Chapter 11
Semantic Web and Geospatial Unique Features Based Geospatial Data Integration

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 and stored in incompatible formats, 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. First, we provide a uniform integration paradigm for users to retrieve geospatial data. Then, we align the retrieved geospatial data in the modeling process to eliminate heterogeneity with the help of Karma. Our main contribution focuses on addressing the third problem. Previous work has been done by defining a set of semantic rules for performing the linking process. However, the geospatial data has some specific geospatial relationships, which is significant for linking but cannot be solved by the Semantic Web techniques directly. We take advantage of such unique features about geospatial data to implement the linking process. In addition, the previous work will meet a complicated problem when the geospatial data sources are in different languages. In contrast, our

proposed linking algorithms are endowed with translation function, which can save the translating cost among all the geospatial sources with different languages. Finally, the geospatial data is integrated by eliminating data redundancy and combining the complementary properties from the linked records. We mainly adopt four kinds of geospatial data sources, namely, OpenStreetMap(OSM), Wikmapia, USGS and EPA, 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).

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; Taheriyan 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

Related Content

Usability Dimensions in Collaborative GIS

Mordechai (Muki) Haklay (2006). *Collaborative Geographic Information Systems (pp. 24-42).*

www.irma-international.org/chapter/usability-dimensions-collaborative-gis/6650

BIM and Cultural Heritage: Compatibility Tests in an Archaeological Site

Cristiana Achille, Nora Lombardiniand Cinzia Tommasi (2016). *International Journal of 3-D Information Modeling (pp. 29-44).*

www.irma-international.org/article/bim-and-cultural-heritage/171612

Quality of Urban Life Index From Location-Based Social Networks Data: A Case Study in Belo Horizonte, Brazil

Rodrigo Smarzaro, Tiago França Melo de Limaand Clodoveu Augusto Davis Jr. (2017). *Volunteered Geographic Information and the Future of Geospatial Data (pp. 185-207).*

www.irma-international.org/chapter/quality-of-urban-life-index-from-location-based-social-networks-data/178805

Discovering Geosensor Data By Means of an Event Abstraction Layer

Alejandro Llavesand Thomas Everding (2013). *Geographic Information Systems: Concepts, Methodologies, Tools, and Applications (pp. 2020-2039).*

www.irma-international.org/chapter/discovering-geosensor-data-means-event/70548

3D Web Services for Visualization and Data Sharing in 3D Cadastre

Ali Zare Zardinyand Farshad Hakimpour (2015). *International Journal of 3-D Information Modeling (pp. 1-15).*

www.irma-international.org/article/3d-web-services-for-visualization-and-data-sharing-in-3d-cadastre/154017