Submitted as book chapter to: *Encyclopedia of Geoinformatics*

Geospatial and Temporal Semantic Analytics

Matthew Perry, Amit Sheth, Ismailcem Budak Arpinar

Large Scale Distributed Information Systems Lab,
The University of Georgia, Athens, GA

1. Introduction

The amount of digital data available to researchers and knowledge workers has grown tremendously in recent years. This is especially true in the geography domain. As the amount of data grows, problems of data relevance and information overload become more severe, and the use of semantic technology is necessary to address these problems. With respect to data and services on the web, this semantics-based approach is embodied in the Semantic Web Activity[1]. More specifically, the Semantic Geospatial Web (Egenhofer, 2002) illustrates the use of geospatial semantics. Both of these approaches use ontologies to provide a shared understanding and conceptualization of relevant aspects of a problem or application area. Independent applications that interpret and process data with respect to these ontologies can achieve a much higher level of interoperability and information sharing.

In addition to information integration and interoperability, other areas of research in semantics and information systems have emerged, namely semantic search and browsing, and semantic analytics and discovery. The majority of the current work in both the Semantic Web and Semantic Geospatial Web communities has focused on issues of application interoperability and semantic search. While these issues are important, we feel that semantic analytics and discovery must not be overlooked, as the metadata created by semantic applications provides a powerful resource for analyzing entities and their relationships, allowing us to go from data to information to knowledge. Research in semantic analytics and discovery in the Semantic Web community has mainly focused on the thematic dimension of information and has largely ignored its spatial and temporal dimensions. At the same time, the Semantic Geospatial Web community has yet to

[1] http://www.w3.org/2001/sw/

integrate thematic knowledge of entities and their relationships with geospatial knowledge for analysis and discovery. It is at the intersection of these two communities that we see enormous potential for next generation geoinformatics applications that will successfully combine knowledge of real world entities and relationships with knowledge of their interactions in space and time. This article reviews background concepts from the Semantic Web community and describes state of the art work in semantic analytics and discovery in the purely thematic dimension. It then discusses our ongoing work in realizing semantic analytics and discovery in all three dimensions of information: thematic, geospatial, and temporal.

2. Background

The Semantic Web has received much attention recently. Its vision promises an extension of the current web in which all data is accompanied with machine-understandable metadata allowing capabilities for a much higher degree of automation and more intelligent applications (Berners-Lee *et al.*, 2001). To make this idea more concrete, consider the statement "The University of Georgia is located in Athens, GA." To a human with knowledge of colleges and universities and the geography of the southeastern United States, the meaning of this statement is clear. In addition, upon seeing this statement, other related information comes to mind such as professors who work at the University. In a Semantic Geospatial Web context, this related information would be GIS data and services, such as road network data and facility locations for the Athens area which could be combined with wayfinding services. The goal of the Semantic Web is to make the semantics of such data on the web equally clear to computer programs and also to exploit available background knowledge of related information. On the Semantic Web this statement would be accompanied with *semantic metadata* identifying an instance of the concept "University" with the name "The University of Georgia". Similarly, the instance of City and State, "Athens, GA," would unambiguously describe the university's geographic location. Note the distinction between semantic metadata describing high-level concepts and relationships and syntactic and structural metadata describing low level properties like file size and format. To create this semantic metadata, we must identify and mark occurrences of known entities and relationships in data sources. This tagging process is known as metadata extraction and semantic annotation. These annotations are especially important for multimedia data, as non textual data has a very opaque relationship with computers. Some examples of annotation of textual and multimedia data are presented in (Dill *et al.*, 2003; Hammond *et al.*, 2002), and (Jin *et al.*, 2005) respectively.

