

Geospatial and Temporal Semantic Analytics

Matthew Perry, Amit Sheth, Ismailcem Budak Arpinar, Farshad Hakimpour

*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. The use of semantics has been proposed to combat these problems (Berners-Lee et al., 2001; Egenhofer, 2002). Semantics refers to the meaning of data rather than its syntax or structure. Systems which can understand and process data at a semantic level can achieve a higher level of automation, integration, and interoperability. Applications using semantic technologies fall into three basic categories: 1) semantic integration, 2) semantic search and contextual browsing, and 3) semantic analytics and knowledge discovery (Sheth & Ramakrishnan, 2003).

This article concentrates on semantic analytics and knowledge discovery. Applications in this category provide tools and techniques for analyzing entities and relationships in semantic metadata. So far, research in this area has mainly focused on the thematic dimension of information and has largely ignored its spatial and temporal dimensions. At the same time, the Geographic Information Systems community has yet to integrate thematic knowledge of entities and their relationships with geospatial knowledge for purposes of semantic analysis and discovery. Next generation geoinformatics applications that can successfully combine knowledge of real world entities and relationships with knowledge of their interactions in space and time will have huge potential in areas such as national security and emergency response. The remainder of 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, spatial, and temporal.

2. Background

In preparation for our discussion of geospatial and temporal semantic analytics, we first review basic concepts of ontologies, Semantic Web and semantic analytics.

2.1. Ontology

Ontologies are central to realizing semantic applications, 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. A number of ontologies describing thematic aspects of data have been developed. 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 been significant work regarding the use of geospatial ontologies in Geographic Information Science. Ontologies in GIS are seen as a vehicle to facilitate interoperability and limit data integration problems both from different systems and between people and systems (Agarwal, 2005). Fonseca et al. (2002) present an architecture for an ontology-driven GIS in which ontologies describe the semantics of geographic data and act as a system integrator independent of the data model used (e.g. object vs. field). Kuhn (2001) claims that, for maximum usefulness, geo-ontologies should be designed with a focus on human activities in geographic space and thus presents a method for constructing domain ontologies based on text analysis of

documents describing the domain (e.g. German traffic code text for the car navigation domain). Kuhn and Raubal (2003) also introduce the concept of semantic reference systems, of which ontologies are a component, as a means to describe the same geographic information from varying perspectives. This includes notions of semantic transformation and projection of ontologies which could possibly be used to present geographic information from different scales and granularities. Frank (2003) goes a step beyond purely spatial ontologies and argues for the inclusion of the temporal dimension and describes a multi-tier ontology with space-time as the fundamental dimension of physical reality. From a Web perspective, Kolas et al. (2005) outline specific types of geospatial ontologies needed to realize 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.

2.2. The Semantic Web

The Semantic Web¹ 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 (Egenhofer, 2002), 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.

To provide ontological 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 Ontology Language

