ORIENTAL JOURNAL OF
COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
www.computerscijournal.org

ISSN: 0974-6471
December 2011,
Vol. 4, No. (2):
Pgs. 261-272

Ontology Based Knowledge Grid in Semantic Web to Discover Knowledge in Distributed Environment

MUQEEM AHMED*, S.Z. HUSSAIN and S.A.M. RIZVI

Department of Computer Science, Jamia Millia Islamia, New Delhi - 110 025 (India).
*Corresponding author: E-mail: muqeem.ahmed@gmail.com

(Received: August 25, 2011; Accepted: October 04, 2011)

ABSTRACT

In spite of various current research and investigations the development of advenced information technology is not the key issue. The different information technologies available now days but the major issue is how to got more advantages and utilization of these technologies for academic purpose in distributed environment where faculty and students communicate with software technology rather than with individual. Knowledge Based Grid was introduced for publishing, managing, sharing and utilizing different amount of knowledge base resources on the semantic web in distributed environment. The Knowledge discovery from heterogeneous information sources available on knowledge Grid environment is a major challenging research and development issue. This paper mainly concerns all aspects of the knowledge discovery, sharing process and integrates grid data resource by ontology server for educational institutes and university in distributed environment to address these issues and challenge.

Key words: Knowledge grid, Discovery, Semantic, Ontology, data mining, Integration.

INTRODUCTION

The development of advanced information technologies are not the key issues. Lots of information technologies are available these days but the major issue is that how to get more advantages of these technologies for academic purpose in distributed environment where faculty and students communicate with software technology rather than with individual. The huge volume of data on web exist in heterogeneous, unstructured and semi structured form that make extraction of relevant information extremely difficult. Semantic Web, envisioned by Tim Berners Lee, is extension of existing web and intends to make it understandable to machines. Mainly Fran Berman[1] introduced the concept of knowledge based grid which is more effective for scientist and academician. Berman also focussed the abstraction and integration knowledge from online information pool through knowledge base grid in distributed environment. The growth in the development of knowledge grid is a need for suitable abstraction. Cannataro and Talia also considerd knowledge based grids techniques based on data mining[2]. Hzhuge consider the future internet and gave definition of knowledge grid, resource semantic grid and social grid[3]. The demand of the knowledge retrieving and discovery increased with next generation of information technology.The

knowledge grid include a new technology as service oriented designed using Java and Corba which consider topic map for discovering and retrieving semantic knowledge and enable queries from grid nodes[4].The knowledge grid enables knowledge discovery services technologies and model such as open grid service ,p2p and ontology which allows educators ,researchers, professionals, students to create and manage knowledge discovery to integrate information ,mining and storage resources in the distributed environment[5]. The experimental view of deducing most of the advantages of Web-assisted or Online Assisted Learning for the number of universities and colleges in typical Indian conditions that does not need large investments in terms of Internet infrastructure is being implemented on the basis of quality metric approach. The education Grid[6] offers the processes involved in the developing, deploying and utilizing E-course for deployment in the State of Kerala in India. It made possible by advanced technology web-servers based on a new architecture to establish effective and well managed learning management and collaboration systems and subject-specific interface which support to enhance the quality of education in distributed environment. The Knowledge Grid based knowledge discovery model is builds on a computational grid which provides dependable and consistent access to computational resources. It used basic grid services and defines a collection of extra additional layers to develop the services of distributed knowledge discovery on the distributed connected resources where each node can be a sequential or a parallel machine. The Knowledge Grid enables the collaboration of scientists that must mine data that are stored in different research centers as well as executive managers that must use a knowledge management system that operates on several data warehouses located in the different company establishments. This represented an initial step for the design and implementation of a grid architecture that integrate different data mining techniques, algorithm and computational grid resources in the distributed environment[7]. The Grid-based data mining and knowledge discovery framework is a efficient problem solving environment which is responsible for the users to express a problem using their own domain specific knowledge to develop their application using data analysis and data mining operations. It provides lightweight distributed approaches to perform large-scale computation in an efficient manner and takes care of the issue of easy, clear representation and manipulation of the knowledge by providing knowledge map layer. While the concept of the knowledge map its structure and implementation provide a novel knowledge and retrieve in heterogeneous distributed grid environment[8].

