Semantic Web Methodologies for Spatial Decision Support

Michael J. Casey — Mark A. Austin**

University of Maryland
Institute for Systems Research and
Department of Civil and Environmental Engineering
College Park
MD 20742
USA

* *caseymj@isr.umd.edu*
** *austin}@isr.umd.edu*

ABSTRACT: *Spatial decision support systems (SDSSs) have emerged from the co-evolution of research in decision support systems (DSSs) and geospatial information systems (GISs). Because of issues of interoperability and scale, these systems face new challenges and opportunities in an Internet-connected world. We present a conventional and semantic agent-assisted view of SDSS capability and discuss current trends in GIS technology relevant to spatial decision support. We introduce and discuss technologies and methodologies being developed for the web-based exchange of data and services using semantic agents. A semantic-web enabled spatial navigation problem is presented and its advantages for decision-time generation of alternatives are discussed. Finally, we discuss extensions and challenges in semantic-web research as they pertain to spatial decision support.*

KEYWORDS: *semantic web, spatial decision support system (SDSS), geospatial information system (GIS), Ontology*

1. Introduction

Decision support systems (DSSs) in the Internet Age will not be individual, interactive systems with specialized modeling features and interfaces, but will instead be combinations of data and services linked over the web for a specific purpose. DSS tools will be modular and reusable. To address a wide range of decision making problems in disparate domains, value will be added to data in the form of expressive metadata models. Intelligent "agents" capable of inferring knowledge from these models and of supporting decisions will be ubiquitous (Hendler, 1999). To arrive at this vision, the widely accepted and highly successful models of DSSs must adapt to the challenges and opportunities of an Internet-connected world.

One area that promises significant new capabilities for decision making is in the use of spatial or geospatial information. Spatial decision support systems (SDSSs) have emerged from the co-evolution of classic DSS research and from research in geospatial or geographic information systems (GISs) (Densham and Goodchild 1989, Densham 1991, Malczewski 1999). A SDSS can be defined as "an interactive, computer-based system designed to support a user or group of users in achieving a higher effectiveness of decision making while solving a semi-structured spatial decision problem" (Malczewski, 1999). Implicit to this definition is that spatial decision problems often combine heterogeneous spatial and non-spatial data, multiple decision criteria defined by multiple stakeholders, and varying degrees of uncertainty. Complex systems are necessary to manage changes to the spatial data, rule-based routines, and interface configuration. Such architecture is often not interoperable with other systems and is not sustainable at the scale of the Internet.

The web has achieved unprecedented success as a medium to freely publish and search documents in the form of web pages. As the volume of documents grows on the web however, we will need an order of magnitude better search capability to retrieve the same quality of search results that are possible today (Hendler, 2001b). Research in the development of tomorrow's web is focused on the exchange of not just documents, but on the exchange of data and services, human-to-machine communication, and machine-to-machine communication. This effective and efficient exchange will depend on our ability to embed meaning or semantics in conjunction with data and services accessible through the web. In order to de-couple a portion of the human interaction with SDSSs, and in order to distribute, and scale SDSS capabilities to the web, new semantic methodologies for spatial decision support are necessary.

The first tenet of this paper is that SDSSs in the Internet Age will need to provide GIS functionality at a level much lower than is common now. In order for SDSSs to interoperate with other systems, they will need to provide a minimal set of open, low level interfaces (e.g., SQL) for access to spatial data. This approach will ensure that the widest possible range of users (both human and other machines) will have access to centralized and decentralized spatial databases and services.

Malczewski (1999) and others have suggested that GIS can be used as a DSS generator whereby the data and model base for the SDSS are created from within the GIS. We argue that this is not the correct approach because is does not address system integration and interoperability issues. To create SDSSs that are scalable and reusable, they should be assembled from modular components where some level of decision support is still possible even without the presence of a particular component.

