptable for cultivation (Yima et al 2004), (Denning 2002). The most common pitfall information is the lack of standardization of the nomenclature and of the data acquisition procedures. The user is not required to know fully the model to interact with the system. The retrieval of a large amount of the same type of data is very efficient, even though the user need not know completely the database schema to formulate the queries. Here the SOM is used in arriving at decisions on what kind of plants can be that grown in soil, based on the domain information given by the user. This effectively supports the decision making problems (Intan & Mukaidono 2002).

Soil ecology blends knowledge of physical, chemical, and biological processes and properties to higher perceive and manage ecosystems, communities, and species’ functions and interactions. The interaction between surface and below ground structure holds nice connectedness to the upkeep of native variety and performance. Here, the state of those interactions and commitment to establish areas wherever larger appreciation for the parallel interests of the two disciplines might advance each restoration successes and soil ecological knowledge (Callaham et al 2008).
Restoration ecologists have long recognized the integral role of soil, notably in its physical and chemical aspects, within the prospering revegetation of degraded sites. However, specific incorporation of Soil Ecological Knowledge (SEK), that acknowledges interactions among the principal elements of the soil system, furthermore as feedback between the surface and belowground system processes, into restoration remains in an exceedingly comparatively early stage of development. Despite early, it makes an attempt to demonstrate the importance of a soil’s perspective for restoration efforts, a recent and helpful review of analysis on restoration ecology makes solely scattered references to soil processes and aggregation. Printed restoration science within the primary literature ordinarily includes soil info related to pre-restoration website assessment and therefore the analysis of specific soil amendments. Recovery of nutrient capital or biogeochemical processes additionally motivates restoration activities, however examples wherever integrated SEK has been used area unit uncommon (Heneghan et al 2008).

Soil database: The database considered for the research consists of 49 different types of plant species and for each plant species the geological taxonomy properties were represented in this database. The 49 different types of plant species represented in this database were Prosopis juliflora, Cyprus sp, Hariyali, Indigo plant, Palmyra, Eucalyptus, Neem, Tamarind, Acacia, Ipomoea, Thespesia populnea, Vagai, Eucalyptus, Vetiver, Banyan, Calotropis, Lantana camera, Ammania baccadera, Astracantha long folia, Cyanodon, Thespesia, Cassia, Acacia Arabica, Grasses, Wetland weeds, Croton sparciiflorus, Gomphrena spp, Pongamia glabra, Tricalli, Jatropha, Bonassus, Agave, Ferns, Rich evergreen shrubs and bushes, Cactus, Tephrosia purpurea, Pungai, Manjanathi, Cassia auriculata, Cyprus rotantus, Dates, Prosophis, Delonix regia, Karuvalam, Argimone, Calotropis, Lantana, Pungam, Sandal. The geological properties suitable for 49 different plant species were represented in the database. They were clay, Granite, Laternite, sand, Western
Ghats, Eastern Ghats. Typically, the taxonomy properties of each plant species were represented here such as Fine, montmorillonitic, isohyperthermic, noncalcareous, Chromic Haplusterts, Typic Rhodustalfs, Mixed, Typic Haplustepts, Clayey, Lithic Haplustepts, Clayey-skeletal, Loamy, Lithic Ustorthents, Typic Ustipsamments, Vertic Haplustepts and montmorillonitic. The various colors of the soils were represented here in this database such as grayish brown, Red, Very dark grayish brown, Yellowish brown, Brownish yellow, Dark yellowish brown, Dark Brown, Reddish brown, Dark reddish brown, Brown, Pale Brown, Light Brownish Grey, Light gray to gray, Grayish brown, Very pale brown, White, Light brownish gray, Dark red, Yellowish red, Light red, Dark grayish brown, Brown to dark brown, Reddish yellow, Brown, very dark brown, strong brown. In this database, the size of the particles that make up the soil such as clay, silt and sand are listed for each color of soil separately. For each color, the depth and the mineral content of soil such as PH, EC, Ca, Mg, Na, K, P2O5, and K2O are represented in this database which will provide nutrients to plants.

