

Integration of GIS Data Sources over the Internet Using Mediator and Wrapper Technology

Leonid Stoimenov, *Student Member, IEEE* Slobodanka Djordjevic-Kajan, *Member, IEEE*, Dragan Stojanovic, *Student Member, IEEE*

Abstract--This paper presents ORHIDEA project which provides intelligent integration of information using mediation technology. ORHIDEA is used for integration of spatial data from GIS databases, alphanumeric data from RDBMS and other data sources.

Index terms --mediator, information integration, GIS legacy databases

I. INTRODUCTION

Geographic information systems are computerized systems for managing data about spatially referenced objects. GIS differ from other types of information systems in that they manage huge quantities of data, require complex concepts to describe the geometry of objects and specify complex topological relationships between them [1]. In addition, GIS data is typically used by various groups of users with different views and needs.

New generations of applications, such as GIS, have much more demands in comparison to possibilities, which could provide legacy information systems and traditional database technology. Most current geographic information systems assume and present the static world. However, information that exists in different spatial database may be useful for many other GIS applications. Also, numerous legacy systems should be coupled with GIS systems, which present additional difficulties in developing end-user applications. Because of that, there is a need to provide communication and collaboration between these applications. Most of large organizations have huge amount of data collected and have invested in many legacy information systems. Nowadays, new applications, such as GIS, can be created combining existing information in a wide information system. In traditional collaborative engineering, information is managed in an ad hoc manner, with data maintained in application specific files and legacy databases. Data

exchange between these systems usually performed by executing scripts and programs that transport data. This approach may work with small projects, but it is inadequate for distributed environments.

Today, the only way to integrate huge amount of available data is to build specialized applications based on mediators [5]. Information mediators provide an intermediate layer between information sources and users. Nowadays, there is a strong trend of information systems integration in chain of systems among public information structures such as Internet. However, no one wants to share his own information with public. This is the reason that integration is interchanging data among the applications without centralized control. Another trend in GIS is publishing maps for World Wide Web community and development of web-based GIS applications.

The goals of our research activities, described in this paper, are defining an architecture for integration of distributed and heterogeneous GIS data sources, identifying the need for mediators and wrappers in telecommunication network administration, and adding the integration technology to the GinisNT [2,3] spatial database. We examine a research whose final goal is to make disparate data sources work together. This work is known as information integration.

II. INFORMATION INTEGRATION

Data integration application begins from a set of pre-existing data sources. These sources might be database systems but more often are unconventional data sources, such as legacy systems, structured files, e-mail files, and www pages [8,9]. Most importantly, a data integration system lets users focus on specifying what they want rather than thinking about how to obtain the answers. As a result, it frees them of combining data from multiple sources, interacting with each source and finding the relevant sources. A data integration system provides a uniform interface to a multitude of data sources.

A middleware system, based on mediator and wrapper technologies, separates the storage of data management facilities, providing common interface for application. Information mediators were originally developed for integrating information in databases [4,5]. Three-layer

Leonid Stoimenov, Slobodanka Djordjevic-Kajan, Dragan Stojanovic
Computer Graphic & GIS Lab
Faculty of Electronic Engineering, University of Nis
Beogradska 14, 18000 Nis
E-mail: leni@elfak.ni.ac.yu

mediator architecture provides transparent view of data sources and independence of data sources and user applications.

Applying the mediator framework to the Intranet/Internet environment solves the difficult problem of gaining access to real world data sources. Internet provides the underlying communication layer and protocols for mediation of distributed systems. A wrapper is a program that is specific to every data source [6,7]. Wrapper extracts a set of tuples from source file and performs translation in the data format.

Mediation concept is a part of the ARPA I3 (Intelligent Information Integration) reference architecture. The I3 reference architecture should be seen as a vision of how vast amount of heterogeneous information can be incrementally pulled into a gigantic, reusable library of information resources [6].

In order to simplify network programming and make component based software architecture real some middleware technologies are developed on paradigm programming level. Two, already existing, distributed object models tend to be standardized. These are DCOM [10] and CORBA [11]. These high-level programming paradigms can be used to support deploying software components, platform-independent, which co-operate using TCP/IP network. These technologies are relatively new and there is so little experience in combining them in order to support huge information integration systems used in enterprise organization.

III. ORHIDEA PROJECT

The research group at the Computer Graphics and GIS Lab at the University of Nis, Yugoslavia has been developing GIS software for eight years now. GinisNT is a scalable, OO environment for the development of GIS applications that is built on top of a relational DBMS the usage of which is made transparent to the user completely [12,13]. The goals of our research activities are defining an open architecture, identifying the need for mediators in telecommunication network management, adding the integration techniques to the GinisNT spatial database and realize the active mediator level which perform integration tasks. We have developed Internet based version of GinisNT, called ORHIDEA. Projects goal in the ORHIDEA is to make it simple for users the use of different data sources in their GIS applications. ORHIDEA is a middleware, mediator system that provides data interchange and access to distributed data sources without changing how or where data is stored.

Our primary goal with ORHIDEA project was to design a platform and methodology that could serve for information exchange over the Internet between distributed and web-based GIS sources. However, the web's browsing paradigm does not readily support retrieving and integrating data from multiple sites. The only way to integrate the huge amount of available data is to build specialized platforms.

The ORHIDEA platform uses agent-wrapper and mediator technology to allow communications between GIS applications over the Internet/Intranet. Wrappers are types of software "glue-ware" that are used to attach other software components. Mediators are "complex", higher level component, that combine data from different data sources. The mediator and agent-wrapper approach also allows maintaining these applications (data sources) and incorporate new sources, as they became available.

