*International Journal of Engineering & Technology*, 7 (3.6) (2018) 94-96

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*

***Research paper***

# An Extensive Research on Knowledge Mining Systems:A Review

**K. Kalaiselvi[1*], J. Sowmiya[2]**

[1]*Assistant Professor & Head, Department of Computer Science, School of Computing Sciences, Vels University, Chennai.*
[2]*Research Scholar, Department of Computer Science, School of Computing Sciences, Vels University, Chennai.*
**Corresponding author E-mail:kalairaghu.scs@velsuniv.ac.in*

**Abstract**

With the huge amount of information available, the analysis over the data is the fertile area of knowledge mining research. Knowledge mining is the recent hot and promising research area. Knowledge mining is defined as the process of obtaining relevant knowledge from the pool of resources. In this review paper, we surveyed about the prior works carried out in the knowledge mining systems. We explore the primitives of knowledge mining systems. Attribute imbalance is the primary issue prevails in the knowledge mining process. In the field of higher education, most of the attributes are shared among the data features. In addition a precise introduction to knowledge mining along with its process is presented to get acquainted with the vital information on the subject of knowledge mining system

***Keywords:*** *Component, formatting, style, styling.*

## 1. Introduction

The rapid advancements of electronic data management systems have lead to an era of the information age. Several mid-ranges and large companies are widely, making use of the powerful database systems. Different data operations are accumulated for its operations, activities and performance. Generally, each datum holds valuable information like trends, patterns, etc. It helps to enhance the decision making process. A recent database has to be analyzed effectively to make better decisional processes. Under different scenarios, an independent attribute is used by the system model to yield accurate results. Henceforth, human effort is more required for data analysis.

The needs of obtaining automatic extraction of knowledge from a larger pool of databases to assist in the development of market analysis and discover the tools. Automated discovery tools have the capability to analyze the raw data and present the extracted high level information about the analyst or decision-maker, rather than find the analyst it for himself or herself. Data representation and attribute selection are the major issues for traditional machine learning algorithms. Knowledge mining is the recent technology, which is used for extracting valuable knowledge from different databases. It also includes the extraction of hidden knowledge from large data sets. And some of the data patterns depict the anomaly detection, predicting future patterns, etc. The knowledge mining techniques are applied to different ranges of marketing trends.

Mining information and knowledge from different databases has been studied by different researchers. The derived knowledge can be applied to information management, query processing, decision making, process control and other applications. Moreover, different knowledge mining techniques help to understand the user behavior, meliorate the service provided and enhancement of business opportunities. In response to the demand, the review paper suggests the database oriented techniques and applicative data mining systems.

The rest of the paper is organized as follows: Section II presents the primitives of knowledge mining systems and the prior works carried out by other researchers. Section III presents the summary of our reviews.

## 2. Related Work

This section represents the prior works in knowledge mining systems.

### Overview of Knowledge Mining Systems

The technologies in knowledge mining systems explore variant research issues which are related to the machine learning, statistics, database systems etc. which resolves the issues posed. Henceforth, the invention of knowledge mining technologies is a mandatory process. In some cases, the knowledge mining has formed an independent new field. Generally, the knowledge mining techniques are classified as follows:

### Kinds of Databases Required to Work on

Relied upon the database types, the knowledge mining system is classified and techniques are applied. Consider an instance, DBMiner that extracts the data from relational databases or any object-oriented systems. In order to derive the knowledge, the following are the types of databases, namely, relational databases, transactional databases, object oriented databases, deductive databases, spatial databases, temporal databases, multimedia databases, heterogeneous databases, active databases, legal databases and internet information bases.

### The Kind of Knowledge to be Mined

With the assistance of data mining, the knowledge can be obtained from association rules, characteristic rules, classification rules, discriminant rules, clustering, evolution and deviation analysis. By applying these rules, the derived knowledge will be in the form of

generalized knowledge, primitive-level knowledge, and multiple-level knowledge. Different data processing techniques provide a different abstraction level of knowledge.

## Kind of techniques to be used

Knowledge mined from different categories based on the data mining techniques. Basic knowledge miners like data-driven miner, query miner and interactive data miner. Furthermore, the data mining approach is based on a generalization based mining, pattern based mining, statistics and mathematical theories and integrated approaches etc.