The semantic component in the knowledge layer provides enables semantic description of grid resources through ontology template. Web ontologies create ontology template for different types of computing resources in the distributed grid system environment. The Globus Toolkit is used to gather grid resource information and Protégé-OWL libraries to create knowledge base of grid resources. Algernon inference engine is most suitable for communicating with the knowledge base to discover suitable resources in the distributed grid environment[9].Knowledge discovery from different heterogeneous data sources available on Grid environment covered all aspects of the knowledge discovery process and integrates this process with service-based gird application supported by agent environment for university domain. The Global Grid Forum, Open Grid Services model, and its associated specifications describe user friendly interfaces through web services that represented different heterogeneous data resources in the discovery process in distributed data resource environment by using groups of multiple task agents[10]as the complexity of problems undertaken by users everyday increases, the requirements surrounding the knowledge Grid have become more complex and demanding. Efforts like the Semantic Grid provide new capabilities to users, and also, as mentioned by De Roure et al. (2005), new research opportunities: semantic service description, smart interaction, autonomous behavior, knowledge technologies, among others, is topics that should be addressed by future efforts. In fact, knowledge technologies for Grid environments are getting more attention from the scientific community. Approaches like the Knowledge Grid presented by Zhuge CRPIT Volume 110 - Conceptual Modelling 201054(2008) proposed a highly

distributed collaborative environment, where explicit knowledge resources are managed to support decision-making processes and cooperative work.

The above discussed tools, techniques and solution of the knowledge discovery in the distributed environment includes the basic concept, for design of ontology based knowledge grid in semantic web to descover knowledge in the distributed environment and suggest basic data mining techniques and algorithms for the effective knowledge discovery in the distributed knowledge grid environment and its application. Therefore little amount of work has been developed on this hot topic. The issue now here is semantic based knowledge discovery that provides quality of education and student decision making for admission in the distributed environment on their own choice. So the quality of education offered is substantially enhanced. The web knowledge grid could offer effective and efficient support to the process of organizational and educational knowledge discovery and integrate grid data resource that provides different services to the student and employee of the university like searching, advisory help desk system, course and resource management, student and faculty communication etc. Web knowledge-based task, content converted into a collection of semantic information. i.e. understanding the context, format, and significance of the data and information called knowledge discovery in the grid based distributed environment. This paper is an attempt to develop a ontology based knowledge grid for knowledge discovery in distributed environment system using data mining tools , techniques and algorithm as development methodology and student information management system as a case study.

Semantic web

Semantic was defined as ' an extension of the current Web in which information is given well defined meaning, better enabling computers and people to work in cooperation.' Semantic web is web of data that is process directly and indirectly by machine.current web is read only web which can not edit and annotate data and information in distributed environment. Semantic web consider the all issues of existing web such as read write and provide a machine to machine communitation that is very helpfull to solve complex problem in the distributed envirement which varies scale and stability

The knowledge based grid

Knowledge based grid is a semantic web based environment that enable machine and human to communicate, coordinate, publish, share and manage different knowledge resources in distributed environment which enhance scaling and stability. It supports on demand, annotative and strong services which is useful for innovation and collaborative work in distributed heterogeneous environment. The Knowledge Grid architecture is design on grid tools and services, i.e. it considers basic grid services to design specific knowledge extraction services [11]. Such types of services can be design in different ways using the different available grid tools and services. Grid architecture based on the Global Tools which is more suitable for knowledge discovery communication, sharing process and integration. It is an environment for providing Grid-based knowledge discovery services. These services allow professionals and scientists to create and manage complex knowledge discovery applications, composed as workflows that integrate data sets, mining tools, and computing and storage resources provided as distributed services on a Grid. It facilities allow users to compose, store, share, and execute these knowledge discovery workflows as well as publish them as new components and services on the Grid. The knowledge grid can be used to perform data mining on very large data sets available over Grids, to make scientific discoveries, improve industrial processes and organization models, and uncover business valuable information. It provides a higher level of abstraction and a set of services based on the use of Grid resources to support all those phases of the knowledge discovery process. Therefore, it allows the end-users to concentrate on the knowledge discovery process they must develop without worrying about Grid infrastructure details.

The Knowledge grid is composed of a collection of services divided in two layers:

´ The core K-Grid layer that contains

metadata and ontologies about data sources and software components and interfaces basic Grid middleware and services. The core layer is a view over the knowledge grid knowledge base.

´ The high level K-Grid layer that offers set of services for design and process of knowledge discovery application by the user interfaces. Knowledge grid work flow environment, discover processes are represented through both concrete and abstract Grid resources. Visual interface shows resources and knowledge discovery to the user and offer mechanism for integrating them in a work flow. The knowledge discovery services can be implemented by high level K-Grid layer using some ontology-based services.

