Design of Framework for Semantic Annotation of Geospatial Data

Preetam Naik, Madhuri Rao, S. S. Mantha, and J. A. Gokhale

***Abstract*—Framework is a key providing for effective development, exchange, and use of geospatial as well as non-geospatial data to meet information needs generated by public domain policy. The development of multiple spatial ontologies, and their representation should be done in such a way that the computer can understand and process. The processing of queries considering these ontologies and the evaluation of results should be based on their semantics.**

***Index Terms*–Geographic information system, ontology, semantic annotation**

I. Introduction

Web becomes an immense repository of geospatial data in different geographic formats like remote sensing images, maps, textual data files etc. The retrieval of these data requires special attention due to geographical distribution of the sources and the heterogeneity of the data [1].

Popularity of GIS (Geographic Information Systems) in government and municipality institutions induces an increasing amount of available information. In local community environment (city services, local offices, local telecom, water and power supply services, etc) different information systems deal with huge amount of available information, where most of data in databases are georeferenced. But, information communities find it difficult to locate and retrieve data from other sources, in reliable and acceptable form. In such systems, geodata reuse is very difficult process, because of poor documentation, obscure semantics, diversity of data sets and heterogeneity of data modeling concepts, data encoding techniques and storage structures.

Data storage and retrieval subsystem are necessary component for a GIS. This subsystem organizes the data, both spatial and attribute, form which permits it to be quickly retrieved for updating, querying, and analysis. Most GIS software utilizes proprietary software for their spatial editing and retrieval system, and a database management system (DBMS) for their attribute storage. Typically, an internal data model is used to store primary attribute data associated with the topological definition of the spatial data. Most often these internal database tables contain primary columns such as area, perimeter, length, and internal feature id number. Often thematic attribute data is maintained in an external DBMS that is linked to the spatial data via the internal database table.

Manuscript received February 22, 2012; revised April 20, 2012.

P. Naik is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (e-mail: author@ boulder.nist.gov).

M. Rao was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar. colostate.edu).

J. A. Gokhale is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp).

II. Background

WebMAPS [1] is a project that aims to provide a platform based on services to formulate, perform and evaluate policies and activities in agro-environmental planning. It involves state-of-the-art research in specification and implementation of software that relies on heterogeneous, scientific and distributed information, such as satellite images, data from sensors and geographic data. The project caters to two kinds of users (farmers), and domain experts, such as agronomers or earth scientists. They can visualize geospatial variables concerning their properties (farmers), or analyze and monitor crop behavior (scientists). Using satellite images, users can search for an image based on content, such as texture or color. Fig. 1 gives an overview of WebMAPS' 3-layer architecture.

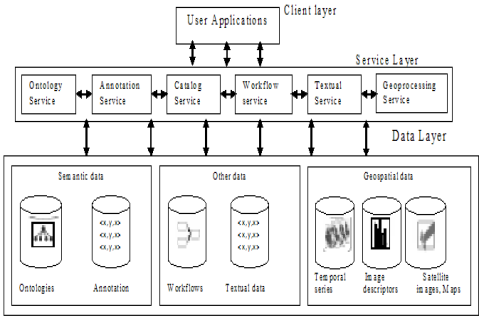


Figure 1 Three layer architecture of WebMAPS

Fig. 1. Three layer architecture of WebMAPS

There are three main layers in the architecture. They are:

Client layer is responsible for processing a user request, forwarding it to be processed by the middle layer and presenting the returned result.

Service (middle) layer provides services such as textual and geospatial data management and ontology management.

Data layer contains primary raw data (productivity data) and derived data (composite images). It also includes geospatial data such as satellite images, region boundaries, crop information etc.

III. Proposed Architecture

The design of proposed framework is shown in the Fig.

2.

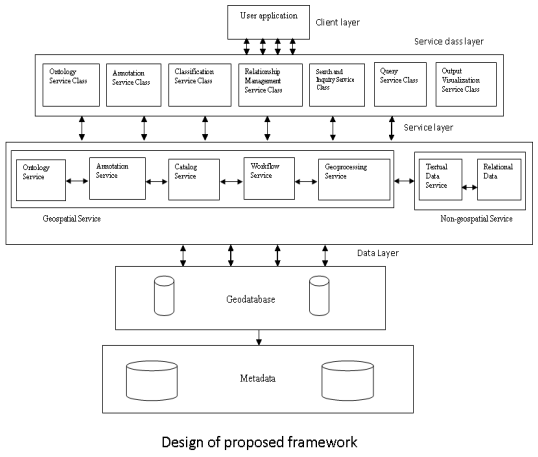


Fig. 2. Design of proposed framework

There are basically four layers: client layer, service class layer, service layer and data layer.