1.3 LITERATURE SURVEY

This section presents a detailed survey of all the approaches and techniques presented in literature to identify an important problem in the emerging field of data mining with knowledge engineering. The research conducted in this area is categorized into four categories, they are (i) Data mining and Query for knowledge extraction, (ii) Decision making and Knowledge Retrieval, (iii) Decision Tree and Soil based Knowledge Acquisition, (iv) Knowledge Representation and Management. In the research based on Data Mining and Query for knowledge extraction, a number of algorithms have been developed and implemented to extract information and discover knowledge patterns that may be useful for decision support. Next, the
research article based on Decision making and knowledge retrieval is integrated with data mining technology for the purpose of knowledge retrieval which provided a knowledge retrieval strategy and supported dynamic learning, mining and adaptability. A detailed survey has been carried out to identify the various research articles available in the literature in all the categories based on data mining and to do the analysis of the major contributions and its advantages. In all the categories of articles Data Mining and DSS are well suited technologies to provide decision support by generating and presenting the relevant information and the knowledge towards quality improvement of Information System (IS). The research article related to decision tree and soil is reviewed in third category of the survey and decision tree approaches for automated building of knowledge based on soil resources mapping is reviewed. The research article related to Knowledge Representation and Management was reviewed in fourth category. A verification mechanism is used to verify knowledge models based on the formal semantics of the knowledge representation. Application of data mining techniques in knowledge management is reviewed. To discuss the applications of data mining techniques in knowledge management the journals from 2007 to 2012 are analyzed and classified.

1.3.1 Classifications

1.3.1.1 Research articles related to Data Mining and Query for knowledge extraction
1.3.1.2 Research articles related to Decision Making and Knowledge Retrieval
1.3.1.3 Research articles related to Decision Tree and Knowledge Extraction from Soil Database
1.3.1.4 Research articles related to Knowledge Representation and Management
1.3.1.1 Research articles related to Data Mining and Query for Knowledge Extraction

Tiple & Dhande (2012), have integrated data mining technology and suggested knowledge retrieval as a new research field. The data mining technology is integrated for the purpose of knowledge retrieval which provided a knowledge retrieval strategy and supported dynamic learning, mining and adaptability. A knowledge retrieval model is proposed here to combine knowledge search with data mining. Here data mining is integrated in to retrieval procedure of query optimizing, searching, results analyzing, and resources constructing. From information sources knowledge retrieval realized deep knowledge discovery and retrieval to raise the level of knowledge retrieval and efficiency. Due to this the rapid development of research and application of data mining will also be promoted. Knowledge retrieval is understood clearly by various approaches. The knowledge retrieval level and efficiency has improved significantly.

Silwattananusarn & Tuamsuk (2012), have reviewed the application of data mining techniques in knowledge management. They have analyzed and classified the journals from 2007 to 2012 to discuss the application of data mining techniques in knowledge management. They have discussed the application of data mining techniques in Knowledge management based on the four types which are (i) knowledge resource, (ii) knowledge types, (iii) data mining tasks, (iv) data mining techniques. Generally data mining is the main part of knowledge management. Some of the assumptions have been made through this research and they are (i) research in knowledge management will increase, (ii) hybridization techniques are used to solve knowledge management problem and tends to support expert decision making. Also they have discussed about the applications of data mining techniques in knowledge management.
Based on grouped similar sentences and word frequencies, Siva Kumar et al (2011), have proposed query-based summarizer. Vector Space Model has been used to find similar sentences to the query and Sum Focus to find word frequency and achieved high Recall and Precision scores. It has allowed the user to get a sense of the content of full-text, or else to know its information content without reading all sentences within the full-text. The data reduction has helped the user to find the required information quickly without waste of time in reading the whole text. A query based document summarizer has been presented here based on similarity of sentences and word frequency. They have used AQUAINT-2 Information-Retrieval Text Research Collections and the obtained summary sentences were evaluated using ROUGE metrics and finally achieved the best average recall, precision and F-score.

Nihalani et al (2009), have proposed an intelligent layer for database which is responsible for manipulating flexible queries. The flexible queries from users are submitted to intelligent layers initially and the intelligent layer converted the flexible queries into a structured SQL query. It has enabled the design of a knowledge based self learning system based on the values obtained from user. The selection of appropriate SQL query would aid when a same flexible query is issued in future. There is an overwhelming need for non-expert users to query relational databases in their natural language. It is using linguistic variables and terms instead of working with the values of the attributes. Therefore most new-generation database applications demand intelligent information management to enhance efficient interactions between database and the users.

