A KNOWLEDGE DEPICTION IN DIGITAL LIBRARY BY INTEGRATING MULTI-AGENT SYSTEM FOR DISTRIBUTED DATABASE GRID

SANSKRUTI PATEL¹, TEJAS THAKKAR², AND PRITI SAJJA³

ABSTRACT: A Digital Library is a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers. Architecture of Digital Library operates in heterogeneous and dynamic information environment can be modeled by integrating the concept of an agent. Multi-agent systems have received much attention in recent years because of their many advantages in complex and distributed environments. Knowledge discovery from heterogeneous data sources available on Data Grid environment is a challenging research and development issue. This paper covers all aspects of the knowledge discovery process and integrates this process with service-oriented grid application supported by agent framework for digital library. The GGF (Global Grid Forum), Open Grid Services Architecture (OGSA), and its associated specifications define consistent interfaces through web services to components of a grid infrastructure. The architecture presented in this paper, shows how different heterogeneous data resources in digital library are integrated via OGSA-DAI framework and how knowledge discovery process is performed in distributed data resource environment by using multiple task agents. Our proposed model provides an agent-based framework which allows Digital Libraries to be considered, designed and developed as information environments being composed of different co-operating agents.

Keywords: Digital Library, Multi-agent Systems, KQML, Knowledge Discovery

1. INTRODUCTION

An agent is a software program that automatically performs tasks on behalf of the user [6]. A multi-agent system is a system composed of multiple interacting intelligent agents. One of the current factors fostering multi-agent development is the increasing popularity of the Internet, which provides the basis for an open environment where agents interact with each other to reach their individual or shared goals. A multi-agent system is a loosely coupled network of problem-solver entities that work together to find answers to problems that are beyond the individual capabilities or knowledge of each entity [7]. The agent paradigm is successfully employed in those applications where autonomous, loosely-coupled, heterogeneous, and distributed systems need to interoperate in order to achieve a common goal like a Digital Library. In a multi-agent system the agents communicate between them in order to fulfill the global goal or their local goals. Also, through the collaboration between different agents, we are aiming to achieve a highly efficient, flexible, customizable system that provides better communication, interaction and management among all users of a Digital Library. In the architecture a Digital Library is seen as a decentralized collection of interacting self-interested agents where an agent represents the knowledge and interests of an individual
user. Each agent is perceived as a personal digital library serving one individual user (or other information source). Software agents have proved to be the best for handling distributed information issues like information retrieval, integrity and navigation assistance. Digital Library is large information repositories built on distributed environment. In this paper, we proposed architecture of a Digital Library by imparting Multi-agent systems.

2. CONCEPT OF DIGITAL LIBRARY
A Digital Library is a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers [11]. The digital content may be stored locally, or accessed remotely via computer networks. A digital library is a type of information retrieval system. An organization, which might be virtual, that comprehensively collects, manages and preserves for the long term rich digital content, and offers to its user communities specialized functionality on that content, of measurable quality and according to codified policies. According to Arms a digital library is a managed collection of information with associated services where the information is stored in digital format and accessible over a network. The digital library federation in the USA defines the digital library as: Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities. To summarize, the Digital Library is:

- Organized collection of multimedia and other types of resources.
- Resources are available in computer process able form.
- The function of acquisition, storage, preservation, retrieval is carried out through the use of digital technology.
- Access to the entire collection is globally available directly or indirectly across a network.
- Support users in dealing with information objects
- Helps in the organization and presentation of the above objects via electronic/digital means etc.

The Internet and World Wide Web provide the impetus and technological environment for the development and operation of a digital library. The Internet provides the TCP/IP and or its associated protocol for accessing the information and web provide tools and technique for publishing the information over Internet through the realization of a Digital Library.