The second tenet of our paper is that in order to support spatial decisions in the future, we will need agents to carry some of the load for us. These agents must be able to query voluminous amounts of loosely structured data and dynamically return decision alternatives that are in-line with the decision maker's objectives and preferences. In the geospatial domain, a core component of SDSSs has been the ability to generate maps to support decisions. Agents cannot easily read maps, however. They lack the human cognitive ability to aggregate spatial and non-spatial data into a one-dimensional form to enable decision making. Efforts are necessary to develop encoding mechanisms for geospatial vocabularies and semantics – to capture the meaning humans derive from maps in a machine-readable form to support spatial reasoning.

2. Spatial Decision Support

The objectives of DSSs are to provide the correct balance of data, processing, and reporting to support individuals or groups for solving semi-structured problems under varying degrees of uncertainty. They are a relatively mature paradigm, and can be used to classify applications ranging from spreadsheets to on-line analytical processing (OLAP) systems (Power, 1997). Although there is some disagreement on what technologically constitutes a DSS, the conceptual structure and processes of a DSS is widely accepted. Simon (1960) and Spargue and Watson (1996) have given the structure, processes, and classification for decision support problems. Malczewski (1999) has mapped the decision support processes of intelligence, design, and choice to the spatial decision support domain.

The capability of GISs to support decision making has been of interest since the first systems were developed. Location/allocation, network routing, and areal clustering are all examples of spatial analysis problems, which, when combined with decision making techniques such as multi-criteria decision analysis (MCDA) form a SDSS. The core SDSS functionality consists of a database of spatial information, a rule-base of problem specific behavior, a dialog/human interface system, and a toolkit of spatial analysis techniques. This architecture, although powerful and widely used, is largely "pre-web." Geospatial data, even if available through the web, must today be manually converted/loaded into a particular GIS system, manipulated to conform with an established spatial referencing system, and filtered to meet the requirements of the current decision context. For this reason, no new data or functionality (services) is available to the decision maker through the SDSS at “decision time” (i.e., in real-time).

3. The Semantic Web

The semantic web is the migration from today’s human readable, document-based web to tomorrow’s machine-readable data-based web. The web has emerged primarily as a medium for people to publish documents (Frauenfelder, 2001). Powerful search engines such as Google (www.google.com) continually index web documents and provide rapid, accurate results to text-based queries. Information can be located on the web on virtually any topic given only a few keywords. But how do we search the web for data or services? The web in its current form has very little metadata and no means to encode *semantics*. The core of the semantic web lies in the ability to express meaning and establish relationships between resources (documents, data, and services) and to process queries based on these relationships. Methodologies borrowed from artificial intelligence, knowledge representation, and many other domain-specific communities will achieve this goal. Frauenfelder (2001) outlines four basic components necessary for building the semantic web: (1) The first component will be that of **machine readable** markup for web content. XML (eXensible Markup Language) is already being used to create “self-describing documents.” New markup languages will be necessary to encode the semantics of web content. (2) The second component will be tools that can read and **index semantic markup**. Just as is the case for today’s web, the semantic web will rely on distributed resources indexed by a centralized knowledge base. Agents will “crawl” the semantic web for metadata – markup that will describe relationships between resources. (3) The third component will be the development of **tools and services** that process semantic information and infer new facts. This component can be viewed as the next generation of travel, news, brokerage, and other service-based sites – ready to communicate with agents, not just humans. (4) The final component is interested in the capabilities of semantic agents to **reason** and to support decision making.

4. Ontology and Metadata

Ontology (uppercase) is defined as “that branch of philosophy which deals with the nature and organization of reality” (Guarino and Giaretta, 1995). An ontology (lowercase) is defined as “a logical theory which gives an explicit, partial account of a conceptualization” (Guarino and Giaretta, 1995). Still another definition appropriate for the semantic web is “a set of knowledge terms, including the vocabulary, the semantic interconnections and some simple rules of inference and logic, for some particular topic” (Hendler, 2001a). Why do we need ontologies? Ontologies capture the model of knowledge for a particular domain. They allow us to describe resources on the web and the relationships between those resources. For example, a geospatial ontology would include the concepts of location, adjacency, containment and way-finding (navigation). A personal web page may contain a postal mailing address written in HTML, but this has no meaning to an agent. If the address is marked-up with instances of a geospatial ontology, however, an agent is able to infer an individual's geographic location both in absolute terms (coordinates) and relative terms (country, county, city, neighborhood, etc.).