To increase the publication rates and interest Ngai et al(2009), have presented a research on the application of data mining. Customer Retention Management (CRM) related articles have been reviewed here which is related to programs such as one-to-one marketing and loyalty. The classification model
has been mostly applied in CRM to predict future customer behavior. To seek profitable segment the underlying characteristics of customer were analyzed by data mining techniques such as neural networks and decision trees. In data mining the two techniques such as neural networks and decision tree are considered as a good choice for non-experts. There has been a lack of aspects such as literature review and a classification scheme in spite of the importance of data mining techniques to CRM.

Shaw et al (2001), have proposed a systematic methodology using data mining and knowledge management techniques. The systematic methodology has been used to manage the marketing knowledge and to support marketing decision. Data mining techniques have been used in several areas such as fraud detection, bankruptcy prediction, medical diagnosis, and scientific discoveries. But the use of data mining for marketing decision support which highlights unique and interesting issues in customer relationship management for real-time interactive marketing, customer profiling and cross-organizational management of knowledge. They understand the deeper need for the use of data mining and knowledge management for marketing decision support.

Padmanabhan & Tuzhilin (2002), have provided methods which generate unexpected patterns with respect to managerial intuition by evoking managers. Here a combined statement has been discovered by the unexpected pattern to resolve problems in domain knowledge. They have addressed the problem to include the combined rules. They have presented a framework for refinement based on a generic knowledge refinement strategy which describes abstract properties of refinement algorithms. Specific instantiations was compared using this refinement algorithm. Also, it described and compared two specific refinement algorithms based on this framework.

To extract knowledge from Gaza city weather dataset Kohail & El-Halees (2011), have applied knowledge discovery process. They have referred
all the knowledge discovery process and applied many data mining techniques like outlier analysis, prediction, classification, association mining and clustering. Meteorological data mining was a form of data mining concerned with finding hidden patterns inside largely available meteorological data in which the information retrieved could be transformed into usable knowledge. Here they tried to extract useful knowledge from weather daily historical data collected locally at Gaza Strip city.

1.3.1.2 Research articles related to Decision Making and Knowledge Retrieval

Yima et al (2004), have recognized knowledge as a new resource in gaining organizational competitiveness. Here, knowledge management suggested a method in managing and applying knowledge for improving organizational performance. The organized knowledge models have enabled decision-makers to understand the structure of the target problem. The basic cause of the above problem was identified which facilitates effective decision-making. The applied data to decision making contains a significant impact on organizational performance than entirely process transactions for data management. The proposed method transforms individual mental models into explicit knowledge by translating partial and implicit knowledge into an integrated knowledge model. The proposed method achieved the strategic goals and objectives of an organization by the facilitated linkage between knowledge management initiatives.

Vohra & Das (2011), have illustrated the activities regarding admissions in the higher institutes where decision support systems are required for taking the admissions. Intelligent Decision Support Systems (IDSS) technologies were suited to provide decision support in the higher education environments because they generated and presented relevant information and knowledge. Decision
Support System consists of a modern approach to the decision making processes in the proposed architecture. In higher education the proposed method provided conceptual frameworks which provide the required decision support. Here the students’ profiles are taken by specific data mining techniques and promising groups are identified for target. The ERP (Enterprise Resource Planning) with IDSS has provided good support for decision making otherwise the limitation of ERP will reduce the effectiveness of the system in decision making.

Ben-Zvi (2010), has suggested that business simulation games is an effective way to engage students in Decision Support Systems (DSS) because they believe that by teaching decision support system using simulation games have important advantage over other approaches. The feedback from task performance is continuously provided by the DSS. In Information Systems (IS) they discussed business simulation games as teaching tools. IS concept is implemented by the spreading of foundational knowledge in traditional teaching method which does not provide optimal platform for students. The performance of the game is improved to allow the students to experience the need of relevant information with little corporate and personal risk. By business games, an effective alternative to traditional teaching methods were provided. The students are more excited, motivated and become actively involved in the decision-making process in the game.

Martinsons & Davison (2007), have examined the IS issues originated from the discovery of the distinctively American, Japanese and Chinese styles of strategic decision making. The global applicability of IS raised doubt for analyzing and conceptualizing strategic decision. It creates a need to know how managers in different parts of the world make decisions, and how decision making system could be supported by computer based information systems (IS). Business leaders from the United States, Japan and China are found to have a
distinctive prevailing decision style that reflected differences in cultural values and the relative needs for achievement, affiliation, power and information.