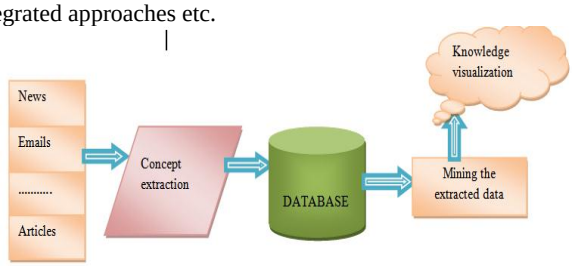


**Fig. 1:** Architecture of knowledge mining

## Prior Works

This section presents the prior works carried out in knowledge mining process. The author, Dewan et al [2012] studied about the knowledge mining process using educational data that depicted the efficiency of the teaching methods. They have discussed web based learning model which slows down the educational content organization. Jiawei Han et al [2012] derived knowledge of the networking models of heterogeneous systems. It was noticed that network modeling to degrade when the performance of the reconstruction phases to occur. Kuldeep Singh Jadon et al [2012] studied about the geographical based information processing systems. The query response to the user's search is not precise. By doing so, the files are not in hierarchical order.

The author Zhi Jun Peng et al [2016] discussed about Bing to search model via clicks based approaches. They depicted the query model for its searching efficiency. Though the author derives the knowledge via Bing model, the relevancy rate of the system is not achieved. The author, Mirghani A. Eltahir [2013] discussed about the knowledge extraction using web server data. The extracted data is generally in the HTML forms. These processes are inconsistent with nature. Clustering process is used for labeling the data. The study in clustering process was extended to the molecular clustering. In specific to, biomedical scientific corpora were used for designing the space. While clustering the molecules, certain dissimilarities occur in the formation of clustering.

The author Xian – Sheng Hua et al [2012] presented the big data concepts using algebraic geometry. It helped to locate the neighboring nodes based on the geometrical properties. But the fact, certainty problems is not solved. The author Xin Chen et al [2013] examined the knowledge system using social media data. The objective of their study was to obtain the learning experiences of the students based on social media data. They analyzed hashtag data to predict the learning experiences which fail to support non-technical researchers. The author Miroslav Hudec et al [2014] discussed about the domain based relationship using educational contents. Each data is associated with semantic annotations. Though it exhibited better accuracy, the assumptions of parametric constant are not precisely defined. The author Rihab Idoudi et al[2016] discussed the fuzzy functional dependencies concepts. Generally, the membership functions are defined for the attributes selected. The fuzzification process associates each data onto its memberships. Based on its relevancy, the data is organized.

The author Kuldeep Singh Jadon et al [2012] presented attribute based knowledge mining systems that specifically deal with Chinese words. Word disambiguation is the process of clearing the uncertainties occurs in the contexts. The attributes are extracted for obtaining the knowledge. Though, they have achieved better accuracy rate, but the window size of attribute is restricted. The author Panagiotis Hasapis et al [2013] presented the association rules based knowledge mining systems which deal out in agricultural machinery. Generally, the faults occur to machinery equipments. Their study intentionally focuses on obtaining knowledge of the faulty elements. They have used Apriori algorithm for controlling the variables which exhibited high storage complexity. A Similar study was conducted by the author Petra VrableCova et al [2016]. They discussed about the time series based clustering approaches. They collected online students' data for identifying the risks of the students towards learning systems. And the data aggregation model used in this study was not organized effectively. The author Tanvir Attaharya et al [2015] discussed about the circuit design based knowledge mining methods. They portrayed a reasoning model where the topology similarities developed for the discovery model. The demerit of this system is that the causal part of the circuit design is not correlated with the framed associative rules.

The author Yaogung Hu et al [2015] presented agent based knowledge extraction systems. They designed Cognitive Domain Ontologies (CDO) model which resolves the Constraint Satisfaction issue. The design models indulged pruning the search space in where certain searching complexity occurs. The author Mirghani A. Elathir et al [2015] discussed about the analog layout model in order to predict the time taken in designing the systems. In the analog layout model, the designing time is reduced. Symmetry and proximity constraints are not focused. The author Po- Hsun Wu et al [2015] described about the tacit concepts. Lattice concepts are used for validating the knowledge. They make use of natural language systems to derive the knowledge, but the knowledge representation is not effective. The author Maamoun Al – mardini et al [2016] discussed about intelligent agent systems to extract the knowledge using abalone dataset. They framed different association rules which depicted several reasoning concepts. By framing so, different rule inferences for the obtained reason are not effective in terms of accuracy.

