Searching for Geographic Data Sources for Land Use Planning

Nancy Wiegand

University of Wisconsin-Madison
B102 Steenbock Library, 550 Babcock Drive
Madison, Wisconsin 53706
wiegand@cs.wisc.edu

Abstract

This paper presents a method to locate geospatial data sets over the Web that are needed to create land use plans. Land use planning is a critical issue for many communities, and current methods of finding data sets are quite cumbersome. We propose a Web Service solution to automatically find the needed data sources for each jurisdiction. The method uses Internet DBMS querying over XML metadata files. Semantic Web research is needed to identify metadata fields that will allow definitive search results. Also, if standard metadata is not used, it is necessary to resolve heterogeneities in the terminology of independently created metadata fields. In addition, values used in the metadata fields must be semantically resolved.

Introduction

This position paper discusses the problem of finding geographic data sources over the Web that are needed for land use planning. Land use planning is an urgent and critical issue for many communities due to intense development pressures. Furthermore, to help insure an orderly growth pattern, many states have recently passed Smart Growth legislation that requires communities to prepare a land use plan. A great deal of new work to create Web resources for land use planning is currently being done (e.g., LICGF 2003). Although there are many issues involved with land use planning, we focus here on the data collection/data discovery problem. Current methods for searching and retrieving data sets for land use planning are extremely cumbersome. This is an area in which semantic Web technologies could be very useful.

For planners to create a land use plan, many types of data sets are needed such as shown in Tables 1 and 2. This information is adopted from (Ohm 1999) for Wisconsin but is likely to apply to other states. A land use planner must gather all the necessary data sets. Currently, obtaining the requisite data sets often involves phone calls and mail delivery. Once obtained, data sets that include spatial references will be spatially integrated using GIS software to help make location-based decisions. Other types of analyses are done to help create a land use plan, but this paper focuses on the data-gathering problem.

We start from the assumption that all relevant data sets are on Web sites as this is likely to be a near-future scenario. The problem, then, for the land use planner is to find the appropriate Web sites and the data sets on the Web sites. This is a confusing and time-consuming task. However, this is a type of problem in which the basic mechanisms for data retrieval are duplicated for each jurisdiction creating a land use plan. Therefore, an automated solution is worth developing. We propose a solution involving metadata and Web services.

Human Made Environment-Public (see Table 2)
Human Made Environment-Private
Population Characteristics
Economic and Employment
Natural Environment

Table 1. General Categories of Data Needed for Land Use Planning (Ohm 1999)

Querying Metadata

Metadata is currently used to locate data sources registered to a particular Web site. Instead, we suggest metadata be published over the Web in general and XML DBMS techniques be used to query metadata for the purpose of finding data sources over the Web.

Metadata

In the geospatial community, metadata forms a separate file from the data source it describes. The Federal Geographic Data Committee (FGDC) has established standards for metadata for geospatial data. Rather than using full FGDC metadata, which may be lengthy and not suitable for a Web search mechanism, minimal metadata fields are typically used in geospatial Web-based information systems to

Type of Information to Collect	Potential Information Sources	Providers
Transportation	Road classif. maps State Highway maps Regional transp. maps	Municipal/County government, Libraries, Railroad companies, RPCs, WisDOT
Sewer Water Drainage Solid Waste	Utilities maps Utilities master plans Capital Improve. Programs Drainage maps USGS maps Field surveys & inventories Interviews with local service providers	Municipal/County government, Libraries, local providers, USGS
Emergency and Public Safety Schools Parks and Recreation Libraries and Public Buildings	Local govern. budget docs Master plans Field surveys & inventories Interviews with local service providers Tourism maps and guides	Municipal/County government, Libraries, local providers, chambers of commerce
Historical/Archeological Resources	Publications Maps Inventories Historical docs	Libraries, State Historical Society of Wisconsin

Table 2. Detail of Categories for Human Made Environment-Public (Expanded from Table 1) (Ohm 1999)

disseminate geographic data sets. Examples are clearinghouses that can be found at various Web sites including those for the FGDC, the Geography Network, WiscLinc, and the Alexandria Digital Library project. Furthermore, there is a new Federal clearinghouse called Geospatial One-Stop. However, currently, there does not appear to be coordination between these geospatial sites as to common metadata standards and terminology. Also, the metadata itself is not available over the Web; it is just used internally by each site.

Even if metadata were now made available over the Web as HTML pages, searching metadata files to locate data sources would have limited usefulness with current keyword search mechanisms. Without being able to pose more query-like expressions using conjunctive "text in context" predicates, too many hits may occur over many irrelevant Web pages.

Determining the appropriate minimal metadata fields is important to enable querying of metadata over the Web. Ideally, an information community would propose a common standard for minimal metadata to adequately describe data sources in their area. For land use planning data sets, we initially propose the following fields:

- jurisdiction type (e.g., city),
- jurisdiction name (e.g., Madison, WI),
- spatial coordinates in latitude and longitude,
- one or more attributes describing a "theme" (e.g., transportation, land use, schools, etc.),
- date,
- a URL to a full FGDC metadata file if one exists,
- and, finally, a URL to the data set itself.

Themes may need subcategories such as transportation-statehighways. Discussions with other scientific communities would be needed to determine the applicability of these fields to other types of data.