Ozbayrak & Bell (2003), have developed a Knowledge-Based Decision Support System (KBDSS). It is for short-term scheduling in Flexible Manufacturing System (FMS) and strongly influenced by the tool management concept. It provides a significant operational control tool for a wide range of machining cells in which a high level of flexibility is demanded. It has the benefits of more efficient cell utilization, greater tool flow control, and a dependable way of rapidly adjusting short-term production requirements. The intelligent management of parts and tools in flexible manufacturing systems for KBDSS is proposed here. KBDSS has provided powerful manufacturing system diagnosis as well as a best tool issue strategy selection mechanism for different cell hardware structure. KBDSS has provided satisfactory results for offline evaluation and opened a promising new direction for applying KBDSS for online dispatching, diagnosing, and strategy selection in an FMS environment.

1.3.1.3 Research articles related to Decision Tree and Knowledge Extraction from Soil Database

Zhou et al (2004), have presented a process based on decision tree approaches for automated building of knowledge based on soil resources mapping. The knowledge bases built by decision tree and Bayesian predictive modeling methods were used by the knowledge classifier for soil type classification of the Longyou area, Zhejiang Province, China using TM (Thematic Mapper) bi-temporal images and GIS data. Due to these methods it was easy to build knowledge bases for automated soil mapping than using conventional knowledge acquisition approach and evaluated the performance of the resultant knowledge bases. Based on field survey the classification results were compared with existing soil map. And showed that the process was
moderately successful in extracting knowledge from the existing soil maps and other GIS data related to soil development.

Yang et al (2003), have presented novel algorithms which suggested actions to change customers from an undesired status to a desired one while maximizing objective function. The technique was applied to industrial problems such as Customer Relationship Management (CRM). CRM was used to change the customers from undesired status to a desired one. But they required human experts to post process the mined information manually. Most of the post processing techniques have been limited to producing visualization tools and interesting ranking. But they don’t directly suggest that action that would lead to increase the objective function such as profit. Most data mining algorithms and tools stop at discovered customer models which produced distribution information on customer profiles. It provided effective solutions to intelligent CRM for enterprises.

Yahia et al (2012), have proposed a formal description of mixed and multimodal decision making in which decision may be made by three possible modes. The three possible modes are individual, collective and hybrid. The proposed system was to make correct decision, individuals, teams and organizations which need both knowledge management and collaboration to make it more effective and efficient. The MDM process have combined the use of proposed model with UML-G profile and elaborated the activity diagram in order to state clearly by step-by-step.

Patil et al (2012), have presented novel algorithms that suggested action to change customers from an undesired status to a desired one when the expected net profit was maximized. The novel algorithm technique applied to industrial problem such as customer relationship management (CRM) were useful in pointing out customers who are likely attractors’ and customers who
are loyal. But to post process the discovered knowledge manually, they require human experts. Most data mining algorithms and tools have stopped discovering customer models which produced distribution information on customer profiles.

Elkaffas et al (2011), have categorized water resources according to the international water quality criteria. From the application of expert system they re-classified the water resource according to a developed decision matrix. It integrated more than one research field and tools which include Decision Support Systems (DSS), Geographic Information Systems (GIS), knowledge base and modeling and visualization enhancement. Here, Knowledge Based System (KBS) for water resources categorizing was implemented and tested for the groundwater analysis. The developed system helped the decision makers to select the suitable water resources for cultivating a specific crop.

Callaham et al (2008), have observed an absence of in-depth discussion of how soils and particularly the ecology of soils. It can be integrated into developed theory of restoration science. Although soils were universally regarded as critical to restoration success, much research has included manipulations of soil variables. They found that better integration of soil ecological principles could still contribute much to the practice of ecosystem restoration.

Qi & Zhu (2003), have developed a knowledge discovery procedure for extracting knowledge of soil-landscape models from a soil map. It consists of four major steps; they are data preparation, data preprocessing, pattern extraction, and knowledge consolidation. Their study pays specific attention to the reduction of representation noise in soil maps in order to recover true expert knowledge from the error-prone soil maps. Decision tree algorithm, Naive Bayes algorithm and artificial neural network were the three algorithm
investigated for a comparison concerning learning accuracy and result comprehensibility. A specific method for sampling pixels based on modes of environmental histograms have proved to be effective in terms of reducing noise and constructing representative sample sets.