The knowledge discovery

Knowledge discovery is a concept of the of computer science that shows the process of automatically searching large volumes of information for patterns that can be considered knowledge. It describe as deriving knowledge from the input data. The main branch of knowledge discovery is data mining, also known as Knowledge Discovery in Databases. It creates abstractions of the input data as many forms of knowledge discovery. The knowledge obtained through the process may become additional data that can be used for further usage and discovery. The important application of knowledge discovery is in the area of software modernization which involves understanding existing software. This process is related to a concept of reverse engineering. Usually the knowledge obtained from existing software is presented in the form of models to which specific queries can be made when necessary. The Grid Miner project[12] at the University of Vienna is the first Grid research work, to cover all aspects of the knowledge discovery process and integrate them as advanced service-oriented Grid application. Object Management Group developed specification Knowledge Discovery Met model defines ontology for the software assets and their relationships for the purpose of performing knowledge discovery of existing code. Knowledge discovery from existing software systems, also known as software mining is closely related to data mining, since existing software artifacts contain enormous business value, key for the evolution of software systems. The software mining focuses on metadata, such as database schemas. Knowledge Discovery and Data Mining may sometimes be used synonymously.

The Proposed System

The proposed system contains web portal interface, ontology server, metadata knowledge directory server, metadata database and knowledge base shown in figure 1. Mainly the user requests through the intelligent web portal interface to the ontology server then a task is generated which is send to the ontology server to generate semantics of web documents[13] and transfer these generated semantics to the knowledge directory server that locates the distributed heterogeneous data bases then transfer it to the grid data .The grid executes the data mining task algorithms and returned the results to the meta data directory server. The knowledge integration integrate effective knowledge which is browsed and discover through the web portal interface to the external user in the grid distributed environment which is given bellow.

The portal Interface

The present web portal interface compiles html tags and displays it as text document. The proposed intelligent interface portal provides group of geographical tools. The client can view the geographical representation of his knowledge using combine visual facilities. Which describe machine understandable and explain the view it provide

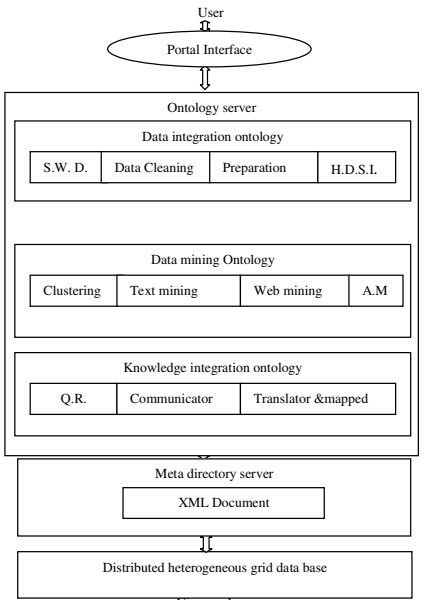


Fig. 1:

different services which help to operate the resources such as automatic suggestion service. The present interface uses synonyms to grid where user automatically suggests search items for example if someone want to access semantic related services. The intelligent portal interface provides semantic web related document facility.

Ontology server

Ontology server is the central server which is responsible for managing and querying the explicit declarative web ontology. It takes the information from the portal interface and generates a task and sends that task to the metadata server. Metadata directory server act as a registry or metadata catalog of knowledge grid. It provides special services called the knowledge directory service which shared information. The ontology server provides services such as data integration ontology services, data mining ontology services and knowledge integration services.

Data integration ontology services

The concept of 'integration' means anything ranging from integration, merges, use, mapping, extending, approximation, unified views and more; sometimes interchanging the words as if all are synonyms, although these concepts are used as homonyms as well. It describes the semantic of web document which bridge the gap between structured and semi structured data bases and facilitate data cleaning, data preparation, heterogeneous data sources integration.

Data mining ontology services

Data mining ontology service suggest a model for different type of data mining algorithms and software which solve a specific problem. The data mining ontology provide functionality class including clustering, classification, prediction, text mining, association mining link analysis and web mining.

Knowledge integration ontology services

Knowledge integration has also been studied as the process of incorporating new information into a body of existing knowledge. This process involves determining how the new information and the existing knowledge interact, how existing knowledge should be modified to accommodate the new information, and how the new information should be modified in light of the existing knowledge. Knowledge integration ontology services maintain set of public ontologies, responding to query, facilitating the communication and knowledge interchange among different knowledge base, mapping between different ontologies, providing on demand services to support problem solving and decision making. It also integrates different knowledge resource of different level to provide problem analyzing and solving.

Meta Data Directory server

The metadata directory server is responsible for maintaining, describing all data, techniques used in the knowledge grid. It acts as a metadata catalog of knowledge grid which identify specific discovery, registration services and also offer mechanism of integrating multiple knowledge base into higher knowledge base to serve domain functional purpose. The metadata information are represented by XML document and are inserted in to knowledge base and metadata database

Metadata database

Metadata stores the distributed heterogeneous data sources, techniques, tools and algorithm which are used for data integration and data mining.

