ORGANIZATION AND MANAGEMENT OF DISTRIBUTED SPATIAL DATA

Hui LIN* Haiyang HAN* Jianya GONG**
*Chinese University of Hong Kong, Hong Kong
Joint Lab for Geoinformation Science
hhy@mail.jlgis.cuhk.edu.hk
http://www.jlgis.cuhk.edu.hk/wwwgis
*Wuhan Technology University of Surveying and Mapping, P.R.China
National Lab for Information Engineering in Surveying, Mapping and Remote Sensing
Jgong@rcgis.wtusm.edu.cn

KEY WORDS: GIS, spatial data, Distribution, Internet. Hypermedia.

ABSTRACT

With the rapid development of Internet, it is possible to share and modify the data and information in the global range. As a result, Internet GIS is becoming a new research branch. Compared to the traditional GIS, the Internet GIS has it's own outstanding feature in architecture, data organization and information publication. On the basis of summarizing the current Internet GIS models, this paper provides a new method for organization and management of distributed spatial data. An Internet GIS software is taken as an example of such an idea.

1 INTRODUCTION

Due to a short history of Internet GIS, compared to traditional GIS, most of Internet GIS software can only provide simple functionality in spatial information publishing and spatial querying, and a systematic theory and appropriate model of Internet GIS are still unavailable [Forte G B. 1997, Limp WF. 1997, Lennon J A. 1997]. Recently, some famous corporations have released their Internet GIS software respectively[OGC. 1997, Oracle Corp. 1997]. The design architecture, construction model and running mode of most Internet GIS can be indicated by those system[Plewe B. 1996, Plewe B. 1996]. Although they are based on different architecture, compared to traditional GIS, some characteristics and advantages can be generalized as follow.

First, Internet GIS, which is based on browser/server architecture, consists of browser, middleware, and database server under Internet/Intranet environment. Browser sends request via HTTP to middleware server. Middleware receives the message and interacts with database[Peng Z, 1997, Peng Z. 1998]. According to the type of requests, middleware accesses and processes spatial data, and then sends back geoprocessing results in server-side.

Second, Architecture of Internet GIS is different from traditional GIS's. Spatial data are located in different places. Data producer is responsible for update and maintenance of spatial database[Strand E J. 1996]. It is not necessary for user to know the location and data format of spatial database server. Middleware, which acts as the bridge between browser-side and database server-side, is responsible for transferring message and geoprocessing in server-side. User can access and operate spatial data by universal graphical interface in browser-side. Generally, simple graphical functionality's such as display, zoom, pan is finished in browser-side. Most of geoprocessing, such as spatial analysis, thematic mapping, statistics chart, query, catalog service, security management, are in server-side. The objective of such a way is to decrease burden in browser-side and share spatial data.

Third, Internet GIS can provide user interactive, hypermedia, distributed spatial information. Combined with hyperlink and VR (Virtual Reality)[Weber J. 1996], spatial information (especially vector-based map) can interact with other hypermedia such as image, video, and text by hyperlink. In this way, it is very convenient for user to browse a vector-based map as the same as web pages by a browser on the Internet. In addition, distribution and interoperability of spatial data, which are based on OpenGIS, are the basis of share, management, and interactivity of spatial information[Wilson JD. 1998], Zhuang V. 1997.

However, there are also many problems for Internet GIS. For example, direct access and geoprocessing of distributed spatial data are not supported by Internet GIS, so it can not meet the requirements of OpenGIS and mountains of raw data. Distributed spatial data are saved in a central database, so it is not convenient for share and update of spatial data. Due to absence of organization and management for multi-source spatial data, no internal relationship among distributed spatial data is reflected, and data from different places is also an information island even in a central database. To represent vector-based spatial data on the Internet is still a problem because browser dose not support vector-based map. Effective security mechanism is not built up to protect spatial data from hacker and malicious user. To deal with those problem as above, the paper puts emphasis on the distributed organization and management of multi-source and hypermedia spatial data, and take Internet GIS-GeoSurfV3.0 as an example in order to explain such an idea.

2 ARCHITECTURE

The architecture of Internet GIS, based on browser/server, is as Figure1. Due to programming in Java language, it is platform-independent. To run it, Java VM-based browser is enough in client-side. Browser dynamically loads and runs Java program in browser-side via HTTP and web server. Afterwards, browser interacts with middleware, loads and displays spatial and hypermedia data via middleware from database in server-side.

