Efficient Intelligent Generic Recommendation Knowledge Graph in Education Domain using Association Rule Mining and Machine Learning

D Sathyanarayanan, M Krishnamurthy

Abstract*: Knowledge graph is used to extract and derive new facts from huge variety of data sources through relationship. An existing Natural language processing tool is specific and performs adaptive learning mechanism through instruction concepts. The specific knowledge graph suffers a problem of finding large collections of new facts with inter domain. This problem is addressed by implementing an efficient model for integrating various domain of interest as a generic knowledge graph. This proposed model has three major phases they are generic data collection, generic relationship establishment and generic deployment for education domain. The data are collected, preprocessed and categorized in to specific subject category by producing integrated data set. The relationship is established based on the pedagogical data with assessment data of leaners are classified in to course list. This generic knowledge graph is compared with the CNN based model and GCN based model. The validation of these models are assessed and deployed into application services for teachers and learners. The main objective of the proposed graph is to organize a generic knowledge graph for deriving huge amount of new facts to the education domain with maximum support and confidence level.

***Keywords*: Knowledge Graph, Association Rule Mining, Semantic Mining, Ontology, Relational Database.**

I. INTRODUCTION

The significant The data are generated by the people who use the internet is grows maximum size because of the latest technologies available in the market. These data are maintained in the huge data center like cloud. The real problem is happens during the data extraction from huge amount of data leads the severe performance problem with minimum reliability. The existing model uses the data mining methods with different approaches which unable to produce the desired result. Association rule mining gives the gain knowledge to the end users with the hyper parameters support and confidence. The attributers are collected and identified for the better relationship. These attributes are arranged in the order in which maximum reliability during the data extraction process. Ontology based organization is complex because of huge information exists in the domain but it lacks in the meaning of the data during the analysis. This problem is overcome by using semantic based approach which is related to page linking model. These problems are overcome by using knowledge graph which provides the gain knowledge to the customer. Knowledge graph is used to access the knowledge from heterogeneous sources and then it produces the new knowledge for further operation. There are various tools exists such as Intelligent Training System (ITS) and MOOC platform for producing adaptive learning model for learners and teachers which uses limited set of attributes and its relationship. Traditional knowledge graph uses instruction concepts related to the domain which leads the performance problem during the analysis and prediction of new data. There are some knowledge graph model support the gain of concepts such as Freebase, Reverb, Google Valut, Microsoft Probase and so on. The education knowledge graph with generic standard needs the student to get maximum knowledge in order to success in their life. Knowledge graph this used to build new relationships based on text information over various entities like customers, nodes and interaction among the nodes. The edges are labeled and link the nodes using semantics with natural language processing tools. The existing specific knowledge graph considered various elements like data, knowledge and information as a graph. It is not suitable or generic domain of knowledge graph [1]. Knowledge Graph method of organizing searching, answering system based on questions, depth reading with semantic mining of subjects and concepts. This model considered the domain for extracting and making relationship among the document which is organized into open network. Multiple sources of specific knowledge make various relationships but it leads the data conflicts problem due to inconsistency. This is restricted to specific domain of interest but it suffers the reliability problem [2]. Topic graph is established over the tweeter which is used to transfer short message to the social networking i.e. tweets. The time lines are mapped with the text in order to identify the recent tweets by the followers. The novel recommendation system has been introduced to integrate the users with their topic of interest. This model of the graph use algorithms which are also uses specific topic which provides minimum knowledge [3].The problem in the knowledge graph is difficulty in use due to the rich contents with flexible structure. The relationships are also leads the model complexity because of various users such as professional and non-professional. The query writing process faces a issues like for simple and complex tables in the database. Simple table takes complex query whereas complex table. Graph uses Query By Example (GQBE) method which captures the data with various attributes and its relationships.

Revised Manuscript Received on August 09, 2019.

D Sathyanarayanan*, Research Scholar, AMET University, Chennai, Tamilnadu, India.

M Krishnamurthy, Professor, Department of Computer Science and Engineering, KCG College of Technology, Chennai, Tamilnadu, India.

