

Design and Development of Virtual Distributed Database for Web-based Library Resource Sharing Network for Orissa Technical and Management Institutions

*Prasanna Kumar Barik, **D. B. Ramesh

***Librarian**

C.V .Raman College of
Engineering, Bhubaneswar,
*Prasanna_barik@yahoo.co.in

****Head, Library**

Institute of Minerals and
Materials Technology (IMMT),
CSIR, Bhubaneswar

Correspondence to:

Prasanna Kumar Barik
parsanna_barik@yahoo.co.in

ABSTRACT

Focuses on designing and developing of a unique virtual distributed database for the web based library network for Private Technical and Management Institutions of Orissa with a hope to set up a scholastic model of web-based library network through a web enabled Virtual Distributed Database provide web-based network services to the member libraries. Possibility of building up library cooperation among private technical and management institutions of Orissa for resource sharing system is also examined.

Introduction

The emerging Information Technologies have brought many changes in libraries. These technologies have facilitated LIS professionals to work together to acquire and share library collection and provide digital library services to the user community. Increase of the literature in all subjects and shrinking of the library budget made the libraries depend upon each other. With the advent of Internet and its associated web technologies in the past decade have significantly influenced the libraries to provide web based services. The dynamic linking capability of the web has provided users an unrestricted access to information resources. Widespread use of converging technologies in the realm of library and information management has augmented the process of collaborative collection development for maximizing information access. There is a paradigm shift in the process of resource sharing from print to electronic media, and libraries in Orissa are coping with the dynamics of changing technology-driven work process. The change is necessitated by evolving consciousness of users about their demand, new information format and the ability of Information and Communication Technologies to serve them efficiently. Library networks have grown mostly during the last thirty years in

different geographical environment in order to cater to the specific needs of the users. Several models have emerged that provided specific services. However, the essential functions should include the promotion of Resource Sharing, creation of resource sharing tools like Union Catalogues, rationalization of acquisition and maintenance of International standards for creation of records uniformly. The growth of library Networks in India can be traced to the initiatives made by NISSAT in establishing CALIBNET in 1986, DELNET in 1988 and other networks subsequently. University Grants Commission (UGC) established INFLIBNET in 1988. DELNET has emerged as the first operational library network in India with the support of the National Informatics Centre.

New challenges and requirements for the management of resource sharing in digital environment require improved data interoperability, data integration and data sharing. The Virtual Database is an example of an integrated information system which aims at integrating different distributed data repositories to build an integrated data federation based on open standards and Internet technology. Data interoperability techniques and methods are developed for integrating and sharing data. An attempt has been made in his paper to describe

the conceptual design and implementation of the Virtual Distributed Database as an integrated information system as it consists of a framework for advanced Web-based retrieval, analysis, and visualization of related data based on the integration of distributed data repositories. The implementation is based on modular and service-oriented software design.

What is OTAMILNET?

In Orissa, the much rhetoric and effort in favour of planned cooperative sharing of resources through web based library network is inevitable. The proposed model is primarily named after OTAMILNET. It stands for ORISSA TECHNICAL AND MANAGEMENT INSTITUTIONS LIBRARY NETWORK. The establishment of the proposed web-based library network among the technical and management institutions in Orissa is highly essential to bring these institutions under an umbrella of a network to provide the web-based library services. It is a co-operative effort of individual libraries to create the proposed library network for the region. It is a source or platform to the participatory libraries to achieve mutual exchange and share of information resources. It is based on co-operation in the creation and management of resources as well as on shared use of resources. It will be a major resource sharing library network among the participating libraries which will cover a wide spectrum of member libraries and their resources. Through sharing and co-operation, this project will facilitate information access and minimize acquisition costs of management and technological periodicals, resulting into increased user satisfaction. There is a great need to consider the entire resources of a single entity that can be made available to the entire user community. Networking or formation of consortia among libraries of the region is the best way to make the resources sharing very effective and also to ensure equitable availability of required information. This regional network must communicate frequently and broadly with local and national level networks.