Swati and Kulkarni (2011), have found a greedy heuristic algorithm to solve both problems efficiently and presented an ensemble-based decision-tree algorithm that use a collection of decision trees, rather than a single tree, to generate the actions. Have taken these results as an input and produced a set of actions that can be applied to transform customers from undesirable classes to desirable ones by the presented novel technique. They predicted a valuable customer who will be an attritor as loyal. Attrition is a phenomenon where there are an increasing number of customers who are switching from one service provider to another. The cost was usually higher than the case when they classify a loyal customer as an attritor. With massive industry deregulation across the world, each customer was facing an ever-growing number of choices in telecommunications and financial services. In addition, a CRM dataset was often unbalanced. The most valuable customer who actually churns can only be a small fraction of the customers who stay. This phenomenon was called customer “churning” or “attrition,” which was a major problem for these companies and makes it hard for them to stay profitable.

Heneghan et al (2008), have proposed that the usefulness of the Soil Ecological Knowledge (SEK) for restoration was best considered in the context of the severity of the original perturbation. They emphasized both on soil organism and ecosystem processes, which were the discipline of soil ecology. Restoration outcome was improved by a body of knowledge. A straightforward manipulation of single physical, chemical, or biological components of the soil system can be useful in the restoration. The aim of the restoration was to overcome constraints on ecosystem recovery through natural processes. It was
to produce resilient ecosystems that were resistant to invasion, capture and use resources efficiently, which contain biological complexity needed to function effectively and provide human-valued services. The results were anticipated and lead to success restoration.

Eldin et al (2012), have introduced Query Expansion (QE) mechanism involved a new expansion process. It was guided by conceptual representation approach using concept mapping tool for expanding the original query in the context of the utilized corpus. The use of concept mapping for a range of IR tasks was presented. Also in particular, their use in the area of query expansion focused on clarifying the effect of linguistic and domain based approach in the retrieved results. A possible direction for further research was applied to the proposed solution in other Information retrieval models. And the effectiveness was compared with other type of QE term selection methods.

Cao & Compton (2005), have addressed the problem to evaluate knowledge acquisition method. Because of the costs of using human expertise in experimental studies evaluation of knowledge acquisition technique was difficult. In building knowledge based system, knowledge acquisition played an important role. Here they addressed the problem for evaluating knowledge acquisition methods. An approach was demonstrated by the evaluated three variants of a knowledge acquisition methodology. The evaluated knowledge acquisition methodology was Ripple Down Rules (RDR). And an analysis was developed for the types of errors generated by human expert in building a Knowledge Based System (KBS).

Amin & Khan (2009), have chosen a structured method using a questionnaire as a knowledge acquisition tool by the presented step by step knowledge acquisition process. The important problem here was the efficient knowledge acquired for the specific problem and also from human experts. The
other was to encode the efficiently occurred problem knowledge to suitable computer format. To develop expert systems, knowledge acquisition from human expert was a major problem. Here the problem domain was represented to evaluate the teacher’s performance in higher education using expert system technology. Complete knowledge acquisition processes have been adapted through structured method by use of questionnaire as a knowledge extraction tool.

Yua et al (2010), have demonstrated the use of decision tree method which classified and predicted building energy demand levels accurately. Significant factors of building Energy Use Intensity (EUI) was identified automatically using decision tree method. When compared to the other the predicting and classification of the categorical value was considered as important advantage using decision tree. It was compared with widely used modelling techniques such as regression method and ANN method. It enabled the user to extract information quickly which lies in the ability to generate accurate predictive models with interpretable flowchart-like tree structures. It was applied to estimate residential building energy performance indexes by modeling building EUI levels. It reported the development of a building energy demand predictive model based on the decision tree method.

