Handbook of Geographic Information Science,
Eds: J. P. Wilson and A. S. Fotheringham, Blackwell Publishing (in print 2004)

Geospatial Ontology Development and Semantic Analytics

I. Budak Arpinar, Amit Sheth & Cartic Ramakrishnan

Large Scale Distributed Information Systems (LSDIS) Lab
Computer Science Department, University of Georgia
Athens, GA 30602-7404
{budak,amit,cartic@cs.uga.edu}

E. Lynn Usery & Molly Azami

Department of Geography, University of Georgia
Athens, GA 30602
{usery,meazami}@uga.edu

Mei-Po Kwan

Department of Geography, Ohio State University
Columbus, OH 43210-1361
kwan.8@osu.edu

Abstract. Geospatial ontology development and semantic knowledge discovery addresses the need for modeling, analyzing and visualizing multimodal information, and is unique in offering integrated analytics that encompasses spatial, temporal and thematic dimensions of information and knowledge. The comprehensive ability to provide integrated analysis from multiple forms of information and use of explicit knowledge make this approach unique. This also involves specification of spatiotemporal thematic ontologies and populating such ontologies with high quality knowledge. Such ontologies form the basis for defining the meaning of important relations and terms, such as near or surrounded-by, and enable computation of spatiotemporal thematic proximity measures we define. SWETO (Semantic Web Technology Evaluation Ontology) and its geospatial extension SWETO-GS are examples of these ontologies. Two enabler for what we term geospatial analytics (GSA) are (a) the ability to automatically and semi-automatically extract metadata from syntactically (including unstructured, semi-structured and structured data) and semantically heterogeneous and multimodal data from diverse sources, and (b) analytical processing that exploits these ontologies and associated knowledge bases, with integral support for what we term spatiotemporal thematic proximity (STTP) reasoning and interactive visualization capabilities. This chapter covers results of our geospatial ontology development efforts as well as some new semantic analytics methods on this ontology such as STTP.

Keywords: Spatiotemporal ontology, thematic ontology, geospatial analytics, geospatial semantics, semantic proximity, spatiotemporal thematic (STTP) proximity, visual analytics.

1. Introduction

Rapid access to and intelligent interpretation of many types of geospatial information require successful information integration and sharing across disparate systems and designs. This is an important subject in current research on meta-data and semantics, which aims to enable schema and instance mapping that, is used to provide a seamless view of all information. New challenges in this area, however, require us to go beyond a thematic-only approach to include the dimensions of space and time. The Geospatial Semantic Analytics (GSA) provides a necessary infrastructure for comprehensive and reliable information analysis, as well as through the use of ontologies and other sophisticated semantics such as implicit complex relationships based on multimodal geographic information. Specifically, GSA provides a framework for developing novel semantic technologies that exploit thematic as well as spatial and temporal information from various domains of knowledge.

Current efforts to integrate geographic information embrace the idea of meta-data standards as the key to information sharing and analysis. These include the Federal Geographic Data Committee (FGDC) and the National Spatial Data Infrastructure (NSDI), GeoSpatial One-Stop, and the U.S. Geological Survey's *The National Map* as well as standards from the International Standards Organization (ISO) for geospatial meta-data. The NSDI attempts to bring together geographical information sources from all levels of government and other organizations into a single point of entry for easier access to data. However, these current efforts lack consideration of the role of space and time as they relate to entities and their relationships across domains of knowledge. Furthermore, traditional analysis of semantic information often uses a purely quantitative approach to represent and infer thematic relations among the entities of interest. This approach has serious shortcomings when dealing with qualitative thematic, spatial or temporal information, which is often incomplete or imprecise. In GSA (Geospatial Semantic Analytics) framework, we develop a new approach that incorporates both the spatial and temporal dimensions, as well as the capabilities to handle qualitative information.

The GSA approach emphasizes the use of semantics to integrate, share, and analyze multimodal geospatial information. Through the use of ontologies and their inherent relationships, GSA enables timely access to unique and powerful knowledge for relevant users and experts. Specifically, sophisticated semantic analysis includes complex relationship discovery that account for the spatial-temporal dimensions, and this enables meaningful interpretation of multimodal information across different domains as they relate geospatially. The multimodal approach will exploit all forms of text-oriented information (unstructured, semi-structured and structured) as well as other digital media (images and maps). Other GSA contributions are the extension of OWL DL as well as notions such as semantic proximity and similarity to include spatial and temporal components. For OWL DL, specific extensions include specifications for space and time concepts for use in ontologies and other knowledge modeling. The notions of semantic proximity and similarity are also expanded to go beyond thematic-only analysis. In particular, GSA can enable discovering complex relations among entities through an integrated spatial, temporal, and thematic reasoning. This approach draws benefit from earlier work on modular reasoning systems [Cutter et al 2003].

In this chapter we exemplify our results in national security domain [NSGIC 2003 & FGDC 2003]. In this domain our approach will benefit every step in the emergency response life cycle, namely preparedness, threat detection, response, recovery and mitigation. In particular, we explore two scenarios: one for detection and another for analysis related to the mitigation component of the emergency response cycle. For detection, we consider analysis of heterogeneous information to detect or discover information such as "what evidence do we have on collaboration of two groups". This is then mapped to the detection of evidence of meetings between members of the two group where the meeting itself is inferred from spatial and temporal collocation of individuals belonging to the groups under consideration, rather than concrete evidence (such as a photograph or sensor data). For the mitigation related analysis, we consider investigations such as "show emergency response activities of each governmental and nongovernmental organizations in the borough of Manhattan immediately after the 9/11 attack." In both cases we assume heterogeneous traditional and geospatial information that cover open source cyber information, internal and/or confidential reports, metadata such as that offered by *The National Map*, and geospatial information held by a UCGIS institution or NASA.

Analytics of the type indicated by these scenarios pose three novel challenges:

(a) the ability to deal with thematic, spatial and temporal information, as well as interactions among these three dimensions,
(b) the ability to capture imprecise relations among different organizations, their members, and their movements, and
(c) the ability to support new analytical techniques and tools for interacting with the GSA system.

These challenges translate to our research which includes:

(a) the adoption of a formal ontology and metadata representation framework that is consistent with the emerging semantic Web ontology representation standard of OWL,
(b) the definition of proximity measures accommodating the three dimensions,
(c) the development of an analytical computation framework incorporating collaborating reasoners that support thematic, spatial and temporal reasoning,
(d) visualization and other tools for developing applications and for helping an analyst to exploit the GSA system, and
(e) demonstrating GSA by exploiting a broad variety of information resources available from the openWeb, *The National Map*, NASA and members of UCGIS.

The rest of the chapter is organized as follows: Section 2 provides background and related work. Section 3 presents geo-spatial ontology development as well as knowledge discovery techniques. Section 4 summarizes some current and future data sources. Finally, Section 5 concludes the chapter.

2. Background & Related Work

2.1 Semantic (Thematic) Analytics

Compared to earlier keyword-based and information retrieval techniques that rely on syntax, there is an increasing role of semantic approaches to information management where meaning is associated with data and terms used in queries [Shah & Sheth 1998]. Some of the most notable aspects of semantic approaches are the development of ontologies and semantic annotation of data. Ontologies [Gruber 1993, Guarino 1998], specifically domain-specific ontologies, are at the heart of most semantic approaches [De Bruijn 2003]. A large number of ontologies have been developed in domains such as biology, and to a lesser extent in geography [Harding 2003, Bennet & Grenon 2003]. Examples of scalable technology for semantic annotation include Semagix's Freedom [Hammond et al 2002] that can perform deep annotation (with a good degree of disambiguation using expressive domain ontologies, often populated with instances [Sheth and Ramakrishnan 2003]) and IBM's Web Fountain that has demonstrated more scalable but shallower annotation (involving broad ontology with limited types of relationships and disambiguation) of over 2.5 billion Web pages [Dill et al 2003].