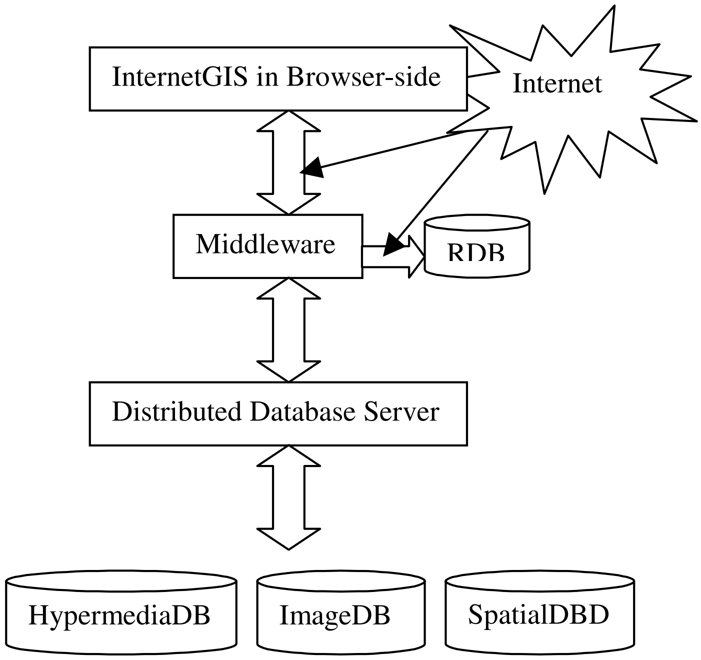

Figure1.The Architecture of Internet GIS

In figure1, browser, web server, meddleware, database server and database are all located in difference places. They communicate by HTTP or socket. Browser loads and runs system via web server, and parts of geoprocessing such as display, zoom, pan can be implemented in browser-side, others, such as thematic mapping, statistics chart, spatial analysis and query, are in server-side. Such a combination of Internet GIS with browser enriches the contents of www. It is possible for vector-based map to interact with other media such as image, text and video. User can access and browse vector-based map as the same as other traditional media. As an interactive channel of browser and server, the objective of middleware is not only to receive and response the request from browser-side; but also takes on parts of geoprocessing tasks. It is responsible for management, organization, access control and logo of distributed database in server-side.

3 DISTRIBUTED GEOPROCESSING FOR MULTI-SOURCE SPATIAL DATA

Under Internet environment, on the one hand, users eager to access more various types of spatial information than simple presentation with text and image. On the other hand, the development of Internet makes it possible to access and save spatial data in multi-machine. Especially, mountains of raw data need to be updated and managed in real time in information age. In traditional way, it is not feasible to update real-time spatial data from difference places in a central database, now, distributed management and organization of spatial data is becoming the trend of Internet GIS. In 1999, a project in relation to managing and organizing distributed spatial data is implemented in Canton province, China. The paper takes it as an example to explain the idea of how to design, and organize distributed spatial data in server-side.

3.1 Data Organization of Multi-scale and Multi-layer Spatial Data

The objective of this project is to publish interactive, bi-directional spatial information on the Internet, and act as GIS service provider to provide services ranging from spatial display, query to spatial decision support in order to meet various requirements of different users from government to research field. So it is necessary to organize and manage all spatial data of Canton province from province-level data to town-level data under Internet /Intranet environment. In addition, hypermedia, including image, text and video, in relation to spatial data interact with spatial data by hyperlink on the Internet. Therefore, distributed data are organized in hierarchical structure. Different thematic layers of a map are in the same level, and maps of different scales are linked by hyperlink. Hypermedia is the attribute of corresponding spatial data, and its responsibility is to represent spatial information from another view with hypertext link.

