Chapter 22

New Discovery Methodologies in GIS:
Improving the Information Retrieval Process

Nieves R. Brisaboa
University of A Coruña, Spain

Miguel R. Luaces
University of A Coruña, Spain

Diego Seco
University of A Coruña, Spain

ABSTRACT

In the last decade, the availability of on-line resources, and also the number of users accessing those resources, has grown exponentially. The information retrieval process, which aims at the improvement of the access to such resources, has been the focus of interest of many researchers. The presence of geographic data in these repositories of information is surprisingly high (for example, note that most of the web pages about business contain information about the locations of their offices). In order to properly manage this geographic data, the information retrieval process has been extended using architectures, data structures, and other techniques developed by the GIS community. This has meant the beginning of a new research field called Geographic Information Retrieval. In this chapter, the authors present a study of the state-of-the-art of this new field, and they also highlight the main open problems that will concentrate efforts during the next years.

INTRODUCTION

The need to manage information has been one of the key factors behind the consolidation of information technology as an essential driving force for the development of our society. Over the years, many system architectures, index structures, and other components have been proposed with the fundamental goal of allowing efficient access to information stored in huge document databases. The research field that focuses on this goal is called *Information Retrieval* (IR) (Baeza-Yates & Ribeiro-Neto, 1999) and it started with the work of Salton (1963). This research field has recently undergone a spectacular development motivated by the growth of the Internet and the need to search the Web. A very important feature of IR is that it deals with the problem of retrieving information by its content rather than its metadata. Thus, there are a number of techniques for retrieving documents of various types: texts, images, sound and video files, etc.

Textual information often includes *geographic references* in the text (for example, press releases usually mention the place where the events happened). Taking these geographic references into account provides added value to classic information retrieval systems. The research on *Geographic Information Systems* (GIS) (Worboys, 2004) has dedicated much effort to study the special features of geographic information and to develop systems able to use and take advantage of them. This field has received much attention in recent years due to recent improvements in hardware that have made possible the development of such systems by many organizations. In addition, two international organizations ISO (ISO/IEC, 2002) and the Open Geospatial Consortium (OGC, 2003) are undertaking a major collaborative effort to define standards and specifications to develop interoperable systems. At the European level, the INSPIRE (Infrastructure for Spatial Information in Europe) directive (European Commission, 2011) has enabled a breakthrough in the field of corporate GIS and it remarks the future importance of geographic information. Thanks to these initiatives, many public organizations are working in the development of spatial data infrastructures (GSDI, 2011) that enable them to share their spatial information.

These two research areas have progressed independently over the years. On the one hand, the index structures and techniques from the IR field do not take into account the spatial nature of geographic references that appear in text documents. On the other hand, spatial index structures are not directly applicable in information retrieval systems. However, users increasingly demand services that allow them to locate the information in its spatial context and even to access this information using queries that take into account the spatial information. These demands have caused that researchers in each area have began to pay attention to the other one resulting in a new research field called *Geographic Information Retrieval* (GIR). The aim of this field is to propose new system architectures, index structures, and other components in order to develop systems to *retrieve documents both thematically and geographically relevant in response to queries of the form <subject, place>*. An example of the type of queries studied in this new field is the following: "*Ph.D. dissertations regarding geographic information systems published in Spain.*" The reader familiar with classic information retrieval systems knows that the relevance of the documents in a textual search engine is based on the frequency of the words that appear in the text of the documents. Therefore, if the word *Spain* does not appear explicitly in a document its relevance will be low with respect to this query. This happens even if the word *Madrid* appears in the document (or any other autonomous region, province or city of Spain) because traditional IR systems are not prepared to take into account the special characteristics of the geographic information space (e.g., the *contained by* spatial relationship between Madrid and Spain). Query expansion techniques in clas-

Related Content

A Theoretical Comparison of Traditional and Integrated Project Delivery Design Processes on International BIM Competitions
Michael Serginson, George Mokhtarand Graham Kelly (2013). *International Journal of 3-D Information Modeling (pp. 52-64).*
www.irma-international.org/article/a-theoretical-comparison-of-traditional-and-integrated-project-delivery-design-processes-on-international-bim-competitions/105906

Reasoning about Space, Actions, and Change: A Paradigm for Applications of Spatial Reasoning
Mehul Bhatt (2012). *Qualitative Spatio-Temporal Representation and Reasoning: Trends and Future Directions (pp. 284-320).*
www.irma-international.org/chapter/reasoning-space-actions-change/66763

Effects of Human Culture Among AEC Professionals Towards Adaptation of Collaborative Technology in Industrialized Project Delivery
Maszura Abdul Ghafarand Rahinah Ibrahim (2020). *International Journal of Digital Innovation in the Built Environment (pp. 36-48).*
www.irma-international.org/article/effects-of-human-culture-among-aec-professionals-towards-adaptation-of-collaborative-technology-in-industrialized-project-delivery/255178

Leveraging the Science of Geographic Information Systems
Rick Bunch, Anna Tappand Prasad Pathak (2013). *Geographic Information Systems: Concepts, Methodologies, Tools, and Applications (pp. 1612-1618).*
www.irma-international.org/chapter/leveraging-science-geographic-information-systems/70525

The State of the Art of Bridge Information Modelling from Conceptual Design through to Operation
Aonghus O'Keeffe (2014). *International Journal of 3-D Information Modeling (pp. 29-39).*
www.irma-international.org/article/the-state-of-the-art-of-bridge-information-modelling-from-conceptual-design-through-to-operation/106850