Chapter 14
Semantically Enriching Geodata

Khaoula Mahmoudi
LTSIRS Laboratory, University of Tunis El Manar, Tunisia

ABSTRACT

The requirement of GIS community in information is in vogue. Indeed, GIS is often extremely useful in many fields such as sciences, health, business, and community services to do everything. This includes creating, querying, overlaying maps and analysis of the spatial and non-spatial information. To carry out all these processing, the GIS refers to the data collected and stored in the GDB. Seen the multitude of GIS users' preoccupations, there is a growing interest and awareness in reaching the valuable and strategic data by enriching the GDB. This is made possible through mining different data sources to complement the native GDB. In this context, this chapter deals with an approach devised to enrich the GDB semantic data by interacting reactive agents to process documents related to the geographic entities. It deals with a modular approach consisting of fragmentation, theme identification, delegation, filtering and eventually results refinement.

INTRODUCTION

Everywhere in the world, the humans in different fields and with different preoccupations are concerned with spatially referenced data. For the decision making, the Geographic Information System (GIS) (Pornon, 2000; Faïz & Krichen, 2013) was proposed as a prominent solution to manage such data by taking into account its semantic and spatial aspects. These data are stored in the Geographic DataBase (GDB). Such dataset is well designed to offer an easy way to reach the intended data. Although, given the multitude of the users' requirements, the GDB is not able to meet all the GIS users' requests.

Indeed, the data-acquisition is the most important and challenging process along through the GIS development workflow. The major time and cost are spent in developing the GDB. This is due to the complexity and the expensiveness of the acquired geodata.

Moreover, given the complexity of the problems to handle with GIS, an exhaustive study of the situation (current geographic object and its surrounding linked objects) is of prime importance. Such dataset is generally beyond the GDB storing capabilities given the reasons cited above.

One illustration of the information requirement in the GIS context, is the disease tracking. Hence, by observing previous cases, we look to predict and minimize the harm caused by the epidemic situations. Such investigation concerns the information related to the social, economic and environmental conditions of the tracked geo-referenced places. This encompasses for instance information about waste management given the adverse effects of waste on health. Citizens' habits are considered as gist information revealing mainly the citizens' interaction with their surrounding natural or artificial entities.

Obviously, such variety of information is no way to be stored in the same GDB.

Moreover, in some situations we look for more detailed information, hence the GDB attributes are so abstract to make the sound decision. By referring to the example above, to be able to undertake the adequate measurement, a detailed meaningful description of the situation at hand is of great importance. Then, in rural areas given the intensive animal husbandry, it is worth informing to know where the animals are kept and how the inhabitants behave regarding the animals' waste, the water and food nature they provide and so on. All these detailed information are to be conveyed by the text documents rather than the GDB attributes.

Hence, mining additional data sources and especially the textual repositories will complement and enrich the dataset stored within the GDB to support the majority of GIS users.

In this context, our contribution is to complement the GDB semantic data by reporting the gist of information in a condensate form by managing the textual material relative to the geographic entities.

The remainder of this chapter is organized as follows. Section 2 provides a background of the semantic enrichment process. In section 3, the overall enrichment workflow is presented. The text fragmentation is described in section 4. Section 5 is dedicated to the theme identification stage. The delegation is reported in section 6. The text filtering is detailed in section 7. The output refinement is to be described in section 8. The simulation and evaluation results are dealt with respectively, in section 9 and 10.

BACKGROUND

The GDB enrichment is becoming a necessity in front to the increasing demand on geographic information. The literature has revealed different efforts towards alimenting the GDB with complementary data to satisfy the majority of requests.

In this context, the GeoNode (Geographic News On Demand Environment) (Hyland et al., 1999, Clifton et al., 2001) operates on news-based information related to some geographic entities. This data source is to be investigated to pinpoint the embedded stories. The latter are explored by the MITRE's Alembic system (Day, 1997) to pick out the named entities. A gazetter is used to geo-reference the places. The story datelines are also captured. The overall extracted data stands for the metadata. The named entities are explored to extract the topic via a clustering process (Clifton & Cooley, 1999). Hence, the co-occurrence of the names related to organizations, persons and places, many times, is considered as the description of a certain topic.

A visualization is offered by the GeoNode exhibiting the stories on ArcView GIS.

PIV (Pyrénées Itinéraires Virtuels) system (Lesbegueries et al., 2006; Marquesuzaà et al., 2008) explores a specific textual repository relative to the territorial aspects of the Pyrenees and its associated land history. By using the Linguastream tool (Frédérik et al., 2006), the corpus is processed to retrieve the most relevant documents with respect to the geographic entity handled by the user. Hence, each document is analyzed to reveal the potential space features. The same processing is performed to the

Related Content

Concept and Approach to Land Management Interventions for Rural Development in Africa
Uchendu Eugene Chigbu (2019). *Geospatial Technologies for Effective Land Governance (pp. 1-14).*
www.irma-international.org/chapter/concept-and-approach-to-land-management-interventions-for-rural-development-in-africa/214476

Demystifying Big Data in the Cloud: Enhancing Privacy and Security Using Data Mining Techniques
Gebeyehu Belay Gebremeskel, Yi Chaiand Zhongshi He (2015). *Geo-Intelligence and Visualization through Big Data Trends (pp. 264-304).*
www.irma-international.org/chapter/demystifying-big-data-in-the-cloud/136108

Spatial-Temporal Analysis of Recent Air Passenger Flows
Changjoo Kimand Hyun Kim (2013). *International Journal of Applied Geospatial Research (pp. 37-57).*
www.irma-international.org/article/spatial-temporal-analysis-of-recent-air-passenger-flows/95193

A Collaborative Virtual Geographic Environment: Design and Development
Jianhua Gongand Hui Lin (2006). *Collaborative Geographic Information Systems (pp. 186-207).*
www.irma-international.org/chapter/collaborative-virtual-geographic-environment/6659

An Assessment of Several Taxonomies of Volunteered Geographic Information
Antony K. Cooper, Serena Coetzeeand Derrick G. Kourie (2013). *Geographic Information Systems: Concepts, Methodologies, Tools, and Applications (pp. 1794-1808).*
www.irma-international.org/chapter/assessment-several-taxonomies-volunteered-geographic/70535