These advances in semantic technologies in general, and the semantic Web in particular, are bringing a new class of applications to reality. These applications, briefly reviewed in [Sheth & Ramakrishnan 2003] include: (a) semantic search and browsing [Heflin & Hendler 2000, Townley 2000, Guha et al 2003], (b) semantic integration [Kashyap 1999], and (c) semantic analytics and discovery [Aleman-Meza et al 2003, Sheth et al 2004]. Of particular interest to GSA are applications such as the *Passenger Threat Assessment* application for national/homeland security [Sheth et al 2004], *Anti-money Laundering* solution [Semagix-CIRAS], and business intelligence applications being investigated by IBM's Web Fountain [IBM-WF]. Additionally, SemDIS team is developing a semantic analysis framework to support discovery of semantic associations (complex relationships) from large amount of semantic annotations (i.e., semantic metadata), but only deals with aspects of the thematic dimensions of GSA approach. It utilizes commercial technology from Semagix, which is based on a technology licensed from the Large Scale Distributed Information Systems (LSDIS) lab at the University of Georgia (UGA) for ontology population and semantic metadata extraction from heterogeneous documents [Sheth et al 2002a]. Key issues addressed in SemDIS include:

(a) definition of semantic associations [Anyanwu & Sheth 2003],
(b) design of a sample homeland security (HS) ontology [Sheth et al 2004],
(c) development of a broader test-bed with a populated ontology (not specific to HS) with large number of instances (approx. 1 million in Jan 2004) and APIs that are being made available openly for non-commercial use in comparing Semantic Web tools and developing benchmarks,
(d) investigations into issues of computing semantic associations over very large metadata sets represented as RDF graphs, and in issues of ranking complex relationships (a search engine returns a ranked set of documents; similarly computing semantic associations would return a set of relationships between objects that would need to be ranked) [Aleman-Meza et al 2003].

The project Web site (http://lsdis.cs.uga.edu/Projects/SemDis/) provides further details.

Semantic technology also has a crucial role to play in the integration of geospatial information sources ([Sheth 1999, Goodchild et al. 2001]). [Fonseca et al. 2002] identifies research questions that pertain to the creation and maintenance of geospatial ontologies and integration of both geospatial sources and ontologies. Recently the Semantic Geospatial Web [Egenhofer 2002] has been recognized as a key UCGIS theme [Fonseca & Sheth 2002]. Use cases for the Semantic Geospatial Web proposed to date involve information retrieval like queries. We however propose to provide support for queries that support *What-if* analysis, hypothesis validation and semantic association discovery. We envision that these sorts of queries will require that the spatial and temporal dimensions are factored in. One aspect of the our work is the incorporation of native support for spatial and temporal reasoning that was lacking in our previous work on thematic dimensions in IScape [Sheth et al 2002b] and Semantic Associations [Anyanwu & Sheth 2002].

2.2 Geospatial Analytics

Special Properties of Geospatial Phenomena

Much of contemporary geospatial analytics is based on the notion of proximity, where space and time provide the necessary link to other potentially interesting factors and to the context that influence the phenomena in question. It is now widely recognized that geographic attributes of phenomena often exhibit the properties of spatial dependency and spatial heterogeneity. Spatial dependency (or spatial autocorrelation) is the tendency for observations that are near each other in space to have similar values, where spatial proximity (or location-based similarity) is matched by value similarity [Anselin 1999]. Spatial heterogeneity refers to the non-stationary nature of most geographic processes, where global parameters do not reflect well the process occurring at a particular locality.

These two properties of spatial phenomena have gained much attention in geospatial analytics in the past two decades or so. While they have traditionally been treated as nuisances in spatial analysis, contemporary research has developed tools that utilize them to gain new insights into geographic phenomena (e.g. local indicators of spatial association developed by [Anselin 1995] and geographically weighted regression formulated by [Fotheringham et al. 2000]). Methods for local analysis attempt to incorporate considerations of geospatial context into the analysis. Besides methods for local analysis, there are other techniques that are also useful for examining spatiotemporal associations and patterns - such as hierarchical or multi-level modeling and space-time clustering techniques [Bailey and Gatrell 1995].

Recent research in geographic knowledge discovery suggests that ignoring spatial autocorrelation and spatial non-stationary nature may affect the patterns derived from data mining techniques [Chawla et al 2001]. For instance, patterns of geospatial-semantic association may shift when one moves from one locality to another, or when one moves across geographical scales (from the metropolitan level to the neighborhood level). Attention to these spatial properties may lead to the formulation of spatially explicit theory or models. A model is said to be spatially explicit when it differentiates behaviors and predictions according to geographical locations, while a spatially explicit theory is a theory whose outcomes depend on the locations of the objects that are the focus of the theory. It follows that one or more spatial concepts, such as distance, location, connectivity, adjacency, or direction, must appear in the theory.

Geospatial Semantics

Conventional geospatial analytics are largely based on metric measurements (i.e. quantitative). But people often express and understand spatial relations through natural language instead of metric measurements. So it is important to establish a geospatial semantics as part of GSA for performing spatial queries using imprecise spatial and temporal references (e.g. *near*, *far*, *around noon*) for analyzing geospatial-semantic associations using textual and other non-metric information. This can also help for effective geographical knowledge discovery that enables quick response to otherwise non-actionable information (especially in light of the fact that information collected by federal intelligence agencies is often vague and imprecise).

Attention to several elements is important when developing geospatial semantics that support effective spatial reasoning. These include the use of qualitative modifiers (e.g. *very*, *a little*, *almost*), proxy place names (e.g. *little Italy*, *short north*), spatial references (e.g. *east side of the city*, *west of the river*), and spatial relation describers (e.g. *near*, *far*). These will be important for extending domain semantics to temporal and geospatial concepts and terminology, and for developing algorithms for computing geospatial proximity and associations. We incorporate geospatial semantics for three major types of geospatial relations into the GSA:

(a) ***Topological relations:*** Topological relations refer to properties like connectivity, adjacency and intersection among geospatial objects. Current topological models are not adequate for handling the vagueness and imprecision in topological relations expressed in natural language (e.g. *intersect with, cross, come through, split, bypass, next to*). They need to be extended through approaches like development of a richer vocabulary of spatial predicates and/or a fuzzy logic approach that better deals with "vagueness" in topological relations (*inside, outside, surrounded, intersect with*.

(b) ***Cardinal direction:*** In daily life, people refer to geographical locations using qualitative describers based on their spatial perception. For example, people use directional describers such as *East*, *West*, *North East* and *South West* to denote relative directions among geographical objects. But this kind of directional reference is imprecise and makes it difficult to identify the exact boundary of geospatial objects. There are different spatial models for handling directional terms. A simple one is to divide an area into eight directions based on an angular division of an area into eight equal sectors (N, NE, E, SE, S, SW, W, NW).

(c) ***Proximity relations:*** Traditionally, geospatial proximity relations refer to the geographical distances among geospatial objects (e.g. *A is close to B*, *X is very far from Y*). Various methods have been developed for modeling proximity relations in the geospatial domain. For instance, [Gahegan 1995] proposes a fuzzy logic model for proximity reasoning, in which each proximity expression such as *close* or *far* has a corresponding fuzzy membership function. Base on this model, the query "Which nuclear power plants are close to R" takes the form of:

Close:O = CloseTo{o:O,{o}, R, {x1, y1, x2, y2}, DistanceMethod, C}

where O is the object type (nuclear power plant), *CloseTo* is a fuzzy set membership function, and *DistanceMethod* is a distance calibration method (e.g. absolute or relative distance). Object o is any object of type O in the study area. R is the reference location, and {x1, y1, x2, y2} defines the size of the area (which is used to represent geographical scale). In addition, geospatial proximity is a contextual relation. Thus the context C is included in the definition and involves factors such as transportation mode. Another example is a context where an obstacle separates two objects. In this particular context, objects can be considered far which are considered near otherwise. Hence, the result of the query above is a set of objects that is denoted as close to R and is of type O.

