Mobile Visualization of Harvested Data to Improve Context Awareness

V. Paelke, T. Dahinden, D. Eggert, J. Mondzech

IKG, Institute for Cartography and Geoinformatics, Leibniz Universität Hannover, Appelstr. 9a, 30167 Hannover, Germany, {paelke, dahinden, eggert, mondzech}@ikg.uni-hannover.de

Abstract:

People often only see what they explicitly look for. To provide mobile users with information on the spatial context of a location or a route we present an approach that gathers context information from freely available sources like Wikipedia and creates visualizations of this data that provide users with cues to increase awareness of their spatial context. By combining web 2.0 technologies and public data sources with filtering and visualization techniques we exploit the browsing capability of humans to provide a service that increases location awareness at arbitrary locations. The approach makes it easy to integrate "authored" content and can incorporate the ever increasing amount of available geo-referenced information.

Keywords: Visualization, tag clouds, location based services, mobile devices.

1. Motivation

Advances in mobile computing and wireless communication technology enable the creation of location based services (LBS) on a variety of mobile devices. As the technical prerequisites for location based services become widely available the creation of the content required for them becomes an increasingly important question.

One possible approach is the development of specific authoring tools that support the creation of location based multimedia content and their integration into user interfaces that consider the constraints of mobile devices.

In this paper we consider a complementary approach: Information is gathered from freely available sources like Wikipedia and a combination of automatic filtering and processing techniques with information visualization techniques based on tag clouds are used to provide the user with a display of the available information.

2. Concept

As illustrated in figure 1 our system starts either by acquiring the spatial position of the user or by planning a route through the environment. Current smartphones feature a GPS unit that can be used for the localization. Using this position it becomes possible to supply the user with information on his spatial surroundings. This functionality can also be exploited to aid pedestrian navigation, by using the context data to improve route descriptions.

The second step of the processing pipeline uses a spatial query to search for spatially referenced articles located in the area. To influence the process the user can interactively modify the query radius in which information should be retrieved. It thus becomes possible to adjust the amount of data to the preferences of the user.

At this stage it is also possible to introduce additional "authored" information into the process. In some cases it is desirable to provide authored or preselected information for specific areas. These are simply added to a spatial data-based and retrieved in the same fashion as the articles. If desired they can be flagged as "authored content" and prioritized in the following processing steps.

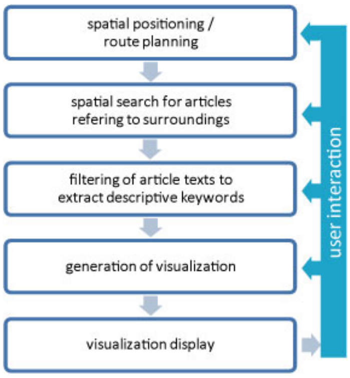

Figure 1: System processing workflow

In the third step the text from the retrieved articles is filtered to remove irrelevant words. Usually verbs and articles are less descriptive than nouns and therefore removed. The filtering can be adjusted by the user and specific filter lists can be supplied to suppress unwanted words from the tag cloud.

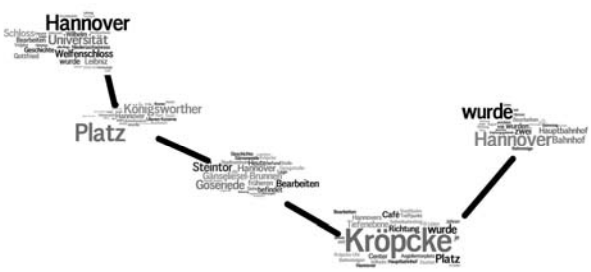

Figure 2: Route visualizations with tag clouds

In the fourth step a tag cloud visualization is created from the filtered text. The user can influence this step by adjusting font type and size, layout criteria and the number of tags to be displayed. A smaller number of tags is easier to browse and therefore the default setting. If a user wants to investigate more deeply or if the information provided in the display is not sufficient the user can choose to display more tags.

The final step is the display of the tag cloud. To improve readability the user can zoom and scroll the display. This is especially useful for route visualizations. In addition the user can “drill down” into the tag cloud to retrieve more detailed information on tags of potential interest: If the user selects a specific tag the system can provide him with a list of the original occurrences in the source articles. Alternatively a web-search for the tag could be initiated. While the tag cloud provides an overview visualization of information of potential interest, exploiting the user’s browsing capability to further filter the information, the possibility to link to the original articles can implement the second function of a LBS, namely to provide detailed information on points of interest in the environment.

3. Related Work

Our work was motivated by previous work on the development and use of conventional LBS using authored content. LBS like tourist guides [4,5] have become popular. Unfortunately, the production effort of such content is high and successful implementations are usually restricted to top tourist destinations.

Central to our approach is the exploitation of the user’s browsing capability using tag cloud visualizations. Tag clouds have become popular through their use on Web 2.0. sites, e.g. Flickr [1]. From a cartographic perspective it is interesting to note the relations between tag clouds and more structured lists of keywords. Lamantia [2] has pointed out the relation between structured lists and tag clouds on one side and maps and cartograms on the other. In recent years the study of tag cloud visualizations has become a research topic in information visualization [6,7]. Our approach relies on the availability of spatially referenced information in publicly accessible sources like Wikipedia [3].