2.1. Digital Library Vs. Traditional Library
A major advantage of digital libraries is that people from all over the world can gain access to the information at any time, as long as an Internet connection is available. This also eliminates the logistical problems inherent in organizing and lending print materials which
are present in traditional physical library. Whereas traditional libraries are limited by storage space, digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain them. As such, the cost of maintaining a digital library is much lower than that of a traditional library. A traditional library must spend large sums of money paying for staff, book maintenance, rent, and additional books. Digital libraries do away with these fees.

3. CONCEPT OF AGENT BASED SYSTEM

Agents are often described as entities with attributes considered useful in a particular domain. There are several types of agents like intelligent agents, information agents, mobile agents, personal assistant agents, and other types of agents. This is the case with intelligent agents, where agents are seen as entities that emulate mental processes or simulate rational behaviour; personal assistant agents, where agents are entities that help users to perform a task; mobile agents, where entities that are able to roam networking environments to fulfil their goals; information agents, where agents filter and coherently organize unrelated and scattered data; and autonomous agents, where agents are able to accomplish unsupervised actions[5].

As part of the study of agent systems, researchers began to develop (purposely or otherwise) terminology for agents. Some of these terms are described as follows [3, 5]:

- Agent Architectures analyze agents as independent reactive/proactive entities. Agent architectures conceptualize agents as being made of perception, action, and reasoning components. The perception component feeds the reasoning component, which governs the agents’ actions, including what to perceive next.

- Agent System Architectures analyze agents as interacting service provider/consumer entities. System architectures facilitate agent operations and interactions under environmental constraints, and allow them to take advantage of available services and facilities.

- Agent Infrastructures provide the regulations that agents follow to communicate and to understand each other, thereby enabling knowledge sharing. Agent Infrastructures deal with the following aspects:
  - Ontologies: allow agents to agree about the meaning of concepts.
  - Communication Protocols: describe languages for agent communication.
  - Communication Infrastructures: specify channels for agent communication.
  - Interaction Protocols: describe conventions for agent interactions.

A multi-agent system is a loosely coupled network of problem-solver entities that work together to find answers to problems that are beyond the individual capabilities or knowledge of each entity [2]. A Multi-agent system (MAS) is a system composed of multiple interacting intelligent agents. Multi-agent systems can be used to solve problems which are difficult or impossible for an individual agent or monolithic system to solve. Multi-agent systems are a
new paradigm for understanding and building distributed systems, where it is assumed that the computational components are autonomous: able to control their own behavior in the furtherance of their own goals. Some of the Agent Applications are education, workflow management, network management, air-traffic control, digital libraries, personal digital assistants, business process engineering, e-mail filtering, information management, data mining, electronic commerce and many more. Agents can share knowledge using any agreed language, within the constraints of the system's communication protocol. Example languages are Knowledge Query Manipulation Language (KQML) or FIPA's Agent Communication Language (ACL).

4. KNOWLEDGE DISCOVERY IN DISTRIBUTED DATABASES (KDD)
Knowledge discovery is defined as “the non-trivial extraction of implicit, unknown, and potentially useful information from data”. Knowledge discovery describes the process of automatically searching large volumes of data for patterns that can be considered knowledge about the data. It is often described as deriving knowledge from the input data. This complex topic can be categorized according to (1) what kind of data is searched; and (2) in what form is the result of the search represented. Knowledge Discovery refers to the overall process of generating patterns from data. The following steps are necessary for this. These steps are known as the “knowledge discovery cycle”. They are: Acquisition, Representation, Storage, Analysis, Visualization, Interpretation and Deployment.

Traditional centralized knowledge discovery in databases (KDD) is reaching its limits due the fact of increasingly high volume of data, big calculation tasks and the geographical distribution of the involved/required (data) resources. The OGSA-DAI based KDD provides services used to extract knowledge from the data stored inside the heterogeneous database Grid. These services will be used both to build high-level knowledge discovery applications, as in the case of the Knowledge Grid, and to enhance existing basic Grid services.

