Data Integration in the Geospatial Semantic Web

Patrick Maué, Westfälische Wilhelms-Universität,Münster, Germany

Sven Schade, Westfälische Wilhelms-Universität,Münster, Germany

EXECUTIVE SUMMARY

Geospatial decision makers have to be aware of the varying interests of all stakeholders. One crucial task in the process is to identify relevant information available from the Web. In this chapter the authors introduce an application in the quarrying domain which integrates Semantic Web technologies to provide new ways to discover and reason about relevant information. The authors discuss the daily struggle of the domain experts to create decision-support maps helping to find suitable locations for opening up new quarries. After explaining how semantics can help these experts, they introduce the various components and the architecture of the software which has been developed in the European funded SWING project. In the last section, the different use cases illustrate how the implemented tools have been applied to real world scenarios.

Keywords: Case Study, Data Semantics, Future of Information Systems, Geospatial Decision-Making, Knowledge Integration, Ontologies, Semantic Annotations, Semantic Matching, Web Architecture

INTRODUCTION

A careful assessment and selection of relevant information retrieved from the World Wide Web (WWW) is crucial for potentially precarious decision making tasks. Geographical (or geospatial) data created by national mapping agencies or globally acting companies is the relevant information in the case of geospatial decision making. Such information is most often served as compiled and sometimes interactive maps. Processing such data, for example to run a geospatial analysis, requires raw data. Using a case example from the mining industry, we are going to explain in this chapter how Semantic Web technologies facilitate the task of finding such geospatial data sets and assessing their usefulness for the intended application. We also show how semantics simplify loading and visualizing of this data using generic desktop clients.

Geospatial decision making requires the acquisition of relevant data sets beforehand (Longley et al., 2005). The notion of Geographic Information Retrieval (GIR) (Larson, 1996) is commonly used to summarize all the tasks needed to acquire geospatial data from the WWW by using search

engines or catalogs implemented especially for managing such data. Most information is captured in form of text-based documents, which makes searching using keywords and text-indexing search engines very efficient. Geospatial data comes with additional dimensions: Space, i.e. the geographical extent of the represented entities, is one apparent characteristic. Special catalogs have been implemented to deal with the spatial component of such data by adding interactive maps to search dialogs. Time is another important aspect (Longley et al., 2005). The data yields information about entities or phenomena in the real world which is continuously moving on in time and changing its state. Hence, making geospatial data available in the Web poses various challenges to the data provider. The major standardization organization in the domain of geospatial applications, the Open Geospatial Consortium (OGC), introduced standardized catalog interfaces which support basic GIR functionalities. Instances of these catalogs are, together with other OGC-conformal Web Services serving or processing geospatial data, embedded in Spatial Data Infrastructures (SDI). In (Nebert, 2004) the basic principles of SDIs are introduced; various components of SDI and the potential benefits of Semantic Web technologies are further discussed later in this chapter.

Standards specifying how to access and use GIR systems are limited to the technical level. They define how to encode the queries, what parameters are allowed, and which operations the catalog should implement. They do not, however, suggest how a user formulates the query. The properties of geospatial data (space, time, and theme) can all be individually queried. But the task of asking the right question to retrieve this information is still a challenging one. It seems to be so simple, though. A user defines a region by drawing a box on a map, selects a point in time using a calendar, and specifies a theme by choosing a category or typing in some search terms. But especially keywords and categories suffers from the information asymmetry (Akerlof, 1970) between the searching user and the data provider. Language is, by nature, ambiguous and imprecise. Semantic conflicts (i.e. heterogeneities) and varying background knowledge impair every discovery task and can decrease the usefulness of the GIR system tremendously. And the success of systems depending on user interaction is measured by its usefulness.

Let us consider a common example. In a mining context, a team of experts is searching for a suitable location for a new quarry (more details about this use case are discussed in the following section). Many regulations constrain the potential areas; the region could be a protected area due to conservation, water protection zones, or national monuments. Before getting to a decision, the experts have to consider this additional data bearing potential conflicts with regulations. Considering relevant information made available in the Web is challenging. Experts looking for potential conflicts due to unknown regulations would probably have no problem to define the spatial and temporal extent of the data they are looking for. Selecting appropriate keywords for discovery is more difficult. Search engines are often limited to combinations of keywords; the so called Boolean search queries (Manning et al., 2008). Too few and too general keywords like "protected area" yield too much results. Specific, but also usually required terms like "ZNIEFF" (or Zones Naturelles d'Intérêt Ecologique Faunistique et Floristique) are unknown to the experts. They are experts in their own domain, but have no knowledge in others. In a next tedious step, they have to extend and reformulate the query until the search results in a manageable number of hits. Next, they have to browse through the remaining results, probably with the result that the requested information is not even included. More sophisticated IR systems are clearly needed to improve efficiency, and the Semantic Web is the promising emergent technology to address this issue.

In addition to the challenging task to find the relevant information, the decision makers have to include the different stakeholders in the process. Results of geospatial decision making can have a severe impact on the living conditions of the local population, especially if it ends

Related Content

Ontology Driven Document Identification in Semantic Web
Marek Reformat, Ronald R. Yagerand Zhan Li (2010). *Progressive Concepts for Semantic Web Evolution: Applications and Developments (pp. 186-220).*
www.irma-international.org/chapter/ontology-driven-document-identification-semantic/41654

Electronic Reputation Systems
Mario Paolucci, Stefano Picasciaand Samuele Marmo (2010). *Handbook of Research on Web 2.0, 3.0, and X.0: Technologies, Business, and Social Applications (pp. 411-429).*
www.irma-international.org/chapter/electronic-reputation-systems/39183

Towards Process Mediation in Semantic Service Oriented Architecture (SSOA)
Tariq Mahmoudand Jorge Marx Gómez (2009). *Handbook of Research on Social Dimensions of Semantic Technologies and Web Services (pp. 780-801).*
www.irma-international.org/chapter/towards-process-mediation-semantic-service/35757

Mobile Ontologies: Concept, Development, Usage, and Business Potential
Jari Veijalainen (2008). *International Journal on Semantic Web and Information Systems (pp. 20-34).*
www.irma-international.org/article/mobile-ontologies-concept-development-usage/2844

Files are Siles: Extending File Systems with Semantic Annotations
Bernhard Schandland Bernhard Haslhofer (2010). *International Journal on Semantic Web and Information Systems (pp. 1-21).*
www.irma-international.org/article/files-siles-extending-file-systems/47106