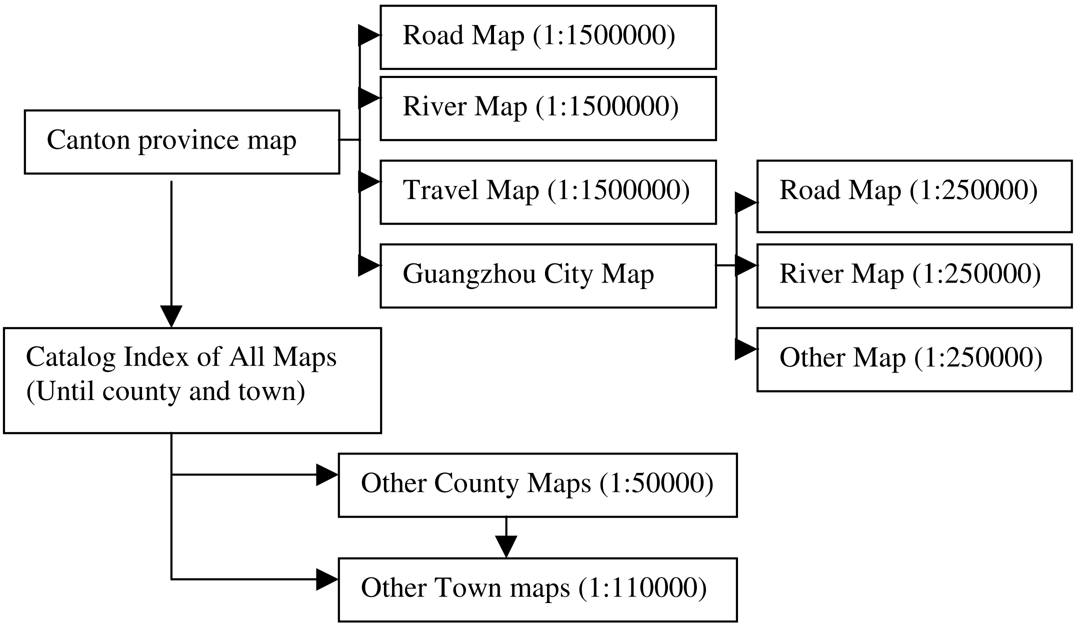

Figure2. The organization of Multi-scale Maps

By organization as above, it is feasible to manage distributed, multi-scale spatial data, and load spatial data layer by layer on the Internet. Especially, at present, Internet speed can not meet the requirement of geoprocessing and transferring mountains of raw data, so it is very important to organize and manage spatial data in an efficient way. At the same time, effective data organization is helpful to construct catalog and hyperlink-based relationship among various spatial data. In this way, it is very convenient for user to browse and access spatial information in a wide range by catalog index and hypertext link.

3.2 Storage and Update of Distributed Spatial Data

The problem of storage and Management of mountains of spatial data has to be solved in order to process spatial data on the Internet. In traditional way, from capturing data to storage in central database, due to limitation of space and time, it is a money and time-consuming work to manage and update spatial data. Sometimes, such spatial data, which costs so much money and time, is outdated and become useless. So it can not provide any help for spatial decision support, spatial analysis or any other geoprocessing. The development of Internet changes traditional way in total. It is becoming more and more feasible and popular to access, process, update spatial data via Internet. In addition, it takes advantages of benefit balance between spatial data producer and end user.

In figure3, data producer need not know who and how to make use of spatial data. His responsibility is to capture and update data continuously. For manager in middleware-side, on the one hand, he requires data producer to provide real-time, needful spatial information. On the other hand, he is responsible for the organization of distributed, multi-source and hypermedia data, security management, access control and other geoprocessing operations. Therefore, middleware is the core of Internet GIS.

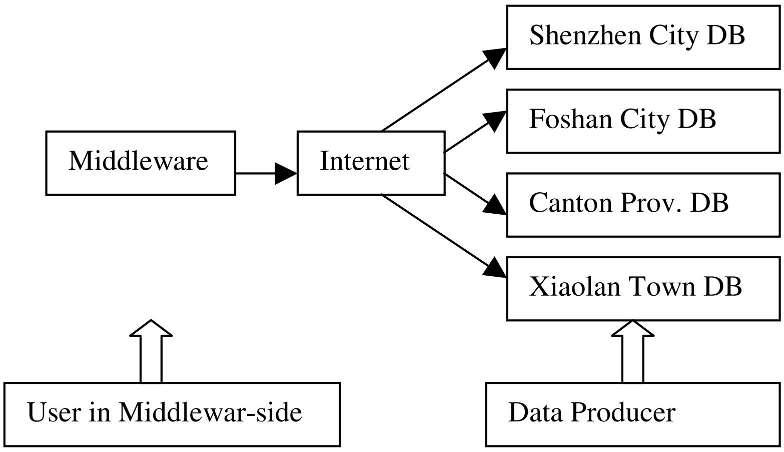

Figure3. the Distributed Management of Spatial Data

3.3 Data Structure of Distributed Organization in Middleware-side

For distributed geography information, databases are located in different places and hypermedia information has to be associated with a feature (FeatureID or OID) such as point, line or polygon in a vector-based map by hyperlink. The data structure is based on HTTP. In middleware-side, hyperlink, linked to each database node, consists of URL, spatial database name, map name, etc. Therefore, from a whole view, distributed database is organized by a catalog service in middleware-side. Catalog service organizes and saves information in relation to distributed database by HTTP-based hyperlink, middleware server can retrieve so all metadata of distributed database. From a local view, the IP address of each vector-based map can be uniquely fixed in a specific database according to its name. With the help of hyperlink and catalog service, storage and geoprocessing of spatial data can be located in difference places.

