Geographical Information Processing for Cultural Resources

Assoc. Prof. Hirohisa Mori,
Department of Geography,
Graduate School of Literature and Human Sciences,
Osaka City University

1. What are the Problems?
-- And Fully Autonomous Distribution

GLOBALBASE is not the first attempt to share geographical information. Various technologies, for instance WebGIS, already exist. All the existing technologies, however, are based on centralized architectures.

Figure 1.-(a) shows the simplest quasi-centralized architecture, where all geographical information is collected and stored on one server. Figure 1.-(b) shows a slightly developed architecture, where the basic geographical information is stored on different servers, and the central server handles only searching and overlapping of target maps. This architecture employs a so-called clearing house method.

Both methods have bottlenecks on the server side. Quantitative bottlenecks, such as CPU processing power and network communication speed, are solved as a matter of course as the technologies are improved. However, we believe that qualitative bottlenecks, which are impossible to solve simply by upgrading the hardware, also exist in these technologies.

Currently, there are many map sites found on the WWW. However, if one sends a request to the administrators of such sites saying "I am interested in stores in the Edo period. Can you place such information on your site?," then, immediately, s/he will probably get a reply along the lines of "Our site focuses only on current information and as far as we know there is no demand for ancient information. We are sorry but we cannot support such information." This is the qualitative bottleneck. We think that there may be many people interested in history in Japan, however, and if such information were available on maps, it might become rather popular.

Problems such as the one outlined above, and others similar problems, will not occur if it is possible for anyone to deliver geographical information of one's own liking on one's own homepage, just like creating a personal homepage.

"I" can deliver information on stores in the Edo period without constraint. Many people interested in history will look at this information and set forth on a romantic journey through history. They can look around many places of historic interest and take a break at a coffee shop near a historical site found on the GLOBALBASE Mapion when tired. Both the current coffee shops and ancient teahouses can be displayed in the same window.

One of the purposes of GLOBALBASE is to solve this qualitative bottleneck. Figure 2. shows an overview of the architecture. Information providers have their own sites and homes, and deliver their own geographical information about their areas freely. Viewers of information enter search conditions according to their interests, and scroll around with the browser to obtain necessary information. There is no restraining bottleneck between the information providers and the viewers. Information will be linked like an amalgam and will develop just like a living organism. People will perform their activities in the GLOBALBASE framework just like they do in the real world.

In this way, another Earth, the parallel world, may be completed.

When we think about it, this is a natural world for the Internet. In the framework of the GIS architectures established so far, however, this could not have been achieved.

2. Data Structure without Reference Coordinate System

In the process of decentralization, the first problem we encounter is the coordinate system. Current GISs have a single reference coordinate system, and all information must conform to it. This approach, however, leads to a qualitative bottleneck.

In GLOBALBASE, all coordinate systems are relative and can be defined individually by the user. In order for many people to be able to deliver geographical information freely and for such information to be shared to create one map of the Earth, GLOBALBASE must be able to handle maps created by normal drawing software as well as maps created by survey institutes in a consistent manner. To achieve this, being able to operate on user-defined coordinate systems is a necessity in order to deal with coordinate systems of different drawing software packages.

All the current GISs employ a single coordinate system that is used as reference. Such a coordinate system prohibits fusion of data from other fields that use different coordinate system concepts; this is one of the primary reason why the GIS technology has not achieved as great a degree of diffusion as expected. The common belief that GIS is difficult to use or requires expertise in geography, surveying etc., has given the technology a negative image.

By implementing user-definable coordinate systems, it becomes possible to capture any information within the same framework, from layouts drawn with drawing tools or architecture CAD software to layouts of rooms and show windows. However, since the only way to connect all this diversified information is through the use of distributed technologies, there must also be a method to integrate all the different fields in order to create one geospace.

3. Concept of Mapping that Connects Coordinate Systems

What will happen if each individual defines an arbitrary coordinate system? We no longer know the relationship between such individual, arbitrary coordinate systems. Fortunately, the technology called Mapping presents a solution to this problem.

Let us assume that a person whose hobby is gardening draws a picture of the layout of his garden in order to present his new garden to other people. He then pastes it onto a Japanese map made available at the International Research Center for Japanese Studies at the location corresponding to his home.