Ontologies are central to realizing the Semantic Web and Semantic Geospatial Web, as they formally specify concepts and their relationships and provide the means to create semantic metadata for objects (documents, data files, databases, etc.). Ontology is classically defined as "a specification of a conceptualization" (Gruber, 1993). In database terms, we can divide an ontology into two parts: a schema and instance data. The schema models a domain by defining class types (e.g. *University*, *City*) and relationship types (e.g. *located_in*). The schema is populated with instances of classes and relationships (e.g. *The University of Georgia located_in Athens*) to create facts representing knowledge of the domain. We have developed a number of ontologies describing thematic aspects of data. Some recent examples include GlycO and ProPreO in the Bioinformatics domain (Sahoo *et al.*, 2006) and more general-purpose ontologies such as SWETO (Aleman-Meza *et al.*, 2004). There has also been significant work regarding geospatial ontologies. Fonseca et al. (2002) propose an ontology-driven GIS. Kolas et al. (2005) outline specific types of geospatial ontologies needed in the Semantic Geospatial Web: base geospatial ontology, feature data source ontology, geospatial service ontology, and geospatial filter ontology. The base geospatial ontology provides core geospatial knowledge vocabulary while the remaining ontologies are focused on geospatial web services. Some examples of geospatial ontologies can be found in (Hiramatsu & Reitsma, 2004; Jones *et al.*, 2004).

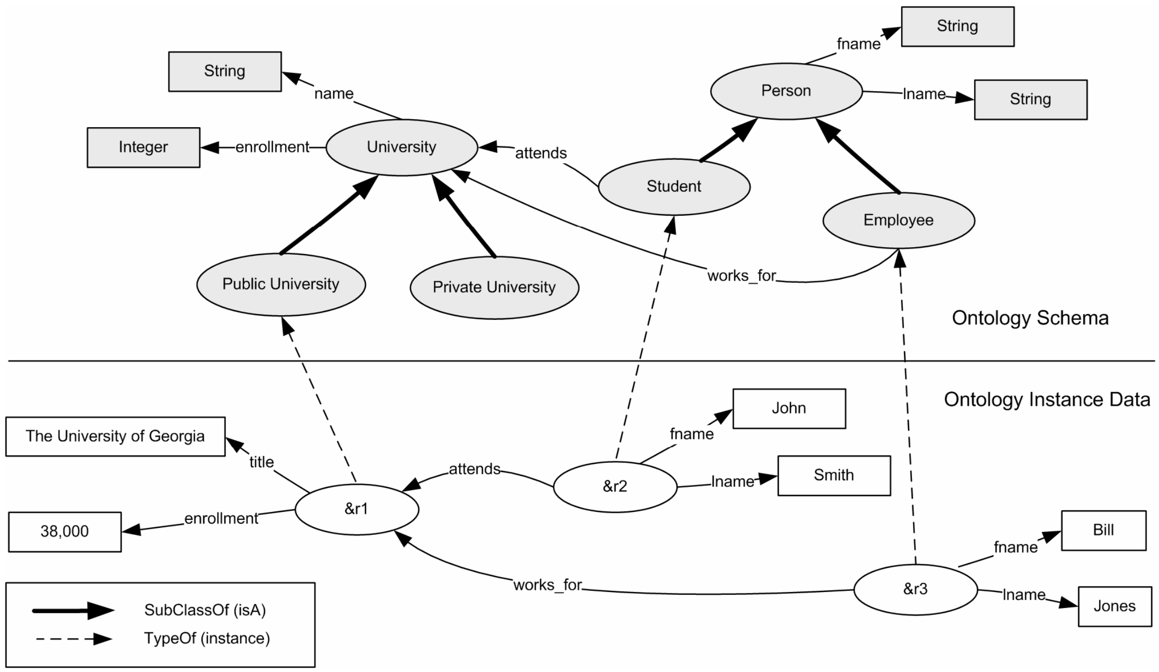

Figure 1. RDF Metadata

To provide this semantic metadata in a machine processable form, a standard way to encode it is needed. The W3C has adopted Resource Description Framework (RDF)[2] as the standard for representing semantic metadata. Metadata in RDF is encoded as statements about resources. A resource is anything that is identifiable by a *Uniform Resource Identifier* (URI). Resources can be documents available on the web or entities which are not web-based, such as people and organizations. RDF also defines literals which are not real-world entities but values (e.g. Strings, Integers) used to define attributes for resources. Relationships in RDF, known as *Properties*, are binary relationships between a resource and another resource or between a resource and a literal, which take on the roles of *Subject* and *Object*, respectively. The *Subject*, *Predicate* and *Object* compose an RDF statement. This model can be represented as a directed graph with typed edges and nodes. In this model, a directed edge labeled with the *Property* name connects the *Subject* to the *Object*. RDF Schema (RDFS)[3] provides a standard vocabulary for schema-level constructs such as *Class*, *SubClassOf*, *Domain*, and *Range*. In addition, the Web

[2] http://www.w3.org/RDF/
[3] http://www.w3.org/TR/rdf-schema/

Ontology Language (OWL)[4] further extends RDFS by defining additional vocabulary for describing classes and properties (e.g. *cardinality*, *disjointness*). Figure 1 shows an example ontology schema and instance data in the RDF model.

