GEOGRAPHICAL INFORMATION APPLICATIONS OVER THE NET

Fiona Lee[1], Cui Yu[2] and Beng Chin Ooi[1]
[1] *National University of Singapore, Singapore*
[2]*Monmouth University, New Jersey, USA*

Keywords: Query processing, Wireless, Internet

Contents

Summary

The catalytic boost of critical mass usage of the World Wide Web (WWW) has brought about the shift in the traditional geographic information systems (GIS) paradigm and engendered new challenges and opportunities. The traditional GIS paradigm involves the use of huge databases and large expensive GIS software packages confined to local disks. The new paradigm, however, enables geospatial data to be shared and accessible in the world’s information system, the Internet. Current states of the art for Web GIS applications are client-side applications and server-side applications. Server-side applications create GIS pages based on a user’s query and request through a data form, while the client-side web GIS allows GIS analysis and data processing to be executed on the web browsers of the user’s local machine. In client-side applications, GIS data and analysis tools are sent as modules to the client upon the user’s request.

With the introduction of Java servlets and Java database connectivity, the development of high performance and interactive online GIS applications has been an easy task. The use of incremental query processing strategy for information retrieval has also proved to be a useful technique for interactive querying or drilling for online GIS operators. Lastly, the convergence of technologies such as the GIS, Internet and wireless communications, location determination, and portable devices have given rise to the new types of information utilities known as location services. A location service, in the broadest

sense, is any service or application that extends spatial information processing, or GIS capabilities, to end users via the Internet and/or wireless network. Geographical information will help revolutionize life styles and the commercial world, emergency services and domestic life with its location services.

1. Introduction

The traditional GIS paradigm involves the use of huge databases and large expensive GIS software packages confined to the local disks. Computer systems with identical and stale copies of the same data do not have the ability to speak with each other, making any form of geographical information integration difficult (see *GIS Interoperability: From Problems to Solutions*). The catalytic boost of critical mass usage of the WWW has brought about the shift in the traditional GIS paradigm.

In the new paradigm, geospatial data will be as accessible as any other data in the world's information system. Free metadata-based catalogs will be available online to help users quickly find specific data sets or geographic information. Interoperability will enable people to work not only within departments but also across departmental and enterprise boundaries. With the aim of interoperability, the Open GIS Consortiums (OCG) of industry and government recently developed the GIS interoperability standards so as to extend geodata and geoprocessing capabilities to regular web users (see *Geospatial Interoperability: The OGC Perspective*). The Consortium is a non-profit organization working toward integration of geospatial capabilities into the world's information system.

The Internet is a global communications network consisting of thousands of interconnected networks. Internet is now a major means for web GIS users to exchange, process and analyze geographical information (for other examples of applications, see *Advanced Geographic Information Systems*). Not only should web GIS have all, or almost all, the functionalities of the traditional GIS, there should not be a need for a user to install GIS data and software into his/her local computer.

Web GIS allow people to gain wider access to GIS information and visualize it in reports without the need to be experts. Experts will not be needed to convert from one GIS format to another or from one spatial reference system to another. People will be able to do their jobs without having to be expert in databases or keep proprietary software up to date; they will only need the knowledge of web navigation. Maintaining and updating servers, databases, and data will be the jobs of specialists.

There are two major web GIS applications for the current state of the art, server-side applications and client-side applications. Server-side applications rely on GIS servers to perform all GIS analysis, while client-side applications perform GIS analysis and processing on the web browser on the user's local machine. The popularity and pervasiveness of the WWW have increased the use of GIS applications because of data sharing and accessibility, but they have also introduced new challenges and opportunities. In this chapter, we shall examine various approaches to enabling GIS over the Internet, the effect on query processing, and various new applications.

2. Web Enabling GIS Processing

2.1. Connecting GIS to the Web

The WWW can be considered a large-scale Internet-based hypermedia information service system. With the popularity of the Web, the demand for large-scale legacy GIS to be Web-enabled is becoming irresistible. However, many of these legacy GIS were not built for WWW access, and are too costly, if not almost impossible, to modify to accommodate these new requirements. Hence, middlewares become a good alternative to provide Web connectivity and capability for access different from conventional full answer-set retrieval. (For more on computer networks, see *Computer Networks under Computer Science and Engineering*.)