Knowledge base

Knowledge base is responsible for storing fetched result of the data mining process i.e. some knowledge gathering from application domain

Implementation of proposed system

Knowledge grid based knowledge discovery in distributed environment involves many researching technologies. At present, we have only considered simplified version of knowledge discovery with Apache Tomcat, protege , and Amaya semantic web based interface open source tool used to create and update document directly on the web. Browsing features are seamlessly integrating with the editing and remote access features in a uniform environment. This follows the original version of the semantic web as a space for collaboration and not one way publishing medium. The main motivation to developed Amaya was to

provide framework that can integrate as many w3c technology as possible. It used to demonstrate these technologies in action while taken advantage of their combination in a single environment. Amaya started as an HTML +CSS style sheet editor. But the now amaya include knowledge discovery domain suggestion automatically which was very challanging. Amaya include collaborative annotation based on recourse description framework (RDF), XLINK, XPOINTER. Amaya software implemented in java and is available for windows. Heterogeneous grid database is in xml/ schema for semantic query and represents the knowledge in semantic acquired from. The heterogeneous database is stored metadata directory services on ontology server.. The greatest strengths of XML lies in separating the data display format and the content using the self-describing method, and creating markup languages in some specific area, thus effectively expresses the data structure and semantic of unstructured or semi-structured documents. XML Schema provides richer data types and a set of mechanisms to enhance the capacity for knowledge description. On the basis of semantic descriptions of knowledge, we use RDF/ RDF [14] Schema for modeling knowledge. Resource Description Framework is a kind of standardized specification on semantic description of metadata, which adopts basic data model composed of Resource, Property and Statement tree objects to establish a framework for the definition and use of meta-data. Thus the metadata stored in the metadata directory server can be effectively translated into machine-understandable information by different kind of data mining functionalities algorithms to solve specified problem. The data mining ontology gives functionality class including clustering, classification, text mining Association mining, web mining [15] being represented in semantic and modeled, the information is submitted to data integration ontology services of the ontology server then users can retrieve required data through the interface.

The proposed system will run through following steps

- User selection of resources and query creating using interface
- The query is send to the ontology server
- The ontology server explains the semantics of the query and transfer to the Meta data directory server
- The Meta data directory server locates the distributed crime case data base, birch Algorithm, clustering algorithm of the query on the grid nodes and returned it to the grid service
- The data grid service execute the data mining task the result is send to the meta data directory server.
- Finally user gets the result.

Content Description example

```
<process: Atomic Process RDF: ID=”DS1">
<Process: hasInput>
    Course information
</process: hasInput>
<process: hasOutput> Details of course.
</process:   has Output>
<process: hasResult> Result of the process
</process: Result>
<Process: hasResultVar>
<process: ResultVar RDF: ID=”id” />
<Process: hasResultVar>
<Process: inCondition>
<Expr: SPARQL-Condition>
<Expr: expressionBody>
SELECT Cid, Cname, desc, Duration, ame, oid
FROM db1. WHERE
(Cid, RDF: type: Course),
(Cid: has, course).
</expr: expressionBody>
<Expr: SPARQL-Condition>
</process: inCondition>
<Process: Result>
</process: hasResult>
</process: AtomicProcess>
```

The domain ontology[16] is constructed by using the traditional machine learning algorithms such as clustering, classification, regression, or active learning to construct data driven ontology for semi automatic system with Ontology Generation. The algorithms given for data mining ontology service implementation[17] are

- Clustering Analysis Algorithm
- Classification Algorithm
- Prediction Algorithm
- Text mining Algorithm

- Link Analysis Algorithm
- Evolution Analysis Algorithm
- Web mining Algorithm

Clustering Analysis Algorithm

Clustering is the task of grouping the objects of a database into meaningful subclasses so that the members of a cluster are as similar as possible whereas the members of different clusters differ as much as possible from each other. Applications of clustering in spatial databases are, e.g., the detection of seismic faults by grouping the entries of an earthquake catalog or the creation of thematic maps in geographic information systems by clustering features spaces. Different types of spatial clustering algorithms have been proposed. The basic idea of a single scan algorithm is to group neighboring objects of the database into clusters based on a local cluster condition performing only one scan through the database. Single scan clustering algorithms are efficient if the retrieval of the neighborhood of an object can be efficiently performed by the SDBS. Note that local cluster conditions are well supported by the neighbors operation on an appropriate neighborhood graph.