¹ <http://www.w3.org/2001/sw/>

² <http://www.w3.org/RDF/>

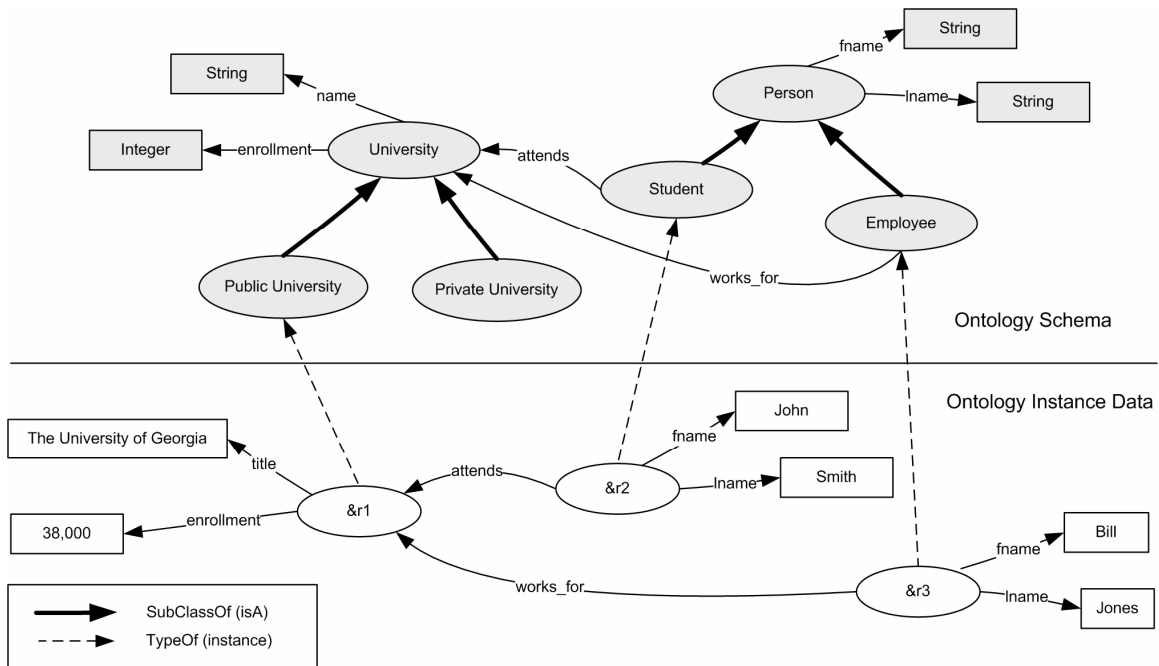


Figure 1. RDF Metadata. Schema-level classes and relationships are shown in gray. Instance data is shown in white. Ovals denote resources, and rectangles denote literals.

(OWL)⁴ further extends RDFS by defining additional vocabulary for describing classes and properties (e.g. *cardinality*, *disjointness*). Figure 1 illustrates the graph data model of RDF for an example ontology schema and instance data.

2.3. 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

³ <http://www.w3.org/TR/rdf-schema/>

⁴ <http://www.w3.org/TR/owl-features/>

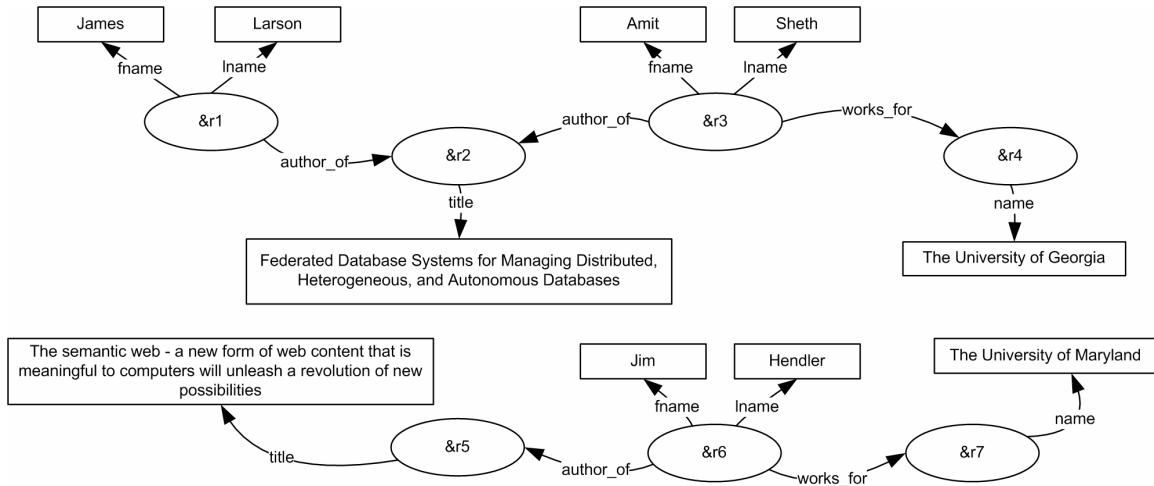


Figure 2. Example semantic associations. James Larson (resource &r1) is ρ -path associated with The University of Georgia (resource &r4) because the two resources are connected by a path in the RDF graph, and resource &r5 is ρ -iso associated with resource &r2 because the two resources are involved in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

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 (2003) introduce the concept of semantic associations to answer such questions. 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 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., 2003). 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 areas such as conflict of interest detection (Aleman-Meza et al., 2006) and searching patent databases (Mukherjea & Bamba, 2004).