3. Semantic Analytics

Semantic analytics is a new form of intelligent information analysis which involves investigation of relationships between entities in populated ontologies and semantic metadata; the latter could be extracted from multiple heterogeneous data sources. It can be seen as a combination of querying and knowledge discovery, but it is fundamentally different from statistical data mining because it involves named relationships with well defined semantics. Semantic Web data models such as RDF/RDFS provide an excellent platform for semantic analytics because relationships are first class objects in these data models, making it very natural to query and analyze data based on these relationships. Our ongoing work in a project titled Semantic Discovery (SEMDIS)[5] addresses key issues in semantic analytics. The SEMDIS project investigates (1) operators for analyzing complex relationships in ontologies/metadata, (2) relevance models for ranking complex relationships, (3) algorithms for discovering complex relationships, and (4) development of large, populated ontologies.

Query operators developed for semantic analytics in the SEMDIS project attempt to answer the fundamental question of "how is entity *A* related to entity *B*?" Anyanwu and Sheth introduce the concept of semantic associations to answer this question (Anyanwu & Sheth, 2003). A semantic association is described as a complex relationship between two resources in an RDF graph. In fact, a semantic association can be defined in terms of connectivity or similarity. Two fundamental types of semantic associations are ρ-path (capturing connectivity) and ρ-iso (capturing similarity). Figure 2 illustrates these two types of semantic associations. Many useful

[4] http://www.w3.org/TR/owl-features/
[5] http://lsdis.cs.uga.edu/projects/semdis/

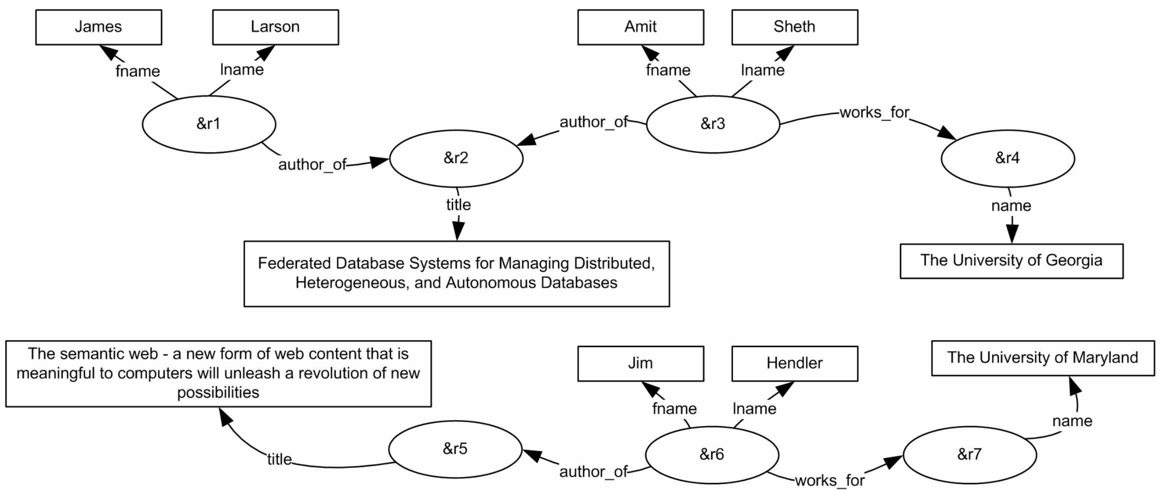

Figure 2. Example semantic associations. James Larson (resource &r1) is ρ-path associated with The University of Georgia (resource &r4) because the two resources are connected by a path in the RDF graph, and resource &r5 is ρ-iso associated with resource &r2 because the two resources are involved in two similar paths: they are both papers authored by university employees.

