

Chapter 12

Semantic-Based Geospatial Data Integration With Unique Features

Ying Zhang

North China Electric Power University, China

Shaowen Liu

North China Electric Power University, China

Chaopeng Li

North China Electric Power University, China

Liming Du

North China Electric Power University, China

Na Chen

*Hebei Vocational College of Rail
Transportation, China*

Zhuxiao Wang

North China Electric Power University, China

Miaomiao Ma

North China Electric Power University, China

ABSTRACT

Since large amount of geospatial data are produced by various sources, geospatial data integration is difficult because of the shortage of semantics. Despite standardised data format and data access protocols, such as Web Feature Service (WFS), can enable end-users with access to heterogeneous data stored in different formats from various sources, it is still time-consuming and ineffective due to the lack of semantics. To solve this problem, a prototype to implement the geospatial data integration is proposed by addressing the following four problems, i.e., geospatial data retrieving, modeling, linking and integrating. We mainly adopt four kinds of geospatial data sources to evaluate the performance of the proposed approach. The experimental results illustrate that the proposed linking method can get high performance in generating the matched candidate record pairs in terms of Reduction Ratio(RR), Pairs Completeness(PC), Pairs Quality(PQ) and F-score. The integrating results denote that each data source can get much Complementary Completeness(CC) and Increased Completeness(IC).

DOI: 10.4018/978-1-5225-8054-6.ch012

1. INTRODUCTION

Geospatial data integration can be used to improve data quality, to reduce costs, and to make data more useful to the public (Auer et al., 2009; Bittner et al., 2009; Brodt et al., 2010; Kuhn, 2002; Su et al., 2012; De Carvalho et al., 2012; Su & Lochovsky, 2010; Ballatore et al., 2014; Buccella et al., 2010; Fonseca, Egenhofer et al., 2002; Malik et al., 2010; Vaccari et al., 2009). However, the large amount of data is produced by a variety of sources, stored in incompatible formats, and accessible through different GIS applications. Thus, geospatial data integration is difficult and becoming an increasingly important subject.

To implement the geospatial data integration, four problems need to be addressed: geospatial data retrieving, modeling, linking and integrating. This paper proposes corresponding approach for each issue. Besides, our work takes advantage of Karma (Szekely et al., 2011; Knoblock et al., 2012; Taheriyani et al., 2012; Tuchinda et al., 2011; Knoblock et al., 2011), which is a general information integration tool. It supports importing data from a variety of sources including relational databases, spreadsheet, KML and semi-structured Web pages, and publishing data in a variety of formats such as RDF. The source modeling work is based on these functions.

1. **Data Retrieval:** This problem concerns data extraction from Web APIs, which provide users with access to the corresponding databases. Diverse APIs have different input requirements. It is hard for users without adequate background knowledge to exploit the growing amount of heterogeneous geospatial data. The objective of our retrieval approach is to enable both API providers and API users to semi-automatically model Web API and invoke the related services to extract the geospatial data.
2. **Source Modeling:** The retrieved geospatial data from various sources are often described according to multiple perceptions, different terms and with different level of detail (Witten et al., 1999; Baeza-Yates & Ribeiro-Neto, 1999; Li & Fonseca, 2014). To overcome the diverse nature of geospatial data and to represent them in a uniform way, recent research has applied the concepts of the Semantic Web to geospatial data integration (Janowicz et al., 2011; Peng, 2005; Arpinar et al., 2006; Oussalah et al., 2013). Semantic Web (Berners-Lee et al., 2001) introduces the ontology languages such as Resource Description Framework (RDF) and Web Ontology Language (OWL) to provide benefits of semantic annotation. By providing a semantic interpretation of the data, RDF and OWL allows software programs to understand structures and meanings of different information sources. In this paper, we align the extracted data in a semantic way. We build a generic geospatial ontology and take advantage of Karma to map and align the extracted geospatial data based on the generic geospatial ontology.
3. **Record Linking:** This problem is about recognizing the same entity from different sources. The task of record linkage is commonly used for improving data quality and integrity by reducing costs and efforts in data acquisition. Our main contribution focuses on addressing this problem. In the previous work, the record linking process has been done by defining a set of semantic rules. These methods missed the unique characteristics of geospatial data such as geospatial relationships directly, which are significant for linking and whose functions cannot be expressed by semantic techniques. We make the best of such unique features and semantic techniques in geospatial data linking. Features refer to characteristics in this work. Further, in addition to the basic functions, the geospatial relationships endow the geospatial data linking with translation function, which can save the translating cost among all the geospatial sources with different languages.

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/semantic-based-geospatial-data-integration-with-unique-features/222902

Related Content

Insurance in Italy: A Spatial Perspective

Giovanni Millo and Gaetano Carmeci (2013). *Geographic Information Analysis for Sustainable Development and Economic Planning: New Technologies* (pp. 158-178).

www.irma-international.org/chapter/insurance-italy-spatial-perspective/69055

Smart GIS-Based Water Information System

Ankur Dumka, Poonam Kainthura and Alaknanda Ashok (2021). *Interdisciplinary Approaches to Spatial Optimization Issues* (pp. 193-205).

www.irma-international.org/chapter/smart-gis-based-water-information-system/279257

Client and User Involvement Through BIM-Related Technologies

Silvia Mastrolemo Ventura and Angelo L. C. Ciribini (2017). *International Journal of 3-D Information Modeling* (pp. 19-34).

www.irma-international.org/article/client-and-user-involvement-through-bim-related-technologies/188401

Geographic Information System Effects on Policing Efficacy: An Evaluation of Empirical Assessments

Yan Zhang, Larry Hoover and Jihong (Solomon) Zhao (2014). *International Journal of Applied Geospatial Research* (pp. 30-43).

www.irma-international.org/article/geographic-information-system-effects-on-policing-efficacy/111099

Semantic Modelling of 3D Multi-Utility Networks for Urban Analyses and Simulations: The CityGML Utility Network ADE

Tatjana Kutzner, Ihab Hijazi and Thomas H. Kolbe (2018). *International Journal of 3-D Information Modeling* (pp. 1-34).

www.irma-international.org/article/semantic-modelling-of-3d-multi-utility-networks-for-urban-analyses-and-simulations/225001