Nie et al (2011), have applied two data mining algorithm to built churn prediction model by the credit card data collected from a real Chinese bank. It was not to provide a new data mining algorithm but they focused on the application of churn prediction. In addition to the accuracy of analytic results they designed a misclassification cost measurement by taking the two types of error and the economic sense into account. It was more suitable to evaluate the credit card churn prediction model. It provided a frame work to understand the knowledge of the card holders’ hidden pattern using the data of Chinese banks.
1.3.1.4 Research articles related To Knowledge Representation and Management

Lai (2007) has proposed a knowledge engineering approach to manage knowledge and the approach was called Knowledge Management through Knowledge Engineering (KMKE). Various types of knowledge was organized and expressed in a unified knowledge representation by using KMKE approach. Then a verification mechanism was used to verify knowledge models based on the formal semantics of the knowledge representation. KM tools was to facilitate the implementation of KM systems and the process of generating, structuring, and sharing knowledge. They combined knowledge engineering with knowledge management to solve the major problems in most KM systems encounter. With respect to the user needs they updated the knowledge to modify knowledge storage. They designed a Knowledge Query Language to enhance the dissemination of knowledge.

Velasquez and Palade (2007), have introduced a Knowledge Base, which consists of a database-type repository for maintaining the patterns, and rules, an independent program that consults the pattern repository. The proposed system was used to improve the relation between the website and its visitors by consulting the knowledge based system. From data originated in website they extracted significant pattern about the visitor’s behavior by the applied web mining tools. The discovered knowledge may become obsolete in a short period of time due to frequent change in the visitor’s interests and as well as in the website itself. The proposed method showed the effectiveness of their approach by testing it with data from a Chilean virtual bank.
Kumar & Rathee (2011), have used a large data base called ‘Fisher’s Iris Dataset’ which contains 5 attributes and 150 instances. They compared the result of classification technique with the result of clustering and classification technique integration by using J48 classifier. It was based on a data mining tool called WEKA. The result provided greatest accuracy rate and robustness even if the data set contained missing values.

Kebede (2010), has advanced knowledge management by showing that KM was a natural and long awaited development in Information System (IS). The main purpose was to call on the members of the IS profession and to take a more proactive and visible role in advanced KM. KM was a natural and long-awaited development in IS and that a number of circumstances have made KM to be an area of emphasis in IS. And Knowledge Management was a logical progression within the knowledge hierarchy framework. The progression was best described as evolutionary and a number of factors have contributed at different times accelerated the progression to take place at this point in time.

Qwaider (2011), has investigated the integration of e-Learning systems and knowledge management technology. It was investigated to improve the capture, organization and delivery of both traditional training courses and large amounts of corporate knowledge. They have illustrated that how knowledge management was enhanced in an organization and provided the benefits of both. Here they presented a model for integration of e-learning system and knowledge management. They conducted a review and defined the common characteristics of knowledge management and e-learning system.

Rašul and Vuksic (2012), have enhanced the organizational performance by creating, accumulating, organizing and utilizing knowledge. They tested the impact of knowledge management practices through structural equation modeling. Knowledge management was a process that transformed individual
knowledge into organizational knowledge. They concluded that three components such as information technology, organisation and knowledge played an important role in a successful adoption of KM. Through indirect influence of IT on knowledge provided better collaboration and motivation of the people in organization. A positive effect of knowledge management practice on organizational performance was confirmed through the result investigated. They argued that knowledge model presented here was a useful starting point to gain a deeper insight into KM elements and their influence to the organizational performance.

Mollahosseini and Barkhordar (2010), have defined knowledge management and supplier knowledge management. The presented conceptual frame work enhanced the supplier knowledge management. Firms to cost switching and raise barriers to entry were created by the supplier knowledge management and applied it to cut production costs, innovating new products and to attract customers and suppliers. Empirical study was conducted to investigate the relation between knowledge management strategies and organizational performance and understood the importance of knowledge management. Hence, knowledge management supports business management to find, choose and keep reasonable supplier network in the best manner in which the firm maintained its competitive advantages.

Using different knowledge-based methods, ThangaRamya and Rangarajan (2011) have addressed the specific aspect of inferring semantics automatically from raw video data. Here, they focused on three technique which were rules, Hidden Markov Models (HMMs), and Dynamic Bayesian Networks (DBNs). The query of publicly available efficient video data was very important. The content based retrieval of video data became challenge and important problem. The flexibility of the integrated system within the content based retrieval system was improved. Tanwar et al, (2010) have presented a
comparison between three representation schemes which were predicate logic, semantic net and frames. Artificial Intelligence was to solve the problem in which user required a knowledge base. It contains information related to problem domain and a method for manipulating the knowledge for finding the solution. In AI various intelligence scheme was represented and each have advantages and disadvantages.