Growth of Private Technical and Management Institutions in Orissa

Most of the Technical and Management Institutions in Orissa are private set up by educational societies/ trusts, philanthropic organizations etc. These are running graduate and post-graduate courses. A large number of these institutions have not seriously thought of developing library systems. In fact it would be interesting to study as to how much percentage of the entire allocation, is invested on this most important activity of all the academic activities over the period of their existence. Networking or formation of consortia among libraries of the region is the best way to make the resource sharing very effective and also to ensure equitable availability of required information.

Growth and status of libraries as presented in the above table in private technical and management institutions of Orissa

Sl. No.	Year of Growth	No. of Technical Institutions	No. of Management Institutions	Total
1	1980 - 1984	2	4	6
2	1985 - 1990	1	Nil	1
3	1991 - 1995	Nil	1	1
4	1996 - 2000	18	5	23
5	2001 - 2005	21	6	27
6	2006 - 2010	80	41	121

The growth and status of libraries as presented in the above table in private technical and management institutions of Orissa is not encouraging especially in the context of the developed countries where user friendly methods and models with a lot of flexibility functions are available effectively. Private Technical and Management Libraries in Orissa are passing through a very crucial period. The present scenario of these libraries in providing web-based services to the users is a matter of mirage. There is no significant and planned development in these libraries in comparison to the libraries of the contemporary in other states of India such as Kerala, Maharashtra, and Tamilnadu. There are one hundred twenty-two technical and fifty-two management institutions approved by AICTE and affiliated under Biju Pattanaik University of Technology (BPUT). Near about fifty thousands of users including students and faculties are the beneficial users of these libraries. It is studied from the websites as well as personal visit to of all the institutions that no library is able to provide web-based services to their users. Many leading institutions have been providing only textual information about their libraries and its holdings through their institutions portals. So many institutions have no library portal and also no information about their libraries. Twenty-seven institutions are the member of INDEST-AICTE consortium to access e-databases which are provided by various e-database vendors. There are many institutions which have no separate library building but the libraries are running in the place designed for class room purposes. Many libraries are not providing Internet facilities and also reference services to their users. The present condition of the libraries in private technical and management institutions are not in the state of providing services to meet the user's requirements. It is expected that the proposed web-based library network will no doubt serve the users by providing web based service in right time at their desktop. Library will be everywhere without confined in the building only.

Distributed Database Management System (DDBMS)

"A distributed database (DDB) is a collection of multiple, logically interrelated databases distributed over a

computer network. A DDBMS is the software system that permits the management of DDBs and makes the distribution transparent to the user."

A distributed database can be defined as consisting of a collection of data with different parts under the control of separate database management systems running on independent computer systems. All the computers are interconnected and each system has autonomous processing capability serving local applications. Each system participates, as well, in the execution of one or more global applications. Such applications require data from more than one site. The distributed nature of the database is hidden from users and this transparency manifests itself in a number of ways. The database is stored across computers data is combined from several servers many computers may help in the processing of one query machines may be both a client and a server.

The idea is that instead of having one , centralized database , we are going to spread the data out among the sites on the distributed network , each of which has its own computer and data storage facilities network , each of which has its own computer and data storage facilities . All of this distributed data is still considered to be a single logical database. When a person or process anywhere on the distributed network queries the database, it is not necessary to know where on the distributed network the data being sought is located. The user just issues the query, and gets the result.

Aim of Designing Virtual Distributed Database

The primary goal of the Virtual Database is to act as a central access point to distributed data resources from the main server of the Orissa Technical and Management Library Network (OTAMILNET). Each institution stores its data in its own database which results in a number of heterogeneous data types and database systems. The Institutional database is automatically merged into the Union database of the hosting server. Hence; the detailed requirements of the Virtual Database are as follows:

- ◆ Establishment of a computing platform which integrate and shares distributed data repositories for combine retrieval, query, analysis, and display.
- ◆ The autonomy of involved institutions and databases must be guaranteed. That is, individual database schemes, maintenance procedures and applications must remain unaffected.
- ◆ In expectation of a long life cycle, the architecture and implementation of the Virtual Database must be extendible and easy to maintain. Extendibility and ease of maintenance are quality aspects of any software. These goals are ensured by a modular software design (Meyer, 1997) and by adhering to service-oriented software design principles (Erl, 2005).