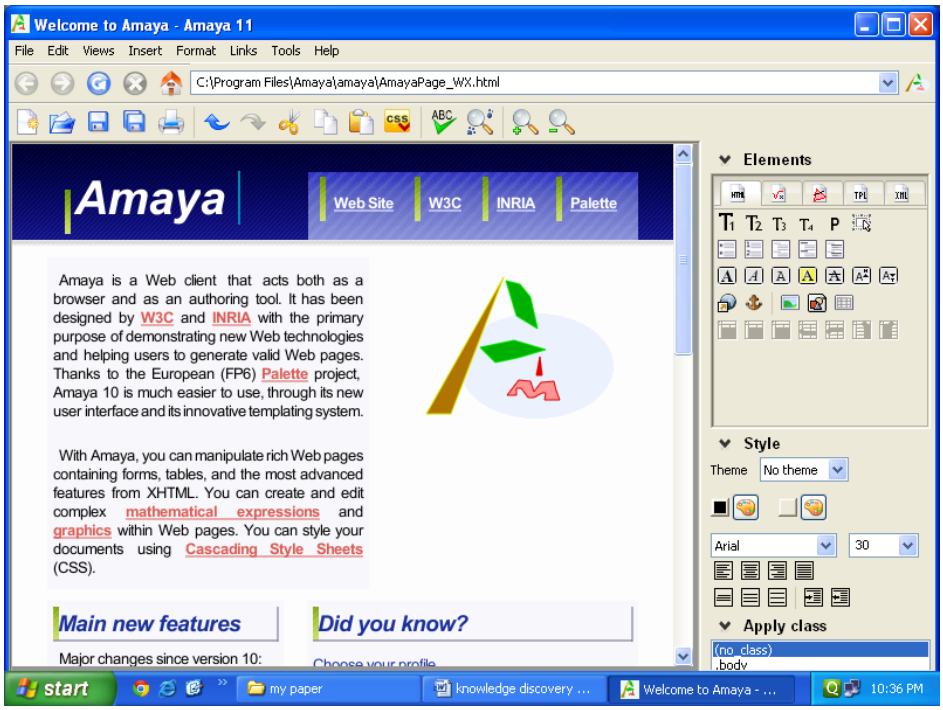
The algorithmic schema of single scan clustering

Single Scan Clustering (Database db; NRelation rel)
Set Graph to create_NGraph (db, rel);
Initialize a set Current Objects as empty;
For each node O in Graph do
If O is not yet member of some cluster then
Create a new cluster C;
Insert O into Current Objects;
While Current Objects not empty do
Remove the first element of Current Objects as O;
Set Neighbors to neighbors (Graph, O, TRUE);
If Neighbors satisfy the cluster condition do
Add O to cluster C;
Add Neighbors to Current Objects;
End SingleScanClustering;

Amaya for the knowledge discovery in the distributed grid environment

Ontology based knowledge grid in distributed environment application

The following figure 2 shows the example of knowledge discovery developed in the grid based distributed system. In many universities, institutes and organization, there can be several different departments such as UG Departments, PG



departments, R & D Center and Career Developments Center. The proposed application has four distributed databases which are included with their sample entities. These databases are stored in distributed heterogeneous environments. If a users or employee and a students need to have an advice that helps to find an optimum choice for the course for getting an admission related decision. There are several courses conducted by university but the information about them is stored in colleges' own databases which are located in heterogeneous distributed environment. So, generally students have to access several databases of the different colleges to get the proper ideas about the courses which are suitable to them. But, by using the university portal, there is no need to access several databases individually because these databases are now connected and formed data resource knowledge grid. They have to just pass out their choice and criteria have to client interface. The interface than choose the particular task to do specific task such as data mining tools, techniques and algorithm. This specific task then accesses the distributed databases via ontology server and submits the result to the Meta data directory server. Finally, the interface presents this result to the students for helping in his decision making process.

Related Work

This gives a brief review of the best-known existing projects in distributed knowledge discovery on the Grid. These emerging frameworks can be roughly classified either as domain-specific or domain-independent. Most of the Knowledge discovery frameworks on the Grid attempt to build domain-independent systems allowing the user to express specific problems.

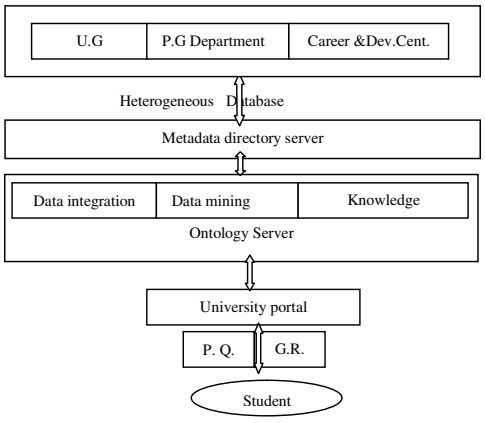


Fig. 2:

Berman F. (2001)

Proposed Tera Grid a knowledge discovery infrastructure combines different resources to create an integrated, persistent computational, and data management on the Grid in the distributed environment. The major challenges addressed by the Tera Grid here are understanding and making scientific contributions of terabytes and peta bytes of distributed data collections using simulation, modeling, and analysis techniques. Tera Grid handles wide range of domains and huge amount of data collections including bio images, proteins, digital maps of the world and the universe, research and organizations sampling data etc. Then, the synthesis of knowledge, through mining and induction for instance, from such huge amounts of data is the most challenging applications on the Tera Grid. This allows the Tera Grid to achieve its potential and enables Tera Grid infrastructure as a knowledge Grid system in the distributed environment to synthesize knowledge from heterogeneous data.