The graph is organized based on the keywords which are exists already and produces new knowledge [4]. Knowledge Graph with structured queries retrieves the maximum knowledge in an efficient manner. In natural language queries are difficult to write by the non-experts because of lack in the domain knowledge. The existing methods used non automatic query processing which leads the time complexity during data retrieval process. The framework has been introduced for embedding data through graph. This framework converts high dimensional graph to low dimensional attributes for generating local graph. This is restricted to only one language, so global knowledge graph will be considered [5]. The problem arises in the relational database design, development and implementation process. Property graph has been implemented with major knowledge management process through specific data model. This method provides better gain and support when compared to existing method of knowledge graph [6]. Vertical knowledge graph is implemented for designing and classifying the knowledge as a graph with feasible solution [7].Data Information-Knowledge (KID) model is developed for decision support system over the digital marketing. This model maintain registry for maintaining processed data as an information then specific knowledge. It consists of reconstructed ontology and rule based mining with corresponding engine for processing. Various algorithms are analyzed with KID knowledge graph such as clustering and classification [8]. Machine learning are more popular method for predicting future data occurrence are integrated with relational database, new facts are generated as a node through training and testing ratio. The machine learning uses vast amount of data and its attributes with relationship. This model improves the observable modeling power with minimized computation cost. Text based machine learning models are integrated by constructing knowledge graph [9]. Uyghur's knowledge graph is constructed for semantic search and automated questing answering system. It is a method for performing task such as information extraction, processing and application development. It also performs operation into various phases such as define the graph, structured the graph, analyze the graph and so on [10]. The association rule mining is applied over the XML data which are accessed from the respective database through dynamic parameters such as support and confidence [13].The existing knowledge graph are restricted to single domain i.e. specific to the domain which leads the performance as well as reliability problem while handling huge amount of data. The main objective of the proposed generic knowledge graph is to gain maximum support for learners and teachers in education domain with all levels. The rest of the paper is organized as follows, Section 2 represents the conceptual diagram of the knowledge graph, section 3 describes that the architecture of the proposed method. Section 4 describes that the algorithm of the proposed method. Section 5 shows that the experimental evaluation and section 6 represents the conclusion and future work.

II. CONCEPTUAL DIAGRAM FOR KNOWLEDGE GRAPH

The concept of the knowledge graph is to establish the relationship among subjects which are related to course of domain in order to regulate the teaching and learning process effectively the knowledge graph with individual subject which provides minimum support shows in figure 1. This is improved by using knowledge graph with more than one subject who provides multi attribute support to the learners and trainers which is shown in figure 2. The existing model is restricted to only one specific domain which suffers in lack of knowledge towards other domain. Nowadays multi-disciplinary learners and trainers get more opportunities than the single domain experts. This is done by implementing generic knowledge graph for handling more than one domain with huge amount of hyper attributes and parameters is shown in figure 3. This model provides better gain and support when compared to specific knowledge graph based model. Figure 4 shows that the examples for proposed generic knowledge graph.

