Abstract

The development of real life problem is not the key issue. Various information technologies available now a days. But the major issue is that how to get more advantage 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 Semantic Web in distributed environment. 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.

Keywords: knowledge grid; ontology; distributing; meta data;

I. INTRODUCTION

The development of real life problem is not the key issue. Lots of information technology available but the major issue is that how to get more advantage of these technologies for academic purpose in distributed environment where faculty and students communicate with software technology rather than with individual. Originally Fran Berman [1] introduced the concept of knowledge based grid which is more effective for scientist and academician. Berman also focus 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 focus knowledge based grids technique based on data mining [2]. Hzhuge focused the future internet and gave definition of knowledge grid, resource semantic grid and social grid [3] 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 education Grid[4] offer the processes involved in the developing, deploying and utilizing E-course for deployment in the State of Kerala in India. This is made possible by advanced technology web-servers based on a new architecture the Info-Space Operating System. Education Grid also help 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 [5], above research consider the basic concept, design of knowledge based grid and suggest related data mining techniques and algorithm which is useful to design knowledge grid and its application in distributed environment mainly. There fore little amount of work is developed on its application. The issue now is to support quality educational and student decision making for taking admission on their choice in distributed environment processes so that 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 on the basis of agent technology and integrate grid data resource which provides different services to the users of the university like searching, advisory and help desk system, course and resource management, student and faculty communication etc. Web knowledge-based task, content is changed into a collection of information, i.e. understanding the context, format, and significance of the data and information called knowledge discovery.

II. THE WEB KNOWLEDGE BASED GRID

Web Knowledge based grid is an 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 support on demand, annotative and strong services which is useful for innovation and collaborative work in distributed heterogeneous environment. The web Knowledge Grid architecture is design on grid tools and services, i.e. it considers basic grid services to design specific knowledge extraction services [6]. 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.

### III. 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 [7] 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 which 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.

### IV. THE PROPOSED SYSTEM

The proposed system contains web portal interface, ontology server, metadata knowledge directory server, metadata database and knowledge base. 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[8] and transfer these generated semantics to the knowledge directory server locates the distributed heterogeneous data bases then transfer it to the grid data .The grid executes the data mining task and returned the result to the meta data directory server .The knowledge integration integrate effective knowledge which is browsed and discover through the web portal interface to the user.

![Figure 1: The portal Interface](image)

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 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 some one 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 as a result of the data mining process. i.e. some knowledge gathering from application domain

IMPLEMENTATION OF PROPOSED SYSTEM

We will design semantic web interface in java for semantic query and heterogeneous database in xml. XML/Schema [9] represent the knowledge in semantic acquired from 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 [10] 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 [11] 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

```xml
<process:AtomicProcess rdf:ID="DS1">
  <process:hasInput>
    course information
  </process:hasInput>
  <process:hasOutput>
    Details of course.
  </process:hasOutput>
  <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, name, oid
        FROM db1. WHERE
        (Cid, rdf:type, :Course),
        (Cid, :has, course).
      </expr:expressionBody>
    </expr:SPARQL-Condition>
    <expr:expressionBody>
      SELECT Cid, Cname, desc., Duration, name, oid
      FROM db1. WHERE
      (Cid, rdf:type, :Course),
      (Cid, :has, course).
    </expr:expressionBody>
  </process:inCondition>
  <process:Result>
    <process:hasResult>
    </process:hasResult>
  </process:Result>
</process:AtomicProcess>
```

The domain ontology is constructed by using the traditional machine learning algorithms such as clustering, classification, regression, or active learning [12, 13] which is also helpful to construct data driven ontology for semiautomatic system with OntoGen. The algorithms useful for data mining ontology service implementation 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 (that is, clusters) 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)

1. set Graph to create_NGraph(db, rel);
2. initialize a set Current Objects as empty;
3. 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;
   end if
4. 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 if
5. end SingleScanClustering

Example of application

The following figure 2 shows the example of knowledge discovery which are under development process. In many universities, there can be several 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 who are a students need to have an advice and help to find an optimum choice for the course for getting an admission. There are several courses conducted by university but the information about them is stored in colleges' own databases which are located in distributed environment. So, generally a student has to access several databases of the different colleges to get the proper idea about the course which is suitable to him. 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 interface. The interface than choose the particular task to do specific task. 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.

SUMMARY AND FUTURE SCOPE

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 which involves understanding of student query for decision making. Through reuse software composition by the building of structure the systematic grid application resource will describe by the use of ontology. We proposed the knowledge work flow knowledge discovery in distributed environment for building large scale distributed knowledge system on grid. Future work is to enhancement in development of architecture model and advance data mining algorithm tools.

REFERENCES

[6] Domenico Talia Knowledge Discovery Services and Tools on Grids 2003 DEIS University of Calabria, 87036 Rende (CS), Italy
[13] [12], M Grcar, M Grobelnik, D Mladenic - Mining Complex Data, 2008 – Springer