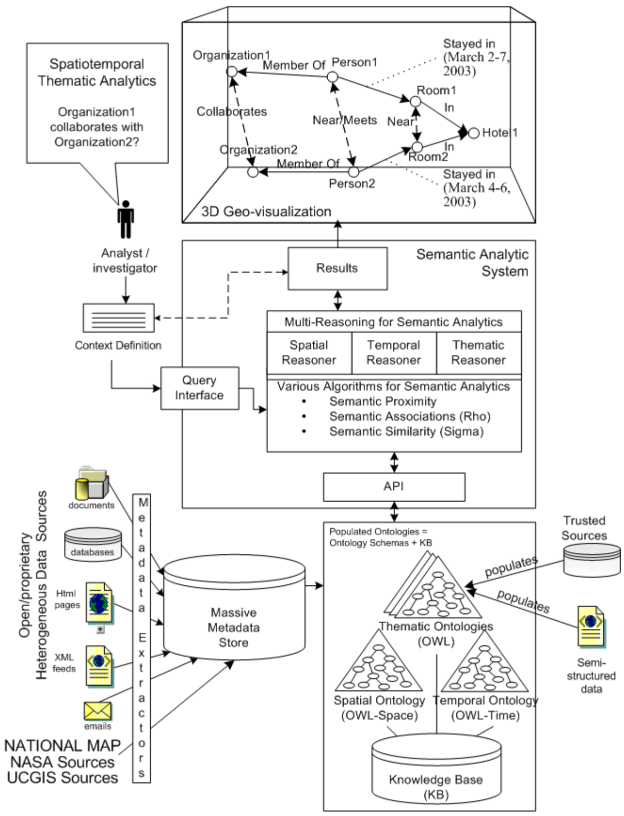

Figure 1: An Overview of GSA System Architecture

3. Semantic Analytics

GSA consists of various components as illustrated in Figure 1 and addresses following research items:

- Development of ontologies that covers the three dimensions of theme, space and time,

- Extraction of metadata from a variety of heterogeneous content/data sources based on the relevant ontologies,
- Proximity definition and computation as the primary component of GSA,
- Tools to support spatiotemporal thematic analytics,
- 3D geo-visualization techniques.

These research items are detailed in the following sub-sections.

3.1 Spatiotemporal Thematic Ontology Development

GSA uses three types of ontologies that capture the thematic, spatial and temporal dimensions in order to support queries and analysis. For the thematic dimension we use a large-scale general-purpose ontology based on the development of the Semantic Web Technology Evaluation Ontology (SWETO, http://lsdis.cs.uga.edu/proj/Sweto) of the SemDIS project.

Development of SWETO

SWETO captures real world knowledge with tens of classes and relationships populated with a growing set of relevant facts. Here "ontology" refers to populated ontology which consists of the schema (description) as well as its knowledgebase (i.e. populated ontology = ontology schema + knowledgebase) (see Figure 1). As part of our research, we have maintained an iterative process that allows the periodic extension of the schema and knowledgebase that is consistent with the concept of emergent semantics [Staab 2002, Kashyap & Behrens 2001]. This largely automated process, adapted from SWETO, includes:

(i) Designing the schema using an ontology design toolkit,
(ii) Identifying knowledge sources (usually public and open sites and databases from governmental, educational and non-governmental organizations) that can be used to populate parts of SWETO without focusing on a specific domain, thus allowing a general purpose evaluation metric (knowledge sources have semi-structured forms such as template based HTML or XML, or structured forms, such as spreadsheets, databases and database driven Web sites),
(iii) Utilizing knowledge extractor agents to periodically and automatically extract parts of knowledge,
(iv) Applying automatic, semi-automatic and manual disambiguating techniques [Mihalcea & Mihalcea 2001, Resnik 1999, Kashyap & Sheth 1996b, Rodriguez & Egenhofer 2003] to extracted concepts when populating the ontology, and
(v) Providing capabilities for exporting the ontology in W3C recommended standards of either OWL [Bechhofer et al 2003] or RDF [Lassila & Swick 1999].

Creation of SWETO requires meticulous selection of data sources. We focused our selection of data sources by considering the following factors:

(i) Selecting sources which were highly reliable Web sites that provide entities in a semi-structured format, unstructured data with parse-able structures (e.g., html pages with tables), or dynamic Web sites with database back-ends. In addition, the Freedom toolkit has useful capabilities for focused crawling by exploiting the structure of Web pages and directories.
(ii) We carefully considered the types and quantity of relationships available in a data source. Therefore we preferred sources in which instances were interconnected.

(iii) We considered sources whose entities would have rich metadata. For example, for a 'Person' entity, the data source also provides attributes such as gender, address, place of birth, etc.
(iv) Public and open sources were preferred, such as government Web sites, academic sources, etc. because of our desire to make SWETO openly available.

To illustrate the ontology-building process, consider the listing of "people" in a computer science department. Typically, they would be listed separately as Faculty, Students and Staff. In such cases we create appropriate classes in the ontology and populate them with instances. In SWETO, the ontology was created using Semagix Freedom, a commercial product which evolved from the LSDIS Lab's past research in semantic interoperability and the SCORE technology [Sheth et al 2002a]. The Freedom toolkit allows for the creation of an ontology, in which a user can define classes and the relationships that it is involved in. Thus, the user is relieved of the burden of serializing the ontology to the OWL syntax. To keep the ontology up to date, the extractors can be scheduled to rerun at user specified time/date intervals (see Figure 1).

As the Web pages are 'scraped' and analyzed (e.g., for name spotting) by the Freedom extractors, the extracted entities are stored in the appropriate classes in the ontology. Additionally, provenance information, including source, time and date of extraction, etc., is maintained for all extracted data. We later utilize Freedom's API for exporting both the ontology and its instances in either RDF or OWL syntax. For keeping the knowledge base up to date, the extractors can be scheduled to rerun at user specified time and date intervals.

Automatic data extraction and insertion into a knowledge base also raise issues related to the highly researched area of entity disambiguation. In SWETO, we have focused greatly on this aspect of ontology population. Using Freedom, entity instances can be disambiguated using syntactic matches and similarities (aliases), customizable ranking rules, and relationship similarities among entities. Freedom is thus able to automatically disambiguate entities as they are extracted. Furthermore, if Freedom detects ambiguity among new entities and those within the knowledge base, yet it is unable to disambiguate them within a preset degree of certainty, the entities are flagged for manual disambiguation with some system help on possible matches. Lastly, there a special cases in which neither the software, nor humans can directly determine if two entities are the same. For example, consider two persons named 'John Smith'. Without metadata attributes, neither the system nor humans can determine what to do by only looking at the entity name. This is a future research direction we wish to follow in which semantic similarity will be used to state with some degree of certainty that these two persons (i.e. 'John Smith'), are in fact the same person. For now, we remove these types of entities from the knowledge base in order to maintain both cleanliness and consistency.

Our aim of achieving a test-bed of over 1 million instances is near completion. The current population includes over 800,000 entities and over 1,500,000 explicit relationships among them. Here we provide initial statistics that illustrate the size in terms of entities and relationships connecting them. Table 1 summarizes a subset of the classes of the ontology that are representative of the majority of instances currently in SWETO ontology.

Table 1: SWETO test-bed ontology initial metrics

Subset of classes in the ontology	# Instances
Cities, countries, and states	2,902
Airports	1,515
Companies, and banks	30,948
Terrorist attacks, and organizations	1,511
Persons and researchers	307,417
Scientific publications	463,270
Journals, conferences, and books	4,256

TOTAL (as of January 2004)	811,819

What makes this work more valuable is in respect to how inter-connected the instances are (this currently is not available in a taxonomy and in most current ontologies that are freely available). As mentioned earlier interconnectedness becomes critical in semantics analytics applications. Table 2 summarizes a subset of the relationships connecting instances in the ontology. Note that some relationships apply to a variety of instances, such as the "located in" relation.

Table 2: SWETO statistics on relationships

Subset of relationships	# Explicit relations
located in	30,809
responsible for (event)	1,425
Listed author in	1,045,719
(paper) published in	467,367

In addition, we provide a graphical user interface for browsing of SWETO ontology (through the use of Touchgraph), the latest version of the knowledge base (instances), our own native API for easy use (alternately tools such as Jena could be used), and a detailed description of the data sources in SWETO Web site.

Spatiotemporal Ontology Development

An extension of SWETO has focused on an aspect of terrorism [Sheth et al 2004]. In GSA, we use of multiple thematic ontologies such as a broader ontology focusing on people, places and organization, or a more focused ontology on transportation and travel. Building upon existing work on representation of geospatial and temporal information, we use an approach similar to the above, to represent and build the spatial and temporal ontologies and populate the corresponding knowledge bases. This starts with the means of representing the modeling primitives for the spatiotemporal domain. Following this we need to define a formal syntax and semantics for these primitives by building on and extending the earlier work. For example, [Randell et al. 1992] proposed an interval logic that can be used to reason about space. In doing so they develop a subsumptive hierarchy of dyadic relations that are all based on the primitive relation *C(x,y)*, read as '*x connects with y*'. Similar hierarchies can be developed that represent geographical regions like continents, countries, regions, states and cities and further extend the subsumptive hierarchy in [Randell et al. 1992] thereby creating the building block of a geospatial ontology. This ontology provides primitives to represent geospatial concepts and the spatial relations between these concepts. Another important characteristic of ontology framework for GSA will be the inter-ontological mappings and their population. A good example of this is the mapping of place names in thematic ontology (e.g., Athens, GA) to geographical location in spatial ontology (e.g., 33:57:39N, 083:22:41W). Although schema level inter-ontology mapping and corresponding multi-ontology query processing techniques have been investigated [Mena et al 1996, Mena et al 2000, Sheth et al 2003bc], we believe GSA may be the first attempt to utilize relevant knowledge (instances associated with the ontologies). Later in our work on proximity computation, it would be possible to substitute spatial proximity with corresponding place name occurrence in metadata to map or associate computation in one dimension to another.