In figure4, the IP address of database is defined by the combination of OID and URL address. Map Level means the level of a vector-based map in a hierarchical structure; for example, China map has a higher level than Canton province map. Various types of Language benefit users of different nations. Data Format means the format of a map in spatial database such as DXF, E00 or MIF. According to specific data format, corresponding geoprocessing component is loaded to access and process spatial data. Hypermedia Link is to reflect multimedia information of spatial data by Image, text and video in a web page.

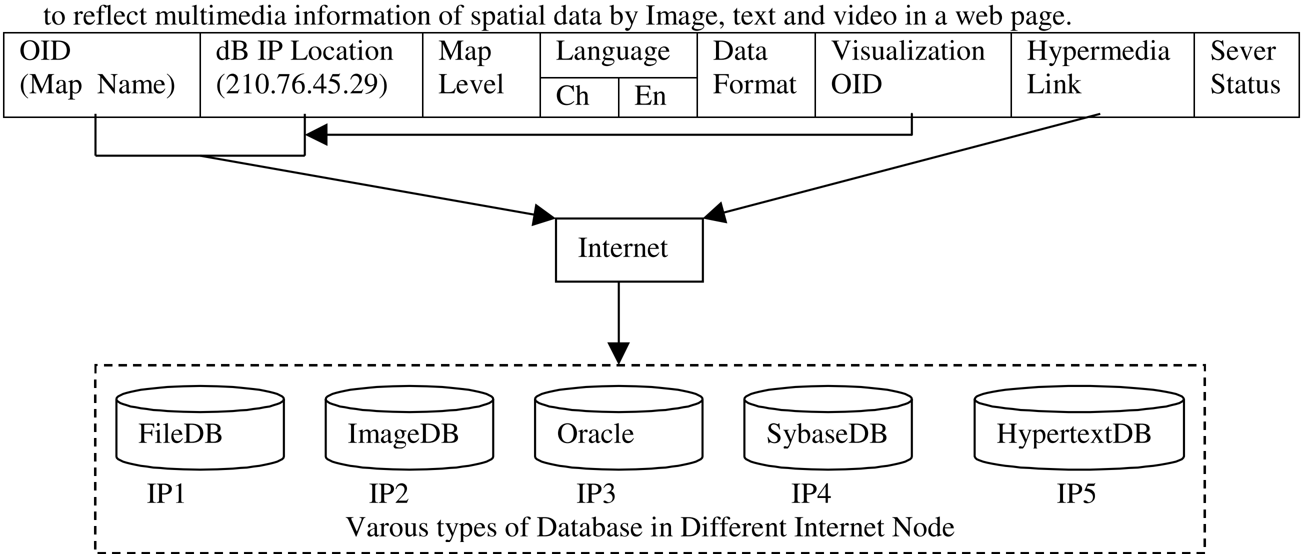

Figure4. Data Structure of Middleware Server and Relationship with Database

3.4 Security and Access Control

It is the most important to ensure the security and access control of database in order to process and manage distributed spatial data on the Internet. Because it is a money and time-consuming work for data producer to capture and update spatial data, data producer is responsible for the maintenance and update of database server. User can login in database server, and browse, query, analyze latest spatial data, but he can not download and save data in his local machine without authorization.

In addition, it is necessary to update Internet GIS in time under Internet environment. In traditional way, when new version software is released, user has to buy and install the updated version in local machine again and again; it is also a time-consuming work. By multi-tier architecture of Internet GIS, server is responsible for the update of Internet GIS, when user connects to the web site, browser dynamically loads and runs Internet GIS without installation. In addition, different users have a different access limitation by authorization of middleware. For example, administrator can modify the system parameter in middleware-side, however, for guest, he does parts of operations such as browsing and querying.

In order to meet requirement of security control, the essential way is to divide users into many security levels from administrator, power user to guest. Under Internet environment, without enough security management, non-authorization access of malicious or mindless user may crack down database server, so it is very dangerous. Therefore, security control is set in middleware-side and browser-side. Principles including: dividing users into different groups, managing database according to its access control, access being accepted for users with authorization, other access denied.

