Mobile Location-Driven Associative Search in DBpedia with Tag Clouds

Bjørnar Tessem, Bjarte Johansen, and Csaba Veres

Department of Information Science and Media Studies,
Postbox 7802, University of Bergen, 5020 Bergen, Norway
bjornar.tessem@uib.no bjarte.johansen@uib.no
csaba.veres@uib.no

Abstract. A primary contextual source for today's context-sensitive mobile phone apps is the user's location. The recent surge in the availability of open linked data can provide location-oriented semantic context, still wanting to be explored in innovative ways. In PediaCloud, the Android tool described here, we show how we can use the associative structure of the Semantic Web at a geographical location, visualize location information with tag clouds, and allow users to follow the associations of the Semantic Web enabled by the tag cloud, with the aim of enabling the users to construct an understanding of the "place" around them. The data we use are found through DBpedia, a project a project aimed to lift the information in WikiPedia into the Semantic Web.

Keywords: tag clouds, mobile, location, semantic web, DBpedia

1 Introduction

Exploiting location context has become a major theme for research on mobile technologies and is widely applied in commercial applications. These new technologies have implications for the realisation of the concept of *place*, which in sociology is understood as not only the location itself, but also the physical surroundings, and a persons attribution of meaning to the surroundings[4]. The location-based mobile tools guide the users to information about the surroundings, and enable new ways of constructing a user's understanding of place. Examples of such tools are Google maps and Google places, which in order to enhance the experience of *places*, connect to located information from among others Wikipedia, and its linked data extraction, DBpedia[1].

A potential technique for presenting located information are tag clouds, which are a means of visualizing natural language information. In a tag cloud a collection of words is drawn in a bounded area, each word with a font proportional to a weight computed from the text collection, often the count of that particular word. The size of the words gives the user an impression of what topics are important in the text collection.

With the PediaCloud tool we aim to show how one can create tag clouds of located information from DBpedia. The goal is to use not only located (primary) resources, but also to collect DBpedia entities that have some semantic link to the located resources

[1] http://dbpedia.org

(secondary resources), and build the tag clouds from this combined set of information. This should give users a richer sense of the semantic aspects of a location, the idea being that both primary and secondary resources contribute to a user's understanding of the place. It is also a goal to create a mobile semantic web application without a dedicated backend for the intermediate organisation of information, instead aiming to use DBpedia as an example of an existing semantic web resource that can be consulted directly.

2 The Workings of PediaCloud

The DBpedia data collected from Wikipedia articles contain features like location, abstract, categories, and other relations to many types of entities. For the PediaCloud application we use Wikipedia articles with a location, and their abstracts, and in addition the abstracts of the linked secondary resources. The collected abstracts are the source of the tag clouds we build. The construction of the tag cloud essentially goes through the following steps:

1. Get user location
2. Get all articles (primary) within fixed radius of location.
3. Get all articles (secondary) that the primary articles link to.
4. Find frequency of each word (wf) in each article's abstract.
5. Weight the wf of the words from the primary articles as a function of distance, scaling linearly with weight 1.0 at the user's location and 0.0 at the radius.
6. Weight the wf of the words from the secondary articles by the cosine similarity to the primary article it is linked from, multiplied by the weight of the primary article.
7. Add the weighted wf for each word together.
8. Select the highest n scoring words.
9. Use the word score to create tag cloud for current area.

Figure 1 illustrates how we weight the words in the tag cloud.

