PERFORMANCE EFFICIENCY VERSUS PRIVACY PROTECTION IN REVERSE GEOCODING

Geocoding is the process of assigning geographic coordinates to locations. The locations can be described by a name or an address and when geocoded, a specific set of latitude and longitude is assigned to them. Based on this, the process of reverse geocoding is the extraction of the textual information from geographic coordinates, namely the name or the address of the location. Studies have investigated the positional accuracy error caused by automated geocoding. In particular the error has been investigated for different geocoding services or geocoding methods and for different areas as well. The main outcome of these studies is the geocoding success rate. Success rate provides awareness to potential researchers about the level of error which may affect the spatial analytical results of their projects.

Though geocoding and reverse geocoding techniques are common in many geo-application scenarios, e.g. in freely available online-based mapping services, previous studies have focused only on the accuracy of the geocoding process. Due to the extensive use of both techniques it is critical to evaluate the reverse geocoding technique as well. Furthermore, with reverse geocoding confidentiality issues come into play; for instance if personal data is mapped, such as a crime database, reverse geocoding facilitates the way in which private information can be accessed.

This work provides an analysis of several freely available web mapping services in regard to reverse geocoding accuracy, usability and privacy aspects. Ten non-profit services are selected and analyzed in terms of their usability based on a defined scoring system. For the scoring system eight objective criteria for the qualitative aspects of use are identified and therefore no user survey is necessary. In addition to that, the textual accuracy of the services when using the reverse geocoding tool is measured. The study area is located in Vienna the capital city of Austria and the dataset is a random sample derived from the crime database of the Austrian police. The accuracy error is measured in three neighborhoods of different residential densities. The scope here is to reveal if any error variation is associated with the underlying residential structure. The last step of our analysis is to show the risk of compromising personal information. For that reason, a freely web-based application is used that allows programming to define specifications of the search engine. These specifications to the programmed search engine return associated names for the given street addresses. Finally, the retrieved names are calculated as probable identities linked to the crimes.

The outcome of this work shows the level of difficulty for users to precisely derive real addresses from discrete point data. Furthermore, by comparing the examined services we identify good practices for better usability and accuracy. The results are also discussed in terms of geoprivacy issues and the risk of compromising personal information is addressed. The project concludes with suggestions on how to avoid the locational privacy disclosure. Moreover, outlines the need to apply masking cartographic practices when sensitive point data are displayed on maps.