An ontology of temporal relations and concepts serves as the third dimension along which our analytical queries will run. Consider the scenario where an intelligence analyst wants to find evidence that supports the hypothesis that two known organizations collaborate. One reason to believe this is that representatives from both organizations met at certain location. This will require knowledge about particular persons, the organizations they belong to and the nature of these organizations (thematic), the location (spatial) and time (temporal) of the meeting. Imagine that the location is a hotel in whose rooms the two individuals had stayed. Also imagine that their respective durations of stay overlapped. Assume that the notion of rooms in a hotel - which is located in a city and the city is in turn located in a country - is represented in the knowledgebase of the thematic ontology. This knowledgebase may look something like this:

Hotel(grand).
Person(x).
Person(y).
LOCATED_IN(grand, cityx).
HASROOM(x, grand).
HASROOM(y, grand).

This gives the information necessary to investigate further and determine if the individuals were indeed in the same hotel at the same time. This notion of "at the same time" can be represented using the primitives described in [Artale 2000]. Treating the duration of stay of both individuals *x* and *y* as intervals *i* and *j* the primitive *overlaps(i,j)* can be used to indicate that there was at least some point in time when both *x* and *y* were at the same hotel and therefore could have met.

[Allen 1991] studied the various relationships between intervals. A good illustrative example of the practical use of this formalism is shown in [Aratale 2000]. The concept Mortal can be represented in terms of another concept LivingBeing as *Mortal = LivingBeing and (sometime (x) (met-by x now) . (at x (not LivingBeing))).* This reads as follows – A Mortal is a LivingBeing such that there is a time instant x at which the LivingBeing ceases to be a LivingBeing. In addition, there have been previous attempts to develop ontologies for both spatial and temporal concepts and relationships. In the temporal dimension, with respect to DAML-TIME led by Jerry Hobbs, effort was made to develop a representative ontology of time that expresses temporal concepts and properties common to any formalization of time. The ontology was formulated as a set of first-order predicate calculus axioms. Another parallel effort for the temporal dimension that was led by Ferguson and Allen, titled "Practical Temporal Ontology and Markup Language" involved collaboration with the DAML-TIME group. This effort captured temporal notions of the following kind: dates of events ("January 3, 2002," "next Wednesday"), durations of activities ("drive for twenty minutes," "wait about an hour"), ordering between events ("wait for Fred then drive to Rochester"), and constraints between events ("don't touch the button while the switch is on," "only one flight can use the runway at a time"), etc. Other examples of work on the spatial ontologies include [Alani 2001] and [USGS-ONT]. Work in DAML can easily be borrowed to define ontologies in OWL-DL as we intend to follow W3C's recommendation for ontology representation language.

Additionally, OpenCyc is a candidate to build a space ontology which includes comprehensive spatial concepts and relations. Related efforts are the OWL-Space initiative (formerly DAML-Space). The Context Broker Architecture project at UMBC[1]) has incorporated four basic space ontologies that define part-whole relations, and other properties for physical objects and describe spatial entities and relations based both on the OpenCyc spatial ontology, and the Region Connection Calculus (RCC). Other work we will build upon is the RDF Geo vocabulary[2].

3.2 Spatiotemporal Thematic Metadata Extraction

Virtually any unstructured, semi-structured and structured content can be analyzed for semantic metadata extraction or semantic annotation. For this, we use metadata extractions of Semagix Freedom as explained earlier. While extracting from semi-structured and structured sources is in part similar, extracting unstructured data involves lexical analysis and automatic classification for 'scraping' and analyzing text (e.g. for name spotting). This process also benefits from a knowledgebase associated with the ontology and is described in detail in [Sheth et al 2002a, Hammond et al 2002]. Extracted entities are associated with the ontology and stored in a meta-base. Additionally, provenance information, including source, time/date of extraction, etc., is maintained for all extracted

[1] http://cobra.umbc.edu/
[2] http://esw.w3.org/topic/GeoInfo

metadata. We later utilize Freedom's API for exporting both the knowledgebase and metadata in either RDF or OWL syntax.

For GSA, *The National Map* also provides data and operational capabilities that include the following:

- High-resolution digital ortho-rectified imagery. Imagery will provide some of the feature information content now symbolized on topographic maps.
- High-resolution surface elevation data, including bathymetry.
- Vector feature data for hydrography, transportation (such as roads, railways, and waterways), structures, and boundaries of government units and publicly owned lands.
- Geographic names, such as those for physical and cultural features.
- Land-cover data classifying the land surface.

GSA content extractors extract data for geographic areas such as counties or watersheds to generate geospatial metadata. The frequency of extraction is customizable using the extraction execution framework of Freedom. At a minimum, metadata meets Federal Geographic Data Committee (FGDC) standards to document the content and characteristics of *The National Map* data, such as lineage, positional and attribute accuracy, completeness, and consistency. Metadata is maintained at the feature level when applicable. To facilitate the tracking of changes over time, the USGS will ensure that *The National Map* content is permanently archived by retaining versions of data sets or feature-based transactional information. The version information is captured in the metadata as needed, along with other provenance information. We also combine metadata extraction from *The National Map* with that extracted from content of other organizations, such as cadastral information from the Bureau of Land Management and demographic data from the Bureau of the Census.

Development of SWETO-GS

SWETO-GS is the geospatial extension of SWETO. It expands the SWETO test-bed making it richer in terms of geo-spatial entities and relations as well as meets some of the research priorities of the University Consortium for Geographic Information Science (UCGIS). UCGIS adopted the Geospatial Semantic Web and creation and management of geo-ontologies as short-term research priorities. The foundations for design, development, storing, maintenance, visualization and querying a geo-ontology are already established in the framework of SWETO. SWETO-GS adds the geographic content on which these techniques can be applied. Development of the schema for SWETO-GS is a bottom-up procedure. A main source of information is first identified then based on the structure of the information and the implicit relationships to existing entities, new classes, entities, attributes and relationships are defined.

The Geographic Names Information System (GNIS) of the USGS is the first source from which geographic data to be extracted. This database contains all the named features collected from every topographic map covering the United States and its protectorates. The data is available as comma delimited text files with each line containing 18 fields. These fields are:

1	State Alpha Code	7	Primary Latitude (DMS)	13	Source Latitude (decimal degrees)
2	Feature Name	8	Primary Longitude (DMS)	14	Source Longitude (decimal degrees)
3	Feature Type	9	Primary Latitude (decimal degrees)	15	Elevation
4	County Name	10	Primary Longitude (decimal degrees)	16	Estimated Population
5	State Number Code (FIPS Code)	11	Source Latitude (DMS)	17	Federal Status

6	County Number Code (FIPS Code)	12	Source Longitude (DMS)	18	Cell Name

The entities being extracted from this data set are State, County, Feature and Cell (or 7.5 minute topographic map - Topomap). The attributes and relationships for each entity look like this:

State – FIPS_code, located_in US, and covered_by Topomap
County – FIPS_code, located_in State, located_in US, and covered_by Topomap
Feature – Geographic_Description cp= Primary Latitude, Primary Longitude (*decimal degrees*), located_in county, located_in State, located_in US, Feature_Type, Elevation, Estimated_Population (*PPL feature type only*), and covered_by Topomap

The State and County FIPS_code attribute and the Featue_Type carry a cardinality of 1. All the others have 0 to *n* cardinality. "Geographic_Description cp= " is the holder for the latitude/longitude point data for any feature where "cp" means center point. In the schema Bounding_Box and Boundary are two other attributes that are defined (following naming conventions of the ISO Geographic Metadata Standards). While not used in the extraction of the data from the GNIS, they are key properties for future geospatial analysis. Figure 2 shows a small part of the ontology.

