BUILDING A SECURE GEOSPATIAL SEMANTIC WEB

Raytheon Corporation and the University of Texas at Dallas

ABSTRACT

Semantic web is a collection of technologies that enable machine understandable web pages. In recent years there has been a lot of research on the semantic web. More recently there is some work on geospatial semantic web and secure semantic web. Our collaborative research attempts to integrate both geospatial semantic web and secure semantic web to develop secure geospatial semantic web.

1. INTRODUCTION

Geospatial semantic web integrates semantic web technologies with geospatial technologies and security technologies. Tim, Berners Lee proposed the semantic web in order to create machine understandable web pages. A semantic web can be thought of as a web that is highly intelligent and sophisticated so that one needs little or no human intervention to carry out tasks such as scheduling appointments, coordinating activities, searching for complex documents as well as integrating disparate databases and information systems. Geospatial data emanates from numerous devices at multiple sites. Such data is complex and heterogeneous in nature. In order to integrate heterogeneous geospatial data sources securely, we need to develop semantic web technologies that handle geospatial data.

In this paper we describe the collaborative research between Raytheon and the University of Texas at Dallas on building a secure geospatial semantic web with crime analysis as our application area. We have focused on three major aspects: geospatial semantic web; geospatial data mining and security. The organization of this paper is as follows. Our example application is discussed in section 2. Our research on geospatial semantic web as well as the DAGIS system we have developed is discussed in Section 3. Our research on geospatial data mining is discussed in section 4. Security considerations including secure interoperability of geospatial data is discussed in section 5. The paper is concluded in section 6.

2. CRIME ANALYSIS APPLICATION

We are particularly interested in Policy Blotter Crime Analysis. Police Blotter is the daily written record of events (as arrests) in a police station which is released by every police station. These records are available publicly on the web which provides us wealth of information for analyzing the crime patterns across multiple jurisdictions. The Police Blotters are available to public or between police departments are generated from legacy systems and may also be published as web documents.

There are major challenges that a police officer would face when he wants to analyze different police blotters to study a pattern (e.g., a spatial-temporal activity pattern) or trail of events. There is no way a police officer can pose a query where query will be handled by considering more than one distributed police blotters on the fly. With the advance of Web 2.0 , there are some mashups of Google Maps with police blotters of some counties. There is not a cohesive tool for the police officer to view the blotters from different counties, interact and visualize the trail of crimes and generate analysis reports. The Blotters can currently searched only by keyword through current tools and does not allow conceptual search, and fails to identify spatial – temporal patterns and connect various dots/pieces. Therefore, we need a tool that will integrate distributed multiple police blotters, extract semantic information from a police blotter and provide seamless framework for queries with multiple granularities. Our research is developing such a tool based on geospatial semantic web, data mining and security technologies.

3. GEOSPATIAL SEMANTIC WEB

Secure geospatial semantic web is an integration of secure semantic web and geospatial semantic web. Follow along the vision of Tim Bernes Lee for the semantic web, we have defined a layered architecture for a geospatial semantic web. At the bottom layer are the protocols for communication. Next we have the GML (Geography Markup Language) and GML schemas layer. We have developed GRDF (Geospatial RDF) to specify the semantics and the GRDF layer lies on top of the GML layer. On top of GRDF we have developed geospatial ontologies and query facilities. Below we give an example of GRDF.

```
<owl:Class rdf:id=http://127.0.0.1/grdf_voc#City />
   <rdfs:subClassOf rdf:resource="http://127.0.0.1/grdf_voc#tplace"/>
   <rdf:Property rdf:about="http:// 127.0.0.1/grdf_voc#boundary">
        <rdfs:domain rdf:resource="http://127.0.0.1/grdf_voc#
CRS_EPSG:6.6:4326/>
    </rdf:Property>
</rdfs:Class>
```

In this example we have defined a City class (or concept in ontology parlance), which has a property that identifies the boundary extent of a particular city. The City class is also a subclass of the Place class and as a result inherits the latter class's properties.

Web Services and the standards provided by OGC (Open Geospatial Consortium) define our approach for a geospatial semantic web. Client queries the Service Requestor Web Service, which handles the GIS Application. The Service Requestor then discovers the required Service Provider through the Service Registry or the Match Maker. The Service Registry selects the Service provider which has already registered with this registry. Service Provider can now bind with the Service Requestor to fulfill the service requestor. The underlying protocol stack is shown in the figure. By this way two different GIS application with different heterogeneities can interoperate with each other using Web Services.