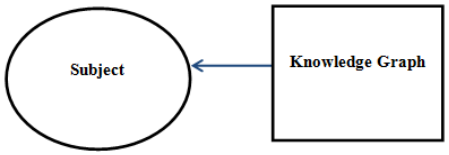

Fig.1 Knowledge graph with single subject

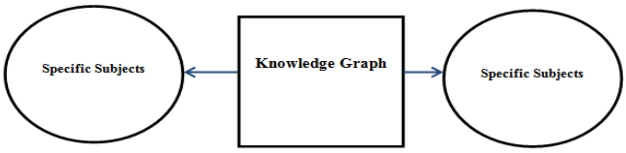

Fig.2 Knowledge graph with Specific subjects

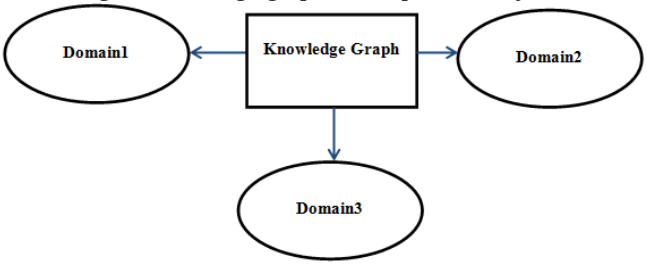

Fig.3 Knowledge graph with Heterogeneous Domain

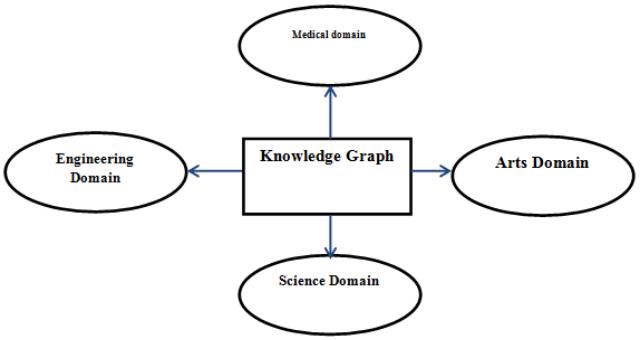

Fig.4 Example for Knowledge graph with Heterogeneous Domain

III. PROPOSED METHODOLOGY

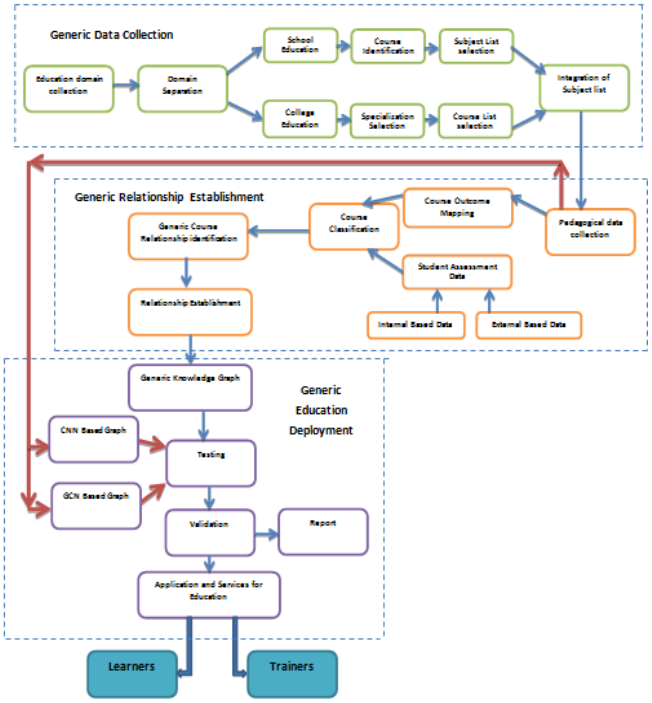

Fig.5 Generic Knowledge graph recommendation model

The data are collected from various education data sources for establishing generic relationship through knowledge graph. Initially, heterogeneous domains are collected for better analysis then domains are separated according to the level of study such as school education and college education. School education domain the courses are identified based on different subjects available in the curriculum. In college education domain various courses and its subjects are identified and maintained in separate list. Both the list of subjects is integrated with the help of classifier which produce integrated list of subjects. Pedagogical data are collected from the list then map the course with the outcome of the students. The attributes are considered as a cluster of subjects in different domain. Student assessment data are also considered with various attributes such as internal assessment as well as external assessment data. The courses are classified from already identified list of attributes in order to make better prediction over the teaching and learning process. Generic course relationships are identified from already classified generic attributes then relationships are established in order to build a generic knowledge graph. This proposed knowledge graph provides the better clarity in the subject's studies by the student and trained by the trainers without any lack in the domain knowledge with high success rate. This model is validated over already existing model such as CNN based and GCN based model. These models are tested over hyper parameters such as attributes with their relationship. The validation report is generated for further analysis. Once the validation process is gets over then the services are deployed as an application for efficient teaching and learning process. Figure 5 shows that the proposed model for generic recommendation using knowledge graphs.

IV. ALGORITHM

Algorithm Generic_Knowledge_Graph_Generation()
Begin
Education domain as ED;
Collect the various domain;
Let D is a database;
For each domain ε ED do
Begin
Separate the domain to SD;
For each domain ε SD do
Begin
If domain='school education' then
Begin
Identify the course list;
Collect the subject from the course ;
Find the frequent items i.e subjects from D with minimum support as L;
Select the subject list;
End
Else if domain=' College education' then
Begin
Identify the specialization list;
Collect the course from specialization;
Find the frequent items i.e subjects from D with minimum support as L;
Select the subject list;
end
end
Integrate the subject list;
S=Find the strong association rules with L and minimum confidence as
End
RE=Relationship_Establishment(S);
ME=Model_Evaluation();
Deploy the application based on RE and ME;
Education service offering to the learners and trainers;
End