3. Geospatial and Temporal Semantic Analytics

The basic goal of geospatial and temporal semantic analytics is an extension of thematic analytics which supports search and analysis of spatial and temporal relationships between entities. In the following, we present an ontology-based model integrating all three dimensions of data: thematic, spatial and temporal. Also, we discuss semantic associations as a means to analyze relationships among the three dimensions.

3.1. Modeling Space, Time, and Theme

Our current approach to modeling theme, space, and time consists of an upper-level ontology defining a general collection of thematic and spatial entity classes and associated relationships connecting these entity classes. We intend for application-specific domain ontologies in the thematic dimension to be integrated into the upper-level ontology through subclassing of appropriate upper-level classes and relationships. We make three distinctions of class types in the thematic dimension: *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

⁵ <http://lsdis.cs.uga.edu/projects/semdis/>

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 simple 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 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.

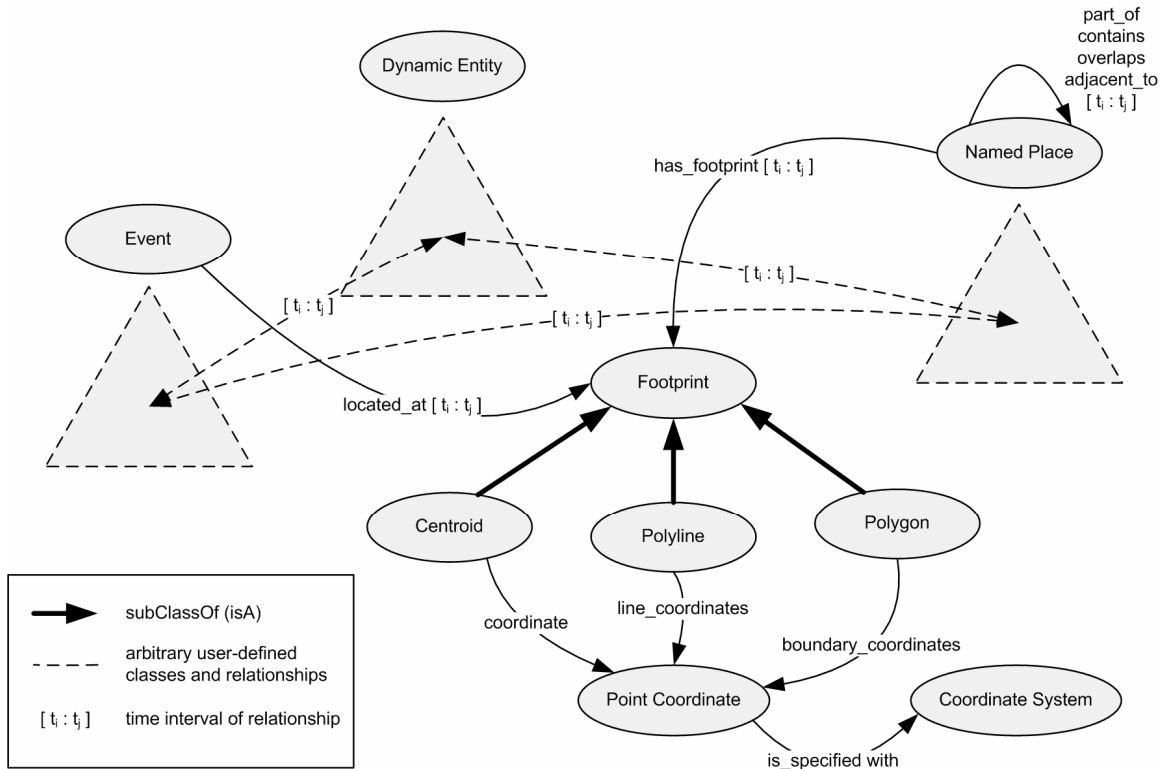


Figure 3. Ontology-based model of space, time, and theme. Events and Named Places are directly linked with footprints which record their geographic location. Temporal intervals on relationships denote when the relationship holds. Triangles represent arbitrary subclasses of Event, Dynamic Entity, and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

3.2. 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*.

