

Automatic Georeferencing and Search on Structured Georeferenced Bibliographic Data

Dragan Ivanović and Marjan Čeh

University of Novi Sad, Serbia
University of Ljubljana, Slovenia
`dragan.ivanovic@uns.ac.rs`,
`marjan.ceh@fgg.uni-lj.si`

Abstract. Georeferencing means associating something with locations in physical space. Geospatial data joined with some structured data can improve searchability of those data. The main open issue described in this paper is whether georeferencing of structured data can be performed automatically using NLP, automatic extraction and data mining techniques.

Keywords: georeferencing, automatic extraction, NLP, Google Maps JavaScript API

1 Introduction

The main objective of the KEYSTONE action is to establish a cooperative network of researchers, practitioners, and application domain specialists working in fields related to semantic data management, the Semantic Web, information retrieval, artificial intelligence, machine learning and natural language processing. Coordination of collaboration among the action members should enable research activity and technology transfer in the area of keyword-based search over structured data sources and should promote the development of a new revolutionary paradigm that provides users with keyword-based search capabilities for structured data sources as they currently do with documents.

Georeferencing means associating something with locations in physical space. GEOPOLO is a software system for georeferencing scientific-research publications, i.e. it is a system which enables joining geospatial data to scientific-research publications. This system is implemented for theses and dissertations defended at the University of Ljubljana (<http://geopolo.fgg.uni-lj.si/>). Currently, librarians manually add geospatial data to publications. Those georeferenced data about publications can be used to visualize distribution of publications by geographic regions. On the other side, those data can be also used to perform search by geographic data. Currently, librarians manually add geospatial data to publications which is a weak point of the system. Automatic extraction of toponyms could significantly improve the system and search of the stored data.

2 Challenges

Besides shared challenges in this field for all Geographic Information Systems (GIS), there are a few specific challenges which can be important for the KEYSTONE action:

1. How NLP and data mining techniques can be used in order to find out toponyms (geospatial keywords) from a structured data source in order to enable automatic georeferencing of those data?
2. Dealing with toponyms which represent homonyms - Moscow is the capital city of Russia or a city in Kansas?
3. Can Google Maps JavaScript API or some similar API be used for automatic georeferencing keywords from a query in order to enable keywords-based search on georeferenced structured data source even though the certain keyword from the query is not mentioned in some data which geospatial data are close to or are in relation superset-subset with geospatial data of the certain keyword?

3 Approaches, Methods, Techniques, Scientific fields

Automatic Georeferencing and Search on Structured Georeferenced Bibliographic Data is interdisciplinary open issue. This issue should be resolved by individuals with expertise in information retrieval, NLP, data mining, manual and semi-automated georeferencing, GIS, collection management, software engineering, protocol standards, data standards, etc.