Georeferencing Works of Literature

Franc Solina, Robert Ravnik
Faculty of Computer and Information Science
University of Ljubljana
Tržaška cesta 25, 1000 Ljubljana, Slovenia
E-mail: franc.solina@fri.uni-lj.si

Abstract. *A pilot web application for collaborative georeferencing of Slovenian works of literature was developed. Problems in georeferencing in general and georeferencing works of literature are discussed. The pilot system can be a model for establishing web sites for georeferencing literature for different languages. The application is based on Google Maps and implemented using PHP, JavaScript and MySQL. The intended users of the system are mainly lovers of literature, travelers and pupils.*

Keywords. Literature, georeferences, KML, collaborative work, web application

1. Introduction

Recently, we can witness a wealth of georeferenced information sources on the Internet which are all the more valuable if they can be accessed on handheld devices such as smartphones. Based on our location we can find restaurants, gas stations, bus stops, automatic banking machines and other sites of interest displayed on a map centered around our position or listed according to the proximity to our standpoint. Georeferencing is in general a good way to make sense of the huge amount of information available on the Internet. Traditionally, web resources were organized into hierarchically organized pre-defined categories but many types of information that we seek on the Internet are closely connected to a physical location in the world. Dedicated navigation devices that can just display our position on a map and give directions to a point are being supplanted by smart devices that can get from the Internet also a variety of other georeferenced information.

One particular kind of information that has so far not been widely georeferenced using modern technology are works of literature although the relationship between written literature and maps seems to be present since the beginning of literature. In the Iliad of Homer we find geographical and cartographic elements. The "ship's catalogue" has been defined as an essay of regional geography of Greece. In the long description of the forging of the weapons made for Achilles, Hephaestus, the god of metals, seems to be also a skilful cartographer.

Most fiction be it novels, short stories or even poetry relate to some geographical location, such as a country, landscape, city or even down to a specific locality or a building. The information necessary to determine a location in a work of literature is sometimes given directly in the works itself, typically in travel books, biography or historic works, often it is more veiled and must be deducted from other circumstances. The information could be found for example in biographical works of the authors or other reference works. Actual places are often a template for places in a work of fiction. In the case of Homer's Odyssey tons of books have been written that try to determine the actual locations and course of his travels, an excellent recent example can be found in [1]. The home ports of major contingents named in the Homeric Catalogue of Ships (Iliad, II.494-760) and locations mentioned in some latin poems can be seen on Google Earth [2]. James Joyce's novel Ulysses chronicles the passage of Leopold Bloom through Dublin during an ordinary day, June 16, 1904. A website dedicated to this famous novel follows Bloom's peregrinations through Dublin street by street [3]. Travel agencies organize trav-

els to visit sites mentioned in Jane Austen's books is just another example of public's interest in literary heros.

To see the geographical locations related to a work of literature on a digital map or to see which works of literature are related to specific geographical locations would be valuable information for travelers as well as for students and researchers of literature. These geolocations could offer further links to information on authors and their works, for example links to related entries on the wikipedia, links to free [4] or payable e-books, and links for purchasing books in online bookstores.

Some further examples of literature links on web maps are for example the Map of Kansas literature [5], Birthplaces of Slovenian writers and poets [6], and Birthplaces of Slovenian writers of historical novels [7]. The last two examples offer for each author further web links, for example to author's pages on Wikipedia or links to their e-books on Wikisources and other open e-book repositories. Birthplaces of authors born before the 20th century shed in general also some light on the contents of their works of literature. Due to higher mobility of people later in time this link between their birthplace and the contents of literature written by an author is much looser. Much more useful is to try to link the actual works of literature to their corresponding geographic locations.

This article describes a pilot project to build a web portal for georeferencing works of literature using all currently available tools and environments. In Section 2 we describe first general issues in georeferencing, then issues specific to georeferencing works of literature. In Section 3 we define the requirements for the project. Section 4 describes the implementation and gives some preliminary results. Section 5 concludes the article.

2. Issues in georeferencing

A georeferenced resource is a piece of information that is relevant to an identifiable subset of the Earth. Any kind of document that is relevant to a particular geographic place is an example of a georeferenced resource. Georeferencing was initially concerned primarily with resources such as maps, aerial photography and satellite imagery whose content is distributed across the Earth.