The architecture of the Virtual Database therefore strongly relies on Web services technology, open standards, open interfaces and the reuse of open source software components.

Functions and Components of Virtual Distributed Database Management Systems

A Distributed Database management System governs the storage and processing of logically related data over interconnected computer systems in which both data and processing functions are distributed among several nodes. It must have at least the following functions and components.

Functions:

- ◆ Application interface to interact with the end user, application programs , and other DBMS within the distributed database
- ◆ Validation to analyze data requests for syntax correctness
- ◆ Transformation to decompose data request components
- ◆ Query optimization to find the best access strategy
- ◆ I/O interface to read or write data from or to permanent local storage
- ◆ Security to provide data privacy
- ◆ Backup and recovery to ensure the availability and recoverability of the database in case of a failure
- ◆ Transaction management to ensure that the data moves from one consistent state to another. This activity includes the synchronization of local and remote transactions as well as transactions across multiple distributed segments.
- ◆ Search for , locate , read and validate the data

Components:

- ◆ Computer workstations (sites or nodes) that form the network system. The distributed database system must be independent of the computer system hardware
- ◆ Network hardware and software components that reside in each work station. The network components allow all sites to interact and exchange data. Because the components - computers, operating systems, network hardware and so on - are likely to be supplied by different vendors, it is best to ensure that distributed database functions can be run on multiple platforms.
- ◆ Communication media that carry the data from one workstation to another. The DDBMS must be communication - media - independent; that is , it must be able to support several types of communications media.
- ◆ The transaction processor, which is the software components, found in each computer that

requests data. The transaction processor receives and processes the application's data requests. The transaction processor is also known as the application processor or the transaction manager

The data processor which is the software component residing on each computer that stores and retrieves data located at the sites. The data processor is also known as the data manager. A data processor may even be a centralized DBMS.

network components.

The design of a distributed computer system involves making decisions on the placement of data and programs across the sites of a computer network, as well as possibly designing the network itself. In the case of OTAMILNET distributed database management system, distribution of applications involves two things: the distributed DBMS software and the distribution of the application programs that run on it. We concern ourselves to the decisions related to the distributed design. It has been suggested that the organization of distributed systems