Curcin V. (2001)

Proposed a service-based computational model for knowledge discovery in the distributed grid environment which is responsible to connect end user online and use some data mining and analysis tools and techniques as data sources which are available on-line. The whole architecture of Discovery Net is composed of knowledge server, resource discovery server discovery meta information server.

Knowledge Servers

The knowledge server contains knowledge and allows the different user to store/ retrieve or publish knowledge

Resources Discovery Server

The resource discovery server publishes service definitions/locations.

Discovery Meta-information Server

The discovery meta information server stores information about each and every type of knowledge.

Discovery Net also provides a composition language called Discovery Process Markup Language representing the graph of services. This framework focuses on remote services composition and has a centralized knowledge representation for each of the composed graph services. This approach is not feasible for large-scale and complex heterogeneous scenarios in the distributed environment.

Cannataro M. (2004)

Proposed Knowledge Grid an intelligent distributed internet application environment which integrates different data mining techniques for collection sharing managing knowledge resources in the distributed environment. In this Knowledge Grid architecture, data mining tools are integrated with in Grid mechanisms and services provided by the Globus Toolkit. Its main aims to handles large volume of datasets available on the Grid for scientific and industrial applications. The Knowledge Grid project was initiated by Cannataro et al. at the University of Calabria in Italy. The architecture of Knowledge Grid is designed in such a way that include data mining and knowledge discovery techniques with lower-level mechanisms and different services provided by globus tool kit which composed of two layers:

Core K-Grid and the High-level K-Grid., the Core K-Grid layer implements basic knowledge discovery services on top of generic Grid services, while the High-level K-Grid layer is intended to design, compose, and execute distributed knowledge discovery over it. The lower layer of knowledge Grid comprises two main services: The Knowledge Directory service, and the Resource allocation and execution management service. Knowledge discovery services manages metadata including data sources and repositories, data mining tools and different algorithms, a distributed execution plans which are basically a graph describing the interactions between processes and the dataflow, and finally the results of the computation, i.e. models or patterns. There are different metadata repositories associated with these services. Resource allocation execution management service is used to map the application, i.e. the graph, into available resources.

The higher layer of knowledge grid includes services for composing, validating, and executing distributed knowledge discovery computations, as well as storing, analyzing, and presenting services. This is offered by four main services: the Data Access Service, Tools and Algorithms Access Service, Execution Plan Management Service, and Results Presentation Service. The first is used for the search, selection, extraction, transformation, and access of the datasets. The second is responsible of searching, selecting, and downloading data mining tools and algorithms. The third service generates an abstract execution plan describing the computation and the mapping onto Grid resources. The last service generates, presents and visualizes the discovered models and patterns. This framework provides a distributed data mining architecture that can benefit from 'standard' Grid services provided by Globus. The algorithmic aspect, i.e. well-adapted approaches for the Grid, is not taken into account. Knowledge Grid does not provide global knowledge sharing, semantic annotation, and overall coordination of the knowledge on the Grid in the distributed environment.

Brezany P. (2005)

Proposed Grid Miner a Grid and Web based architecture for distributed knowledge discovery based on the open grid service architecture. Each and every service of Grid Miner is implemented as a Grid service specified by Open Grid Services Architecture. The data mining process within Grid Miner is supported by several Grid services that are able to perform data mining tasks. This also supports a workflow system that controls the Grid services composition provided as a Dynamic Service Control Language document by the Dynamic Service Control client. The Grid Miner uses open grid service architecture and data integration as a standard middleware implementation of its Grid Data Service for supporting access and integration of data within the Grid. Grid Miner has a graphical user interface which provides a user friendly front-end for the end-user and the system administrator. It is similar to Knowledge Grid but the actual difference is that the Knowledge grid framework is based on a non-open grid service architecture version of the Globus Toolkit (Version 2). This work is also an architecture-

oriented approach and does not provide any algorithmic aspect on the Grid as the high-level knowledge representation and discovery in the globus distributed environment which is a major challenge for the relevant knowledge discovery.

Cannataro [18]

Proposed the ontology for the data mining domain which simplify the development of distributed knowledge discovery services on grid .It is also describe how ontology enhancing the distributed data mining applications in the grid environment.