Algorithm Relationship_Establishment (Association Rule as S)
Begin
Collect the pedagogical data as PD;
PD is mapped with course and its subjects;
Collect the student assessment data as SAD;
For each (data ε SAD and rule ε S) do
Begin
If data_type == 'Internal' then
Begin
Collect the assignment data as AD;
Collect the Unit test data as TD;
Collect the attendance data ATD;
Cumulate the AD, TD and ATD;
Generate the cumulative data as ICD over rule;
End
If data_type == 'External' then
Begin
Collect the theory based data;
Collect the Practical based data;
Cumulate the both of the data;

Generate the cumulative data as ECD over rule;
End
End
Classify the course based on ICD and ECD;
Identify the course relationship with classified course data;
Establish the relationship for generic education data;
Built the knowledge graph based on the generic relationship with S as GIR;
Return GIR;

End

Algorithm Model_evaluation()

Begin

Proposed knowledge graph with relationship;
Collect the relationship and knowledge graph from CNN based model;
Collect the relationship and knowledge graph from GCN based model;
Test the models in an individual manner;
Integrate the model for validation as IMV;
Return IMV;

End.

V. RESULT AND DISCUSSION

There are various knowledge graph has been analyzed and compared the model with proposed generic model. Search engine with recommendation system provides the knowledge from the web log files with associated attributes through freebase [11]. Pattern analysis are predicted over different topics of interest are implemented over minimum set of relations and entities. Relationship is established for pattern which are very close to each other using similarity matrix with inter linked data sets using Reverb knowledge graph [12]. Association rule mining has been introduced by converting transactions into bit vectors using anti-mirroring technique [14]. Google Vault is used to import, export and search data from various application in order to generate knowledge graph. It is restricted to only specific domain of interest which is not suitable for wide knowledge. Microsoft probase is used to make systems can understand human knowledge with semantic in nature but it is applicable to single domain. It applies to natural language processing, voice analysis and complexity in languages. These problems are addressed by using generic knowledge graph with better prediction and generation of new knowledge. Figure 6 shows that the comparison of support and confidence assessment of proposed algorithm. Figure 7 and figure 8 shows that the classification of courses and subjects in school level and college level respectively. The integrated course and subject lists are specified in figure 9. The relationships of the subjects and topic of interest are shown in figure 10. Comparison of various knowledge graphs with proposed generic knowledge graph is shown in figure 11.

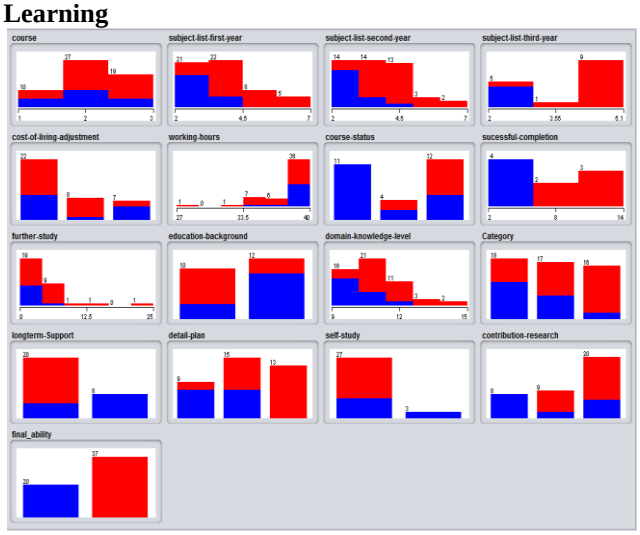

Fig.6 Proposed Attribute support and confidence comparison

Minimum support: 0.2 (20 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 16
Generated sets of large itemsets:
Size of set of large itemsets L(1): 22
Size of set of large itemsets L(2): 182
Size of set of large itemsets L(3): 56

Best rules found:

1. a1=false a5=false 24 ==> class=c0 24 <conf:(1)> lift:(1.52) lev:(0.08) [8] conv:(8.16)
2. a5=false a8=false 24 ==> class=c0 24 <conf:(1)> lift:(1.52) lev:(0.08) [8] conv:(8.16)
3. a5=false a6=false 23 ==> class=c0 23 <conf:(1)> lift:(1.52) lev:(0.08) [7] conv:(7.82)
4. a8=false class=c1 22 ==> a5=true 22 <conf:(1)> lift:(1.79) lev:(0.1) [9] conv:(9.68)
5. a5=false a7=true 21 ==> class=c0 21 <conf:(1)> lift:(1.52) lev:(0.07) [7] conv:(7.14)
6. a5=false a9=false 21 ==> class=c0 21 <conf:(1)> lift:(1.52) lev:(0.07) [7] conv:(7.14)
7. a3=false a5=false 20 ==> class=c0 20 <conf:(1)> lift:(1.52) lev:(0.07) [6] conv:(6.8)
8. a6=false class=c1 20 ==> a5=true 20 <conf:(1)> lift:(1.79) lev:(0.09) [8] conv:(8.8)
9. a2=false a5=false 27 ==> class=c0 26 <conf:(0.96)> lift:(1.46) lev:(0.08) [8] conv:(4.59)
10. a4=false a5=false 23 ==> class=c0 22 <conf:(0.96)> lift:(1.45) lev:(0.07) [6] conv:(3.91)