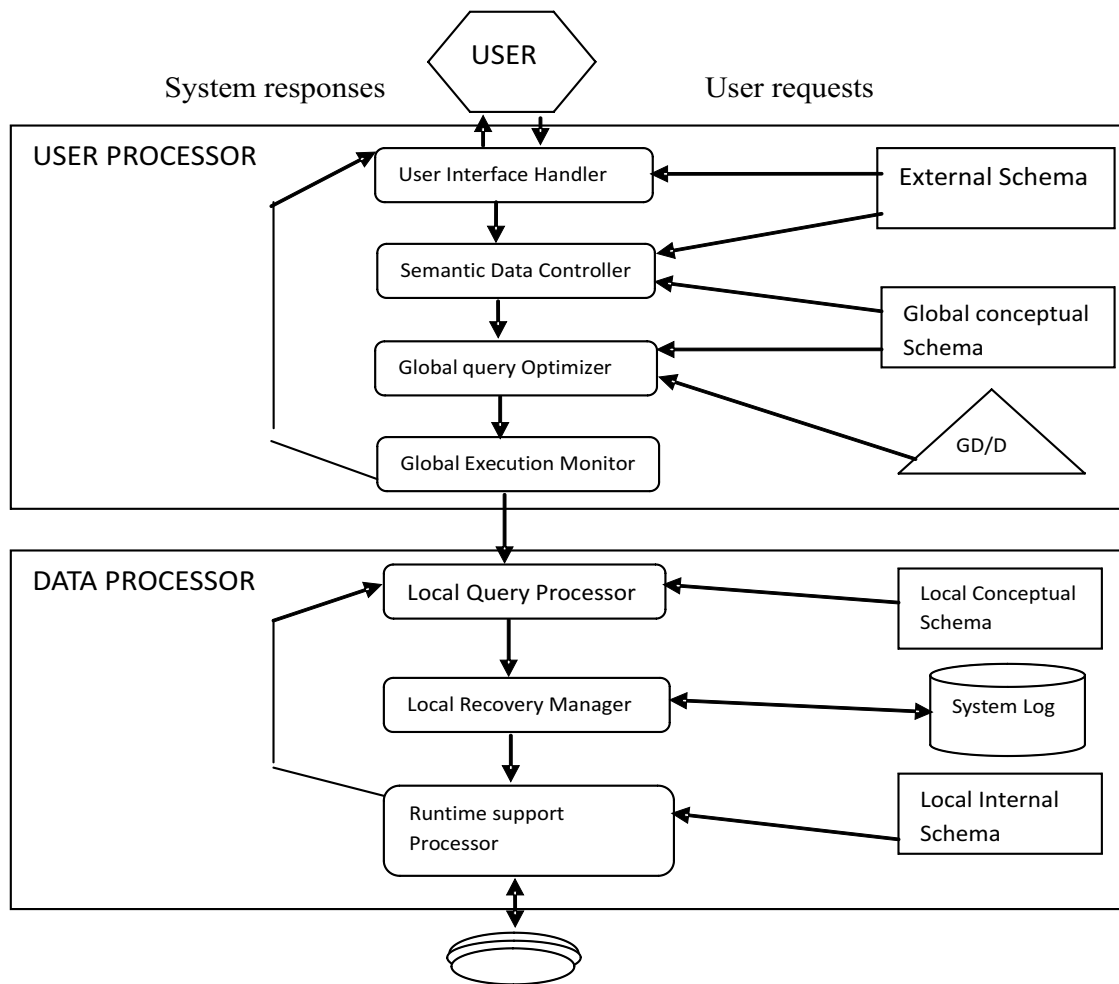


Fig. 1: Components of Distributed Database Management Systems

Conceptual Design of the Virtual Distributed Database

The architecture of a system defines its structure. This means the components of the system are identified, the function of each component is specified and the interrelationships and interactions among these components are defined. This general framework also holds true for computer systems in general and software systems in particular. This architecture is a particulate type of distributed computing networks which is very popular and among the most widely used today for business computing networks. This architecture concerns how processing activity is distributed over the network, regardless of the topological arrangement of the

can be investigated along three orthogonal dimensions - Level of sharing, Behavior of access patterns and Level of knowledge on access pattern behavior

The primary goal of the Virtual Distributed Database is to bring all technical and management institutions under an umbrella of a web based network to share the distributed data resources among these member institutions. Each institution stores its data in its own database which automatically merged into the main database of the central server of the hosting institution. Three Tier architectural systems is proposed to implement for the virtual database.