Seidel [19]

Described the motivation and architecture of the Grid Lab project. The Grid Lab provides users and application developers with a simple and robust environment which enable them to produce applications which can exploit the full power and possibilities of the Grid. It brings together computational scientists from different application areas to design and implement a grid application toolkit , together with a set of Grid services, in a production grid environment which will provide functionality through a designed set of generic high-level APIs, through which an application will be able to catch the grid services and will demonstrate the benefits of the Grid Application tool by designing and implementing real application scenarios, illustrating wild, exciting, new uses of the Grid. Grid Lab is a balanced program with the development of a range of Grid applications along with infrastructure development, working on transatlantic test of different computing resources terminal and clusters. This experimental view ensures that the designed software enables easy and efficient use of Grid resources in a real user environment, tested by several closely related user communities.

Smirnov [20]

Discussed techniques, supporting procedures/ tasks used for implementation of knowledge logistics systems based on the KSNet approach. It is the state-of-the-art of the Knowledge Management areas, Knowledge management technologies (Knowledge fusion operations and an ontology-based methodology), a knowledge repository, multi-agent architecture of the system. It is a complex set of relations between people, processes, and technology bound together with the cultural norms, like mentoring and knowledge sharing. Knowledge management includes knowledge entry, capture of tacit knowledge, knowledge fusion etc. for knowledge discovery, knowledge representation (Knowledge base development, knowledge sharing and reuse, knowledge exchange, etc.), identifying Knowledge services, indexing knowledge, making knowledge accessible. There are many strategies and tools developed based on the algorithms of data search and retrieval in large databases, technologies of data storage and representation, etc.among them these pointed out: Microsoft Share- Portal, Knowledge Server/ Search Server, Text-To-Ontology. (Knowledge searching and retrieving from different types of documents); Disciple (Knowledge acquisition from experts and tacit knowledge revealing); Onto Edit, Protégé, onto lingua (ontology engineering); these do not consider whole problem solution but solve particular tasks. Possible applications of Knowledge logistics constitute to

- Large-scale dynamic systems with distributed operations in an uncertain and rapidly changing environment, where the information collection, assimilation, integration, interpretation, and dissemination are needed.
- Focused logistics operations and/or web-enhanced logistics operations addressing sustainable, transportation, and end-to-end rapid supply to the final destination. In this area the distributed information management and real-time information/ knowledge fusion to support continuous information and knowledge integration and exchange between all participants of the operations are Markets via partnerships with different organizations, where the dynamic identification and analysis of information sources and providing for interoperability between market participants in a semantic manner are needed. For all the above areas management systems can be described as an organizational combination of people, technologies, procedures and information/ knowledge. Knowledge logistic is based on individual user requirements, available content analysis in the information grid

environment. Hence, systems operating in this area must react dynamically to unforeseen changes and unexpected user needs, keep up-to-date resource value assessment data, support the execution of complex operations, and deliver personalized results to the customers. Proposed Knowledge logistics is based on the Knowledge fusion technology and therefore assumes integration of knowledge from different sources (probably, heterogeneous) into a combined resource in order to complement insufficient knowledge and discover new knowledge from the grid based distributed environment.

Cannataro [21]

Proposed software architecture for geographically distributed parallel distributed knowledge discovery systems is built on top of a grid which provides dependable, consistent, and pervasive access to high-end computational resources known as knowledge grid. IT USED the basic grid services and defines additional layers to implement the services of distributed knowledge discovery on worldwide connected resources where each node can be a sequential or a parallel machine. The Knowledge Grid enables the collaboration of researchers that must mine data and stored in different research centers as well as analysts that must use a knowledge system to operate on several data warehouses located in the different company establishments in distributed environment. He designed the basic principles to motivate the architecture design of a grid based parallel distributed knowledge discovery is

- Handling of large data and heterogeneity data
- Integration algorithm and independence,
- Grid infrastructure compatibility
- Openness,
- Scalability
- Data privacy and Security

Cannataro [22]

Proposed ontology based web knowledge and ontology based services for system management, applications and an initial implementation in the knowledge grid environment for the design and execution of geographically distributed high-performance knowledge discovery applications. He identified main requirement of future grid to face the enhancing complexity of Grid.

- knowledge discovery and knowledge management functionalities, for both user's needs (intelligent exploration of data, etc.) and system management;
- semantic modeling of user's tasks/needs, Grid services, data sources, computing devices (from ambient sensors to high-performance computers), to offer high level services and dynamic services finding and composition;
- pervasive and ubiquitous computing, through environment/context awareness and adaptation;
- advanced forms of collaboration, through dynamic formation of virtual organizations;
- Self-configuration, autonomic management, dynamic resource discovery and fault tolerance.