5. A PROTOTYPE MODEL OF AGENT-BASED DIGITAL LIBRARY
The overall architecture of the Agent based digital library system is presented in Figure 1. We have followed the FIPA agent management specification (FIPA SC00023K 2004), which is one of the most widely adopted agent management standards available. The Foundation for Intelligent Physical Agents (FIPA) was formed in 1996 to produce software standards for heterogeneous and interacting agents and agent-based systems. FIPA is a non-profit organization who dedicates their efforts to the standardization of agent-based technologies and multi-agent systems. FIPA specifications represent a collection of standards which are intended to promote the interoperation of heterogeneous agents and the services that they can represent. As shown in Figure 1, the system consists of seven major components: Directory Facilitator Agent, Agent Management Service Agent, Interface Agent, Search Agent, Publishing Agent, View Agent, and User Authentication Agent. All those agents interact and work together to provide different services to the end users, related to Digital
Library. In Figure 1, the Agent Management Service Agent and the Directory Facilitator Agent are used from the FIPA agent management specification. According to FIPA agent management specification, the Agent Management Service Agent is responsible to manage basic operation of an agent platform (FIPA SC00023K 2004) like creation and/or deletion of agents, query platform profile, authentication of agents, and the registration and/or deregistration of agents. In short it provides the basic control of agents for the multi-agent system. According to FIPA Agent Management Specification, a Directory Facilitator (DF)

![Figure 1: An Architecture of Agent Based Digital Library Accessing Distributed Database Grid](image-url)
Agent is an optional component of the Agent Platform, but if it is present, it must be implemented as a DF service. The DF provides yellow pages services to other agents. Agents may register their services with the DF or query the DF to find out what services are offered by other agents. Multiple DFs may exist within an Agent Platform and may be federated. The DF is a reification of the Agent Directory Service in [FIPA00001]. At any time, and for any reason, the agent may request the DF to modify its agent description. An agent may search in order to request information from a DF. The Interface Agent is used to connect end users or external systems together with the multi-agent system. It reacts to different request made by the end user or external system, and translates these commands into agent understandable requests, and sends them to appropriate agents [4]. The ContentView agent provides the representation of the content through multi-page document. It also provides functions for marking passages, selecting existing markup, zooming and flipping through the document pages. The publisher Agent is responsible to interact between the publisher and the authors or the consumers of a Digital Library. Search Agent is available for database queries, pictorial search, full text search and advanced search (on metadata and annotations) [8].

The database for the digital library may not store centrally. Rather, each component has its own location to store the database. So, in such scenario, the data resources are stored on different locations and they can be heterogeneous or homogenous. We are proposing OGSA-DAI framework for deployment a database grid environment to integrate all the distributed heterogeneous or homogeneous databases for digital library. On the client layer, there are two types of agents: interface agent, which interacts with individual user; task agents, which are responsible to carried out different problem-solving tasks and produces different functionalities. Users first have to contact interface agent in order to contact task agents. Different task agents facilitate the user by providing various facilities.

6. KNOWLEDGE DISCOVERY PROCESS IN DISTRIBUTED DATA RESOURCE GRID ENVIRONMENT FOR DIGITAL LIBRARY

In digital library architecture, the content repositories may store in distributed environments and are now connected and formed data resource grid. To access the system, a user has to just pass out his request to an interface agent. These interface agent than choose the particular task to do a specific task. This specific task agent then accesses the distributed databases via OGSA-DAI environment and submits the result to the interface agent. Finally, the interface agent presents this result to the user for helping in knowledge depiction process. In order to communicate with one another, a specific agent remains idle while no messages arrive. When a specific message is arrived from any other agent, it interprets the message and act accordingly by using a message queuing mechanism. All agents in the system work on this principle and remain idle while no messages arrived. Also, each agent can send a message to any other agent meanwhile. Web services are used implement the different task
agents. Separate web services can be created to justify task of particular agent. Also, these web services can be created using any language platform which is supported the creation of web services.