The decimal minute latitude/longitude is chosen for the ease of conversion and use in algorithms. Within digital boundary files, such as those available from the US Census, state and county FIPS codes are used as indexes for the data sets. One of the key entities being extracted from this first source is the cell name or Topomap quadrangle. It is a crucial link for expansion of the geospatial part of this ontology. The naming conventions of the 7.5 minute Topomaps are used by the USGS digital raster graphic (DRG) files which are, in turn, linked, through the file's meta-data, to other image and digital files available from the USGS.

The approximately 800,000 features that will be extracted from the GNIS database are all point data. While some actually cover large surface areas or are linear in nature, the information identifying their location is a single set of latitude and longitude points. Geospatial information comes also in linear and polygon forms. The next step in the further population of SWETO will be capturing polygon data of cities and counties.

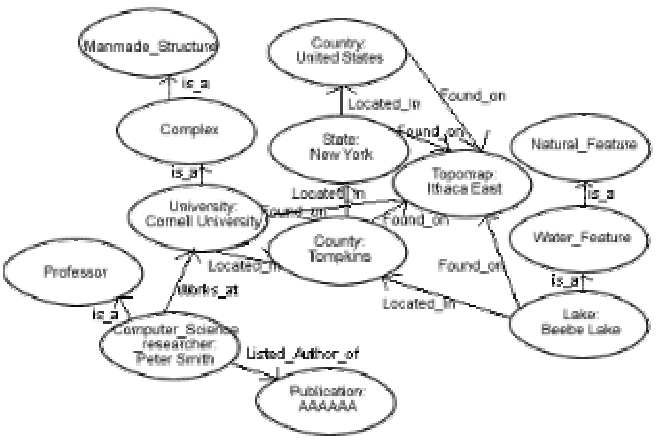

Figure 2: SWETO-GS Ontology (a fragment)

One of the issues of deciding which boundary files to use is that of coordinate systems used to create the file. The US Census Cartographic Boundary and GNIS files use latitude and longitude, and the values are in the decimal degree format. Using these boundary files will aide in consistency and ease of data management within the ontology.

Capturing linear features entirely across borders and at different scales is another challenge for the SWETO-GS. A long term issue is to develop methods of conducting geospatial analysis using the data and relationships within the ontology. With enough entity instances and relationships defined, conducting various analyses, such as point-in-polygon, or distance algorithms, would be possible.

3.3 Spatiotemporal Thematic Proximity (STTP)

GSA provides for reconciliation of the thematic, temporal, and spatial perspectives and uses it as a step towards knowledge discovery and correlation across multiple points in a multidimensional GSA space. As illustrated in Figure 3, the points in the GSA space correspond to the movements of entities in space and time. This multimodal space also captures how entities thematically relate to each other in space and time and continuity of those relationships (i.e. when and where they exist or do not exist). One important criterion in understanding the relationships among entities is their proximity to each other. Spatial proximity captures the physical distance among the entities and helps in drawing conclusions about their nearness to each other. On the other hand, nearness can not be verified without the existence of temporal proximity of entities of interest. Finally thematic proximity, i.e., a notion capturing how entities closely or similarly relate to each other, can benefit from these spatial and temporal proximity definitions.

For example in Figure 3, the points in GSA space correspond to the movements of entities as understood or interpreted in thematic context in space and time. There are three spatial objects on the left side of the diagram, namely two rooms in the same hotel. On the right side, the spatial objects are two buildings separated by an obstacle such as a river or a border. The movements of entities (i.e. two persons) are depicted with lines progressing in the time dimension. In this figure space is assumed to be two-dimensional. However the approach can be generalized for three dimensions easily. Interpretation of objects is obtained by associating metadata objects with corresponding dimensions. That is, the thematic dimension is captured by associating the objects to the thematic ontology. For example, a theme represents objects participating in a semantic relation (e.g. Person1 meets Person2). These relations correspond to RDF(S) and OWL statements and can be represented through a graph model.

In Figure 3, spatial proximity captures if Person1 and Person2 come near to each other (e.g. if they stay in the same hotel). Temporal proximity captures the temporal relationships among their stays (e.g. if their stay overlaps). Finally thematic proximity captures the nature of relationships between these persons (e.g. if they work for some collaborating organizations). Different possibilities for these types of proximities are exemplified in Table 3. Spatiotemporal Thematic Proximity (STTP) is a combination of these measures to draw conclusions about entities involved, as well as connected to other entities in terms of their proximity in each of the three dimensions. In GSA space, a set of entities and relations among them evolve semantically and move to different locations in a certain time period. Many intelligence analysis subjects including terrorist actions can be represented in this way in a multi-dimensional GSA space.

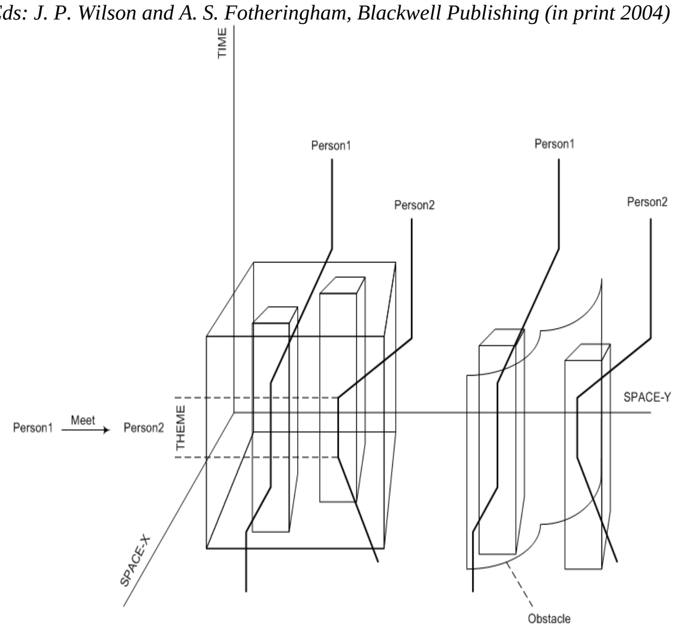

Figure 3: The Multi-dimensional GSA Space

Spatiotemporal Thematic Proximity (STTP), which is a new semantic proximity notion addressing time and space dimensions in addition to the theme dimension, becomes the basis of controlled movement in the GSA space and facilitates knowledge discovery. STTP can be measured in terms of contextual distance between the entities in space, time, and theme.

Between the same two points in the GSA space, we can have different proximity results for different contexts. This distinction based on different contexts is fundamental to a person's proximity judgments and, therefore, natural. For example, two persons may be similar in the context of their financial transactions, but not in the context of their reading habits (i.e. contextual thematic proximity). In this regard, context can be defined in terms of user selected elements or subspaces of available ontologies [Aleman-Meza et al 2003]. However, the context in which two buildings are separated by an obstacle such as a border or river (Figure 3) that prevents their nearness and affects their spatial proximity requires a more advanced context specification mechanism (see last row in Table 3).

We build the STTP concept based on prior definitions of semantic proximity [Kashyap & Sheth 1996b]. Yet, the introduction of space and time, as well as the more comprehensive modeling of the domain from which context is created, make the computation of STTP a very challenging research issue.

Table 3: Spatiotemporal Thematic Proximity and Some of Possible Inferences

Metadata Extracted				
Spatial Proximity	**Temporal Proximity**	**Thematic Proximity**	**Context**	**Possible Inference**
NA	NA	Person1 *meets* Person2 Person1 *is member of* Organization1 Person2 *is member of* Organization 2	NA	Organization1 *collaborates with* Organization 2 (?)
Person 1 *is in* Room1 Person 2 *is in* Room2 Room1 & Room2 *are in* Hotel1	March 2-7, 2003 ⊃ March 4-6, 2003	Person1 *is member of* Organization1 Person2 *is member of* Organization 2	NA	Room1 *is near to* Room2 Person1 *is near to* Person 2 Person 1 *meets* Person2 (?) Organization1 *collaborates with*

				Organization 2 (?)
Person 1 *is in* Room1 Person 2 *is in* Room Room1 & Room2 *are in* Hotel1	March 2-7, 2003 ∩ January 4-6, 2003 = ∅	Person1 *is member of* Organization1 Person2 *is member of* Organization 2	NA	*not* (Person 1 *near* Person2) *not* (Person 1 *meets* Person2) (?) *not* (Organization1 *collaborates with* Organization 2) (?)
Person 1 *is in* Building1 Person 2 *is in* Building2 Building1 *is near* Building2	NA	Person1 *is member of* Organization1 Person2 *is member of* Organization 2	Obstacle *separates* Building 1 *and* Bulding2	not (Person 1 *near* Person2) *not* (Person 1 *meets* Person2) (?) *not* (Organization1 *collaborates with* Organization 2) (?)