Akhavan et al (2010), have determined several important weakness factors which affected the implication of knowledge management. Organizational strategy, information over crowds, content management, portals project management were some of the weakness factors of KM. To perform seventeen factors in to six issues they performed a factor analysis. Financial and information security, Technology and management, senior management support and strategy, Acceptance, User's motivation and culture, Project management, Change management and training were the six major problems found in KM. Here, they identified the challenging factors of the implementation of knowledge management portals for Iranian organizations. They said that it was harder to gather information from limited number of organization because the knowledge management was a brand new concept of Iranian organization.

Debeljaka et al (2012), have used the methodology of relational data mining which can take into account several coexistence measures at the same time. It predicted gene flow from Genetically Modified (GM) to non-GM maize fields under multi-field crop management practices at a local scale. The feasibility of coexistence between GM and non-GM crops was explored by basic empirical and modeling studies. It focused on pair-based interactions between fields while multi-field studies build upon them. Because of this it attempted to consider the complexity of gene flow under crop management practices. A set of preventive coexistence measures was applied to ensure coexistence with the production of conventional maize which satisfies the 0.9%
EU threshold. It was for adventitious presence of authorized genetically modified (GM) material in conventional (non-GM) maize crops.

Lange et al (2007), have computed networks of protein knowledge which interconnected metabolic, disease, enzyme and gene function data to motivate possible application. Here, they presented a method for knowledge extraction in life science databases. It prevented the scientists from screen scraping and web clicking approaches. Among measured experimental facts and the combined biological knowledge networks generated computed networks enabled a holistic relationship. Interpretation of measured molecular facts was supported by the considered biological context of the system. The developed method was to extract knowledge from distributed networks from heterogeneous life science databases.

Ras and Rech (2009), have presented an approach based on Web 2.0 technologies. It was used in order to improve the knowledge acquisition during experience reuse. Practical knowledge was built in higher education and implemented experiential learning in higher education for student provided by reuse experience. The approach was evaluated by the conducted control experiment and showed an important improvement for knowledge acquisition. They assumed that the results could be generalized for German computer science students of the Net Generation. It provided best solution for collaboratively capturing, organizing, and distributing emergent knowledge.

1.3.2 Summary

A detailed survey has been performed for all the works of data mining available in the literature. The collected articles are divided into four major categories based on data mining methods and reviewed about the application in data mining techniques. The review analyzed the use of data mining for marketing decision support which highlights unique and interesting issues in
customer relationship management for real-time interactive marketing. Application of data mining techniques in Knowledge management was discussed based on the four types which were (i) knowledge resource, (ii) knowledge types, (iii) data mining tasks and (iv) data mining. The use of data mining for marketing decision support system highlighted unique and interesting issues in customer relationship management. The new research field called knowledge retrieval was suggested here for knowledge search with data mining. Data mining was the main part of knowledge management. The knowledge management was combined with knowledge engineering to solve major problems in knowledge management system. The knowledge bases built by decision tree and Bayesian predictive modeling methods were used by the knowledge classifier for soil type classification of the Longyou area, Zhejiang Province, China using TM (Thematic Mapper) bi-temporal images and GIS data. Generally data mining was the main part of knowledge management.. The data mining technology was integrated for the purpose of knowledge retrieval which provided a knowledge retrieval strategy and supported dynamic learning, mining and adaptability. Integrated data mining technology and suggested knowledge retrieval as a new research field and Evaluation of KA tools and methodologies remains problematic.

1.4 MOTIVATION FOR THE RESEARCH

Knowledge management refers to transforming the available knowledge into useful information. When knowledge management is applied to Edaphology domain, it would lead having higher productivity and yield. Here, the data is collected form Edaphologists, soil and plant researchers, domain experts and this data is transformed to useful information. This useful information will have huge hand in increasing the plant yield. Edaphology consist of plant details and its related soil features. For having useful information, proper knowledge representation and efficient retrieval algorithms
have to be there. By making the system automated, we can also evade human errors and inconsistency.