Consider the example of analyzing historical entities and events of WWII as an illustration of the use of STT contexts. Suppose we want to know which Military Units came in close spatiotemporal proximity of the 3rd Armored Division during November 1944. The following STT context connects *Military Unit* to *Footprint*: *(Military Unit, participates_in, Battle, located_at, Footprint)* with a time interval covering November 1944. Starting from the 3rd Armored Division, we traverse instances of this context to retrieve the associated geospatial regions. Next metric or topological spatial relationships, such as intersect, are used to find close spatial regions from which instances of the battle participation context are traversed to find those Military Units in close spatiotemporal proximity.

3.3. Related Approaches

There is no shortage of spatiotemporal data models for temporal GIS. (Pelekis et al., 2004) identifies 10 distinct spatiotemporal data models. Of these, the three domain model (Yuan, 1994, 1996) is most relevant. This model represents semantics, space and time separately. To represent spatiotemporal information in this model semantic objects are linked via temporal objects to

spatial objects. This provides temporal information about the semantic (thematic) properties of a given Spatial Region. This is analogous to temporal `located_at` and `has_footprint` relationships in our model. In the three-domain model there is a one-to-one mapping from semantic and temporal objects to spatial objects and from spatial and temporal objects to semantic objects. This is the key difference with our approach because we incorporate non-spatial entities into the semantic domain and provide the notion of a thematic context to link these entities to the spatial domain in a variety of ways. Our Semantic Web style approach has the potential to incorporate a larger amount of non-spatial information by utilizing indirect connections to the spatial domain, and it allows the direct application of existing thematic analytics techniques.

4. Conclusions and Future Directions

In this article, we discussed the emerging field of semantic analytics and our thoughts on 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.

5. References

- Agarwal, P. (2005). Ontological considerations in GIScience. *International Journal of Geographical Information Science*, 19(5), 501-536.
- 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*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.

- Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A., et al. (2006). *Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- Anyanwu, K., & Sheth, A. P. (2003). *r-Queries: Enabling Querying for Semantic Associations on the Semantic Web*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- Anyanwu, K., Sheth, A. P., & Maduko, A. (2005). *SemRank: Ranking Complex Relationship Search Results on the Semantic Web*. Paper presented at the Fourteenth International World Wide Web Conference, 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-+.
- 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*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- 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.
- Frank, A. (2003). *A Linguistically Justified Proposal for a Spatio-Temporal Ontology*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- 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*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- 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.
- Jin, Y., Khan, L., Wang, L., & Awad, M. (2005). *Image Annotations by Combining Multiple Evidence and WordNet*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- Jones, C. B., Abdelmonty, A. I., Finch, D., Fu, G., & Vaid, S. (2004). *The SPIRIT Spatial Search Engine: Architecture, Ontologies, and Spatial Indexing*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- Kolas, D., Hebler, J., & Dean, M. (2005). *Geospatial Semantic Web: Architecture of Ontologies*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- Kuhn, W. (2001). Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science*, 15(7), 613 - 631.
- Kuhn, W., & Raubal, M. (2003). *Implementing Semantic Reference Systems*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- Mukherjea, S., & Bamba, B. (2004). *BioPatentMiner: An Information Retrieval System for BioMedical Patents*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- Pelekis, N., Theodoulidis, B., Kopanakis, I., & Theodoridis, Y. (2004). Literature Review of Spatio-Temporal Database Models. *The Knowledge Engineering Review*, 19(3), 235-274.
- 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*. Paper presented at the

- Conference Name|. Retrieved Access Date|. from URL|.
- Sheth, A. P., & Ramakrishnan, C. (2003). Semantic (Web) Technology In Action: Ontology Driven Information Systems for Search, Integration and Analysis. *IEEE Data Engineering Bulletin*, 26(4), 40-47.
- Worboys, M., & Hornsby, K. (2004). *From Objects to Events: GEM, the geospatial event model*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- Yuan, M. (1994). *Wildfire Conceptual Modeling for Building GIS Space-Time Models*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
- Yuan, M. (1996). Modeling Semantical, Temporal and Spatial Information in Geographic Information Systems. In M. Craglia & H. Couclelis (Eds.), *Geographic Information Research: Bridging the Atlantic* (pp. 334-347): Taylor & Francis.

6. 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.