Finally, STTP provides for new analytical methods where thematic proximity and analytics are supported by temporal and spatial proximities. In general, inferring a thematic relation among entities can be accomplished by using their STTP proximity. For example, in the detection aspect of national security it is very informative to verify if two organizations are collaborating. If STTP measure allows us to draw conclusions among nearness of entities belonging to these organizations (e.g. their members) in space and time then we can infer thematic proximity (e.g. collaboration) of respective organizations. Some of the possible inferences for this example are outlined in Table 3.

3.4 Spatiotemporal Thematic Analytics

Our approach extends conventional geo-computational framework through an integration of quantitative geo-computation with Spatiotemporal Thematic Analytics (STTA). The term geo-computation refers to an array of activities involving the use of new computational tools and methods to depict geographical variations of phenomena across scales [Longley et al 1998]. It encompasses a wide range of techniques, including expert systems, fuzzy sets, genetic algorithms, cellular automata, neural networks, fractal modeling, visualization, and data mining. Among these methods, our effort mainly focus on extending a particular type of GIS-based geo-computational algorithms recently developed by Kwan and her associates (e.g. [Kwan 1998], [Kim and Kwan 2003]). Originally developed for the evaluation of individual space-time accessibility, these algorithms are based on set-theoretical framework and use measurable attribute of geographical areas (e.g. the number of land parcels). Since these geo-computational algorithms define space-time proximity in terms of a relational notion of nearness (i.e., in terms of the time-dependent relative position of geographical objects), they can be a basis for integrating the analysis of spatiotemporal co-location and thematic proximity within a geo-semantic computational framework. For example, suspicious behavior of individuals or movements of objects may be first identified in the space-time domain. Then inferences from such spatiotemporal co-location analysis may inform analysis of thematic proximity as explained in the previous section.

A major advantage of extending conventional geo-computation framework in this manner is that it allows the STTA to be performed within (or closely coupled with) a massive geospatial database that accounts for the highly complex geospatial properties of real-world environments and phenomena (e.g. the topological structure of a real transportation network). Further, an important task in the development of this framework is the development of a set of multi-scalar spatiotemporal thematic operators. This is necessary because geospatial events and phenomena often cross geographical scales and a STTA that does not take this into account may fail at critical junctures. For example, the consequence of a major terrorist attack may span a large area that includes several metropolitan regions and cities. Effective emergency response needs to be supported by geospatial queries and routing operations that are suitable for a specific scale. For instance, the transportation network can be used to compute the distance between two cities, but a 3D network is needed for computing the distance between a fire station and a room in a building hit by plane ([Kwan 2003], [Kwan and Lee 2004b]). In the latter case, the shortest path operator needs to take into account additional factors such as entry-point uncertainty and possible delays within the traffic conduits inside a building.

3.5 3D Visualization for Analytics

3D geo-visualization is the process of creating and viewing graphical images of data with the aim of increasing human understanding [Hearnshaw & Unwin 1994] (see Figure 4 for an example). It is a powerful means of geospatial knowledge discovery, especially when dealing with large and complex datasets [Kwan 2000]. This is so because conventional inferential techniques and pattern recognition algorithms may become less successful when there are a large number of heterogeneous data items, and when many categories and attributes are involved [Gahegan 2000] - as in most analyses of potential homeland security threats. Scientific visualization methods, however powerful, omit an important element that is addressed only in recent research: the geographical context within which events and movements unfold. Adding geography and creating visualization environments that integrate the spatial, temporal and thematic dimensions is a major innovation under the rubric of geo-visualization (or visualization of geographic information, GViz).

Figure 4: 3D Geo-visualization of Human Movements in Space and Time

Geo-visualization is the use of concrete visual representations and human visual abilities to make spatial contexts and problems visible ([Gahegan 2001], [MacEachren et al 1999]). Through involving the geographical dimension in the visualization process, it greatly facilitates the identification and interpretation of spatial patterns and relationships in complex data in the geographical context of a particular area. Although geo-visualization can be performed using conventional geographical information systems (GIS), introducing the temporal dimension call for 3D GIS capabilities that allows for the visualization of events and movements in space-time. This is achieved through adding the third dimension (z) to the 2D geographical space (x,y) of conventional GIS, and the resulting environment greatly enhances the power of our spatiotemporal thematic analytics, particularly in knowledge discovery pertinent to the complex interactions among space, time and thematic attributes of dynamic events (e.g. movements of vehicles and spatiotemporal co-location of people from particular organizations).

An important recent development in this area is the use of 3D geo-visualization for identifying and extracting critical patterns in massive spatiotemporal databases. For instance, [Kwan 2000)] used 3D GViz for identifying complex human activity patterns in space-time with a large activity diary dataset. Using time-geographic devices such as space-time paths and aquariums, movements of individuals over space and time are represented as continuous lines connecting various destinations. The method is particularly powerful in revealing associations and patterns in space-time ([Kwan 2001], [Kwan & Lee 2004a]). It can also be used to visualize real-time space-time data such as those collected by global positioning systems (GPS) or radars (e.g. flight paths of hijacked planes and trajectories of transcontinental missiles). Part of the project will seek to extend this 3D GViz method through adding the capability in handling vague and imprecise spatiotemporal references.

[Kwan 2000] specifically identifies three major advantages of this method. First, since GIS has the capability of handling massive amount of geographic data, it can generate far more realistic representations of any area and greatly facilitate the task of geospatial knowledge discovery. Second, unlike quantitative methods, there is no need to reduce the dimensionality of the data in the analysis since space, time and thematic attributes can be analyzed at the same time. Third, with navigational capabilities such as zooming, panning and dynamic rotation, and the ability to generate 3D animation series such as “fly-bys”, 3D GViz can create a virtual world with very high level of realism [Batty et al. 1998].

4. Data Sources for Geospatial Semantic Analytics

Information resources for the GSA include standard geospatial datasets available from NASA, USGS[3], and UCGIS member institutions. These sources include metadata prepared in accordance with FGDC metadata standards. Such metadata can be automatically extracted to build a test-bed of geospatial data characteristics that can be queried and used in semantic analysis. Also, *The National Map* of the USGS can be used to assemble and manage information about various USGS resources [USGS] as explained earlier. *The National Map* with its eight data themes and multiple levels of resolution and security access provides detailed information about national, regional, state, and local situations. *The National Map* will be the foundation of a USGS enterprise-wide geographic information system to which other organizations can add or reference their information, such as land use data, school district boundaries, or wildlife census information. Geospatial One-Stop, the current administration’s spatial component of e-government, provides another source and the portals to these data archives are linked. Currently, many geospatial datasets, including *The National Map*, lack adequate ontologies to support GSA. For example the orthographic image base for *The National Map* consists of 1-ft resolution digital color photographs from the Homeland Security Infrastructure Program (HSIP). While HSIP has defined features of interest, these features are neither annotated in the images nor in the other datasets of *The National Map*. Work is underway to develop an ontology for *The National Map.* This internal USGS effort is focused on developing complete specification of all features, attributes, and relationships on a layer basis. A feature ontology and the ability to use a multidimensional feature approach to The National Map is being developed [Usery 2004].

Also, data the National Imagery and Mapping Agency (NIMA), now the National Geospatial-Intelligence Agency (NGA has been examined [Usery et al. 2003]. The approach examined optimization of coastal zone databases using multimodal data and developed feature extraction methods for littoral warfare database features from overhead remotely-sensed images including multi-spectral and panchromatic images (Landsat TM, SPOT, Ikonos and Quickbird), lidar, radar, color photos, maps and other digital spatial sources. The development and implementation of a complete ontology with coded-domains for feature attributes was a part of the research. The work included determining optimum image sources for specific features and optimum viewing scales. The developed knowledge base includes feature, image, and map database components with prototype extraction and map generation capabilities.

UCGIS member institutions also provide a wealth of data and many serve as NSDI nodes. Again, most of the geospatial datasets of these institutions contain FGDC-compliant metadata. Included with the data are temporal stamps of data creation. Thus, from the metadata, geospatial characteristics including geographic location of the dataset with associated time of creation is available. This meta-

[3] http://geography.usgs.gov/www/products/1product.html

data can be used to devise a semantic association for a given location and the actual geospatial data including specific location of geographic entities, adjacency of other entities, and connectivity.