Ontologies are metadata for the web. But how is metadata relevant for decision support? DSSs can provide the means for generating a series of decision alternatives for comparison and evaluation of different objectives. In order for DSSs to carry out their prescribed functions in the future, they will rely on metadata to describe attributes, objectives, context, and constraints and will therefore be ontology-driven. Research is necessary to develop ontologies and tools which can encode the semantic nature of decision making given logical, mathematical, or spatial criteria.

5. The Semantic Web Layer Cake

There is no holistic view of the semantic web and few examples exist to demonstrate the farthest goals of its functionality. Researchers in the semantic web community have adopted a "layer cake" model to classify capabilities of technology where each successive layer relies on the layers below it. Figure 1 depicts the various layers that comprise the semantic web.

The bottom-most layer contains the underpinnings of the web itself. Unicode (via HTTP) makes text and images universally readable by computers everywhere. Universal resource indicators (URIs) are globally unique addresses to resources on the web and are a superset of the more familiar URL addresses. This layer is where the majority of today's web can be assigned. The next layer contains the rapidly growing technologies for adding structure to documents on the web; XML, name spaces, and XML schema. These technologies combine to structure documents as hierarchical trees based on domain-specific tags. A XML schema, similar to a relational database schema, formalizes the structure and allows documents to be machine-read and parsed.

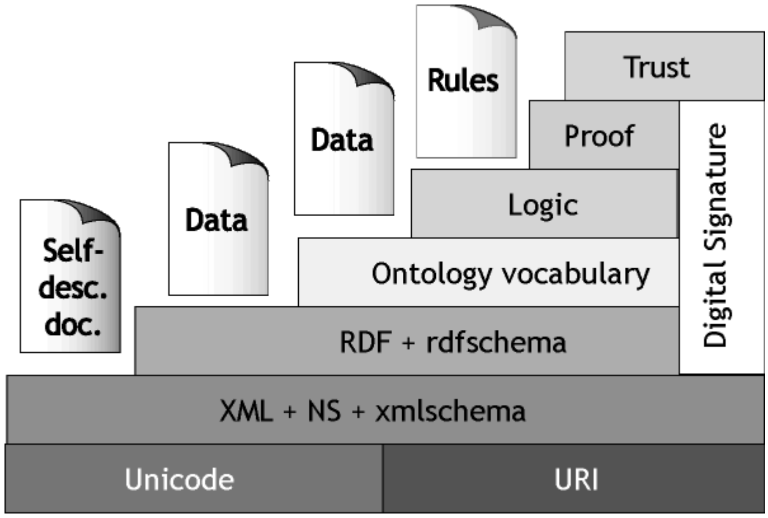

Figure 1: *Semantic Web "layer cake" presented by Tim Berners-Lee at XML2000 conference*

The RDF and RDF schema layer build upon the XML layer by defining a metadata description language. RDF (resource description framework) is a simple knowledge representation language for expressing predicate logic. It consists of loosely coupled subject-verb-object "triples" that allow metadata creators to establish relationships between resources spanning the entire web. For example, "Maryland is located in the United States", "I am working on Project A", or "water flows downhill" are conceptual statements that can be expressed given links to the appropriate subjects (e.g., "Maryland), verbs (e.g., "is located in"), and objects ("the United States"). Instances of the appropriate subjects, verbs, and objects (collectively resources) are available via URIs which globally and uniquely identify them. RDF, then, allows us to describe data (images, sounds, terrain databases, etc.) in a serialized, machine readable manner.