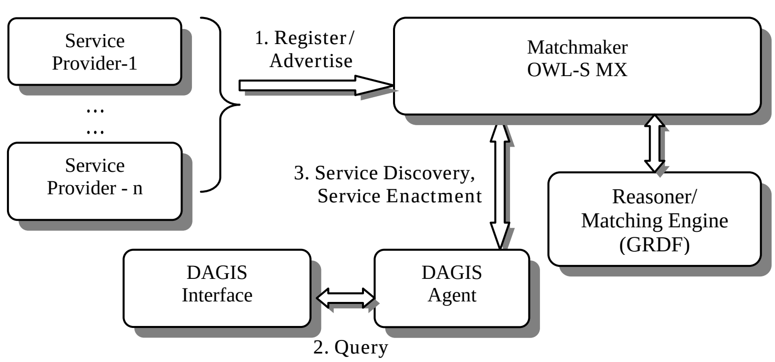


Figure 1. DAGIS System

We have developed a system called DAGIS (Discovery of Annotated Geospatial Information Services) that reasons with the ontologies and answers queries (see Figure 1). It is a framework which provides a methodology to realize the semantic interoperability both at the geospatial data encoding level and also for the service framework. DAGIS is an integrated platform that provides the mechanism and architecture for building geospatial data exchange interfaces using the OWL-S Service ontology. Coupled with the geospatial domain specific ontology for automatic discovery, dynamic composition and invocation of services, DAGIS is a one-stop platform to fetch and integrate geospatial data. The data encoding is in GRDF and provides the ability to reason about the payload data by the DAGIS or client agents to provide intelligent inferences. DAGIS at the Service level and GRDF at the data encoding layer provide a complete unified model for realizing the vision of geospatial semantic web. The architecture also enhances the query response for the client queries posed to DAGIS interface.

4. GEOSPATIAL DATA MINING

Geospatial data mining will have applications in crime analysis as well as in developing and refining ontologies. We are utilizing a multi-step approach for geospatial data mining for crime analysis, border control as well as for suspicious activity detection. In this section we will describe our initial research on developing data mining techniques for the classification of remote sensing data. Such classifications can be used to determine whether there are suspicious activities in particular regions.

Land cover information can be derived from various remote sensing systems, such as images from Landsat 7 ETM+, SPOT HRV/HRVIR, Terra ASTER and AVIRIS. The remote sensing images can have different spatial resolutions and spectral resolutions. Classification on the pixel level cannot reveal semantic concepts at higher levels, and the semantic concepts at high levels can be crucial for security protection, environment evaluation and urban open space research. For instance, if a pixel or a few neighboring pixels are classified as water body, the location can be a pool in a residential area, a pond in an urban area, or a lake in a park or open rural area. Similarly, a group of buildings can be for public service in an urban area, for residential purpose in a residential area, or for highly confidential military use in a desert. It is of great need to develop high level concepts and distinguish them so that the semantic meanings of pixel classes become clear and accesses to some confidential concepts become controllable for security consideration.

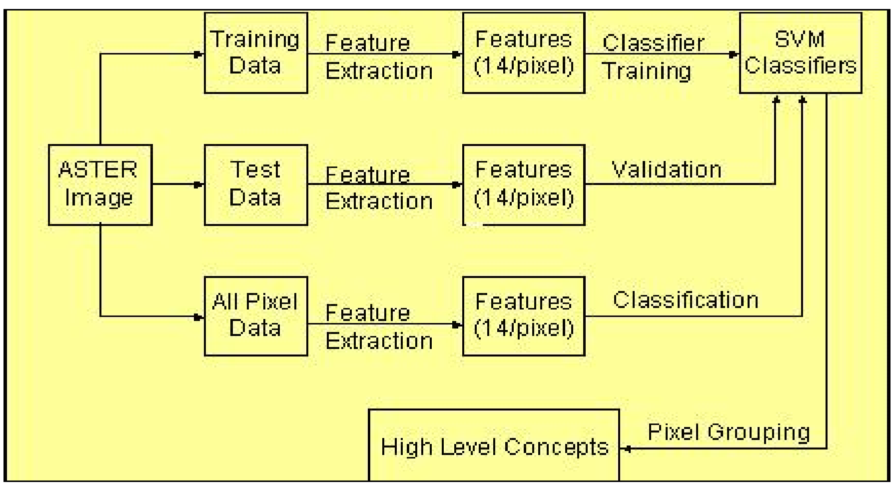


Figure 2. Geospatial Data Mining

Our proposed approach (illustrated in Figure 2) classifies data combined from different resolutions and forms high level concepts by grouping and re-evaluating classes of pixels. The classification is performed by using support vector machine (SVM) classifiers (see Table 1 & 2), which have been successfully demonstrated to outperform Maximum Likelihood (ML) and artificial neural network (ANN) classifiers.