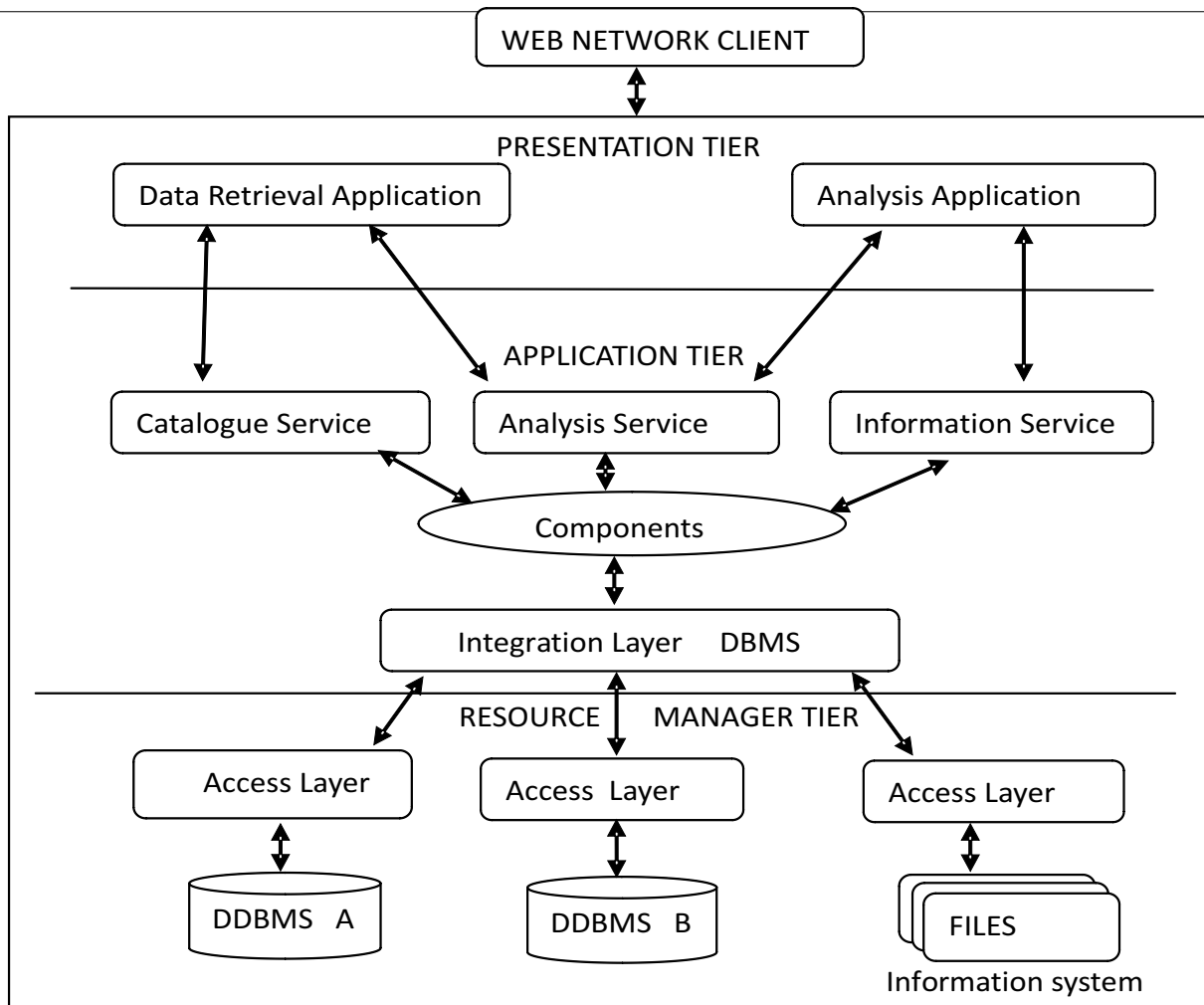


Fig. 2: Three-tier architecture of the Virtual Database

The above picture shows the so called three-tier architecture of the Virtual Database consisting of a presentation tier, an application tier, and resource manager tier with loosely coupled software modules. This design pattern is commonly used in Web applications and suits our needs for distributed processing.

Development of Web-based Resources Sharing Strategy:

Members of the participating libraries in the networking and resources sharing systems must have personal computer with CD drives and at least Internet facilities to the users. High speed and large storage capacity computer may be installed to support networking requirements. The every member of the OTAMILNET should have kept the following factor in mind for the effective management of the network.