The relationship between the layout of the garden and the Japanese map only needs to describe where the four corners of the garden are situated on the Japanese archipelago. A mapping is a list of such correspondences between the four coordinates of the four points in the "garden" coordinate system and the corresponding coordinates in the "Japanese archipelago" coordinate system. By providing this mapping, the garden layout is connected to the Japanese archipelago, i.e., users can now find the garden by zooming in on the map of the Japanese archipelago.

A mapping plays the role of a bi-directional link. Please refer to Figure 3. Coordinate systems are indicated by rhombs and mappings defining the relationships between coordinate systems are indicated by ovals. Mappings and coordinate systems can be placed on any server, because the mappings work as links beyond servers.

Currently, mapping processing must be manually edited by information providers, but in the future, mapping editors and drawing editors supporting GLOBALBASE will become available. When that happens, it will become possible to paste one's own "garden" onto a map of the Japanese archipelago on other server just by dragging and dropping a control point in the drawing software.

4. Distributed Search Engine

When an infinite number of information providers define one coordinate system after another, the maps become cluttered and we can no longer know what is where. That is, if one zooms to a location, all sorts of information may be displayed all at once. Obviously, from the point of view of users who search for some specific information, it is necessary that such information can be displayed in an orderly manner so that only the necessary information is obtained.

This problem is similar to the problem that occurred in the WWW in its early stage. One decade ago, a list of all WWW sites in the entire world was found at the NCSA site. It was possible to view all the WWW information in the world in one day by accessing those sites sequentially. At that time, there were less than 10 Japanese sites in existence.

However, after that time, the Internet has been commercialized and the number of servers has exploded.

Finding necessary information quickly evolved into an insurmountable task. For this reason, much research has gone into efficient search engines that can help solving these issues.

Similar problems may occur to GLOBALBASE as well. If too much geographical information flows into the browser, it may cease to be usable. Search engines for GLOBALBASE, naturally, become necessary. Search engines for WWWW, however, are of a centralized search type. In the GLOBALBASE architecture, they will be decentralized.

It is almost impossible to decentralize the WWW search engines. Fortunately, in GLOBALBASE, all types of information have coordinate systems, albeit relative, as well as geographical locality. Using these features, it is possible to create an infinite number of small search engines that gather local information. This new technology is called local search engines and works as follows.

A browser connects to a local search engine in the area it is currently displaying. When the user enters certain search conditions, such as a time period, the browser searches for geographical information of the surrounding areas. The local search engine, then, returns all geographical information that match the conditions from a slightly wider area than the displayed area. The browser caches the information and displays it as new information when the user searching for information moves to the left or right. When the displayed area changes, the browser searches for a new local search engine corresponding to the new displayed area and searches for information again. The browser repeats this process regularly, thus moving from one local search engine to another.

Figure 4. shows an example of states of local search engines. Each server collects geographical information in its vicinity by tracking mappings regularly. In this example, it can be seen that server A never overlaps with server C, and server C never overlaps with server A. Since a server only accumulates local information in this

manner, each local search engine does not need to have a large capacity, and can be very simple.

5. Searching for Mapping Paths Auto-Configured Routing Protocol

A browser may collect and store necessary geographical information, but all of this information is based on relative coordinate systems. This means that information on different servers cannot be superimposed as is. It is necessary to check the physical relationship between two or more coordinate systems.

Furthermore, it is not guaranteed that two coordinate systems among the searched geographical information are connected with one mapping. This is because mappings created by information providers are arbitrary. The coordinate system of the garden mentioned before and the coordinate system of the Japanese archipelago are connected with a mapping. Now suppose that there is also a map of parks in the Kanto area connected to the Japanese archipelago, and the garden map and the map of parks in Kanto area are not directly connected, but it is required to overlap these two maps only.

In this case, it is necessary to convert the coordinate system of the garden to the coordinate system of the Japanese archipelago, and further convert this to the coordinate system of the map of parks in Kanto area. In order to do so, it is first necessary to search for and link between the arbitrarily specified mappings.

We found a hint for the solution to this search problem in an unexpected field, namely network routing technology. Let us try to apply the Internet framework to a network of coordinate systems connected by mappings. Coordinate systems can be considered to correspond to hosts on the Internet and mappings correspond to networks; then GLOBALBASE can be considered as the Internet as is. On the Internet, two arbitrary hosts can exchange packets by specifying the IP address of the destination, no matter how they are connected. IP addresses are Internet addresses that can be used for routing. This implies that it is possible to know how two coordinate systems are connected via mappings if addresses that can be used for routing can be assigned to the coordinate systems.