CONCLUSION

Knowledge discovery is a concept of the of computer science that shows the process of automatically searching large volumes of information for patterns that can be considered knowledge. It describe as deriving knowledge from the input data. The main branch of knowledge discovery is data mining, also known as Knowledge Discovery in Databases. It creates abstractions of the input data as many forms of knowledge discovery. The knowledge obtained through the process may become additional data that can be used for further usage and discovery. The important application of knowledge discovery is in the area of education modernization involved understanding of student query for decision making. The proposed paper is an attempt to design the semantic based knowledge discovery system in distributed environment using data mining tools ,techniques , algorithms and knowledge base repository as a server for the client in the distributed environment on grid and developed and test student information management application as a case study for verification . Future work is to enhancement in development of architecture model and advance data mining algorithm tools.

REFERENCES

1. F. Berman from Tera grid to knowledge grid, Commun. ACM **44**(11): 27-28 (2001).
2. SalmanAlSairafi, Filippia-Sofia, Emmanouil, Moustafa Ghanem,Nikolaos Giannadakis, Yike Guo, Dimitrios Kalaitzopoulos, Michelle Osmond, Anthony Rowe, Jameel Syed, Patrick Wend,el." The design of discovery net :towards open grid services for knowledge discovery" The *International Journal of High Performance Computing Applications,* **17**(3): 297–315 (2003).
3. Cannataro M, Talia Ds. "Knowledge Grid: architecture for distributed knowledge discovery" (2003).
4. Zhuge H. The knowledge grid, world scientific publishing co, Singapore (2004).
5. Aiello, M.Mango, Furnari, A.Masoorti, S.Brandi, V.Caputo and V.Barone."An experimental ontology server for information grid environment IJPP **34**(6): (2006).
6. Axel korthaus,tabias hildenboard"enabling a java and corba based knowledge grid using topic maps (2003).
7. Demenico Talia "enabling knowledge discovery services on grid" sprinter (2004).
8. K.R.Srivathsan, Manjul Sahai and Abhay Singhal Education Grid (2002).
9. M.Cannataro, D., Talia, P. Trunfio. "Knowledge Grid: High Performance Knowledge Discovery Services on the Grid." Proc. GRID 2001, LNCS, pp. 38-50, Springer-Verlag (2001).
10. Domenico Talia." Knowledge Discovery Services and Tools on Grids 2003 DEIS University of Calabria, 87036 Rende (CS), Italy.
11. Thamarai Selvi Somasundaram, R.A. Balachandar, Vijayakumar Kandasamy, RajkumarBuyya, Rajagopalan Raman3, N.Mohanram and S.Varun. Semantic-based Grid Resource Discovery and its Integration with the Grid Service Broker" (2005)
12. Sanskruti Patel, Tejas Thakkar, Priya Swaminarayan & Priti Sajja. "Development of Agent-Based Knowledge Discovery Framework to Access Data Resource Grid" Advances in Computational *Sciences and Technology*. **3**(1): pp. 23-31 (2010).
13. P. Brezany1, Janciak1, A. W¨ohrer1, and A M. Tjoa2I. Grid Miner: A Framework for Knowledge Discovery on the Grid - Design and Implementation (2004)
14. Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation Standard,http://www.w3.org/TR/1999 syntax-19990222//REC-rdf
15. Grcar, M., Mladenic, D., Grobelnik, M., Fortuna, B., Brank, J. Ontology Learning Implementation. Project report IST-2004-026460 TAO, WP 2, D2.2 (2006)
16. M Grcar, M Grobelnik, D Mladenic. Mining Complex Data, 2008 -Springer
17. XindongWu Vipin Kumar · J. Ross Quinlan Joyd. 'algorithms in data mining' Springer-Verlag LoData Mining and Knowledge Discovery in Real Life Applications, Book edited by: Julio Ponce and Adem Karahoca,
18. Ed Seidel a, Gabrielle Allen a, André Merzky b, Jarek Nabrzyski c." GridLab—a grid application toolkit and testbed" *Future Generation Computer Systems* **18**: 1143-1153 (2002).
19. Lamine Aouad, An Le-Khac and Tahar Kechadi. "Knowledge Discovery on the Grid" 438, (2009).
20. Alexander Smirnov, Mikhail Pashkin, Nikolai Chilov, Tatiana Levashova." Knowledge logistics in information grid environment" *Future Generation Computer Systems* **20**: 61-79 (2004).
21. Cannataro M."Knowledge Discovery and Ontology- based services on the Grid" (2003).
22. Mario Cannataro and Domenico Talia." knowledge grid An Architecture for Distributed Knowledge Discovery (2002).