# Spatial delineation of geographic objects using an online knowledge repository

Places or regions may have an obvious and definite boundary, e.g. land parcels. But, there are also areas with fuzzy boundaries, e.g. mountain ranges. For certain use cases it is necessary to define also for those areas a precise boundary, e.g. for the application of mouse over interaction. The existence of maps showing the boundaries of landscapes (such as the map *Landschaften* of *BKG*) also shows the need for well defined boundaries of geographical objects.

We present a method to delineate geographical regions by the usage of the online knowledge repository Wikipedia. The method is based on the coordinates that are related to the articles. Additional, the coordinates are classified according to the link and page structure. The resulting point cluster is used for an estimation of the location of the regions.

Articles in Wikipedia concerning geographical objects (geo-) reference these objects with single points (footprint). As shown by (Hecht and Raubal, 2008) and (Dahinden, 2009) there is a relation between the spatial extend of an object and the footprints of the linked articles in the Wikipedia article about this object. The existence of a footprint in a linked article is an indication or rather an accidental occurrence that represents the object under investigation.

By comparing the distribution of the footprints (of the linked articles) of several objects, the dominating object at a certain position can be evaluated. By analysing the relations between the articles, the distribution of the footprints may be refined and the estimation of the location of the object may be improved.

The link from one article to another expresses a relationship between these articles, they are neighbours. The relation is not necessary spatial but may be semantically. Two neighboured objects my build a horizontal or vertical neighbourhood. Horizontal means that the objects use a comparable level of granularity (e.g. counties). Vertical indicates that one object is inferior/superior to the other (e.g. a capital city and a country).

An analysis of the links of all articles shows that some articles build up cliques. This means, each article in the clique links to each other. These cliques represent a kind of neighbourhood. By comparing the structure of the articles in a clique it is possible to determine the type of neighbourhood (horizontal, vertical, and other).

To determine the location of a certain object we collect the footprints of inferior objects and compare their distribution with the distribution of footprints of inferior objects of the horizontal neighbours. On these distributions, we apply home range estimation methods, such as Delaunay-triangulation, Thiessen-polygons, kernel density estimation and so on.

The analysis was calculated using several language versions of Wikipedia. Because there are obvious differences in the content in these versions, the results differ. A discussion of these differences is necessary.

Dahinden T (2009) Localization of uncertain and fuzzy-bordered areas by geocoded articles of a knowledge repository. In: Garcia-Huidobro CJV (eds) Proceedings of the International Cartographic Conference, Santiago de Chile.

Hecht B, Raubal M (2008). GeoSR: Geographically Explore Semantic Relations in World Knowledge. AGILE Conf., Springer: 95-113.