- ◆ The participating libraries should recognize that other libraries should develop an external programme to meet their requirements effectively and efficient mechanism will have to be developed to respond to their request or materials

- ◆ Every participating library must have a core collection to meet the requirement of its own client.
- ◆ The participating library should recognize its challenging role as members of the network by participating continuously in co-operative programme which may strengthen the resources of the member libraries collectively.
- ◆ The participants will build up their resources on co-operative basis so as to ensure participation by each member libraries to share available resources.
- ◆ Every participating library will have mandate from its own authority to participate in the networking programme.
- ◆ The participating library in the network will maintain the spirit of "give and take" policy.

Implementation of Virtual Distributed Database

Having described the general architecture of the Virtual Database, we are now going to outline how the design was mapped to an actual implementation. We chose the Java 2 Platform and the corresponding Web technologies for

developing the component-based three-tier application, since it offers the necessary tools and guarantees platform independence. Particular APIs as described below are applied for solving data access and service incorporation.

1. Presentation tier

The presentation tier contains applications that handle user interaction and display. These applications are designed as Web applications which use the services exposed by the application tier. Both applications may be built on the same services and thus may use the same logic. Applications of the presentation tier must be accessible from remote locations and from different types of hardware and software. The presentation tier is implemented as a Java Web application server. For processing user requests and controlling site navigation we use the Web technology. There are two modules at the core of the application: one module for data retrieval, and another module for analysis. The user interfaces are based on Web pages using the Java Server Pages technology (JSP, java.sun.com/products/jsp/) responsible for dynamically creating the content of the Web pages and accessing the business logic in the background. The data retrieval module includes tools for browsing lists of data sets, full-text searching, and filtering lists by predefined criteria. The analysis module contains a table of content on the left, an interactive map in the middle and a tool box on the right. Depending on user interaction, page elements may change dynamically. For instance, attribute tables or forms for specifying queries are embedded on demand. Elements such as table of content, and attribute tables depend on the requested data. Similar to the implementation of the business logic, the generation of Web content is encapsulated in JavaBeans.

2. Application tier

The application tier contains the business logic of the Virtual Database. It is in charge of applying incoming requests to the data. Data are represented by the integration layer which collects and integrates the data retrieved from the access layers in order to make data distribution transparent. The business logic operates on the combined set of data ignoring possibly different origins. It is encapsulated in software components. These components are robust, reusable software modules which can be used in an arbitrary context. Components for a certain purpose may be bundled in an application server with a published application program interface (API). This API describes the functionality and corresponding methods offered by the business logic. According to the requirements of the Virtual Database, the application tier must offer business components for retrieval, query, analysis, and display of data. Retrieval of data does not only include access to the data but also to metadata describing the characteristics, content, and quality of the data. Components related to data query include logic for accessing data from the integration layer

according to given query clauses. Since every institution contributing data to the Virtual Database has already developed analysis methods tailored to their specific needs, analysis components must concentrate on providing the most important and valuable spatial analysis methods. In our opinion, these are methods related to spatial overlays combined with buffering capabilities. Components are accessed by corresponding services within the application tier. The application of Web services offers a solution in this respect.

2.1. Integration layer

The integration layer fetches data from the repositories by querying the access layers. These data are either XML documents describing the content and structure of the data, or GML documents containing the spatial data together with attribute data. Data are merged according to their XML schemas and written into a local Oracle database which uses the Spatial Database Engine (SDE) for storage and fast access of spatial data.

2.2. Application logic

The business logic implemented in components is either encapsulated in JavaBeans components (java.sun.com/products/javabeans/) or delegated to an external application server. Components for a catalog service, analysis service and other information service must be implemented. However, since the Virtual Database is of dynamic nature, the use of the static configuration file is omitted and the dynamic features of configuration are applied.

3. Resource manager tier

The resource manager tier consists of distributed data repositories with different types of data and storage systems. These repositories are located at various institutions spread over all district of Orissa. Residing on top of each data repository there is a so called access layer, which handles data access according to the available database system. Standardized interfaces specify how data are served and how data can be queried. The Virtual Database makes use of the Web service to the member institutions. Member institutions have their own database to upload and store their resources which merged into the main database of the hosting server. Nodal institutions have no authority to change or any administrative access to the main server. They can only access the resources of the main database and have the user login to up-date their database.