Many approaches for interfacing databases from the WWW have been proposed over the last few years. These approaches include:

- *Common gateway interface (CGI).* This is a standard for external gateway programs to interface with servers. In this case, the input to this application program is simply an SQL query. The application initiates a connection to the underlying database management system, submits the query for evaluation, and returns the answer to the Web server, which forwards it to the client. It usually uses the C or Perl programming language.
- *Server-side include (SSI).* This uses a server that knows how to cope with database commands. This is implemented by extending the web server. Netscape's LiveWire, for instance, allows direct SQL connections to Oracle, Sybase and Informix databases as well as ODBC-connections to databases that are ODBC-compliant.
- *Database extended with HTTP-functionalities.* This would allow a database to be called directly from a browser, but would not be useful in the case of legacy databases that are already committed to a database management system (DBMS) platform.
- *Client side-include (CSI).* Plug-ins, Java applets, and Java scripts are all examples of recent innovations that allow client functionality to be extended. For example, an HTML document can be used to trigger the downloading of a Java applet. Once the applet has been loaded, it initiates and maintains a connection with a remote database using JDBC (Java database connectivity) without having to go through any Web server.

These approaches are really useful for web database applications. However they are not able to fully realize requests such as interactivity, high performance, and platform independency. For example, we need to modify our CGI programs when we want to move them from one platform to another. Each time the HTTP server receives a request for a CGI application, it must start a new process to run the application, which means less interactivity. Recently, Java servlets were introduced. Coupled with JDBC, they can be used to build high-performance, interactive and platform-independent web database applications. With Java servlet techniques, we an create high performance applications for information retrieval (IR), especially for incremental information retrieval over the WWW.

2.2. Progressive Query Processing

The mass of information and limitation of bandwidth of the WWW have brought about a shift in the paradigm of query retrieval. To the average computer user, 'searching' now means using IR-based systems to find information on the WWW or in other collections. The WWW, on the other hand, offers great potential for information exchange on an unprecedented scale. With growing demand for information exchange; more and more corporations are attempting to make corporate databases, or at least part of them, available on the WWW. Some of these legacy databases are terabytes in size, and it is common for queries issued against these databases to return gigabytes of data. There are two major sources of inefficiencies for information retrieval of legacy databases through the WWW. The first bottleneck lies with the network bandwidth: the transmission of a large document across long-haul networks is not only wasteful but also time-consuming. The second source of inefficiencies is the deprivation of the limited resources at the client side: a document delivered to a client has to be stored and read into main memory by the web browser. Consequently, a large document imposes significant strain on the client's resources and often results in sluggish performances.

These inefficiencies have in turn precipitated radical shifts in the way decision makers view information. Instead of identifying all information relevant to a decision, decision makers frequently take small 'nibbles' in the search for information. Anecdotal evidence suggests that the 'nibbling' approach to information consumption is effective for at least two reasons. First, decision makers are only boundedly rational and are increasingly incapable of coping with the sheer amount of information available. Under these circumstances, there is no point in collecting information that cannot be digested fully or made use of. By obtaining information in small chunks, the decision maker can be sure that the marginal cost expended in information retrieval can remain close to the marginal utility obtained from processing the additional information. In other words, taking small nibbles is an efficient strategy here.