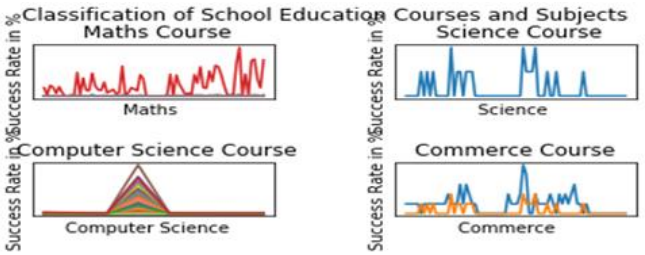

Fig.7 School education classification

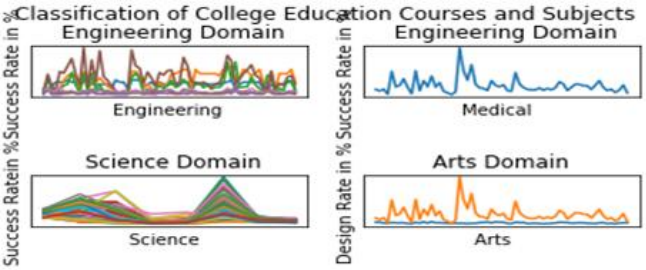

Fig.8 College education classification

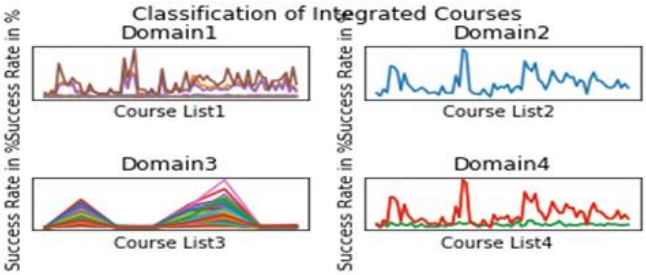

Fig.9 Integrated course classification

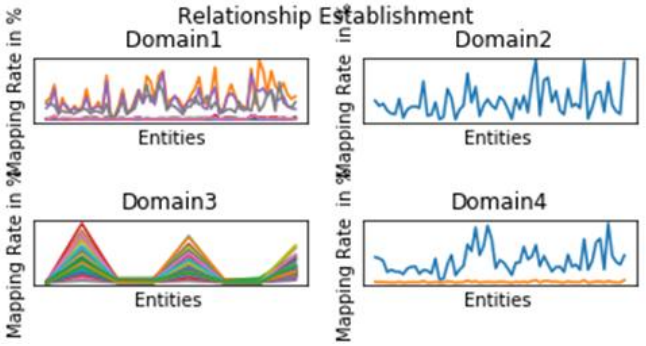

Fig.10 Relationship Establishment

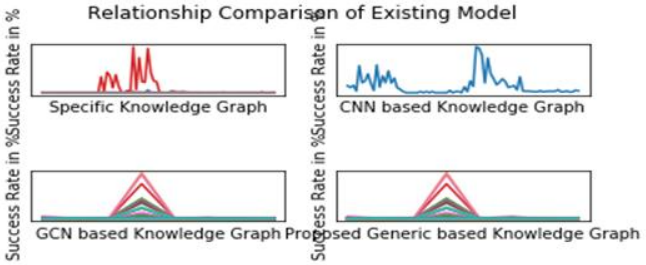

Fig.11 Relationship Comparison

Generic knowledge graph is evaluated based on the hyper parameters like rank score, Mean Reciprocal Rank (MRR), Mean Rank (MR) and hits_at_n_score. Rank score labels with list of score and positive in nature. MRR computes the vector of ranking with reciprocal property . MR is used to computes the mean of the vector of ranks. Hits_at_n_score returns the top rank in the list of rank which is related to entities and relationship. Figure 12 shows that the generic knowledge graph with rank score metrics. Figure 13 represents the generic knowledge graph with MR score. Figure 14 signifies the generic knowledge graph with MRR score. Figure 15 illustrates the generic knowledge graph with hits_at_n_score. The proposed knowledge graph provides maximum new knowledge when compared to existing graphs.

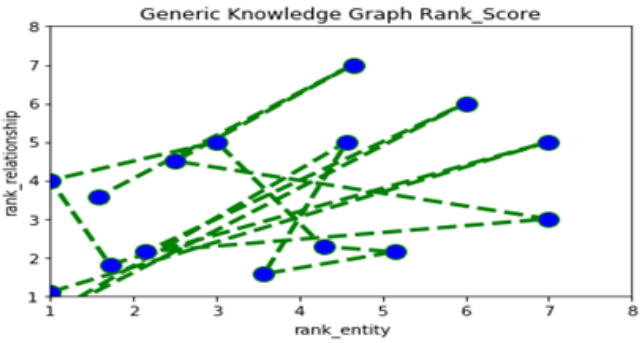

Fig.12 Rank score comparison

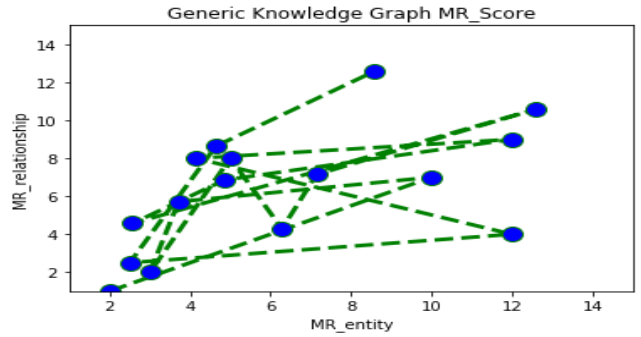

Fig.13 MR score comparison

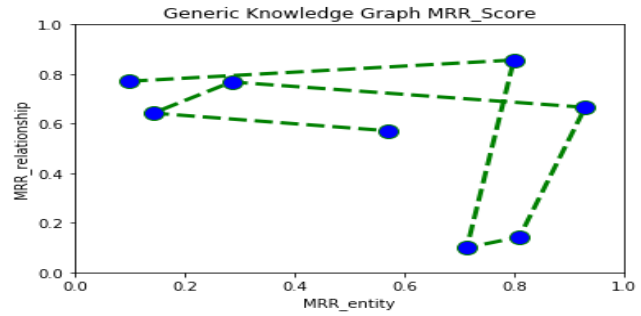

Fig.14 MRR score comparison

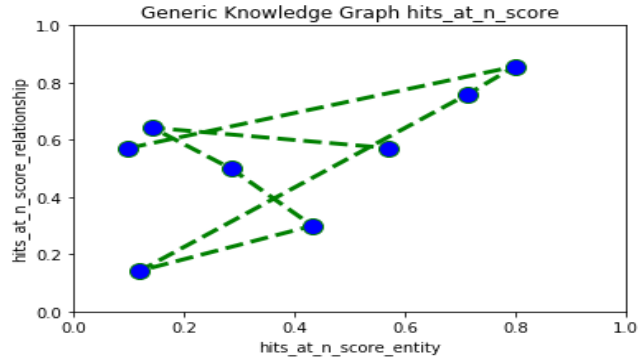

Fig.15 hits_at_n_score comparison `

VI. CONCLUSION

Knowledge graph uses semantic information for providing better facts to the customer and end user i.e. trainers and learners in education domain. Various graphical model has been analyzed which are restricted to specific domain of interest, so the knowledge generation is minimum. Predefined models are maintained by software industries for producing relevant information related to the search query.