Client Layer is used to interact with user and the service class. With the help of it, user can select different services as per the requirements.

Service Class Layer is an intermediate layer. It lies between client layer and the service layer. It contains several necessary functions such as high performance GIS services, multi-access to other services, high stability and reliability (e.g. load balance) and high security.

It provides seven types of service classes:

Ontology Service Class
Annotation Service Class
Classification Service Class
Relationship Management Service Class
Search and Inquiry Service Class
Query Service Class
Output Visualization Service Class

The functionality of each service class is explained in detail as following:

1) Ontology Service Class is used for accessing, browsing, searching and traversing ontologies. It can perform semantic matching of queries to return results that are logically related to user's request, rather than searching only for exact information using attribute-value matching. Then it combines the results obtained by ontology service with keyword matching [1].
2) Annotation Service Class provides the template for annotation of different kind of digital contents such as textual documents, images and multimedia documents etc. Instead of manually maintaining consistency among the entire set of files, only the base file would need to be maintained since the derived file (source files, class files, deployment descriptor) are generated.
3) Classification Service Class is a procedure of identifying a set of features as belonging to a group and defining patterns like raster-based GIS, vector-based GIS. Raster data represent features as a matrix of cells within rows and columns in continuous space. These cells are formed by pixels of a specific dimension size, and can be described as either "cell-based" or "image-based" data. Vector data comes in the form of points and lines that are geometrically and mathematically associated. Points are stored using the coordinates, for example, a two-dimensional point is stored as (x, y) [2].
4) Relationship Management Service Class describes spatial relationship between two or more objects like distance, direction, topology etc. Objects are created by connecting points with straight lines some systems allow points to be connected using arcs of circles areas are defined by sets of lines. GIS topology is a set of rules and behaviors that model how points, lines and polygons share geometry [3].
5) Search and Inquiry Service Class allows searching and inquiring data from the service layer. Usually, the search for these data and methods is done by their syntactic content, focusing primarily in keyword matching. This can lead to the retrieval of irrelevant data, disregarding relevant files.
6) Query Service Class determines a statement or logical expression that selects geographic features based on location or spatial relationship. It is a service that allows a user to find, display, and/or isolate attributes records linked to map features located within a defined area of interest - window, circle, polygon or trace.
7) Output Visualization Service Class deals with displaying information regarding geospatial as well as non-geospatial data.

Service Layer aims at providing value-added services and applications at the highest application layer. Geospatial service has further six types of services:

1) Ontology Service is responsible for handling ontologies. It provides wide range of operations to store, manage, search, rank, analyze and integrate ontologies. Generally geospatial services have no semantic service description. Ontologies are used within the context of geospatial data infrastructures to denote a formally represented knowledge that is used to improve data sharing and information retrieval. So it is very difficult and time consuming to invoke a geospatial service correctly. To solve this problem, ontology is explicitly used for semantic service description [4].
1) Annotation Service semantically annotates different kind of geospatial data, such as satellite images, maps and graphs. In geographic applications, annotations should also consider the spatial component, since geographic information associates objects and events to localities, through places and geographic object names, spatial relationships and standards.
2) Catalog Service is responsible for management and publishing of the produced annotations. Annotation generation requires accessing several data sources, including external data. The desired data will be discovered through metadata catalogs. Catalog service helps interfaces to discover, browse, and query metadata about data, services, and other potential resources.
3) Workflow Service executes and manages workflow of annotations. It selects the annotation workflow to be performed based on the nature of the content to be annotated. Then it comprises the execution of the selected workflow [5].
4) Geoprocessing Service manipulates and analyzes geographic information. It supports a limited set of input and output data types.

Non-geospatial Service contains data in textual form, relational form etc. Textual Data Service is responsible for all operations performed on textual data, like input and query processing. Relational Data Service contains data in

the form of rows and columns.

Geodatabase is a database with extensions for storing, indexing, querying, and manipulating geographic information and spatial data. It is the common data storage and management framework for GIS. It combines geo (spatial data) with database (data repository) to create a central data repository for spatial data storage and management. It allows to store GIS data in a central location for easy access and management. It stores a rich collection of spatial data in a centralized location. It maintains the integrity of spatial data with consistent and accurate database.