semantic associations involve some intermediate entities. Relationships that span several entities are very important in domains such as national security, because they may enable analysts to see the connections between seemingly disparate people, places and events. The result set of a semantic association query can involve an extremely large number of paths leading to information overload. To address this problem, we have researched relevance models for ranking semantic associations (Aleman-Meza *et al.*, 2005; Anyanwu *et al.*, 2005). We have also considered subgraph discovery as a complementary solution to discovery and enumeration of semantic associations (Ramakrishnan *et al.*, 2005). In this work, we are interested in finding dense subgraphs containing the “best” set of associations between two resources. The concept of semantic associations has been successfully applied in national security problems such as insider threat analysis (Aleman-Meza *et al.*, 2006b), and in other areas such as conflict of interest detection (Aleman-Meza *et al.*, 2006a) and searching patent databases (Mukherjea & Bamba, 2004).

4. Geospatial and Temporal Semantic Analytics

So far, the majority of our work in semantic analytics has focused on the purely thematic dimension of data, but we are now extending this work to incorporate the space and time dimensions. We believe that this will significantly enhance the power of semantic analytics operations. It will allow the integrated analysis of thematic relationships, geospatial relationships, and temporal relationships. In fact, it has been argued that relationships are at the heart of semantic technologies (Sheth *et al.*, 2003). At a high level, we want to use thematic and temporal knowledge to extract spatial or spatiotemporal regions and to influence spatial computations. Similarly, we want to use spatial and temporal knowledge to select portions of thematic space and influence thematic analysis. In the following, we discuss an ontology-based model integrating all three dimensions and a means to semantically connect entities from different dimensions based on semantic associations.

4.1. Modeling Space, Time, and Theme

The first challenge in realizing geospatial and temporal semantic analytics is that of modeling all three dimensions of data: thematic, spatial, and temporal. We are using an ontology-based approach to build this model which incorporates existing research in modeling dynamic geographic processes (Worboys, 2005; Worboys & Hornsby, 2004), temporal extensions to RDF (Gutierrez *et al.*, 2005), and ontology-based modeling of geographic data (described in Section 2). The thematic aspects of information are modeled with domain ontologies. However, we make three distinctions of class types in the thematic ontology: *Dynamic Entities*, *Named Places*, and *Events*. *Dynamic Entities* represent those entities with non-stationary (e.g. people, automobiles) or undefined spatial properties (e.g. abstract entities). *Named Places* are those entities with static spatial properties and clear spatial extents (e.g. a manufacturing plant, an apartment building, a city, etc.), and *Events* are special types of entities which represent occurrences in space and time (e.g. a car accident or a business meeting). To explain the rationale behind the division of thematic entity classes we must first discuss the geospatial aspects of the model.

Both qualitative relationships (topology, cardinal direction, etc.) and quantitative relationships like distance should be modeled to support geospatial analytics. These are common requirements for geospatial ontologies, so we can utilize existing work on geospatial ontologies from the Semantic Geospatial Web community for this purpose. For the remainder of the discussion we will consider a portion of the geo-ontology from (Jones et al., 2004), but it should be noted that our multidimensional model could use any geo-ontology capturing similar basic concepts. This ontology models a number of important concepts. Two of the main ones are *Geographical Place* and *Footprint*. *Geographical Place* represents a geographic feature (man made or natural) and is analogous to *Named Place* described previously. *Footprint* models a spatial element; it is a georeferenced point, line, or polygon. To give geospatial properties to