Building upon the expressive capabilities of RDF are ontology vocabularies which describe the specific associative, causal, and functional relationships between resources of a particular domain. The semantics of one ontology can vary greatly with respect to another ontology, especially for homonymic terms. For instance, the term "field" in a sports ontology might refer to a playing surface while in a physics ontology it might refer to an electromagnetic property. Expressive ontology languages are necessary to encode the relevant semantics with higher order relationships such as cardinality and Boolean operations. One example of such a language is the DARPA Agent Markup Language combined with the Ontology Interchange Layer (DAML+OIL, see www.daml.org).

The three uppermost layers of the cake represent the vision of the semantic web's true capabilities. Semantic agents driven by indexed ontology markup will perform basic reasoning on facts encoded in metadata. Ontology languages such as DAML+OIL will describe the rules applicable to resources from a particular domain. Methodologies borrowed from the artificial intelligence and knowledge representation communities (i.e., logical programming and theorem provers) will enable semantic agents to reason based on rules and facts given a problem's semantic context. An example of a semantic web reasoning facility is CWM (pronounced as in the Welsh term for valley). CWM is capable of reading and parsing RDF syntax and performing forward reasoning using simple horn rules.

The model of the semantic web layer cake is consistent with the processes used by DSSs (intelligence, design, and choice). Capabilities realized at the highest levels of the layer cake will allow semantic web agents to gather data based on ontological markup, filter it based on problem attributes and objectives, reason, then put forward alternatives based on the decision maker's preferences. The semantic web offers the opportunity for decision support in an implicitly distributed, heterogeneous environment where semi-structured decision problems can be solved *efficiently* and *effectively* with the aid of semantic agents.

6. Semantic web-Enabled Spatial Navigation

To see how semantic web-enabled navigation and spatial decision support are enhanced by semantic web methodologies, we provide a simple example of a tourist trying to navigate from one location to another using an "advanced" hand-held device. It is not unreasonable to assume that in the next five years a significant portion of the world's population will be carrying next generation PDAs which will integrate today's mobile phone, schedule/address book, GPS, and wireless web device. Taylor (2000) has referred to this device as the "fist." Imagine a tourist is in an unfamiliar city and is trying to reach a place he or she knows only by name. How could a collection of semantic agents help the tourist make decisions to navigate to their destination? Figure 2a shows a map of Central Park south of 86^{th} Street in New York City – the locale of our navigating tourist example.

The tourist enters Central Park near East 66th street and is trying to reach The Great Lawn, somewhere in the park. Let's assume that no map or signage is available to guide the tourist to this destination and that he or she must rely on their "fist" (the map in Figure 2a is for the benefit of the reader). A typed or spoken query into the tourist's fist would be something like "Help me find the Great Lawn." Even if a device such as the fist existed, this type of query on the web of today would return at best documents describing Central Park attractions and possibly map images. As we have stated before, agents cannot interpret maps. Internet Map Servers and location-based services could help, but as we will see, they lack the capability to assemble new information at decision time.

