Tag-Cloud Based Location Description

Tobias Dahinden, Daniel Eggert, Juliane Mondzech, Monika Sester
Institut für Kartographie und Geoinformatik, Leibniz Universität Hannover

Initial situation

A typical application of location based services is to provide additional information about a certain place to a user. A possibility to display such information is the use of Tag-Clouds. They can be generated e.g. using the tags and coordinates of Flickr or the articles with coordinates of Wikipedia. But experiments show that mainly superior administrative units are displayed in the Tag-Cloud (cf.[1]). If someone is moving inside such a unit the information becomes stale. Thus we currently develop a system that overcomes this problem.

Current Prototype

To test this approach we developed an evolutionary prototype of this system. The prototype uses an *Android* mobile device (Linux OS, Java runtime environment, and an integrated GNSS-receiver) and Wikipedia. The current position of a user is received from the GNSS or given with a touch on the screen using Google-Maps. Afterwards the user has to declare a range of interest.
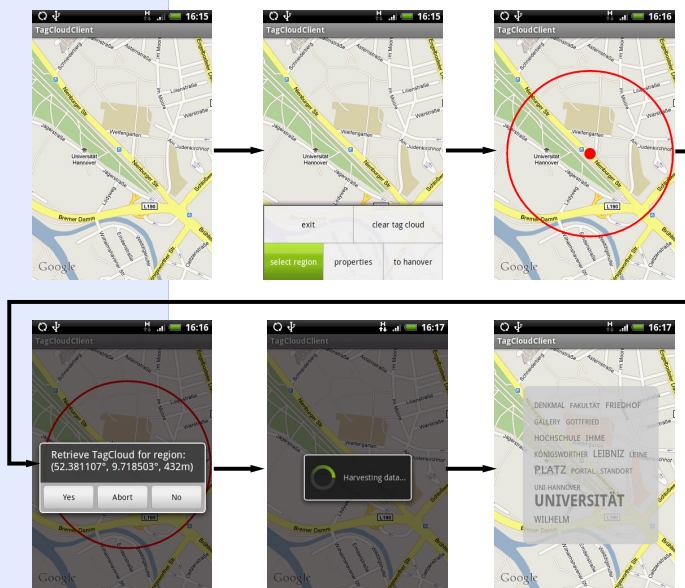

Figure 1: Screenshots of the Android phone during the creation of a Tag-Cloud.

In the next step all articles from Wikipedia in the search range are determined. This can be done with a request in the Wikipedia-World gazetteer. All these are requested from Wikipedia and processed. For this processing we use OpenCloud Library. Figure 1 contains the screen sequence of this process.

Filtering of frequent words

In the Tag-Cloud we only want to display characteristic expressions about a location. Thus we have to filter useless words. For the filtering process we determined the occurrence of each word in Wikipedia (O_g). It contains the typical distribution dominated by stop words (the, a, of, and, …)

This leads to a first list of (un)important words in articles. A similar distribution is obtained if we determine the occurrence of each word in articles with coordinates (O_k). If we determine the normalized relation of O_k to O_g a second list is obtained. It contains popular expressions in articles with coordinates (Tab. 1), e.g. in the German Wikipedia the word "Einwohner" (inhabitant) is 4.6 times more often in articles with coordinates than in all articles.

Wort	Anzahl	Verhältnis
Einwohner	15087	4.61
Bürgermeister	10806	4.50
Fläche	10479	4.23
Gemeinde	30171	4.02
Wappen	19488	3.90
Höhe	14893	3.80
Fluss	10626	3.71
Stadt	67570	3.44
Ort	22729	3.43
...		
so	12426	0.668
sie	20800	0.661
man	10792	0.626
ISBN	14340	0.619
dass	17543	0.582
werden	29905	0.582
nicht	26500	0.552
oder	19623	0.399
er	17288	0.378

Table 1: Popularity of words (1st column) in articles with coordinates occurring more than 10,000 times (2nd column). The 3rd column is the ratio of the occurrence of the word in articles with coordinats compared to the occurrence in all articles.

To avoid the problem of detecting superior administrative units we use the principle of negative feedback. This can be done by determing the Tag-Cloud of places in the neighbourhood and weight the terms of these Tag-Cloud negative in the Tag-Cloud of the place under examination (Fig. 2).
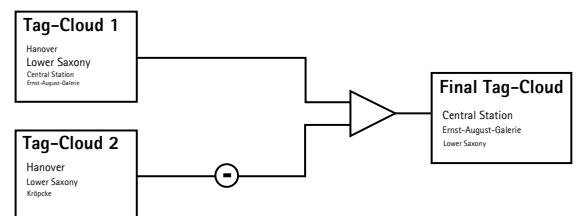

Figure 2: Schema of the lateral inhibition for two Tag-Clouds of nearby locations.

Future Work

The design of the Tags as a collection in a Tag-Cloud is very bold. Maybe it could be useful to present the Tags in a same kind like points of interest. But we have to keep in mind that the expressions are abstract and not directly related to a certain place. On the other hand points of interest shown on the map should be excluded from the Tag-Cloud. Concerning route description, it should be possible to describe where the user is driving through, in a manner like "Now, you drive through Annenriede (Delmenhorst)."

Further, we have to concern about the distribution of articles. While the method works quite well for Europe or the East of USA there are much too few articles about the rest of the world (Fig. 3).
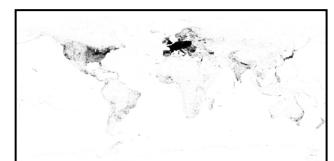

Figure 3: Distribution of articles with coordinates in Wikipedia.

Literature

[1] Hollenstein L Purves RS (2010) Exploring place through user-generated content: using Flickr to describe city cores. In: JOSIS (in discussion)