Furthermore, by using technology of digital signature, spatial data is encrypted with a public key and an encrypted signature and spatial data are sent to user. If data is modified, then the signature can not open the data. Any user receiving data can restore encrypted data with private key. Therefore, user without authorization can not access encrypted spatial data. Combined security control of web server with digital signature, the problem of access security will be solved effectively.

4 IMPLEMENTATION OF DISTRIBUTED GEOPROCESSING IN BROWSER-SIDE

Internet GIS in browser-side is largely different from traditional GIS in objective, implementation, information content. The architecture of Internet GIS has changed from traditional c/s to web-based browser/server. Recently, browser has become the door of Internet and more and more people publish and access information by browser. So Internet GIS can take advantage of browser for accessing and geoprocessing of spatial information.

4.1 Distributed Access of Heterogeneous Spatial Data

Data organization of heterogeneous spatial data in middleware-side is directly reflected in browser-side. Maps of multi-source data are organized in hierarchical structure. User in browser-side can loads map of any level from catalog tree. When loading heterogeneous spatial data in middleware, according to its data format, middleware imports spatial data into its internal data format by calling data transferring component, and sends results to browser-side via Internet. The advantage of this way is to recognize and transfer various data format in real time. Now, due to limitation of Internet speed, it is more feasible for online application under Intranet environment. For example, in spatial decision system, data producer in server-side can directly publish latest spatial data and hypermedia document on the Internet, and decision-maker runs Internet GIS by browser capture spatial information necessary in order to support his decision.

4.2 Implementation of Hypermedia and Interactivity

In traditional way, the representation of multimedia information is one part of GIS software. However, with the advantages of Internet, browser and hyperlink, browser is the main way of representing hypermedia on the Internet. Therefore, hypermedia information is independent to InternetGIS, and hyperlink is the bridge of them. On the one hand, Internet can provide mountains of hypermedia information. On the other hand, the efficiency of Internet GIS will be improved largely without programming for hypermedia.

In figure5, when Canton province map is loaded, the related hypermedia introduction is displayed in another browser window. At present, browser supports many types of media including video, image, text, VRML and voice, so it is easy for common user to grasp such a way to operate Internet GIS.