An early and influential initiative for creating a georeferenced digital library was the Alexandria Digital Library (ADL) Project [8] which started in 1994. The project went through several stages and identified several key issues in building and developing georeferenced digital libraries. The first two and most important issues are how can georeferenced resources be discovered and gazetteer integration. The traditional approach to discovery of georeferenced resources in digital libraries is to use text-based information retrieval techniques. This would mean basing discovery on matching of textual place names. Problems that need to be addressed with this approach is that place names change in time due to political and historical changes [9]. The same city can have different names in different languages, for example Trst in Slovenian and Croatian, Trieste in Italian and Triest in German. Spelling of a particular geographic name in the same language can also change over time. Some types of place names, especially names of streets, are particullary prone to change after large political upheavals. For example, when searching for references for Ljubljana, the capital city of Slovenia in Google books, one has to try beside the modern spelling at least the following searches: Luwigana, Laybach, Laibach and Lubiana. Place names can exist also in duplicates that is the same name can refer to different geographical locations. There is on the other hand a whole range of georeferenced resources that have no associated place names, mostly imagery gathered with sensors which carry only the source metadata such as the coordinates of the camera obtained automatically with the built-in GPS sensor. Coordinates are necessary also for display of georeferenced sources on a map. Matching place names with a specific coordinates is therefore important in both ways. For a given place name we need the coordinates or a range of coordinates to be able display the place name

on a map and for a given coordinates we need to know a place name that is the most relevant description in terms of human spatial cognition. A gazetteer is a special dictionary that supports the translation between place names and physical coordinates on the Earth's surface [10].

The problem of georeferencing can be better understood in the general problem of accessing and managing the explosively growing multimedia content. Extensive research efforts have been dedicated to automatic multimedia analysis and processing in the last decades, such as categorization, annotation and indexing. Despite huge efforts in the past decades it is evident that the semantic gap cannot be solved automatically in the near future. People must be included in the loop and the power of collective human efforts on the Web shown excellent results (i.e. Wikipedia) and that combining human intelligence and automated computer processing can solve the problems of the semantic gap in a collaborative manner (i.e. the ESP game that helps Google in image search).

Georeferencing works of literature typically starts with text-based information. Due to various efforts the share of digitized works of literature is increasing and therefore opens the possibility of using automatic retrieval techniques. Complete automation, however, is for works of literature not possible. First, because georeferences are in works of literature often hidden, sometimes obscure and an object of biographical and literary research. Second, because works of literature are often set in a historical context bringing up all the above mentioned problems of spelling, changing and use of several place names concurrently. Therefore we believe that the only practical way of addressing the task of georeferencing works of literature is in using automatic tools as much as possible but to rely on human selection and interpretation of computer results.

3. Requirements

Our application for enabling georeferencing of literary works should be accessible and simple to use. This is nowadays achieved by a web application that requires just a browser on a personal computer or a smart phone and connection to Internet. Our goal is to provide a service similar to Panoramio where users can upload their georeferenced photos. Users can see their own photos as location tags on Google Earth and/or Google Maps. Some photos are selected according to some qualitative criteria by an editorial team for inclusion into Google Earth and a subset also into Google Maps to be seen by all users of Google Earth and Google Maps. Instead of photos a user would in this application provide a short description of a literary work including primarily links to other web resources related to the work such as wikipedia pages of the work, its author, to the digitized version of the work (in public domain or in payable versions), to the work available in online bookstores etc.

Each description is georeferenced using the Keyhole Markup Language (KML). The KML file (Fig. 2) specifies a set of features (place marks, images, polygons, 3D models, textual descriptions, etc.) for display in Google Earth, Maps and Mobile, or any other 3D earth browser (geobrowser) implementing the KML encoding [11]. The coordinates must be provided by the user by interactively selecting them in Google Earth or Google Maps. Google Earth/Maps strives to match coordinates for the most current place names so that any older place names currently not used anymore must be put into the correspondence with the modern ones by the user.

Beside individual locations a particular literary work could be connected also to a geographical path. Travel works, biographies and any other works with a specific time line could benefit from such geographical representation.