The general approach to mediation in ORHIDEA is based on our previous work on the GinisNT architecture with GinisNT Active Mediator layer. The GinisNT Active Mediator layer provides an intermediate level between information sources and user GIS applications.

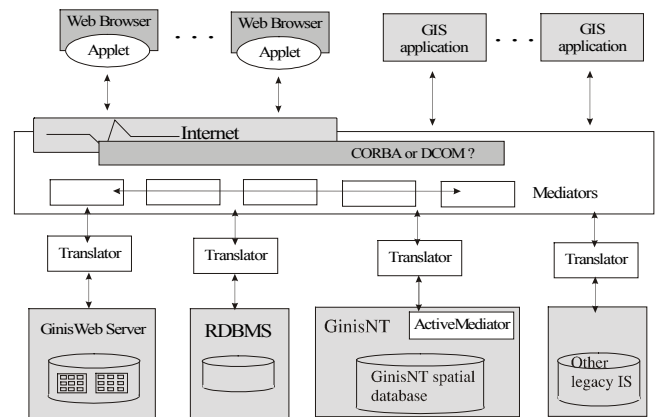


Figure 1. ORHIDEA platform

The architecture of this system is shown in Figure 1. ORHIDEA is set of distributed software components, which provide GIS information integration. System provides "virtual" database schema, and user application are operative system independent. Client application can be placed on Web or stored on any workstation. Data sources, which should be recognized by our system are spatial GinisNT database (and any other commercial spatial database), ActiveMediator and GinisWeb interface for GinisNT, relational databases and other legacy information resources. For every data source in heterogeneous information system there is a translator (or wrapper) which converts logically base data objects to common information model. The set of information agent/wrappers implement the middle-layer linking services. This layer perform mediator functions which include transformation of data and mapping between data models. In order to make this logical translation, translator converts query's to requests through information from common model. These requests can be executed by data source. Translator also converts data returned from data source to common model.

Mediators are above translators. Mediator is a system, which filters information from one or more data sources. Mediator builds in knowledge, required for processing specific information. Mediators should hide existence of

various (heterogeneous) data sources; they should abstract, collect data for users and translate data and schemas.

Important system features are that there is no global database schema and that mediators can work independently. For mediator design it is necessary to know only which data sources it will access and format of data returned by source. Complete understanding of used data source features is not necessary.

IV. CONCLUSION

This paper focuses on the components of the ORHIDEA platform that are needed to support the mediating functions in this architecture over the Internet/Intranet. In this architecture we are used new technologies, such as mediators, wrappers, and agents.

Project ORHIDEA, described here, uses intelligent information integration using mediators as link between data source and user applications. System should provide actualization of client/server applications, using Internet and Web technologies as under-layer for network service and integration of distributed data sources. Mediators provide easy information access and no need for creating new database exists, "virtual" database schema, which gives user data stored in existing databases is created. Mediators are efficient and cheap way of integrating data from heterogeneous information systems. Logical consequence is that mediators provide evolving system development and use existing software investments in software and databases.

V. REFERENCES

- [1] D. Mitrovic, S. Djordjevic-Kajan, "An open GIS architecture for inexpensive hardware platforms", Proceedings, AM/FM European Conference X, Heidelberg, October 18-20, pp. 185-189, 1994.
- [2] D. Mitrovic, A. Mitrovic, S. Djordjevic-Kajan, D. Rancic, "A GIS solution for developing countries", Proc. AURISA-95 Melbourne, pp 262-271., 1995
- [3] D. Mitrovic, A. Mitrovic, S. Djordjevic-Kajan, "A scalable, object-oriented GIS Framework", In Proc. of ISPRS Int. Workshop "Hardware and Software Aspects of GIS", Milan Italy pp. 20-21, 1996.
- [4] G. Wiederhold, "Mediators in the architecture of future information systems", IEEE Computer, 25, pp.38-49, 1992.
- [5] G. Wiederhold, "Weaving Data into Information", Database Programming & Design, Vol. 11, No. 9, pp.22-29, September 1998.
- [6] *Reference Architecture for the Intelligent Integration of Information*, ARPA, Prepared by the Program on Intelligence Integration of Information (I3), version 1.0.1, May 19, 1995.
- [7] M.T. Roth, P. Schwarz, "Don't Scrap It, Wrap It! A Wrapper Architecture for Legacy Data Sources", Proceedings of the 23rd VLDB Conference, Athens, Greece, pp.226-275, 1997.
- [8] Cohen, W. "Integration of heterogeneous databases without common domains using queries based on textual similarity", SIGMOD'98, Seattle, WA, USA, pp.201-212, 1998.
- [9] Miller, R.J., "Using Schematically heterogeneous Structures", SIGMOD'98, , Seattle, WA, USA, pp.189-200, 1998.
- [10] *The Component Object Model Specification*, Microsoft, 1995
<http://www.microsoft.com/oledev/olecom/title.htm>
- [11] *The Common Object Request Broker: Architecture and Sepcification*, Revision 2.0, July 1995,
<http://www.omg.org/corba/corbaiiop.htm>
- [12] Stoimenov L., Mitrovic A, Djordjevic-Kajan S "An Object-Relational Environment For Developing Object-Oriented GIS Applications Based On Mediator", 16th IASTED International Conference on Applied Informatics, Garmisch- Partenkirchen, Germany, pp.168-171, 1998.
- [13] Stoimenov L., Mitrovic A, Djordjevic-Kajan S., Mitrovic D., "Bridging Objects and Relations: a Mediator for an OO Front-End to RDBMSs", prihva}en za publikovanje za: Information and Software technology, Elsevier Science, Vol. 41, Issue 2, pp.59-68, 1999.