Incremental information retrieval is a process of finding information bit by bit or bucket by bucket. Several approaches to retrieving information from legacy database systems have been proposed. A cursor-based approach at the application level is very inefficient for it still needs to submit the entire query. An approach at the DBMS level can be achieved by adding a STOP AFTER clause to SQL. However, the approach is costly and time-consuming. Alternatively, a strategy of using incremental query processing can be achieved by incorporating a progressive processing strategy to an existing database system in a nonintrusive manner. This is accomplished by rewriting a user query into distinct subqueries, each of which returns answers of a fixed size. The advantage of this method is that an existing optimizer can be reused without modification, resulting in a shorter development cycle and lower cost. This incremental information retrieval is however mainly suitable for interactive querying or drilling. (For comments on query processing issues, see *Spatial Query Languages.)*

To avoid batch processing due to blocking operations, online GIS operators are typically non-blocking operations that allow the user to begin seeing partial results immediately, while refining and improving the quality of the results progressively. As a means to speed up retrieval, approximate information is retrieved when small errors can be tolerated. As an example, approximate nearest neighbors are returned to a query asking for K nearest neighbors of a given point, and they are progressively refined as more nearer neighbors are found.

-
-
-

Bibliography

Bressan S., Bonnet B. and Granveaud B. (1996). The World Travel Editor: a geographical search engine for the Internet. *International Conference on Extending Database Technology EDBT 1996*, Avignon, France, 1996. [The paper demonstrates the World Travel Editor, an Internet geographic search engine in the cartographic paradigm, using Java and spatial database components.]

Hill L.L, Frew J. and Zheng Q. (1999). Geographic names: the implementation of a gazetteer in a georeferenced digital library. *D-Lib Magazine*, **5**(1). [The paper describes the development of the gazetteer component of the Alexandria Digital Library.]

Lee F., Bressan S. and Ooi B.C. (2001). Hybrid transformation for indexing and searching web documents in the cartographic paradigm. *Information Systems* **26**(2): 75-92. [The paper presents the design and implementation issues of Global Atlas, a geographical web search engine, and evaluates various surface-fitting techniques for map calibration.]

Tan K.L., Goh C.H. and Ooi B.C. (1999). On getting some answers quickly, and perhaps more later. *Proceedings of the International Conference on Data Engineering*, Sydney, Australia, pp. 32–39. [The paper presents studies on how a query can be rewritten into sub-queries so that users can obtain the answers to the first sub-query quickly.]

Biographical Sketches

Fiona Lee is a Teaching Assistant in the Computer Science department of the School of Computing at the National University of Singapore. She is also currently pursuing a M.Sc. degree in Computer Science. Fiona holds a B.Sc. (Hons.) degree in Computer Science (1999). At NUS, Fiona has been the tutor for courses on management science, digital logic and computer organizations. Her research interests include web-based geographical information systems and registration methods for Internet maps. Fiona is currently working on a project called Global Atlas: a global WWW search engine based on spatial and semi-structured information. She has had articles published in *Proceedings of the WISE Conference* and the *Information Systems Journal*.

Cui Yu is Assistant Professor at the Monmouth University. She obtained her B.Eng. in Computer Science, Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, China in 1995. She obtained her Ph.D. in Computer Science, Department of Computer Science, National University of Singapore in 2001. Her research interests include database management systems, spatial databases, multimedia databases, multi-/high-dimensional databases, data warehouses, parallel and distributed databases, Internet databases, query processing and optimization, indexing, clustering techniques. She has had articles published in *ACM PODS*, *VLDB* and *VLDB Journal*, and she is the author of the monograph entitled High-Dimensional Databases, Lecture Notes in Computer Science #2341, Springer-Verlag.

Beng Chin Ooi is Professor at the National University of Singapore. He obtained his B.Sc. (Honors) and Ph.D. from Monash University, Australia, in 1985 and 1989 respectively. His research interests include database performance issues, indexing techniques, multimedia databases, high-dimensional databases and Internet applications. He has served as a PC member for SIGMOD‘94, VLDB‘96-97, 99-02, ICDE‘02, EDBT‘98, 02, DASFAA‘93-01, ACM-GIS‘98-01, and SSD‘93-99, and was ICDE‘00 Vice PC Chair, SSD‘93 PC Chair, FEGIS‘93 Workshop Chair, and MDM‘02 Conference Chair. He is an editor of *GeoInformatica, the Journal of GIS*, ACM SIGMOD Disc and the *VLDB Journal*.