Conceptual models are analyzed and identified attributes towards the association. The data extraction process is done by using association rule mining with the hyper parameter such as support and confidence. The proposed model has been implemented for generating generic domain data and its relationship. There are three phases are considered namely data collection, relationship identification and deployment. The algorithms are compared with the existing algorithms in order to achieve maximum support. Machine learning algorithms are used to classify and make the generic facts to the end users. The main objective of the proposed graph is to organizing the relevant facts to produce the high quality fact with maximum support to the society. In future this model can be extended to real time scientific application through deep learning method.

REFERENCES

1. Yucong Duan, Lixu Shao, Gongzhu Hu, Zhangbing Zhou, Quan Zou and Zhaoxin Lin,"Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph", DOI: 10.1109/SERA.2017.7965747,2017,Pp 1-12.
2. Lv Qingjie, Xu Lingyu, Yu Jie, Wang Lei, Xun Yunlan, Shi Suixiang and Liu Yang, "Research on domain knowledge graph based on the large scale online knowledge fragment", IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA),DOI: 10.1109/WARTIA.2014.6976259,Ottawa, ON, Canada, 2014, Pp 1-6.
3. Danae Pla Karidi,"From user graph to Topics Graph: Towards twitter followee recommendation based on knowledge graphs", IEEE 32nd International Conference on Data Engineering Workshops (ICDEW),DOI: 10.1109/ICDEW.2016.7495629, IEEE, 2016,Pp 1.8.
4. Nandish Jayaram, Arijit Khan, Chengkai Li, Xifeng Yan and Ramez Elmasri,"Querying Knowledge Graphs by Example Entity Tuples", IEEE Transactions on Knowledge and Data Engineering, Volume: 27 , Issue: 10 ,2015,Pp 2797 - 2811.
5. Ruijie Wang, Meng Wang, Jun Liu, Siyu Yao, Qinghua Zheng,"Graph Embedding Based Query Construction Over Knowledge Graphs",IEEE International Conference on Big Knowledge (ICBK), DOI: 10.1109/ICBK.2018.00009, 2018, Pp: 1 - 8.
6. Zuopeng Justin Zhang,"Graph Databases for Knowledge Management", IT Professional, Volume: 19 , Issue: 6 , 2017,Pp 26 - 32.
7. Fu Leijie, Bai Yv, Zhong Zhenyuan,"Constructing a vertical knowledge graph for non-traditional machining industry",IEEE 15th International Conference on Networking, Sensing and Control (ICNSC),DOI:10.1109/ICNSC.2018.8361341 2018, ,Pp 1-6.
8. Jiandong Li, Atsushi Sato, Runhe Huang, Dongming Chen,"A Rule-Based Knowledge Discovery Engine Embedded Semantic Graph Knowledge Repository for Retail Business",International Conference on Advanced Cloud and Big Data (CBD),DOI: 10.1109/CBD.2016.024, 2016,Pp 1-8.
9. Maximilian Nickel, Kevin Murphy, Volker Tresp, Evgeniy Gabrilovich,"A Review of Relational Machine Learning for Knowledge Graphs", Proceedings of the IEEE, Volume: 104 , Issue: 1 , 2016, Pp 11 – 33.
10. Lirong Qiu, Huili Zhang,"Review of Development and Construction of Uyghur Knowledge Graph",IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), DOI: 10.1109/CSE-EUC.2017.181, IEEE, 2017, Pp 1-6.
11. Xiao Yu, Hao Ma et al,"On building entity recommender systems using user click log and freebase knowledge", WSDM '14 Proceedings of the 7th ACM international conference on Web search and data mining ,2017, Pp 263-272.
12. Feichen Shen, Hongfang Liu, Sunghwan Sohn, David W. Larson, and Yugyung Lee,"Predicate Oriented Pattern Analysis for Biomedical Knowledge Discovery",Intell Inf Manag. volume 8, Issue 3, Pp 66–85.
13. D Sathyanarayanan, M. Krishnamurthy,"An Efficient Association Rule Based Dynamic Support CountAdaptation Model for XML Databases Using XQuery",International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 6 (2015) pp. 16129-16147.
14. D Sathyanarayanan, M Krishnamurthy,"Association Rule Mining Using Frequent Itemsets Generation by Anti- Mirroring Of Bit Vectors",International Journal of Pure and Applied Mathematics,Volume 118 No. 20 2018, Pp.5033-5044.