There are also several valuable, real-time data sets that are available from NASA. *MODIS*[4] is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites. Terra's orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days. There are 44 standard MODIS data products[5] that scientists are using to study global change. MODIS data is publicly available via the EOS (Earth Observing System) Data Gateway[6]. This interface allows users to search for data products using a variety of constraints, including geo-spatial constraints where we can specify particular geographic regions and also time ranges for the data products.

A second, very valuable NASA data source is the *QuikSCAT* scatterometer[7]. The *SeaWinds* instrument on the QuikSCAT satellite is a microwave-radar that measures near-surface wind speed and direction under all weather and cloud conditions over Earth's oceans. QuikSCAT data is made accessible through the Physical Oceanography Distributed Active Archive Center[8] (PO.DAAC). PO.DACC offers a variety of tools, interfaces, and protocols for accessing the winds data. Amongst other sources, the QuikSCAT data is also available through the EOS Data Gateway, thus having access to the EOS Data Gateway provides us with many of the important data products and simulation models we need from NASA.

Finally we must also mention on going activities at NASA that make accessing, searching and integrating earth science data possible using semantic web technology. The Earth Science Markup Language[9] developed for this very purpose is being used to mark up many NASA earth sciences related data sets. Such mark-up directly facilitates semantic descriptions of both the data as well as meta-data about the datasets (and data resources), and further facilitates search, access and integration by semantic Web based applications. There are several NASA data products and datasets for which the data and meta-data are already described using ESML, and we believe ESML is a strong candidate for data and meta-data descriptions of data and resources relevant to homeland security applications as well.

5. Conclusions

The role of semantics for interoperability and integration of heterogeneous data, including geospatial information, has been long recognized. The idea of a *Semantic Web* proposes "a Web of data that can be processed directly or indirectly by machines," bringing a higher degree of automation in exploiting data in a meaningful way. Semantics is captured by associating formal descriptions to provide well defined meaning to data and other Web resources so that information processing (retrieval or integration) can be based on meaning instead of on mere keywords. The W3C Semantic Web Activity Working Group has been working on a series of standards. Ultimately, ontologies can be an important tool in expediting the advancement of related sciences, and they can reduce the cost by improving sharing of information and knowledge.

The Geospatial Semantic Analytics (GSA) framework specifically looks for better support for geographic information that the basic Semantic Web research has not addressed. In particular, we develop thematic and spatial ontologies (SWETO-GS) and try to capture interactions among them. In

[4] http://modis.gsfc.nasa.gov
[5] http://modis.gsfc.nasa.gov/data/dataproducts.html
[6] http://edcimswww.cr.usgs.gov/pub/imswelcome/
[7] http://winds.jpl.nasa.gov/missions/quikscat/index.cfm
[8] http://podaac.jpl.nasa.gov/
[9] http://esml.itsc.uah.edu/index.jsp

this context, SWETO a large-scale test-bed is being developed and being extended with the data from several geospatial data sources outlined in Section 3. Various knowledge discovery techniques including spatiotemporal thematic proximity can use these test-beds for evaluation purposes.

References

[Alani 2001] Alani, H. Spatial and Thematic Ontology in Cultural Heritage Information Systems, PhD Thesis, University of Southampton, 2001.

[Aleman-Meza et al 2003] B. Aleman-Meza, C. Halaschek, I. B. Arpinar, and A. Sheth, "Context-Aware Semantic Association Ranking," Proceedings of Semantic Web and Databases Workshop, pp. 33-50, Berlin, September 7-8 2003.

[Allen 1991] J.F. Allen. Temporal Reasoning and Planning. In J.Allen, H. Kautz, R. Pelavin and J. Tenenberg, eds., Reasoning about Plans, M. Kaufmann, San Mateo, 1991.

[Anselin 1995] Anselin, L., Local indicators of spatial association - LISA. *Geographical Analysis* 27, 1995.

[Anselin 1999] Anselin, L., Interactive techniques and exploratory spatial data analysis. In P.A.Longley, M.F. Goodchild, D.J. Maguire, and D.W. Rhind (eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications*. New York: Wiley, 253-266, 1999.

[Anyanwu & Sheth 2002] K. Anyanwu and A. Sheth, The ρ Operator: Discovering and Ranking Associations on the Semantic Web," *SIGMOD Record,* Vol. 31, No. 4, December 2002.

[Anyanwu & Sheth 2003] K. Anyanwu, and A. Sheth, "The ρ-Operator: Enabling Querying for Semantic Associations on the Semantic Web. The 12th International World Wide Web Conference Budapest, Hungary, May 2003.

[Baader et al. 2003] F. Baader, D. Calvanese, D. McGuiness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook-Theory*, Implementation and Applications. Cambridge University Press, 2003.

[Bailey & Gatrell 1995] Bailey, T.C., and A.C. Gatrell., *Interactive Spatial Data Analysis*. Harlow, Essex: Longman, 1995.

[Batty et al 1998] Batty, M., Dodge, M., Doyle, S., Smith, A., Modelling virtual environments. In: Longley, P.A., Brooks, S.M., McDonnell, R., MacMillan, B., (Eds.), *Geocomputation: A Primer*, pp. 139-161, John Wiley & Sons, New York, 1998.

[Bechhofer et al 2003] S. Bechhofer, F. Harmelen, J. Hendler, I. Horrocks, D. McGuinness, and P. Patel-Schneider, OWL Web Ontology Language Reference. W3C Proposed Recommendation, 2003, http://www.w3.org/TR/owl-ref/.

[Chawla et al 2001] Chawla, S., S. Shekhar, W.L. Wu, and U. Ozesmi, Modeling spatial dependencies for mining geospatial data: An introduction. In H.J. Miller and J. Han (eds.) *Geographic Data Mining and Knowledge Discovery*. Taylor and Francis, 2001.

[Cutter et al 2003] Cutter, S.L., Richardson, D.B, and Wilbanks, T.J., *The Geographical Dimensions of Terrorism*. London, UK: Routledge 2003.

[Dill et al 2003] Dill, SemTag and Seeker: Bootstrapping the semantic Web via automated semantic annotation, The 12th International World Wide Web Conference Budapest, Hungary, May 2003.

[Egenhofer 2002] Egenhofer, M., Toward the semantic geospatial web, Proceedings of the tenth ACM international symposium on Advances in Geographic Information Systems, MaLean, Va, 2002.

[FGDC 2003] "Homeland Security and Geographic Information Systems, "FGDC-NSDI briefing paper June, 2003. http://www.fgdc.gov/publications/homeland.html.

[Fonseca & Sheth 2002] Fonsenka F., and Sheth A., The Geospatial Semantic Web, UCGIS White Paper, 2002. http://www.ucgis4.org/priorities/research/2002researchagenda.htm

[Fonseca et al 2002] F. Fonseca, M. Egenhofer, P. Agouris, and G. Câmara, Using Ontologies for Integrated Geographic Info. Sys. - Trans. in GIS, vol. 6, no. 3, pp. 231-257, 2002.

[Fotheringham et al 2000] Fotheringham, A.S., C. Brunsdon., and M. Charlton. *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage, 2000

[Gahegan 1995] Gahegan, M., Proximity operators for qualitative spatial reasoning. In *COSIT '95 Proceedings: Spatial Information Theory: A Theoretical Basis for GIS*. Eds. A. U. Frank and W. Kuhn. Berlin: Springer-Verlag 1995.

[Gahegan 2000] Gahegan, M., The case for inductive and visual techniques in the analysis of spatial data. *Journal of Geographical Systems* 2(1), 77-83, 2000.

[Gahegan 2001] Gahegan, M.,Visual exploration in geography: analysis with light. In H.J. Miller and J. Han, eds. *Geographic Data Mining and Knowledge Discovery*. London: Taylor & Francis, 2001.

[Goodchild et al. 2001] Goodchild, M., et al., eds. *Interoperating Geographic Information Systems*, Kluwer Academic Publishers: Boston1999,

[Gruber 1993] Gruber, T., Towards principles for the design of ontologies used for knowledge sharing. Technical report, Technical Report KSL93-04, Stanford Univ., Knowledge Systems Lab, 1993.