DBMS XML Web Querying

We suggest minimal metadata files be published in XML over the Web along with the actual data sources. XML tags will describe each metadata attribute (e.g., <jurisdiction_type> *city* </jurisdiction_type>). Then, Web XML search engines can be used to perform "text in context" queries over these metadata files (Figure 1). Furthermore, metadata files could have identifying extensions, such as "file$_1$.meta_xml". If so, the search engine could search just metadata files, rather than all XML files. An example query may specify:

jurisdiction_type = "*city*" and jurisdiction_name = "*Madison, WI*" and theme = "*transportation-roads*" and year >= *2000*.

The intent is to have the metadata fields and user-selected values from the query uniquely identify a metadata file. From that file, the URL for the data source is obtained.

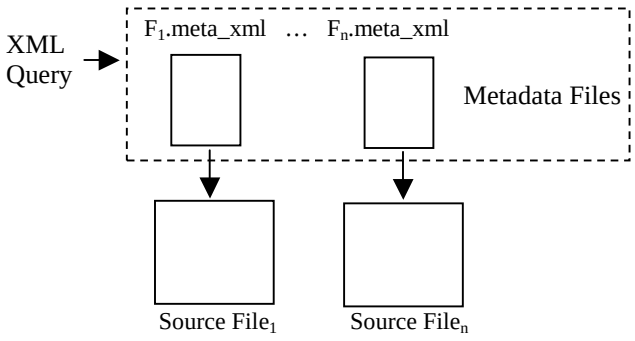

Figure 1. Querying Independently Published Metadata Files that Describe Source Files

Web Service to Find Land Use Planning Data

As stated, a land use planner from a particular jurisdiction needs to gather many data sets to form a land use plan. We envision a Web Service to automatically retrieve the needed data sources using our method of Web-published metadata and Internet DBMS querying. The user locates the Web Service from a directory by specifying the need to

"locate land use planning data sources". The Web Service is designed such that a land use planner first specifies either spatial coordinates or jurisdiction information to identify an area or jurisdiction. The Web Service then supplies a list of potential types of data sources needed for land use planning, as per Tables 1 and 2. As the user checks each type of data, the Web Service formulates a query and activates an XML DBMS Web query engine. The query engine processes the query over XML metadata files and returns the needed URL.

Semantic Issues

There are various semantic problems that need solutions for this automated method to work. One problem is determining the minimal metadata. This requires an in-depth analysis of the characteristics of the data sources with respect to application needs and user concepts. However, even if standard metadata fields are determined and used by data publishers, the contents or values used within the fields are likely to vary. A method would be needed to interpret and compare values.

In addition, data publishers may not use standard metadata fields even if they are available. Data publishers may just choose their own fields and use their own terms to represent those fields. Again, interpretation and mappings would be needed. Although the task of resolving these semantic differences may seem difficult, the geospatial community shares many common concepts.

We have worked on a related problem regarding data sets containing land use information (Cruz et al. 2002; Wiegand, Zhou, and Cruz 2003). For that work, we resolved schema level heterogeneities for land use data sets and resolved value level heterogeneities for various land use coding systems. We integrated an ontology subsystem into an Internet query engine. Some of the methods from that project may be useful in developing the Web Service proposed here.

Discussion

The method presented here is promising for future data set discovery over the Web. Although some research has been done to automatically find types of data sets using values found in the data itself (e.g., Kang and Naughton 2003), such methods may need many sample cases and also do not definitively find a data source. Instead, we propose adding metadata to the search environment of the Web and using XML Web query technology to search the metadata to locate data sources. We believe asking data publishers to additionally publish a small metadata file will not be a burden.

XML querying is now being deployed over the Web (e.g., Naughton, DeWitt, and Maier 2001). However, future work could extend such systems to include spatial processing such as in (Egenhofer 2002). Adding spatial search capabilities would greatly enhance the data search process.

As far as we know, the idea of independently publishing metadata files on the Web for the purpose of querying them to find data sources is new. We believe this is a workable approach that can take advantage of the availability of XML Internet query engines.

Acknowledgement

This work was partially supported by the Digital Government Program of NSF, Grant No. 091489.

References

Cruz, I. F.; Rajendran, A.; Sunna, W.; Wiegand, N. 2002. Handling Semantic Heterogeneities Using Declarative Agreements. In Proceedings *of the Tenth ACM International Symposium on Advances in Geographic Information Systems*, 168-174. McLean, Virginia: ACM Press.

Egenhofer, Max. 2002. Toward the Semantic Geospatial Web. In *Proceedings of the Tenth ACM International Symposium on Advances in Geographic Information Systems*, 1-4. McLean, Virginia: ACM Press.

Kang, J., and Naughton, J. F. 2003. On Schema Matching with Opaque Column Names and Data Values. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 205-216. San Diego, California: ACM Press.

LICGF. 2003. Community Planning Resources Web site of the Land Information and Computer Graphics Facility, University of Wisconsin-Madison, www.lic.wisc.edu.

Naughton, J. F.; DeWitt, D.; Maier, D.; and others. 2001. The Niagara Internet Query System, *IEEE Data Engineering Bulletin* 24(2): 27-33.

Ohm, B. 1999. *Guide to Community Planning in Wisconsin*. Cooperative Extension Publishing. http://www.uwex.edu/ces/pubs.

Wiegand, N.; Zhou, N.; and Cruz, I. F. 2003. A Web Query System for Heterogeneous Geospatial Data. In Proceedings of the Fifteenth International Conference on Scientific and Statistical Database Management (SSDBM), 262-265. Cambridge, MA: IEEE Computer Society.