It is, however, not practical to make the information providers themselves assign addresses for which IP routing is possible to each coordinate system. For this reason, a technology to automatically assign addresses that can be routed to coordinate systems is required. This technology is called the ACRP (Auto-Configurated Routing Protocol) and was originally invented by the leader of this project as a protocol for control of Local Area Networks used in intelligent buildings. It is now applied as a technology for sharing information in the GLOBALBASE architecture. It is safe to say that GLOBALBASE was made possible because we had this technology.

6. Browsing Method

The last technology to be explained here is the method for overlapping maps and changing the display of geographical information according to the instruction by information searchers, based on geographical information and mapping information thus gathered. As shown in Figure 5., collected coordinate systems are overlapped by converting mappings. The area indicated by the vertical cube is the area displayed on the browser. If the user moves this display area to the left or right, some coordinate systems go outside the cube and others come inside the cube. Those outside the cube are excluded from overlapping, while those coming into the cube are included in the overlapping by searching through the associated mappings. By repeating this process, the browser is able to connect maps seamlessly, and is eventually able to scroll across the entire Earth.

Figures

Figure 1.-(a) Centralized Type 1 / fig1a.jpg

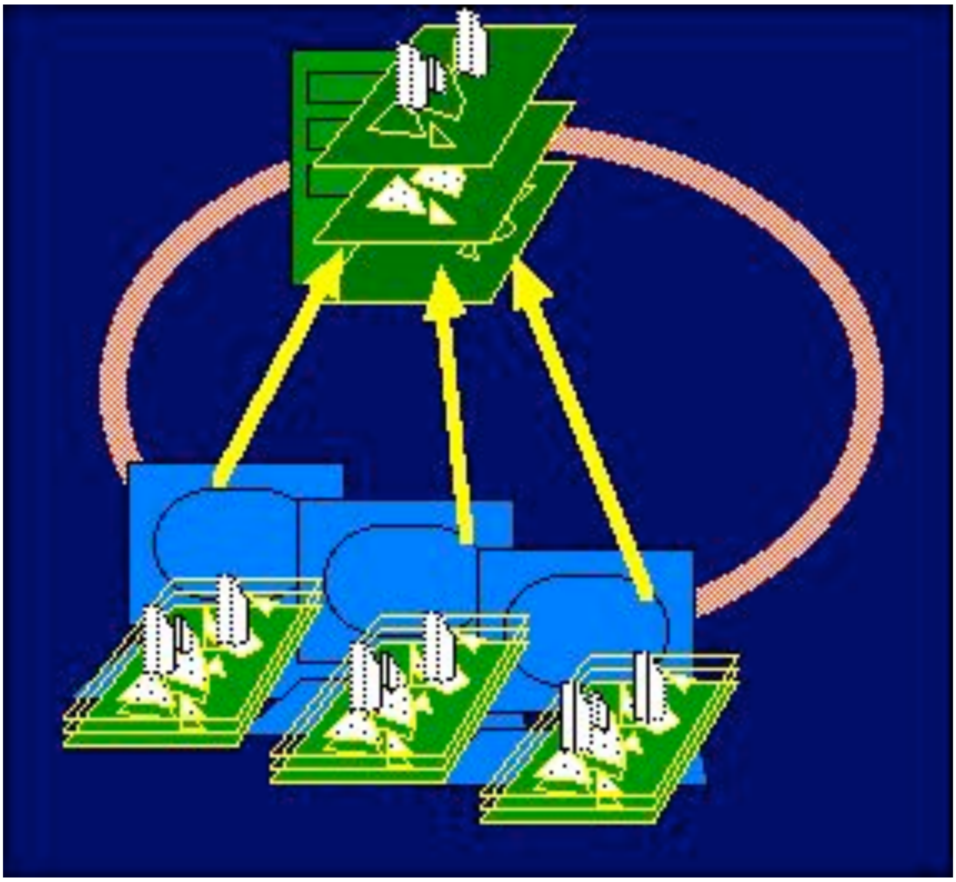

Figure 1.-(b) Centralized Type 2 / fig1b.jpg

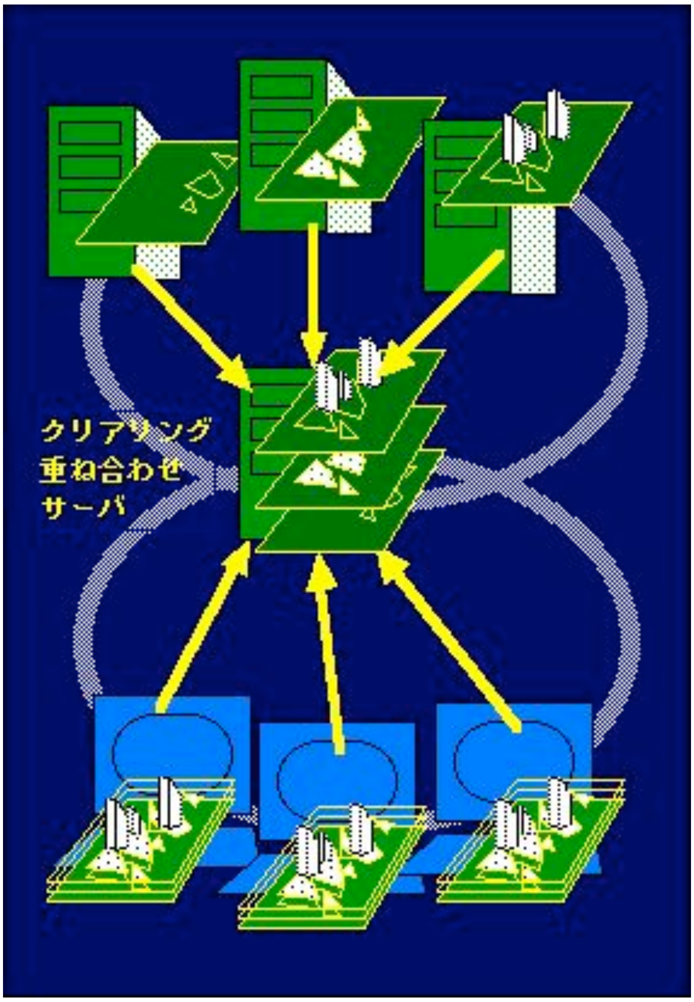

Figure 2. The Fully Autonomous Distributed Architecture Pursued by GLOBALBASE / fig2.jpg