Metadata is a summary document providing content, quality, type, creation, and spatial information about a data set. It can be stored in any format such as a text file, Extensible Markup Language (XML), or database record.

IV. Related Work

Ontology-Based Geospatial Clustering

Geospatial clustering is an important topic in knowledge discovery and geospatial information systems (GIS). It can be used to find natural clusters (e.g., extracting the type of land use from the satellite imagery), to identify hot spots (e.g., epidemics, crime, traffic accidents), and to partition an area based on utility (e.g., market area assignment by minimizing the distance to customers). The framework for ontology-based geospatial clustering is shown in Figure 3 as:

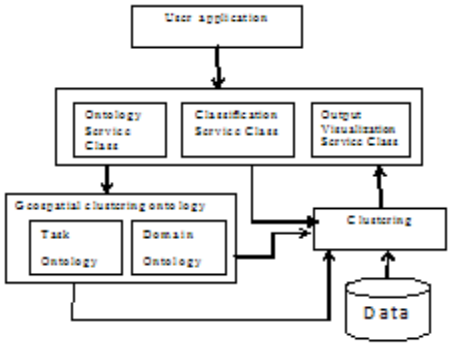


Fig. 3. Framework for ontology-based geospatial clustering.

Ontology Service Class will be used to support rule-based reasoning for spatial clustering ontology.

Classification Service Class will allow users to effectively explore and select appropriate data relevant to user's task.

Geospatial Clustering Ontology (ontology service) component is used when identifying the clustering problem and the relevant data.

Within this component, the **task ontology** specifies the potential methods that may be suitable for meeting the user's goals, and the **domain ontology** includes all classes, instances, and axioms in a geospatial domain.

V. General Applicability

GIS is used for various types of applications such as agriculture, transportation, environment, health, disaster management, banking, telecom etc.

In agriculture, for example, GIS is used to manage crop yields, to monitor crop rotation technique, and to project for soil loss for individual or entire agricultural regions as well as for spatial analysis, the study of geographic features, and the relationships that exist between them. Following are agriculture GIS based applications:

- Farm asset allocation
- Combining agronomic and economic data sets
- Decision support for farm management
- Farm asset allocation

In this way, GIS is playing an increasing role in agriculture production throughout the world by helping farmers increase production, reduce costs, and manage their land more efficiently.

Acknowledgement

P.N. Author thanks to project guide Ms. Madhuri Rao for her constant encouragement, intellectual and valuable suggestions, and family members for providing their support during project work.

References

[1] C. Geovana, N. Marcario, C. Bauzer Medeiros, "Specification of a framework for semantic annotation of geospatial data on the web" *IEEE International Conference on Challenges in Environmental Science and Computer Engineering,* pp. 27-32, 2009.

[2] W. Shanming and S. Jianjing: "Ontology-based framework for geospatial web services," *IEEE International Symposium on Information Science and Engineering,* pp. 107-110, 2008.

[3] S. Roberto de Sousa: "A semantic approach to describe geospatial resources", ER 2009 Workshops, LNCS 5833, pp. 327-336, 2009.

[4] G. Z. Pastorello Jr, R. D. A. Senra, Claudia Bauzer Medeiros: "A standards-based framework to foster geospatial data and process interoperability," IEEE International Conference on Advanced Computer Control, Singapore, pp. 22-24, 2009.

[5] C.B Medeiros, J. Perez-Alcazar, L.Digiampietri, G.Z.Pastorello Jr, A .Santanche, R. S. Torres, E. Madeira, and E. Bacarin: "WOODSS and the Web: annotating and reusing scientific workflows" of the ACM SIGMOD Record, vol. 34, pp. 18-23, 2005.



Preetam Naik has completed her B.E. in Computer Engineering from Pune University, INDIA in the year 2003. She is pursuing her M.E. in Computer Engineering from Mumbai University, INDIA. Presently she is working as Assistant Professor in Shah and Anchor Kutchhi Engineering College, Mumbai, INDIA. Her research interest includes Database Management System, Data warehousing and Mining and Geographic Information system.



Madhuri Rao has completed her B.E. in Computer Engineering from Shivaji University, INDIA, in the year 1996. Also she has completed her M.E.in Computer Engineering from Mumbai University, Mumbai, INDIA in the year 2003. Presently she is working as an Assistant Professor in IT department in Thadomal Shahani Engineering College, Mumbai, INDIA. Her areas of interest are Data Mining, GIS and Software Engineering.