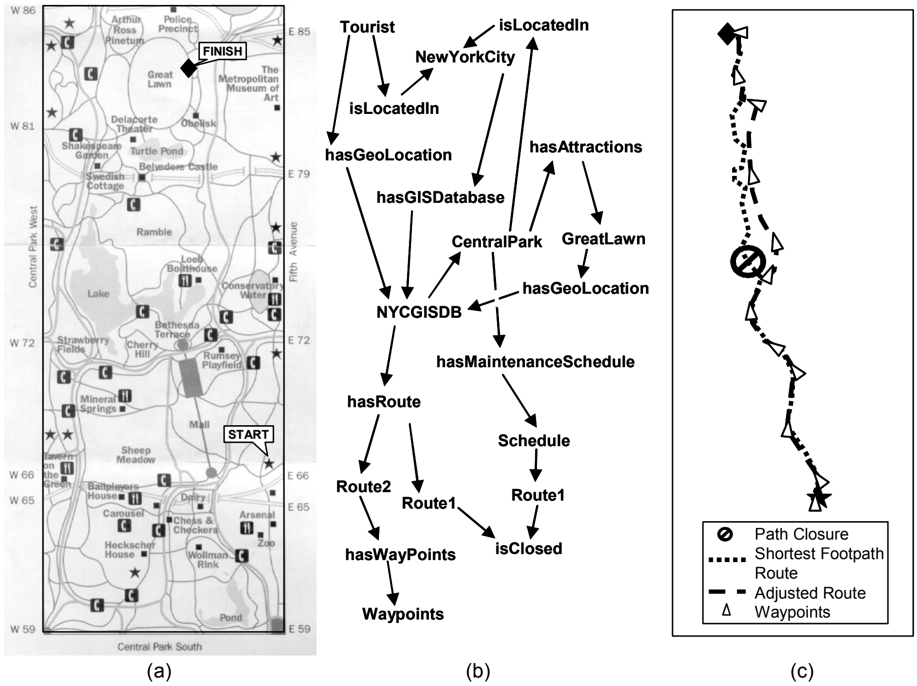

Figure 2: *Example of spatial navigation problem in New York City's Central Park with (a) background map (b) directed graph of linked ontologies, and (c) calculated routes and waypoints*

On the semantic web, the tourist's query would be accepted and processed by a geo-location service (agent). Figure 2b is a directed graph of the processing steps executed by the geo-location agent. Submitted with the query would be the tourist's geographic coordinates. Using semantic web methodologies, the geo-locating agent would begin to break down the tourist's query based on ontological references. Based on the submitted coordinates, the agent would quickly determine that the tourist is in New York City, specifically Central Park. An ontology for place names in Central Park could be located at the Central Park website (http://www.centralparknyc.org) rendered in an ontology language such as DAML+OIL. The ontology would include both a description of The Great Lawn and its geographic coordinates.

The agent would next recognize that a database of geospatial information (NYCGISDB) is available for querying. Using the tourist's coordinates and the coordinates of The Great Lawn as input, the agent would be able to determine the shortest route along a combination of Central Park footpaths using a separate network routing service. At the time the geo-locating agent was negotiating with the routing agent for service, it simultaneously identified in the Central Park ontology an instance of a maintenance schedule effective at the time of the tourist's query. The agent compared the optimal route computed by the routing agent (Route1) with the list of footpaths scheduled for closures. Because Route1 included a closed footpath, the geo-location agent repeats its routing request to the routing server with this new information. Given this network "stop", the routing server quickly re-computed an alternative route (Route2).

Our tourist is not interested in a map, only in point-to-point directions along the shortest route. The final processing step executed by the geo-location agent is the translation of Route 2 into waypoints corresponding to directions along the route. Figure 2c shows the initially computed route (Route1), the footpath closure, the re-computed Route2, and its associated waypoints. The tourist receives the waypoints on their fist which provides turn-by-turn directions directly to The Great Lawn.

If IMS applications such as MapQuest and location-"aware" devices exist today, why is the semantic-web enabled spatial navigation approach better than current methods? Semantic web methodologies have assisted the tourist in the following ways:

1. **Minimal interface to GIS/SDSS**. The geo-location agent has eliminated some of the human-computer interaction necessary on today's web. The tourist specified their query having no direct connection to a GIS, spatial database, or routing server. The geo-location agent identified the appropriate ontologies, geospatial data, and routing service (agent) and delivered the waypoints of the best path to the user via a minimal interface with their fist.

2. **Decision-time reasoning.** At the time the query was processed, the agent discovered the ontology for the maintenance schedule. It could have just as easily identified a schedule of events or other features of the park which would have effected computation of the route. *A pre-existing interface to this information was not in place and explicit interfaces to new information need not be created.*

3. **Disaggregation of spatial data**. The tourist was not interested in a web-page or document, only the turn-by-turn waypoints (i.e., data). The geo-location agent was able to disaggregate and deliver the minimum amount of spatial data preferred by the user.