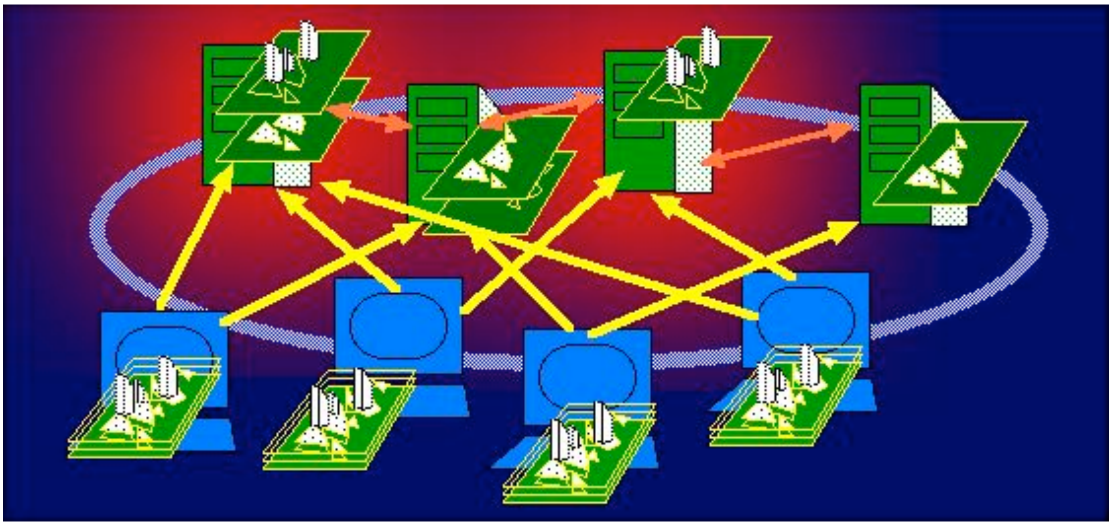

Figure 3. Coordinate Systems and Mappings / fig3.jpg

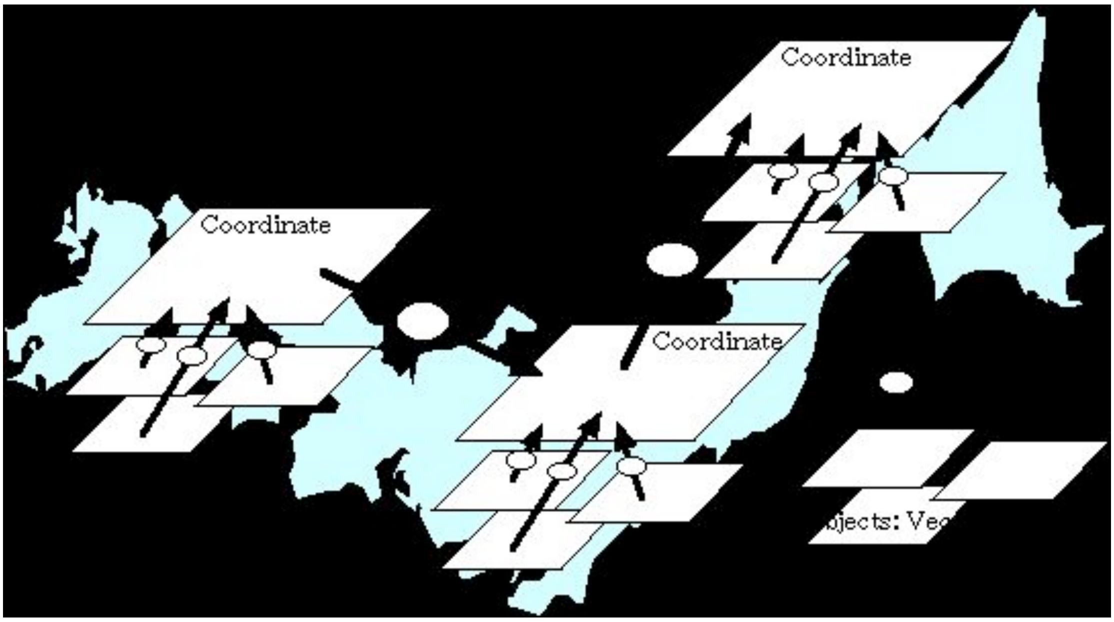