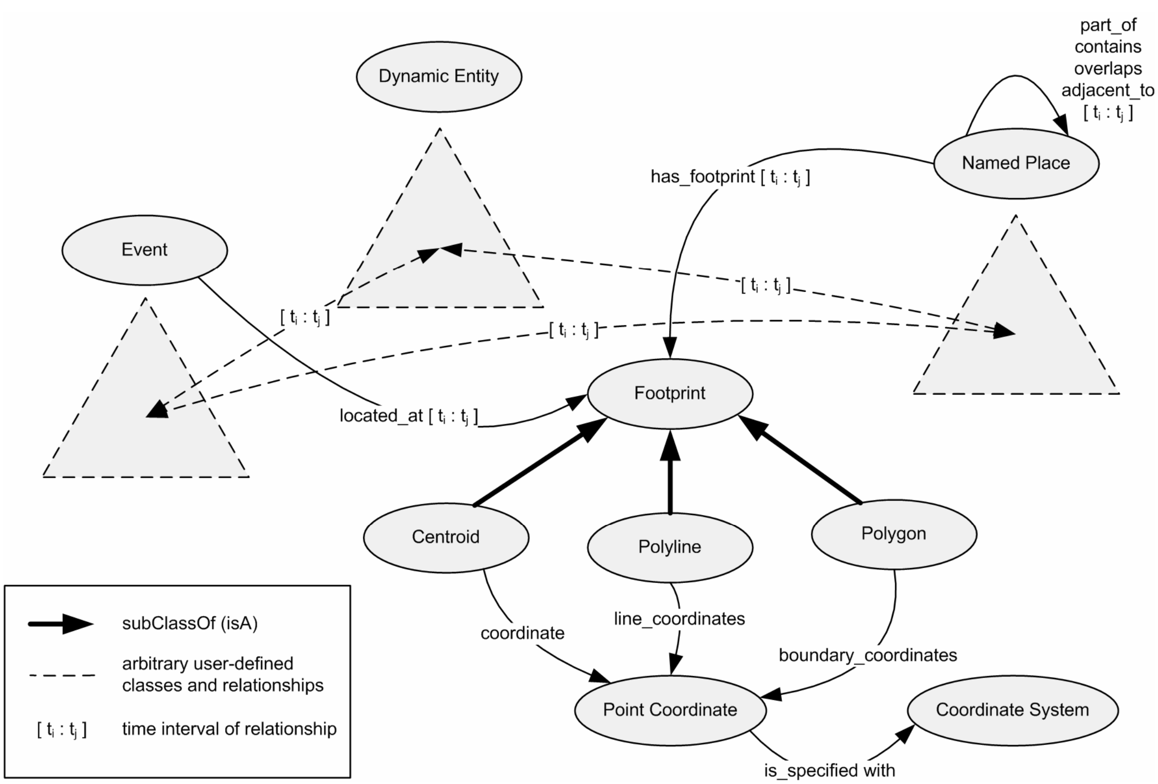

Figure 3. Ontology-based model of space, time, and theme. Events and Named Places are directly linked with footprints which record their geographic location. Temporal intervals on relationships denote when the relationship holds. Triangles represent arbitrary subclasses of Event, Dynamic Entity, and Named Place.

thematic entities we adopt the notion of a spatial setting from (Worboys & Hornsby, 2004). We can define a spatial setting as an instance of the class footprint. To link the thematic ontology with the geo-ontology, we define relationships to connect *Event* to *Footprint* and to connect *Named Place* to *Footprint* (*located_at* and *has_footprint,* respectively).

The remaining aspect of this model is temporal information. In (Gutierrez et al., 2005) the authors present a framework to incorporate temporal reasoning into RDF. Their framework defines a temporal label for an RDF statement. The label denotes the times that the statement or fact holds. We use this framework to give temporal properties to all relationships in the ontologies. This provides temporal settings for relationships in thematic ontologies and provides spatio-temporal settings for relationships connecting thematic ontologies to geo-ontologies. Figure 3 illustrates this model.

4.2. Spatiotemporal Thematic Contexts – Inter-dimension Connections

A context intuitively specifies a template for connecting entities in different dimensions: it defines the relevant classes, entities and relations for making the connection.

You may notice that *Dynamic Entities* have no direct relationship with *Footprints* in this model. However, *Dynamic Entities* participate in various thematic relationships with *Named Places* and *Events*, so it is through these connecting thematic relationships that *Dynamic Entities* can obtain different geospatial and temporal properties. The thematic relationship connecting the *Dynamic Entity* provides a *context* for the connection and allows us to analyze the geospatial and temporal properties of *Dynamic Entities* with respect to different thematic contexts. For example, a person's spatiotemporal properties with respect to employment relationships (where he works and when he works there) will differ from his spatiotemporal properties with respect to residence (where he lives and when he lives there). Note that this allows us to semantically query geospatial information because we are selecting regions of space using explicit semantic relationships between entities.

We can build the notion of context around semantic associations and time intervals. A thematic context is a schema-level ρ-path association representing the type of an instance-level ρ-path association. The type of the association is determined by the class types and property types of the entities and relationships on the path. For example, the association *Bob* - *works_for* - *The University of Georgia* - *located_in* - *Athens* would have the type (*Person*, *works_for*, *University*, *located_in*, *City*). A temporal context is represented by a time interval. An association matches a temporal context if all statements (i.e. RDF triples) making up the association are valid at some time *t* within the given time interval. A spatiotemporal thematic (STT) context is a temporal context in combination with a thematic context containing the class *Footprint*.