4. Implementation

In our current implementation we focus on the use of smartphones as the hardware platform to access the location based information in a mobile use situation. Compared to complete route visualizations (as illustrated in figure 2) the display space on these devices is severely limited. It is therefore necessary to combine the tag cloud display with mechanisms for zooming and panning.

Current smartphones with multitouch displays are well suited to implement such an approach. A second limitation is the limited processing power and storage space available. We have therefore implemented a client server approach in which the mobile device (the client) handles only the visualization of the tag cloud and the user interaction while the harvesting and filtering of the data is offloaded to a server that is better suited to handle large data volumes.

Our current client component is implemented on an HTC Hero smartphone using the Android SDK v1.5. The server component is also written in Java. Communication is achieved through the exchange of shared tag cloud objects, that are serialized and exchanged using the OpenCloud Java library.

On the client side we have extended the MapView and MapOverlay classes from the Android Google API to handle the visualization and interaction. The original MapView class provides a map view based on GoogleMaps and we use the MapOverlay class to render the tag cloud on top of the map. Various output filters (e.g. alphabetic or score-based ordering) can be applied to the tag cloud to manipulate the tags shown in the final display.

The server adresses four tasks:

(1) network communication
(2) harvesting data
(3) data filtering
(4) tag cloud creation

Network communication:
The network communication module established a TCP/IP connection. Following this the client node transmits the coordinates and region buffer values for the desired tag cloud. After the tag cloud has been calculated, the server sends it back to the requesting client.

Harvesting data:
The server module responsible for harvesting the data for the desired location/region first queries the locally stored Wikipedia World Place database with the given region. As the result all article names for this region of interest are retrieved.

Based on the article names the corresponding articles are then retrieved online using the Wikipedia API and harvested for keywords.

Data filtering:
Various filter mechanisms are supported to extract descriptive keywords from the retrieved articles, e.g. filtering by word length, word type, explicit blacklists or semantic filtering.

Tag cloud creation:
The next module creates a tag cloud from the previously harvested Wikipedia data. The results is a tag cloud object with a predetermined number of tags that is transmitted to the client for visualization.

5. Examples

The following examples show the approach applied to areas in the city of Hannover where reference data for points of interests (POI) was available for comparison. The system can be used in arbitrary locations, however our current filtering techniques are tailored to German texts and the system is therefore best used in areas that are well covered by content in the German wikipedia. An extension to wikipedia articles in other languages can be realized by adapting and extending the filtering algorithms correspondingly.

Figure 3 shows the selection of an area of interest. This function enables users to specify a location and a radius of interest on a Google map display. Alternatively the current location of the user could be used to specify the query area.

Figure 3: Area selection

Once the region of interest has been determined the client contacts the server and initiates the harvesting of data in this area. After processing and filtering the server returns a tag cloud object.

This tag cloud object is then rendered as an overlay on the map and displayed to the user. As illustrated in Figure 4 both individual tag clouds (e.g. to provide context information for the current location) or multiple tag clouds (e.g. to cover a route) can be displayed.

Figure 4: Visualization of tag cloud and route

6. Results and Outlook

Initial results and experiences with the system are encouraging. In areas well covered by spatially annotated wikipedia articles like the central areas of Hannover between 40 and 80 percent of the tags rendered are meaningful landmarks. Since tag clouds support fast browsing of information this is usually sufficient to provide users with pointers to potential objects of interest. Initial test users got quickly used to the system and were positive about the results delivered.

Even test users who knew an area well were often surprised by the detail of information available. The intention of the first prototype described in this paper has been to validate the viability of idea. In the future we aim to address several areas for refinement:

First, we aim to experiment with more advanced filtering algorithms to increase the signal to noise ratio in the tag cloud and to extend the system to other languages, notably English.

Second, we intend to extend the data used for harvesting tags to other sources of geo-referenced data beyond wikipedia, e.g. by incorporating Flickr tags.

We are also working on more refined rendering and adaptive layout strategies for the tag clouds. Other open questions include the best size, the optimal tag count and the best visualization style to combine the cloud display with maps.

Thus, while the use of public data sources seems to be promising to fill the gaps in spatial coverage in current LBS there remain many open questions. We have found that the use of tag clouds as a visualization technique that exploits the browsing capability of humans to add a further “filtering stage” is useful to make this content available to users but further refinements are required.

7. References

[1] Flickr, 2009: http://www.flickr.com/photos/tags/ (accessed 28 Oct. 2009)

[2] Lamantia, J. 2009: http://www.joelamantia.com/ideas/cartograms-tag-clouds-and-visualization/ (accessed 28 Oct. 2009)

[3] Wikipedia, 2009: http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Geographical_coordinates (accessed 28 Oct. 2009)

[4] Zipf, A., 2002. Location aware mobility support for tourists. Trends & Controversies. *IEEE Intelligent Systems*. Nov./Dec. 2002. pp. 57-59.

[5] Baus, J., Cheverst, K. and Kray, C., 2005. A Survey of Map-based Mobile Guides. *Map-based Mobile Services*, Springer, pp.197-213.

[6] Halvey, M. J. and Keane, M. T., 2007. An assessment of tag presentation techniques. *Proc. ACM WWW '07*, Banff, Canada, pp. 1313-1314.

[7] Rivadeneira, A. W., Gruen, D. M., Muller, M. J., and Millen, D. R., 2007. Getting our head in the clouds: toward evaluation studies of tagclouds. Proc. ACM CHI '07, San Jose, California, USA, April 2007, pp. 995-998.