Further, the study was extended to the Library Management Systems (LMS). The author Jun Xie etal [2013] discussed about the tacit knowledge from the University Libraries. Correlation analysis is used for depicting the knowledge. In the library perspectives, sharing is the most predominant factors. Any received information should be shared among the readers and the technicians. The objective of their study was to predict the barriers between the readers and technicians. By doing so, they have achieved a better efficiency, but the data used for analysis is low. The author Liyang et al [2015] depicted the clustering based knowledge mining systems in which transition state concepts were defined. Using transition state concepts, the size of the cluster is restricted which automatically leads to lesser hold of clustering subsets. The author Dong Sihui and Xu Xueguo [2015] presented the web server based, bilingual knowledge mining systems. In order to study about the bilingual model for collecting web data, the data points are sampled out for aggregated data points. By aggregating the data points, the coordination between those points is not effectively organized. The author Ashish Dutt et al [2017] discussed about the ontology based association rule mining systems. Medical datasets are used for deriving ontology based data mining systems. A hierarchical cluster form is not achieved properly in terms of cluster size and its prediction speed. The author Jui- Long Hung et al [2017] discussed about the molecular clustering for biomedical scientific corpora. It is perceived that the molecules of the cluster size are restricted to extraction phase. The files are effectively arranged in molecular clusters. The author Guozhong Cao et al [2016] presented the key technologies in Patent knowledge mining systems. Functionality matrix is annotated for every query. Their method fails in large scale application systems. The author Beat A. Sshwendimann [2017] discussed about novel and automatic acquisition of learning contents. Based on the input types, i.e discrete or continuous variable, the quality of the results is achieved. The author Dewan et al [2012] discussed about review of educational data mining

systems. They have reviewed the importance of clustering concepts for deriving relevant knowledge. The author Jiawei Han et al [2012] did a dashboard research in order to understand the learning experiences of the online students. The disadvantage of this approach is that our search might not have captured papers that did not explicitly use the term 'dashboard' or ascribe to the fields of Learning Analytics and Educational Data Minings.

## 3. Conclusion

Knowledge mining and data discovery model are related to each other in the fields like machine learning, statistics and databases. The fundamental objective of the knowledge mining systems is to obtain relevant context of developing methods and mining techniques. It usually involves heterogeneous sorts of data into understandable, abstract and relevant knowledge. Henceforth, this paper reviews for the importance of knowledge mining systems. We have explored the concepts introduced by other researchers. And also, we have discussed about the demerits of their systems. From the reviews, we have found that Attribute Imbalance is the predominant issue to be focused on knowledge mining systems. Most of the real-time attributes are shared among nature. In this framework, there exist dilemmas in deriving the knowledge from a large pool of resources. As our proposed work, we intend to resolve the 'attribute imbalance issue' using the concepts of swarm intelligence based decisional algorithms.

## References


[1] Farid DM & Sarwar H, "Knowledge mining for effective teaching and enhancing engineering education", *7th International Conference on Electrical & Computer Engineering (ICECE)*, (2012), pp.354-357.

[2] Han J, Sun Y, Yan X & Philip SY, "Mining knowledge from data: An information network analysis approach", *IEEE 28th International Conference on Data Engineering*, (2012), pp.1214-1217.

[3] Hua XS, Ye M & Li J, "Mining knowledge from clicks: MSR-Bing image retrieval challenge", *IEEE International Conference on Multimedia and Expo Workshops*, (2014), pp.1-4.

[4] Jadon KS, Maheshwari S & Dixit M, "Competent Searching for Geographic Information Gathering Using Knowledge Mining", *Fourth International Conference on Computational Intelligence and Communication Networks (CICN)*, (2012), pp.968-972.

[5] Hasapis P, Ntalaperas, D, Kannas CC, Aristodimou A, Alexandrou D, Bouras T, Georgousopoulos C, Antoniades A, Pattichis CS & Constantinou A, "Molecular clustering via knowledge mining from biomedical scientific corpora", *IEEE 13th International Conference on Bioinformatics and Bioengineering*, (2013), pp.1-5.