The connections created from STT contexts allow a number of ways to analyze information in three dimensions. This starts with examining an entity's spatial and temporal patterns with respect to an STT context. For example, viewing all spatiotemporal regions connected via the employment context with an open time interval will allow us to see an entity's employment history along with spatial and temporal information of the employment. Also we can look at relationships between space time regions of different contexts. We can see if a person lives close to where he works and how this pattern changes over time or how this pattern compares to others in the knowledgebase. In addition, we can examine which entities fall in close spatial and temporal proximity. For example, who does this person come in contact with on his commute to work? Who works in neighboring offices? These spatiotemporal proximity relationships can help us determine thematic relationships like *knows* or *collaborates with*. Thematic relationships can also affect spatial proximity. For example two close entities (e.g. persons or buildings) can be considered quite separated (or far) if there is natural barrier or international border between them. In addition, using STT contexts to constrain semantic association searches can be very useful. For example, an intelligence analyst trying to identify possible members of Al-Qaeda operating in the area of a suspected terrorist attack can benefit from semantic association searches between this organization and people which are connected to the attack location through various STT contexts. An analysis that supports STT contexts can also provide powerful tools to utilize Lifelogs (Gemmell *et al.*, 2003) which are a record of everything that a person does using video, audio, and other sensors.

5. Conclusions and Future Directions

In this article, we discussed the emerging field of semantic analytics and our ongoing work in extending semantic analytics from the purely thematic dimension to all three dimensions of theme, space, and time. We presented background work from the Semantic Web community and described how these new technologies can provide a means for semantic analysis of geospatial and temporal information. Semantic analytics have applications in general areas like document search and analysis and in more focused domains like Bioinformatics. In the future, we see semantic analytics that span the thematic, geospatial, and temporal aspects of information gaining more popularity. Diverse applications of such a capability include emergency and natural disaster response, homeland and, national security, and education and training.

6. References

Aleman-Meza, B., Halaschek-Wiener, C., Arpinar, I. B., Ramakrishnan, C., & Sheth, A. P. (2005). Ranking complex relationships on the semantic web. *IEEE Internet Computing, 9*(3), 37-44.

Aleman-Meza, B., Halaschek, C., Sheth, A., Arpinar, I. B., & Sannapareddy, G. (2004). Sweto: Large-scale semantic web test-bed, *Sixteenth International Conference on Software Engineering and Knowledge Engineering (SEKE2004): Workshop on Ontology in Action* (pp. 490-493). Banff, Canada.

Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A., et al. (2006a). Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection, *Fifteenth International World Wide Web Conference*. Edinburgh Scotland.

Aleman-Meza, B., Sheth, A. P., Palaniswami, D., Eavenson, M., & Arpinar, I. B. (2006b). Semantic analytics in intelligence: Applying semantic association discovery to determine relevance of heterogeneous documents. In K. L. Siau (Ed.), *Advanced topics in database research* (Vol. 5): Idea Group Publishing.

Anyanwu, K., & Sheth, A. P. (2003). ρ-queries: Enabling querying for semantic associations on the semantic web, *Twelfth International World Wide Web Conference* (pp. 690-699). Budapest, Hungary.

Anyanwu, K., Sheth, A. P., & Maduko, A. (2005). Semrank: Ranking complex relationship search results on the semantic web, *Fourteenth International World Wide Web*