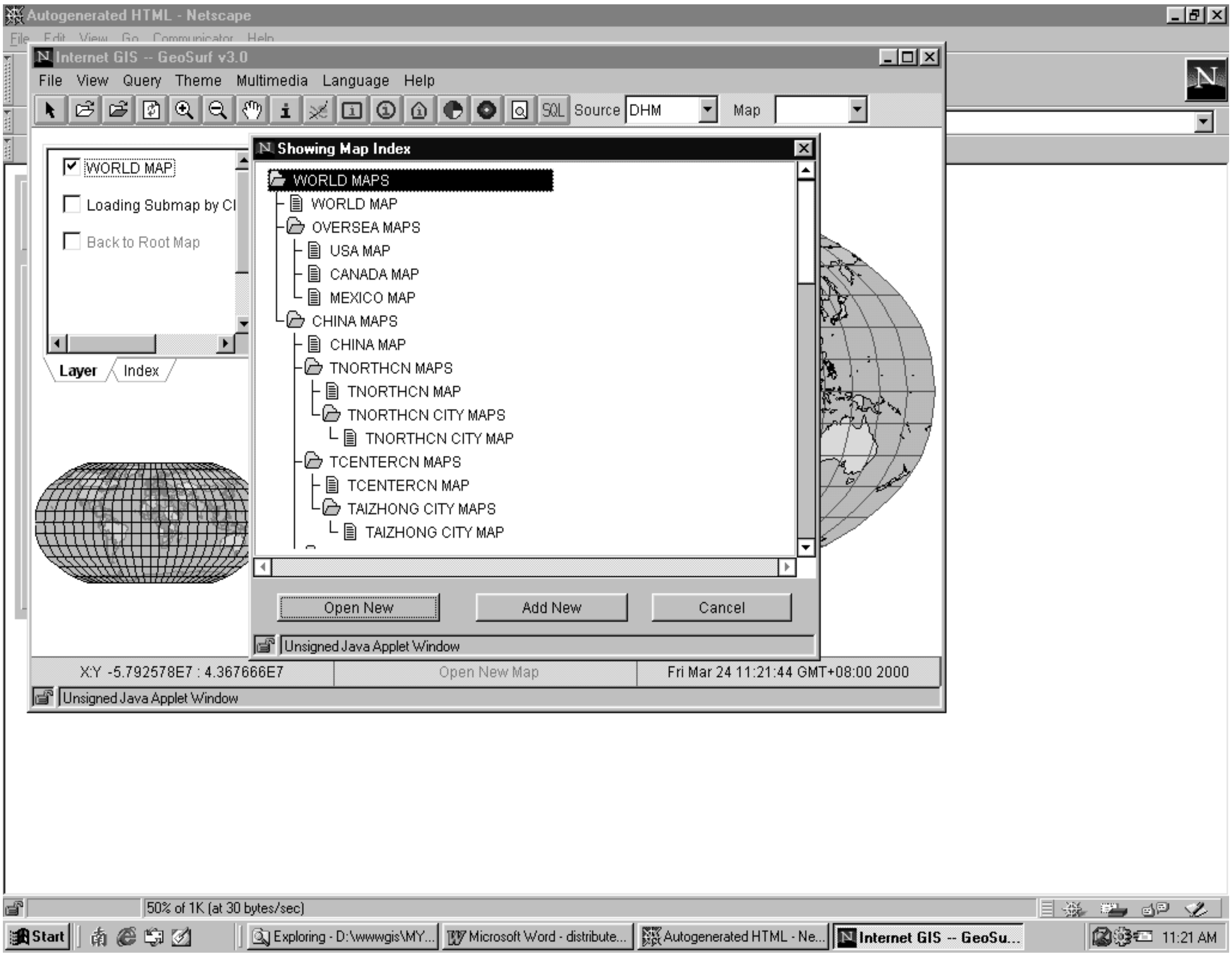

Figure5. The Browser-side Show of Internet GIS-GeoSurf

In addition, it solves the problem of transferring mountains of data on the Internet. Multi-window and multi-thread are helpful to transfer spatial data, and improve the efficiency of loading data.

5 CONCLUSION

Figure5 is the representation of Internet GIS software-GeoSurf which is based on ideas as above. The objective of GeoSurf in server-side is to organize distributed spatial data, construct the relationship among spatial data, and provide catalog service. In figure5, catalog tree reflects the organization of distributed data in middleware-side. User can load, process spatial data by choosing map in catalog tree. Each map has its security level and access denied for illegal user.

In the near future, global information share has to be faced with the problem of distributed organization and management of multi-source spatial data. The paper provides an idea of organizing and managing distributed spatial data by catalog service and hyperlink. However, how to process distributed spatial data in middleware-side is also a problem need to be further probed and solved. There are also a long way to go to organize and process distributed data on the Internet.

ACKNOWLEDGMENTS

We are grateful to the support from HKSAR Government for the RGC Grant No.CUHK 4334/98E and the ISF Grant No. AF/169/98.

REFERENCE

Forte G B. 1997.The GIS Book, 4th Ed Santa Fe: On Word Press

Limp WF. 1997 Weave Maps across the Web.GIS World, 10(9): 46-55

Lennon J A. 1997,Hypermedia Systems and Applications: World Wide Web and BeyondBerlin, New York: Sprinters.

OGC. 1997.The OGC Technical Committee Technology Development process, Open GIS Consortium, Way-land, Massachusetts.

Oracle Corp. 1997.Oracle 7 Spatial Data Option Release 7.3.3: Spatial Solution Center Training Class Notes.

Peng Z, Nerber D D. 1997.An Internet –Based GID Data Access Sydtem.Journal of Urban and Regional Systems, (1): 20-30

Peng Z. 1998, An Assessment of Internet GIS.Department of Urban Planning, University of Wisconsin Milwaukee.

Plewe B. 1996,9. A Primer on Creating Geographic Services.GIS World, (1): 56-58

Plewe B. 1997,GIS Online: Information Retrieval, Mapping and the Internet.Santa Fe, NM: Onward Press.

Strand E J. 1996,9.Java-Enabled Browser provides Geographic Services.GIS World, (1): 38

Strand E J. 1996,9. Vector Maps Weave Their Way across the Web.GIS World, (1): 32-33

Weber J. 1996.Spatial Edition using java.2nd ed. Indianapolis, In: Que Corporation.

Wilson JD. 1998,11.Interoperability: Open the GIS Cocoon.GIS Worl, (3): 58-60

Zhuang V. 1997,10, Spatial Engine Drive Web-based GIS.GIS World, (10): 45-58