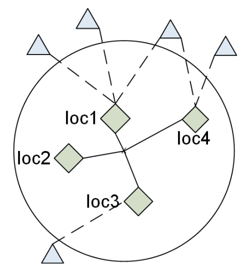
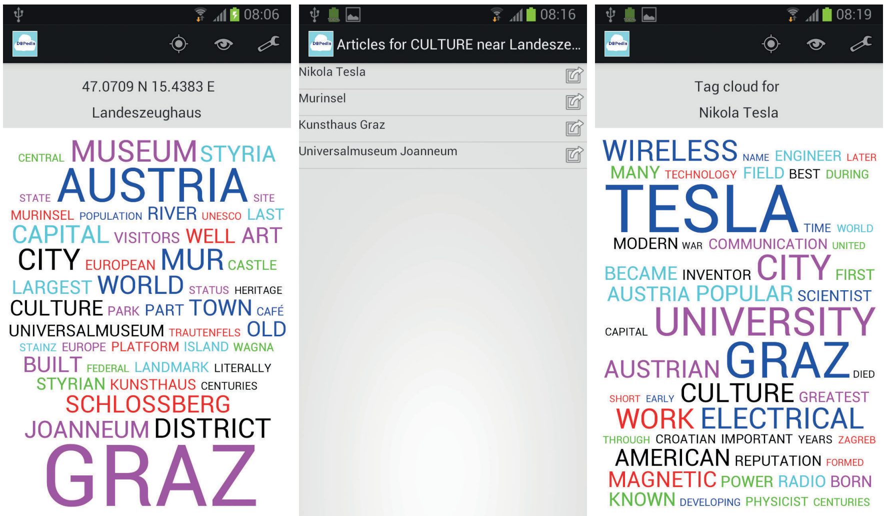


Fig. 1. The centre indicates the user's position, and the located articles are illustrated with diamonds at their locations loc1, loc2, ... Light gray triangles outside the circle indicate secondary resources. Located articles contributes according to their distance, secondary according to the distance of their located origin multiplied by the cosine similarity (indicated by dotted lines).



2a 2b 2c

Fig. 2. a: Tag cloud for a user located at the Graz Main Square. The closest located Wikipedia article is about Landeszeughaus. **b**: Ranked Wikipedia articles with the tag “CULTURE”. **c**: Tag cloud for “Nikola Tesla”.

The tag cloud for the Main square of Graz is shown in Fig. 2a. The user can select any word in the tag cloud, for example ”CULTURE”. This will show the list of WikiPedia articles that contain the word ”culture” (Fig. 2b). Now, the user may choose to look at the Wikipedia article for a resource, but may also select one of these articles as a focused resource (as an alternative to the user’s location) in a new tag cloud. This is shown in Fig. 2c where “Nikola Tesla” was chosen as a focal point. Note that the selected article no longer has a spatial component. We use the same collection of articles as for the previous tag cloud. However, the weight of the different abstracts will only be based on the cosine similarity to the focused resource. The word weights in the tag cloud thus depends on the choice of focused resource, giving the user a sense of what words are most prominent in the collection given this special focus.

The effect of using cosine similarity combined with word frequency to weight the words as opposed to using only word frequency is shown in Table 1, where we show the ten most prominent words and their weights in two tag clouds generated for the “Nikola Tesla” resource. We notice that words relating to Nikola Tesla’s achievements (ELECTRICAL, WIRELESS) get a place in the list when we use a cosine similarity approach as opposed to when we use word frequency only. The word frequency approach mainly results in an emphasis on general geographical words and nationalities. With the use of cosine similarity, larger fonts are given to words that are more informative.

The query we send to DBpedia’s sparql endpoint[2] is a single SELECT with UNION gathering data for both primary and secondary resources at the same time (steps 2 and 3 in the process). From the returned data the tool is able to compute the tag cloud for

[2] http://DBpedia.org/sparql

Nikola Tesla: Word frequency		Nikola Tesla: Word freq. * Cos.sim.	
GRAZ	57.0	TESLA	9.00
UNIVERSITY	44.0	GRAZ	6.78
CITY	30.0	UNIVERSITY	5.42
AUSTRIA	24.0	CITY	4.83
AUSTRIAN	21.0	ELECTRICAL	4.00
FIRST	16.0	WIRELESS	4.00
KNOWN	13.0	WORK	3.81
CROATIAN	13.0	CULTURE	3.60
CAPITAL	12.0	AUSTRIAN	3.32
WORLD	12.0	AMERICAN	3.23

Table 1. Weights and ranking for the 10 highest ranked words in a tag cloud for the same resource ("Nikola Tesla"), but with different weighting approaches.

the mobile screen. A positive effect of doing one single query is that it saves response time due to less network connection time.

We would have preferred to run the data gathering as a single CONSTRUCT in order to store all relevant triples locally, possibly on a triple store, but we ran in to a memory limit at the DBpedia endpoint when sending the query, so we had to go for the SELECT version. We considered installing a triple store with a query engine on the device to support the CONSTRUCT version, but discovered early that even though many of the triple stores and query engines are written in Java they are currently not working on the Android platform.