Conference (pp. 117-127). Chiba Japan.
Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web - a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American, 284*(5), 34-43.
Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R. V., Jhingran, A., et al. (2003). *Semtag and seeker: Bootstrapping the semantic web via automated semantic annotation.* Paper presented at the Twelfth International World Wide Web Conference, Budapest, Hungary.
Egenhofer, M. J. (2002). Toward the semantic geospatial web, *Tentch ACM International Symposium on Advances in Geographic Information Systems*. McLean, VA.
Fonseca, F. T., Egenhofer, M. J., Agouris, P., & Camara, G. (2002). Using ontologies for integrated geographic information systems. *Transactions in GIS, 6*(3), 231-257.
Gemmell, J., Lueder, R., & Bell, G. (2003). The mylifebits lifetime store, *ACM SIGMM Workshop on Experiential Telepresence*. Berkeley, CA.
Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199-220.
Gutierrez, C., Hurtado, C., & Vaisman, A. (2005). Temporal rdf, *Second European Semantic Web Conference*. Heraklion, Crete, Greece.
Hammond, B., Sheth, A., & Kochut, K. (2002). Semantic enhancement engine: A modular document enhancement platform for semantic applications over heterogeneous content. In V. Kashyap & L. Shklar (Eds.), *Real world semantic web applications* (pp. 29-49): Ios Press Inc.
Hiramatsu, K., & Reitsma, F. (2004). Georeferencing the semantic web: Ontology based markup of geographically referenced information, *Joint EuroSDR/EuroGeographics workshop on Ontologies and Schema Translation Services*. Paris, France.
Jin, Y., Khan, L., Wang, L., & Awad, M. (2005). Image annotations by combining multiple evidence and wordnet, *Thirteenth ACM International Conference on Multimedia*. Singapore.
Jones, C. B., Abdelmonty, A. I., Finch, D., Fu, G., & Vaid, S. (2004). The spirit spatial search engine: Architecture, ontologies, and spatial indexing, *Third International Conference on Geographic Information Science*. Adelphi, MD, USA.
Kolas, D., Hebeler, J., & Dean, M. (2005). Geospatial semantic web: Architecture of ontologies, *First International Conference on GeoSpatial Semantics*. Mexico City, Mexico.
Mukherjea, S., & Bamba, B. (2004). Biopatentminer: An information retrieval system for biomedical patents, *Thirtieth International Conference on Very Large Data Bases* (pp. 1066-1077). Toronto, Canada.
Ramakrishnan, C., Milnor, W. H., Perry, M., & Sheth, A. P. (2005). Discovering informative connection subgraphs in multi-relational graphs. *SIGKDD Explorations, 7*(2), 56-63.
Sahoo, S. S., Thomas, C., Sheth, A., York, W. S., & Tartir, S. (2006). Knowledge modeling and its application in life sciences: A tale of two ontologies, *Fifteenth International World Wide Web Conference*. Edinburgh Scotland.
Sheth, A. P., Arpinar, I. B., & Kashyap, V. (2003). Relationships at the heart of semantic web: Modeling, discovering and exploiting complex semantic relationships. In M. Nikravesh, B. Azvin, R. Yager & L. A. Zadeh (Eds.), *Enhancing the power of the internet studies in fuzziness and soft computing*: Springer-Verlag.
Worboys, M. (2005). Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science, 19*(1), 1 - 28.
Worboys, M., & Hornsby, K. (2004). From objects to events: Gem, the geospatial event model, *Geographic Information Science* (pp. 327 - 344). Adelphi, MD.

7. Key Terms

Ontology: A specification of a conceptualization consisting of a hierarchy of class types and non-hierarchical relationships between classes.

Resource Description Framework (RDF): A Framework for describing resources on the web. RDF makes statements about resources consisting of a *Subject*, *Predicate*, and *Object* which translates to a directed, labeled graph.

Semantic Analytics: Analyzing, searching, and presenting information using explicit semantic relationships between known entities.

Semantic Annotation: Identifying and marking occurrences of ontological entities and relationships in raw data (e.g. documents, images, and digital geographic data).

Semantic Association: A complex relationship between two resources in an RDF graph. Semantic Associations can be a path connecting the resources or two similar paths in which the resources are involved.

Semantic Geospatial Web: The application of Semantic Web concepts and technologies for the sharing and reuse of geographic data and services on the web.

Semantic Metadata: Metadata that describe contextually relevant or domain specific information about content based on a shared metadata model (e.g. ontology).

Semantic Web: A framework that allows data on the web to be shared and reused across application, enterprise and community boundaries. The framework is realized through metadata annotations serialized using standard representations like RDF.

Spatiotemporal Thematic Context (STT Context): A specification of the type of ρ-path semantic association used to connect thematic entities to geospatial footprints. It is specified using a schema-level semantic association in combination with a time interval.

Uniform Resource Identifier (URI): Strings that uniquely identify resources on the web.