Figure 4. Local Search Engines / fig4.jpg

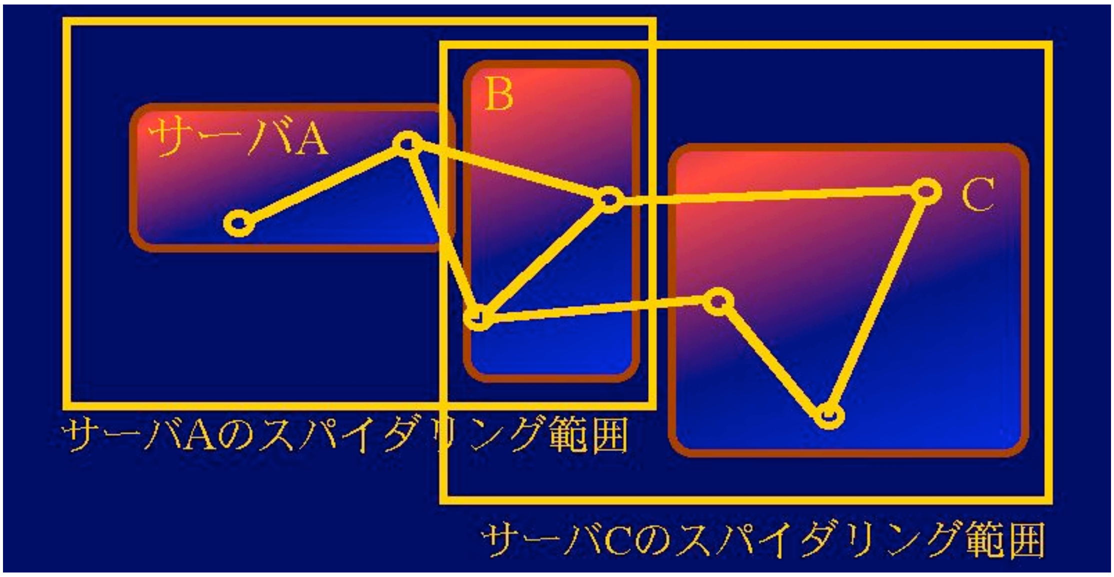

Figure 5. Movement of Browser / fig5.jpg