[Guarino 1998] Guarino, N., Formal ontology and information systems, In Proceedings of the 1st International Conference on Formal Ontologies in Information Systems, IOS Press, Trento, Italy, 1998.

[Hammond et al 2002] B. Hammond, A. Sheth, and K. Kochut, Semantic Enhancement Engine: A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content, in Real World Semantic Web Applications, V. Kashyap and L. Shklar, Eds., IOS Press, 2002.

[Harding 2003] Harding, J., Geo-ontology Concepts and Issues, Report of a workshop on Geo-ontology, Ilkley UK, September, 16–17, 2003.

[Heflin & Hendler 200] J. Heflin and J. Hendler. Searching the web with shoe. In Artificial Intelligence for Web Search. Papers from the AAAI Workshop. WS-00-01, pages 35--40. AAAI Press, 2000.

[Hernshaw & Unwin 1994] Hearnshaw, H.M., Unwin, D., (Eds.) *Visualization in Geographical Information Systems*. John Wiley & Sons, Chichester, England, 1994.

[Heuer 1999] R. J. Heuer, Jr., "Psychology of Intelligence Analysis", Center for the Study of Intelligence, Central Intelligence Agency, 1999.

[IBM-WF] IBM WebFountain.

[Karvounarakis et al 2002] G. Karvounarakis, S. Alexaki, V. Christophides, D. Plexousakis, M. Scholl, RQL: A Declarative Query Language for RDF, *WWW2002*, May 7-11, Honolulu, Hawaii, USA, 2002

[Kashyap 1999] V. Kashyap, Design and creation of ontologies for environmental information retrieval, Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management (KAW'99), Banff, Canada, October 1999

[Kashyap & Sheth 1996b] V. Kashyap, and A. Sheth. Schematic and Semantic Similarities between Database Objects: A Context-based Approach, pp. 276 – 304.VLDB Journal 5 (4), 1996.

[Kashyap & Behrens. 2001] V. Kashyap and C. Behrens. "The Emergent Semantic Web: A Consensus approach for Deriving Semantic Knowledge on the Web", Proceedings of the International Semantic Web Working Symposium, Stanford, USA, July 2001

[Kim & Kwan 2003] Kim, H.-M. and Kwan, M.-P, Space-time accessibility measures: a geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration. *Journal of Geographical Systems*, 5: 71-91, 2003.

[Kwan & Lee 2004a] Kwan, M.-P. and J. Lee, Geovisualization of human activity patterns using 3D GIS: A time-geographic approach. In *Spatially Integrated Social Science*, 48-66, Michael F. Goodchild and Donald G. Janelle (Eds). Oxford University Press, 2004.

[Kwan & Lee 2004b] Kwan, M.-P. and J. Lee, Emergency response after 9/11: The potential of real-time 3D GIS for quick emergency response in micro-spatial environments. *Computers, Environment and Urban Systems*, forthcoming, 2004.

[Kwan 1998] Kwan, M.-P, Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis* 30(3): 191-216, 1998.

[Kwan 2000] Kwan, M.-P, Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set. *Transportation Research C*, 8:185-203, 2000.

[Kwan 2001] Kwan, M.-P. Analysis of LBS-Derived Data for Social Scientists: Prospects and Limitations. LBS specialist workshop, 2001

[Kwan 2003] Kwan, M.-P Intelligent emergency response systems. In *The Geographical Dimensions of Terrorism*, 111-116, Susan L. Cutter, Douglas B. Richardson and Thomas J. Wilbanks (eds). New York: Routledge, 2003

[Lassila & Swick 1999] Ora Lassila and & Ralph R. Swick: "Resource Description Framework (RDF) Model and Syntax Specification", W3C Recommendation, Cambridge (MA), February 1999

[Longley et al 1998] Longley, P.A., Brooks, S.M., McDonnell, R., MacMillan, B., (Eds.), *Geocomputation: A Primer*. John Wiley & Sons, New York, pp. 139-161, 1998.

[MacEachren et al 1999] MacEachren, A.M., Wachowicz, M., Edsall, R., Haug, D. Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization and knowledge discovery in database methods. International Journal of Geographical Information Science 13(4), 1999.

[Mena et al 1996] E. Mena, V. Kashyap, A. Sheth, A. Illarramendi, "OBSERVER: An Approach for Query Processing in Global Information Systems Based on Interoperation Across Pre-existing Ontologies", Intl. Conf. on Cooperative Info. Systems (CoopIS'96), Brussels, June 1996.

[Mena et al 2000] E. Mena, A. Illarramendi, V. Kashyap and A. Sheth, "OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies", DAPD, pp. 223-271, Vol. 8, No. 2, April 2000.

[Mihalcea & Mihalcea 2001] R. Mihalcea, and S. I. Mihalcea: Word Semantics for Information Retrieval: Moving One Step Closer to the Semantic Web. ICTAI 2001: 280-287.

[NSGIC 2003] "Appropriate GIS Coordination Actions to Improve State Response to Acts of Terrorism, Sabotage and Natural Disasters," at http://www.nsgic.org. June, 2003.

[Randell et al. 1992] D. Randell, Z. Cui, and A. Cohn. A spatial logic based on regions and connections. In Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning, pages 165-176. Morgan Kaufmann, 1992.

[Resnik 1999] P. Resnik, "Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language", Journal of AI Research, 1999.

[Rodriguez & Egenhofer 2003] M. Rodriguez, and M. Egenhofer, Determining Semantic Similarity among Entity Classes from Different Ontologies, IEEE Trans. on Knowledge and Data Eng., 15(2), April 2003.

[Semagix-CIRAS] Anti-Money Laundering, Semagix, Inc.

[Shah & Sheth 1998] K. Shah and A. Sheth. Logical Information Modeling of Web-accessible Heterogeneous Digital Assets. Proceedings of the Forum on Research and Technology Advances in Digital Libraries, (ADL'98),pp. 266–275, Santa Barbara, CA. 1998.

[Sheth 1999] A. Sheth, "Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics", in Interoperating Geographic Information Systems. pp. 5-30, M. F. Goodchild, M. J. Egenhofer, R. Fegeas, and C. A. Kottman (eds.), Kluwer, 1999.

[Sheth & Ramakrishnan 2003] A. Sheth and C. Ramakrishnan, "Semantic (Web) Technology In Action: Ontology Driven Information Systems for Search, Integration and Analysis," In IEEE Data Eng. Bulletin, Special issue on Making the Semantic Web Real, Dec 2003.

[Sheth et al 2002a] A. Sheth, C. Bertram, D. Avant, B. Hammond, K. Kochut, and Y. Warke, "Semantic Content Management for Enterprises and the Web", IEEE Internet Computing, July/August 2002.

[Sheth et al 2002b] A. Sheth, S. Thacker and S. Patel, Complex Relationship and Knowledge Discovery Support in the InfoQuilt System, the VLDB Journal, 12 (1), January 2003.

[Sheth et al 2003bc] A. Sheth, S. Thacker, and S. Patel, Complex Relationship and Knowledge Discovery Support in the InfoQuilt System, VLDD Journal, 2003

[Sheth et al 2004] A. Sheth, B. Aleman-Meza, I. B. Arpinar, C. Halaschek, C. Ramakrishnan, C. Bertram, Y. Warke, K. Anyanwu, D. Avant, F. S. Arpinar, and K. Kochut. Semantic Association Identification and Knowledge Discovery for National Security Applications. In *Special Issue of Journal of DB Management on DB Technology for Enhancing National Security*, Eds: L. Zhou and W. Kim (submitted) 2004.

[Staab 2002] S. Staab: Emergent Semantics. IEEE Intelligent Systems 17(1), pp. 78-86, 2002.

[Townley 2000] J. Townley, The Streaming Search Engine That Reads Your Mind, Streaming Media World, August 10, 2000.

[Usery 2003] E. L. Usery, R. Welch, S. Fleming, and T.R. Jordan. Database and Symbology Issues Associated with Littoral Map Generation, Presented paper, American Society for Photogrammetry and Remote Sensing, Anchorage, AK, 2003.

[Usery et al 2004] E. L.Usery. Multidimensional Data Modeling for Feature Extraction and Mapping, Presented paper, American Congress on Surveying and Mapping, Nashville, TN, 2004.

[USGS] U.S. Geological Survey. http://www.usgs.gov

[USGS-ONT] http://www.geog.buffalo.edu/ncgia/i21/i21ontology.html