Table 1. Training and Test Pixels and Accuracy

	Water	Barren Land	Grass	Tree	Building	Road	House	Total	Accuracy(%)
Training	1175	1005	952	887	1041	435	1584	7079	99.8
Test	1898	1617	1331	1479	768	648	1364	9105	89.25

Table 2. Confusion Matrix for Independent Test Data

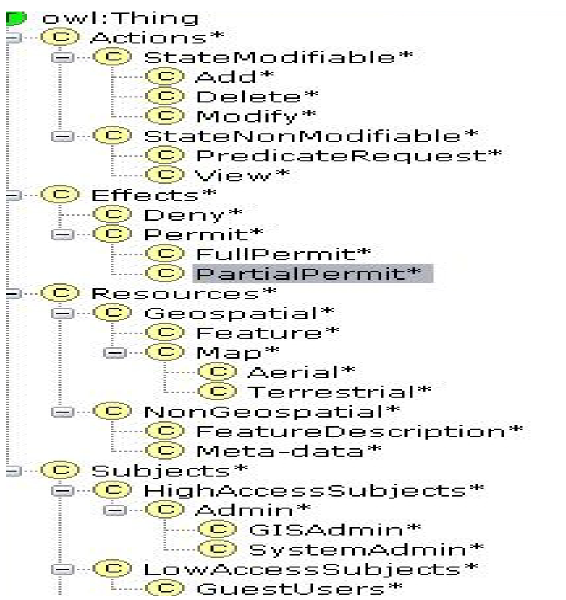
Class	Predicted						
	Water	Barren Land	Grass	Tree	Building	Road	House
Water	1898	0	0	0	0	0	0
Barren Land	0	1225	216	0	143	33	0
Grass	0	15	1175	54	69	0	18
Tree	0	0	0	1454	0	0	25
Building	0	1	0	0	578	189	0
Road	0	0	0	0	143	500	5
House	0	0	0	0	9	59	1296
Accuracy	100.00	75.76	88.28	98.31	75.26	77.16	95.01

To generate high level concepts for a group of neighboring pixels, we will exploit ontologies. An ontology is a collection of concepts and their inter-relationships that collectively provide an abstract view of an application domain. We will develop domain-dependent ontologies as they provide for the specification of fine grained concepts while generic ontologies provide concepts in coarser grain.

5. SECURE GEOSPATIAL SEMANTIC WEB

While organizational resources can be protected with a semantic access control system, geospatial data protection in a distributed environment can present many challenges beyond providing or denying access. Geospatial data is unique in that the same piece of data has varying level of granularity depending on the context. For instance, raster images could be processed in different resolutions, scale and accuracy. Even vector data is available at differing scales depending on the particular data collection agency. For this reason, when data from multiple agencies are integrated, access control of the aggregated geospatial data becomes a potential security trap. For instance, if a multi-layered view is presented to users with each layer belonging to a different source provider, it is not clear how a user from one particular agency will see the aggregated view. The special traits make it infeasible for generic semantic access control such as the one we have developed for DAGIS to govern over geospatial data.

In our research, we have defined two types of constructs so far. First type provides alternative abstract elements for vector data and the second type constitute ontology for subject and action roles. Subjects are classified based on their functional criteria. Currently defined top level classes for various categories of subjects are ‘Administrator’, ‘GISAdmin’, ‘SystemAdmin’, ‘Manager’, ‘Regular Professional’, ‘Facility Personnel’, and ‘Guest’. The actions defined so far are ‘Read’, ‘Write’, ‘Save’, and ‘Execute’. In an ideal circumstance, all parties in a distributed system have an agreed-on set of measures to combine their policies or resolve them in case of a failure. However, the pre-arrangement is not always possible and in such cases, our security constructs would allow a semantic access control processor to interpret the role and action definitions and combine the corresponding policies. Our research is focusing on policy integration algorithms. We have developed a technology for developing partial responses at the same time ensuring security. We have defined our ontology using SWOOP [1] ontology editor. A snapshot of the ontology hierarchy is given below:



6. SUMMARY AND DIRECTIONS

In this paper we have discussed our initial research on developing a secure geospatial semantic web. Our major contributions so far include the following:

- Development of Geospatial RDF for specifying geospatial semantics and ontologies
- Geospatial data mining for classification of remote sensing data
- Policy integrator for geospatial data interoperability

Our goal is to apply the technology for security applications including crime analysis and border control. While we continue with our collaborative research on building a secure geospatial semantic web, we will also enhance the DAGIS system into a fully functional prototype that will answer complex queries and help in decision making. In particular, our prototype will include the following components:

- Semantic Search Browser for Police Blotters
- Tools for Generating Crime Analysis Concepts from Blotters
- Map based Visualizing Tools and Semantic Dashboard