7. Extensions and Challenges

Our "Walk in Central Park" example is presented simply as a means of demonstrating how the semantic web and location-based services can work together to provide useful navigation guidance that steps beyond today's practices (i.e., just reading the maps and signs in Central Park). Of course our real interest in this area is to understand the extent to which the logic, proof and trust elements of the semantic web layer cake can enhance the timeliness and quality of decision making support in much more sophisticated real-world scenarios. One can imagine, for example, the semantic web playing a role in the real-time control and customized vehicle-level routing of city road traffic, based upon geospatial map and weather information, integrated with local events, accident and incident information.

The implementation of such a semantic web-based system will require an ability to gather, synthesize, learn and reason efficiently with large quantities of heterogeneous data (not all relevant), derive conclusions for plausible solutions and deliver them to end-users (using terminology they can understand) through a wide array of information devices. As pointed out by Greiner (Greiner et al., 2001) there are numerous difficult problems in dealing with just the reasoning component of the problem. For example, we assume that ontologies in the semantic web may be connected in arbitrarily complex ways and may have borders that are unknown (i.e., they are open-ended). This means that reasoning and decision making will often need to operate in the presence of "partial knowledge". Research is needed to understand how combinations of logic- and probability-based reasoning systems should be implemented in a semantic web setting. A second key issue is balancing quality of service (QoS) with computational effectiveness. Everyone wants decision making that is correct, precise, expressive and efficient. In practice, however, any system that wants to guarantee efficient reasoning must sacrifice something -- usually, correctness and expressiveness in decision making.

But for decisions that are non life-threatening, the consequences of this trade-off may not matter. For example, in our "Walk in Central Park" there may be no need to provide directions with more precision than "turn left," "turn right," or "go straight." Applying the same degree of precision to the flight of an airliner is almost guaranteed to result in disaster. Trade-offs in the implementation of decision making computations need to be controlled, in part, by the end-use of that decision. This points to the need for "end-use" ontologies and/or profiles connected to the computational aspects of decision making.

8. References

BERNERS-LEE T, HENDLER J, LASSILA O. (2001) "The Semantic Web." *Scientific American.* 284 (5): May 2001.

DENSHAM P., GOODCHILD M. (1989) "Spatial Decision Support Systems: A Research Agenda." *Proceedings of GIS/LIS'89.* Orlando, FL., pp. 707-716.

DENSHAM P. (1991). "Spatial Decision Support Systems." In: D.J. Maguire, M.S. Goodchild, and D.W. Rhind (eds.), Geographical information Systems: principles and applications. Harlow, Essex, England: Longman, pp. 403-412.

FRAUENFELDER, M. (2001) "A Smarter Web." *MIT Technology Review*. November 2001.

GREINER R. ET AL. (2001) "Efficient Reasoning" *ACM Computing Surveys.* 33(1). March 2001, pp 1-30.

GUARINO N., GIARETTA P., (1995) "Ontolgies and Knowledge Bases: Towards a Terminological Clarification." In N.J.I. Mars (ed.), Towards Very Large Knowledge Bases, IOS Press.

HENDLER J. (1999) "Is There an Intelligent Agent in Your Future?" *Nature.* March 1999.

HENDLER J. (2001a) "Agents and the Semantic Web." *IEEE Intelligent Systems*. April 2001.

HENDLER J. (2001b) "The Great Debate: Dr. James Hendler and Dr. Ben Shneiderman on the Future of the Internet." American Society for Information Science and Technology 2001 Annual Meeting." Washington, DC. 16 November 2001

MALCZEWSKI, J. (1999) GIS and Multicriteria Decision Analysis., Wiley, John & Sons., New York, USA.

OPEN GIS CONSORTIUM (2001). http://www.opengis.org (30 November 2001).

POWER T. (1997) "What is a Decision Support System?" *DS* (Online Journal for Data Intensive Decision Support)*. October 21, 1997. Vol. No. 3.

SIMON H. (1960). The new science of management decision. Harper & Row Publishers, New York.

SPRAGUE, R., WATSON H. (1996). Decision Support for Management. Prentice Hall, Upper Saddle River, New Jersey.

TAYLOR, C. (2000) "4 We Will Be Perfectly Productive." *Time Digital Magazine (On Monthly)*. May 2000.