[6] Xie J, Chen Z, Xie G & Lin TY, "Knowledge mining in big data-A lesson from algebraic geometry", *IEEE International Conference on Granular Computing (GrC)*, (2013), pp.362-367.

[7] Eltahir MA & Dafa-Alla AF, "Extracting knowledge from web server logs using web usage mining", *International Conference on Computing, Electrical and Electronics Engineering (ICCEEE)*, (2013), pp.413-417.

[8] Chen X, Vorvoreanu M & Madhavan K, "Mining social media data for understanding students' learning experiences", *IEEE Transactions on Learning Technologies*, Vol.7, No.3,(2014), pp.246-259.

[9] Hudec M, Vučetić M & Vujošević M, "Synergy of linguistic summaries and fuzzy functional dependencies for mining knowledge in the data", *18th International Conference System Theory, Control and Computing*, (2014), pp.335-340.

[10] Duan J & Fu Y, "Attribute knowledge mining for Chinese word sense disambiguation", *International Conference on Asian Language Processing (IALP)*, (2015), pp.73-77.

[11] Atahary T, Taha T, Webber F & Douglass S, "Knowledge mining for cognitive agents through path based forward checking", *16th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, (2015), pp.1-8.

[12] Hu Y, Guo Z, Wen J & Han J, "Research on knowledge mining for agricultural machinery maintenance based on association rules", *IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, (2015), pp.885-890.

[13] Chemchem A & Drias H, "From data mining to knowledge mining: Application to intelligent agents", *Expert Systems with Applications*, Vol.42, No.3,(2015), pp.1436-1445.

[14] Yang L, Wang Y & Xu Y, "Tacit knowledge mining algorithm based on linguistic truth-valued concept lattice", *10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, (2015), pp.121-127.

[15] Eltahir MA & Dafa-Alla AF, "Extracting knowledge from web server logs using web usage mining", *International Conference on Computing, Electrical and Electronics Engineering (ICCEEE)*, (2013), pp.413-417.

[16] Wu PH, Lin MPH & Ho TY, "Analog layout synthesis with knowledge mining", *European Conference on Circuit Theory and Design (ECCTD)*, (2015), pp.1-4.

[17] Sihui D & Xueguo X, "Research on tacit knowledge mining of university libraries based on data mining", *13th International Conference on Service Systems and Service Management (ICSSSM)*, (2016), pp.1-4.

[18] Vrablecová P & Šimko M, "Supporting semantic annotation of educational content by automatic extraction of hierarchical domain relationships", *IEEE Transactions on Learning Technologies*, Vol.9, No.3,(2016), pp.285-298.

[19] Al-Mardini M, Hajja A, Clover L, Olaleye D, Park Y, Paulson J & Xiao Y, "Reduction of hospital readmissions through clustering based actionable knowledge mining", *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, (2016), pp. 444-448.

[20] Peng ZJ, Peng JL & Jiang YX, "Research and Implementation of Large Scale Bilingual Knowledge Mining Algorithm Based on Web", *International Conference on Robots & Intelligent System (ICRIS)*, (2016), pp.292-295.

[21] Idoudi R, Ettabaa KS, Solaiman B & Hamrouni K, "Ontology knowledge mining based association rules ranking", *Procedia Computer Science*, (2016), pp.345-354.

[22] Cao G, Luo P, Wang L & Yang X, "Key Technologies for Sustainable Design Based on Patent Knowledge Mining", *Procedia CIRP*, (2016), pp.97-102.

[23] Dutt A, Ismail MA & Herawan T, "A systematic review on educational data mining", *IEEE Access*, Vol.5, (2017), pp.15991-16005.

[24] Schwendimann BA, Rodriguez-Triana MJ, Vozniuk A, Prieto LP, Boroujeni MS, Holzer A, Gillet D & Dillenbourg P, "Perceiving learning at a glance: A systematic literature review of learning dashboard research", *IEEE Transactions on Learning Technologies*, Vol.10, No.1,(2017), pp.30-41.

[25] Hung JL, Wang MC, Wang S, Abdelrasoul M, Li Y & He W, "Identifying At-Risk Students for Early Interventions-A Time-Series Clustering Approach", *IEEE Transactions on Emerging Topics in Computing*, Vol.5, No.1,(2017), pp.45-55.