3.1. Access layers of the resource manager tier

Currently the resource manager tier comprises different remote data repositories with heterogeneous data management systems. The implementation of the access layers is based on the Web Feature Service (WFS). Although open source implementations of the WFS interfaces exist, we implemented our own WFS modules due to the high complexity of the database structures and the mixed case of database and file structures for the database. The individual WFS modules are implemented as Java servlets which run in a Web container on top of the data repositories.

Application Scenario

The implementation is currently in the state of a working prototype ready to be used by the institutions involved in the project. Being at this state of the software development process, we cannot yet report on a current project that makes extensive use and capabilities for providing web based library service to the private technical and management institutions of Orissa. However, we can only outline an application scenario that gives an impression of the additional value added by the chosen approach.

Conclusions

The Virtual Database is designed for data management and analysis of data stored in distributed data repositories. It provides techniques and methods related to data interoperability which increase the capability to cope with new challenges and requirements for the management of data. The architecture for the management of distributed data with a simple model for sharing resources in distributed database systems is presented in this paper. An experimental implementation of this model and the associated support facilities required were discussed. Many of the ideas described in this paper are related to design and develop a virtual distributed database which provides opportunities to the member users to go ahead under an umbrella of a web based library network with cooperation and coordination to bring the ideas in fact.

REFERENCES

1. Alonso, G., Casati, F., Kuno, H., Machiraju, V., 2004. *Web Services. Concepts, Architectures and Applications*. Springer, Berlin/Heidelberg.
2. Alpdemir, N, Mukherjee, A Paton, N and Watson. P Service-Based Distributed Querying on the Grid. In *Proceedings of International Conference on Service Oriented Computing (ICSOC)*, December 2003.
3. Anderson, G., Moreno-Sanchez, R., 2002. Building web-based spatial information solutions around open specifications and open source software. *Transactions in GIS* 7(4), 447-466.
4. Apache Software Foundation. Jakarta Lucene. <http://jakarta.apache.org/lucene/>.
5. Kazakos, W., Briesen, M., Lehfelt, R., Reimers, H.-C., 2005. Using ISO 19115 metadata and web services to facilitate data access, visualization and processing. In: *Proceedings of the 19th International Conference Informatics for Environmental Protection*, September 2005. Masaryk University in Brno, Czech Republic, pp. 775-782.
6. Kim, J Balsoy, O Pierce, M and Fox, G Design of a Hybrid Search in the Online Knowledge Center. In *Proceedings of the IASTED International Conference on Information and Knowledge Sharing*, November 2002.
7. Kim, J and Fox, G. A Hybrid Keyword Search across Peer-to-Peer Federated Databases. In *Proceedings of East-European Conference on Advances in Databases and Information Systems (ADBIS)*, September 2004.
8. Maguire, D.J., Longley, P.A., 2005. The emergence of geoportals and their role in spatial data infrastructures. *Computers, Environment and Urban Systems* 29(1), 3-14.
9. Ozsu, T. and Valduriez, P. *Principles of Distributed Database Systems*. Prentice Hall, 1999.
10. Pallickara, S and Fox, G. C Narada Brokering: A Distributed Middleware Framework and Architecture for Enabling Durable Peer-to-Peer Grids. In *Proceedings of International Middleware Conference*, June 2003.
11. Peng, Z., Tsou, M., 2003. *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. John Wiley & Sons, New Jersey.
12. Sheth, A and Larson, J Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. *ACM Computing Surveys, planetary-scale network services*. In *Proc. of the 1st NSDI (Mar. 2004)*, pp. 253-266.
13. Stoebraker, M., Aoki, P. M., Devine, Litwin, W., and Olson, M.: A new architecture for distributed data. In *Proc. of the Tenth ICDE (1994)*, IEEE Computer Society, pp. 54-65.