The plants demand varying quantities of diverse nutrients at different stages of growth, the preservation of fertility at the appropriate level in the soil and the selection of suitable vegetation type for the soil are especially vital for cropping. Therefore, in taking care of plants the knowledge of deficiency/excess of the nutrients in the soil is very significant. The large quantity of data and the multiple areas of expertise that are indispensable for soil exploration generate a massive volume of knowledge. These factors highlights the need for designing an efficient system to adjust, standardize, manage, retrieve and process soil information in order to attain improved productivity in agriculture. Hence, an efficient system which can find the right plants suited to the soil characteristics and knowing the right plant for the given soil characteristics can have substantial increase in crop output.

1.5 OBJECTIVE

Developing a proficient intelligent system to extract helpful information from the soil data is a puzzling and motivating problem. The helpful data can benefit the Edaphologists and agriculturists in knowing the details of the plant and its growing conditions demanded by the plant with respect to soil characteristics and features. The data can also aid researchers and the practitioners. One of most important problems faced is converting the obtained vast data into an organized structure so that the retrieval becomes easier and faster. Gathering of data and converting them to structured data can be done with help of data structures and other techniques and this would enhance the system performance. This is due to the fact that system performance and retrieval process efficiency largely depends on the structure to which the raw data is transformed.
The data size used in modern applications is increasing at an alarming rate which cause increase in dimensionality and as a result, degrading the performance of the system. During retrieval, one of the common problems faced is that the fact that data retrieval becomes tough when searching the plants in unknown vectors. In case of multidimensional database, processing time incurred for the searching operation would be very high even if the search happens in known vectors. Hence, generally information retrieval methods need huge amount of time. Hence, efforts have to be put to reduce the data size and also the time incurred for information process. This would a marked effect on the performance of the system.

The knowledge management related to Edaphology is under progress and stands in a premature phase of development. The most important problem faced in the area is the fact that there is no definite standardization with respect to the nomenclature of features and of the data gathering methods. Normally, the systems in Edaphology are modeled using information gathered from Edaphologists. This may vary and contradict each other which cause some concern in the modeling. It also faces the problem that information given by the different Edaphologists may be different. Lack of standard defined information about soil characteristics linked to the plants also is another problem.

This work is aimed to build effective intelligent systems in Edaphology domain which concerns with plants and its related soil features. By building these systems, the aim is to provide the Edaphologists and agriculturists useful information about the plants and soil features. This is of huge importance as plant growth and yield are directly dependent on the soil features.
1.6 CONTRIBUTION

Initially, the raw data about the plants and soil features linked to it collected from various sources are converted to standard data structures and subsequently, based on the user input query, the plant list is retrieved efficiently by the use of retrieval algorithms. The work employs developing different systems of knowledge storage and information retrieval. In the first system, a tree-based approach is developed which is an efficient representation of knowledge so that the information can be easily retrievable. In the second system, SOM, one of the AI techniques is incorporated into the knowledge retrieval so that more appropriate plants can be retrievable and in the third system, Incremental clustering technique is used to handle the large quantity of soil data and also, an algorithm is effectively designed for retrieval purpose. Furthermore, in the fourth system, flexible knowledge retrieval was done with the help of Fuzzy concept along with the reusable knowledge storage with XML. Finally, a Quad-tree based system is developed to reduce the amount of soil data stored in disk space and to improve the memory usage.

1.7 THESIS ORGANIZATION

This Chapter provides introduction to the knowledge system, literature survey of various research articles related to data mining, Query for knowledge extraction, decision making, knowledge retrieval, Decision tree, knowledge extraction from soil database, knowledge representation and management and also the scope and objectives of the research work,

- Chapter 2 gives the design, implementation and experimentation results of efficient tree-based system for Knowledge management in Edaphology.
• Chapter 3 gives the design, implementation and experimentation results of Knowledge management in Edaphology using Self-Organizing Maps and Localized matching model for plant prediction using Incremental clustering.

• Chapter 4 gives design, implementation and experimentation results XML and Fuzzy-based two various Knowledge retrieval methods in Edaphology.

• Chapter 5 gives design, implementation and experimentation results of KSR-QUADTREE: An Intelligent knowledge storage and retrieval using quadtree structure.

• Chapter 6 gives the performance analysis and comparison of all the systems using common input queries.

• Chapter 7 gives the Conclusion and future enhancements.