

Individual Theme Route Planning on User-generated Content

Karsten Pippig*, Dirk Burghardt, Nikolas Prechtel

*Institute of Cartography, Technische Universität Dresden,
Helmholtzstr. 10, 01062 Dresden, Germany*

**Corresponding author, email: karsten.pippig@tu-dresden.de*

In the recent years, several approaches have been developed concerning an automated touristic tour planning and a determination of theme-related locations based on user-generated content. But yet (to the authors' knowledge) there is no approach which enables an individual thematic route planning in a way that the user can define a start plus destination point and gets the most thematically relevant route in return.

We suggest that an individual thematic route planning could be ensued by a semantic similarity determination between a user specified Wikipedia concept and other georeferenced Wikipedia concepts. The semantic similarity determination is carried out via the *Explicit Semantic Analysis* (ESA). We introduce two methods how we can computed theme-related routes on a navigation graph derived from road data of Openstreetmap according to the semantic similarity values.

The first method is called continuum method. By using the kernel density estimation, we derive pseudo density continua of geo-tagged photos from Flickr and Panoramia which are tagged or titled like a geo-referenced Wikipedia concept. Hereby, we can determine a perceptive extent of any geo-referenced Wikipedia article within a test area. The continua can be multiplied with the semantic similarity values of the related Wikipedia concept to derive theme-related continua. These ones, in turn, can be merged to a theme continuum defining the (perceptive) spatial extent of a specific theme. This theme continuum is the basis for the derivation of the impedance values of the navigation graph.

The second method (spot sequence method) is based on a ranking of the semantic similarity values of Wikipedia articles. The articles are ranked according to relevance of the theme. Starting with the highest ranked article, all articles with their geospatial location are iteratively inserted into the routing as intermediate points. The optimal sequence of intermediate points is computed in accordance with the traveling salesman problem. If the examined route length within each sequence exceeds a maximum user-defined length, the last inserted intermediate point is excluded from the computation. To minimize the computation time, the user can also define a threshold value for the theme relevance. This threshold represents the minimum correlation value between a theme and a geo-referenced Wikipedia article. It defines whether an article will be considered in the computation or not.

An evaluation of both methods has shown that the continuum method provides routes which are closer to the shortest route. The impedance values for the route computation can be modified in a way that the route is guided through theme-related areas nearby the shortest path. For the determination of the impedance values, a power function with high a scaling parameter has to be used as a base. The routes generated by this method are particularly suitable for users who want to visit as many theme-related sites along a direct path, without any guarantee to see the most relevant places. However, the spot sequence method allows the determination of routes, which consider the

most thematically relevant sites by specifying a maximum route length. The computed routes differ significantly from the shortest routes. These routes are particularly suitable for users, in which the thematic aspect is superior. By combining both methods, the route determination can be optimized with respect to the thematic relevance. Here, the theme-relevant sites will be connected through theme-related pathways.