To run OGSA-DAI, we need to install java 1.4 or 1.5, Jakarta tomcat 5.5 or 5.0, and apache 1.6 or above. The OGSA-DAI client toolkit consists a JAVA API that provides the basic building blocks for OGSA-DAI client development. Using these APIs, a developer can construct anything from a basic query client to a complex distributed data integration client with relative ease. The Client Toolkit minimizes the specialist knowledge required to interact with OGSA-DAI services and provides some protection from future changes to the data service interfaces. It also operates transparently with OGSA-DAI data services compliant with both WS-I and WSRF. The main concept of the Client Toolkit is that of an activity. An activity dictates an action that is to be performed by a data service resource. OGSA-DAI provide many different types of activity to perform operations such as SQL queries, XSL transformations and FTP data delivery. The Client Toolkit provides a simple mechanism for assembling requests describing multiple activities, then submitting the request to data service resources for processing. To deploy a new activity onto an OGSA-DAI server do the following. If the activity class is not already in a JAR on the server then we should create a temporary directory containing this JAR [12].

For example:

$ mkdir tmpActivityDir
$ cp my-activity.jar tmpActivityDir

Now, to deploy the new activity run the deployActivity command this takes the following arguments: [1] dai.activity.jar.dir - location of directory containing the activity implementation class. If the activity implementation class is already on the server then this value can be omitted. [2] dai.activity.id - activity ID. Unique ID by which the activity is known on the server. [3] dai.activity.class - name of activity implementation class. [4] dai.activity.description - human-readable description of the activity. This value is optional.

Then after, if using OGSA-DAI Axis on Tomcat run by following commands:

$ ant -Dtomcat.dir=/PATH/TO/TOMCAT \  
-Ddai.activity.jar.dir=ACTIVITY-JAR-DIR \  
-Ddai.activity.id=ACTIVITY-ID \  
-Ddai.activity.class=ACTIVITY-CLASS \  
-Ddai.activity.description=ACTIVITY-DESCRIPTION deployActivity

This takes the following arguments:
tomcat.dir - location of OGSA-DAI Axis deployment on Tomcat
For example:

$ ant -Dtomcat.dir=/home/user/tomcatAxis \
   -Ddai.activity.jar.dir=tmpActivityDir - \
   Ddai.activity.id=MyActivityID \
   -Ddai.activity.class=org.my.ActivityClass \ 
   -Ddai.activity.description="This is my new activity" deployActivity

Also, the OGSA-DAI can be directly implemented by the Globus Toolkit, which is an open source software toolkit used for building Grid systems and applications. The Globus Alliance and many others all over the world are developing it. Globus allows people to share computing power, databases, and other tools securely online across corporate, institutions, and across geographic boundaries.

7. CONCLUSION

This paper is aimed at applying Multi-agent approach to provide different services in Digital Library. We have proposed Agents Based Knowledge Discovery Framework Accessing Data Resource Grid for digital library that can be used to provide generalized and flexible solutions applicable to all digital library units. Besides giving advantages like effective retrieval and presentation of information, identifying user need, evaluation of user responses, and providing advisory, help and searching facilities, here we proposed a new library environment by integrating agents with OGSA-DAI and Data Grid environment. As future enhancement, we can use this system by adding Library Ontology in the system to generate the Knowledge Grid. The combined use of Knowledge Grid, Ontologies and Multi-agent technologies will enable the sharing of heterogeneous, autonomous knowledge sources in a capable, adaptable and extensible manner.

REFERENCES

A Knowledge Depiction in Digital Library by Integrating Multi-Agent System....


[9]. http://www.dlib.org/dlib/July95/07birmingham.html

[10]. www.fipa.org


Sanskruti Patel
Lecturer, Faculty of Computer Science & Applications, Charusat-Changa
E-mail: sanskrutipatel.mca@charusat.ac.in,

Tejas Thakkar
Lecturer, Computer Science Department, NVPAS, V V Nagar
E-mail: ththakkar@gmail.com,

Dr. Priti Sajja
Associate Professor, Computer Science Department, Sardar Patel University, V V Nagar
E-mail: Priti_sajja@yahoo.com