We are getting interesting results with the current implementation, but there are weaknesses that we would like to fix. One problem is that some words describing large enclosing areas (like "GRAZ" and "AUSTRIA") often get very high weights, but may not be very informative in the sense of getting a cultural and historical overview of a location. We believe that we can reduce the weight of these words by modifying the weighting algorithm we use, for instance by using tf-idf in conjunction with the already implemented cosine similarity.

3 Related Work and Conclusion

DBpedia has been used as a source of information in DBpedia mobile which is a tool presenting DBpedia resources close to the user at a map [2]. Ruta et al. [7] also use DBpedia data, and combine semantic similarity with location closeness to give DBpedia sources an overall match to a search criteria. MapXplore [9] is a tool that uses DBpedia to present classification and other factual data about points of interest, to users. MapXplore uses a category browser for locating relevant points of interest for users, in order to give them an overall impression of the important concepts at a place. For example, Bergen in Norway features prominently with the concept "mountain", whereas Dubai features with the category "skyscraper". van Aart et al. [8] use data from a variety of sources and connect non-located resources to a location through links to a located resource. Mäkelä et al. [5] do the same and put these data into a backend store and make them accessible through a mobile application. Paelke et al. [6] and Baldauf et al. [1] both use tag clouds on mobiles to present information from resources that are tagged with location data. Dörk et al. [3] also use tag clouds in a web application allowing location-based exploratory web searches with visualization tools.

PediaCloud integrates ideas from these related projects as it focuses on located information from DBpedia, and further, its visualisation through tag clouds. A particular feature of PediaCloud is the use of secondary resources in constructing the tag cloud, and that the tag cloud changes depending on the user's choice of focus, either the user's location or a particular DBpedia resource. The weighting of tag cloud words are computed from word counts combined with cosine similarity and geographical distance, resulting in a higher emphasis on the more informative tags. PediaCloud also does not depend on a dedicated backend. The tool gathers information from a main Semantic Web endpoint, and computes the visual presentation locally.

References

1. Baldauf, M., Fröhlich, P., Reichl, P.: The ambient tag cloud: A new concept for topic-driven mobile urban exploration. In: Proceedings of the European Conference on Ambient Intelligence. AmI '09, Berlin, Heidelberg, Springer-Verlag (2009) 44–48
2. Becker, C., Bizer, C.: Exploring the geospatial semantic web with DBpedia Mobile. Web Semantics: Science, Services and Agents on the World Wide Web **7**(4) (2012)
3. Dörk, M., Williamson, C., Carpendale, S.: Navigating tomorrow's web: From searching and browsing to visual exploration. ACM Trans. Web **6**(3) (October 2012) 13:1–13:28
4. Gieryn, T.F.: A space for place in sociology. Annual Review of Sociology **26**(1) (2000) 463–496
5. Mäkelä, E., Lindblad, A., Väätäinen, J., Alatalo, R., Suominen, O., Hyvönen, E.: Discovering places of interest through direct and indirect associations in heterogeneous sources — the travelsampo system. In: Terra Cognita 2011: Foundations, Technologies and Applications of the Geospatial Web, CEUR Workshop Proceedings, Vol-798 (2011)
6. Paelke, V., Dahinden, T., Eggert, D., Mondzech, J.: Location based context awareness through tag-cloud visualization. In: Advances in Geo-Spatial Information Science. CRC Press (2012) 265–273
7. Ruta, M., Scioscia, F., Di Sciascio, E., Piscitelli, G.: Location-based semantic matchmaking in ubiquitous computing. In: Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 03. WI-IAT '10, Washington, DC, USA, IEEE Computer Society (2010) 124–127
8. van Aart, C., Wielinga, B., van Hage, W.R.: Mobile cultural heritage guide: location-aware semantic search. In: Proceedings of the 17th international conference on Knowledge engineering and management by the masses. EKAW'10, Berlin, Heidelberg, Springer-Verlag (2010) 257–271
9. Veres, C.: MapXplore: Linked data in the app store. In Sequeda, J., Harth, A., Hartig, O., eds.: COLD. Volume 905 of CEUR Workshop Proceedings., CEUR-WS.org (2012)