In addition to georeferencing entries could be also time referenced so that by changing some display parameters only entries for a particular time period could be seen. There should be two different time periods, first is determined by the time the entry was written, that is by the time period when the au-
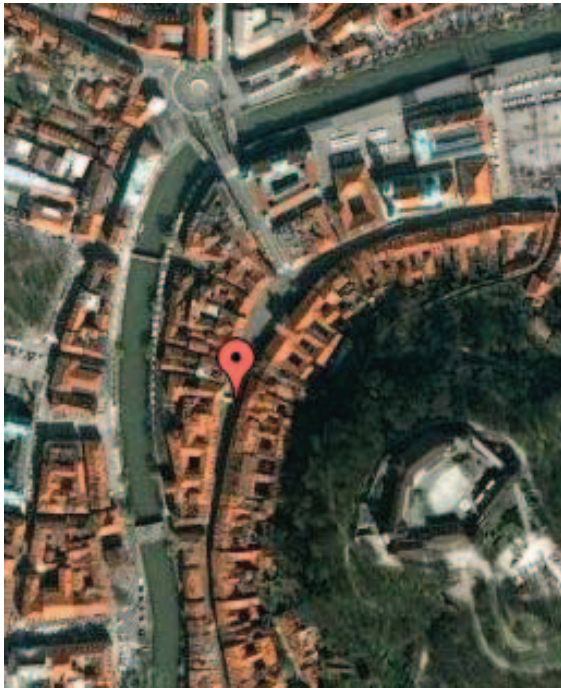
Figure 1: The precise location (Stari trg) in Ljubljana where the first slovenian ballad published in 1830 written by France Prešeren was supposedly taking place.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://earth.google.com/kml/2.1">
  <Document>
    <name>France Prešeren: Podvodni mož</name>
    <description>First slovenian ballad published in 1830
      written by France Prešeren.
    </description>
    <Placemark>
      <name>Prešeren, France: Povodni mož</name>
      <description>Na Starem so trgu pod lipo zeleno
        trobente in gosli, in cimbale pele,
        plesále lepote 'z Ljubljane so cele
        v nedeljo popoldan z mladenči vesele;
        bila je kraljica njih Urška brhka,
        plesáti ni dolgo nje volja bila.
      </description>
      <Point>
        <coordinates>14.5063, 46.0476</coordinates>
      </Point>
    </Placemark>
  </Document>
</kml>
```

Figure 2: KLM encoding for location in Fig. 1. Included is the part of the ballad with the georeference to Stari trg in Ljubljana.

thor lived, the second optional one by the purported time when the narrative in the work is taking place.

Implementation for popular hand held devices such as Apple iPhone will be developed. Beside the map-based view of links a list of links ordered by the proximity to the user's stand point.

Finding georeferences in works of literature can be an interesting hobby or a useful homework for pupils. The input of information into the system can therefore be most easily achieved by a collaborative effort. Initially the application would require an administrator who would approve entries, later with a larger community a wikipedia like community based administration could be established.

Literature is closely connected to languages and not all literature is available in translation and even if it is translated then often only to some major languages. Therefore, in practice, each language with a substantial body of literature would require its own instantiation of the system which would generally be interconnected, much like wikipedia pages in different languages.

4. Implementation and preliminary results

We are implementing a demonstration system for georeferencing of Slovenian literature works. This pilot system makes georeferences to works which are in the public domain and available on the server `http://lit.ijs.si/`.

A user makes a georeference to a literary work using Google Maps where he determines a location on the map (Fig. 1). The coordinates of a location can be exported with a Google Maps API. In a data entry window the user than writes the contents with possible further web links The data is stored in a MySQL database. Entries including the coordinates can be saved in KML format (Fig. 2). The application is implemented using PHP and JavaScript.

5. Conclusions

We argue for a web based application that would enable a collaborative user based georeferencing of works of literature. Based on the experience of the described pilot project for Slovenian literature more

complete requirements for such project could be defined. The project could serve also as the visualization part for automatically derived georeferenced information using computer assisted text analysis [12].

References

[1] R. Bittlestone, J. Diggle, and J. Underhill. Odysseus Unbound: The Search for Homer's Ithaca. Cambridge University Press, 2005.

[2] Geolocating Literature using Google Earth/Maps, *www.haverford.edu/classics/* courses/omni/geolocation.php

[3] The Internet Ulysses by James Joyce, *www.robotwisdom.com/jaj/ulysses/*

[4] Project Gutenberg, *www.gutenberg.org*

[5] Map of Kansas literature, *www.washburn.edu/reference/cks/ mapping/index.html*

[6] Rojstni kraji slovenskih literatov, *www.geopedia.si/#T105*

[7] Zgodovinski romanopisci, *www.geopedia.si/#T944*

[8] G. Janee, J. Frew, L. L. Hill. Issues in Georeferenced Digital Libraries. D-Lib Magazine 10(5). Retrieved from *www.dlib.org*

[9] D. A. Smith, G. Crane. Disambiguating Geographic Names in a Historical Digital Library. In P. Constantopoulos, I. T. Solvberg (Eds.): Proc. 3rd European Conf. on Research and Advanced Technologies for Digital Libraries, LNCS 2163, pp. 127–136, 2001.

[10] L. L. Hill. Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints. In J. Borbinha, T. Bakerf (Eds.): Proc. 4th European Conf. on Research and Advanced Technologies for Digital Libraries, LNCS, pp. 280–290, 2000.

[11] S. Udell. Beginning Google Maps Mashups with Mapplets, KML, and GeoRSS: From Novice to Professional. Apress, Berkeley, 2009.

[12] M. Stubbs. Text and Corpus Analysis: Computer-Assisted studies of language and culture. Blackwell, 1996.