A Web Platform for the Evaluation of Vernacular Place Names in Automatically Constructed Gazetteers

Florian A. Twaroch
Cardiff University
School of Computer Science
Cardiff, Wales, UK
+442920876058

f.a.twaroch@cs.cf.ac.uk

Christopher B. Jones
Cardiff University
School of Computer Science
Cardiff, Wales, UK
+442920874796

c.b.jones@cs.cf.ac.uk

ABSTRACT
Vernacular place names pose a research challenge in geographic information retrieval. There is a long standing demand from investigators for a reference collection to train their methods and evaluate their models and data. However no large collection of informal place names associated with type and footprint data is currently available to the GIR community. The present contribution discusses the implementation of a web platform to collect such an evaluation data set. Design considerations of the user interface are addressed and we present first results of a nationwide attempt to collect the vernacular place names of Great Britain. Our result will aid further research in automatic gazetteer construction, considering vernacular place names.

Categories and Subject Descriptors
H.3.7 [**Digital Libraries**]: Collection

General Terms
Measurement, Design, Human Factors

Keywords
gazetteer services, vernacular place names, evaluation

1. INTRODUCTION
Informal place names such as the 'South of France', 'the English Midlands', or the 'city centre' are often used as references to locations in everyday communication. Although the extent of these places is perceived as vague there is often some common agreement about the functionality of these places and their spatial relations to other places. Due to vernacular place names' inherent vagueness they are often omitted from gazetteers. As gazetteers form the knowledge base of contemporary information services, search engines still have problems in handling vernacular place names in their queries and responses. The research interest into how to consider these vague and indeterminate geographic objects in information retrieval has therefore been growing in recent years [1,2].

2. VERNACULAR PLACE NAMES
Research methods to elicit people's knowledge about vernacular place names need to be designed carefully and are usually labour intensive. Evans and Waters [3] have designed a spray can tool that allows users to provide vague regions on a web mapping platform. Users can express the vagueness of the geographic object through intensities in their contributed spray can patterns. The spray can tool is very interesting but takes too much time to use and also has a certain learning curve, which is why we are seeking for simpler methods. The aim is to keep the interface as simple as possible while involving a mass of users in a short simple interrogation (cf. [4]).

3. THE WEB PLATFORM
We are interested in a very broad approach to collect informal place names and ideally want to attract a very high number of participants, where each participant contributes a few valid place names together with location and type information. Here it is not important that the subject has an actual interest in geography or place names. The latter being exploited in other web platforms where a credit system is utilized to reward their most active participants (kudos). On the contrary we want to avoid the envisioned vernacular place name resource being biased towards a small but active user group. In order to gain a broad mass of responses our survey is kept very short and concise. It can be found under http://www.yourplacenames.com. As an incentive to attract as many users as possible, we give away prizes in monthly prize draws. We have foreseen statistics on IP addresses, timestamps and users email addresses to detect mass polling and avoid multiple votes.

A viable method to acquire vague geographic information is based on the interaction with a machine [4]. We developed a web platform where we ask users to contribute as many place names as come to their mind for a familiar location. We decided to provide three options to provide this location:

Firstly users can enter a full UK postcode in association with their listed names. On average 15 households in the UK hold the same postcode. Thus the provided postcode can be converted into a set of coordinates using the Ordnance Survey CodePoint database (OSCP).

Secondly users can point to a location on an interactive map or satellite image that corresponds to a provided name. On the one hand this method might be biased by the names already shown on the map; on the other hand suggesting a place name for a location does not mean that a user necessarily agrees and selects this name to describe a location. It even helps to create a debate about the extent of vaguely cognized regions such as neighbourhoods. Also the utilized maps do not contain any landmark information.

With a third purely textual interface we allow users to provide place name pairs. Such as 'Edinburgh' is also known as

'the Athens of North'. Here we use the left hand part of the user's response to geo-code the provided location with administrative data sources.

Figure 1. www.yourplacenames.com with three entry points.

The three mentioned interfaces (see figure 1) serve as an entry point or login to a variety of further tools to discuss, analyse and inform about vernacular place names in Great Britain. To allow the user to interact with the web platform, we immediately geocode the user's response. Then the user enters into a feedback loop where vernacular place names contributed by other users are presented to the user for evaluation. The user can verify whether a name provided by other users is indeed a place name and express an agreement to the location and type (contributed by other users) of that name. We implemented the principles of a *direct manipulation interface* here. The user receives immediate feedback by using AJAX technology and server-sided scripting for the implementation of the web platform. The implementation allows amendment of previous actions by revising location (postcode, position on the map) and place names provided, which will have immediate effect on the place names in the feedback loop. The names presented to the user are limited to 10 names using an adaptive geometric buffer. If there are more than 10 names within the buffer we select randomly 10 place names to be presented in the feedback loop. Thus we avoid an amplification of popular names and also get feedback on less popular, but commonly used local place names. With a manually created stop word list we filter inappropriate contents that might have been entered. We also record a number of personal sociological variables that allow an evaluation of vernacular place name use. For example, some place names might be used only by older users, some by young people.

4. FIRST RESULTS
A first survey has been carried out in Cardiff University and had 1206 participants. 237 participants had to be excluded due to not completing the survey or multiple votes. This gives the promising success rate of 80.35 % for the first pilot. The remaining 969 provided 4751 place name contributions or 2536 distinct place names. Histograms have been calculated to generate confidence measures for the provided place names. A place name that is provided by several users independently and whose associated points form a spatial cluster is given a higher measure of trust than place names whose points do not cluster or are contributed

by fewer users. However more work needs to be carried out to enhance these confidence measures with web retrieval methods for place name detection.

Our current data include a variety of features and place names such as: 'town', 'The Rec' (recreation grounds), 'city centre', 'The Village' (a pub), 'Victoria Park', 'Pontcanna' (neighbourhood) and many others. We aim to identify 1) point features such as landmarks, 2) linear or curved features such as shopping streets or valleys and 3) regional features such as neighbourhoods, by applying robust estimation techniques for the representation of the possible geometric extent of a provided place name. Here we use methods that can maintain the vagueness of the objects when visualizing (e.g. buffers). For regions we utilize kernel density estimation (KDE) to convert point sets into surfaces [3]. We calculate three different thresholds (20%, 50% and 80% confidence interval) to present possible vague extents of vernacular place names in a web mapping application. We feed these objects back into a web mapping application, again to allow users to vote and debate on the extents and correctness of the resulting features and for us to calculate confidence measures.

5. DISCUSSION AND CONCLUSION
Although the pilot was just carried out in a single location, the results are distributed widely over Great Britain. The proposed framework is critically dependent on successful dissemination of the web platform to attract as many users as possible. This requires a promotion of the tools via all possible dissemination channels and broad support of academic and public institutions in Great Britain. The proposed framework is applicable in any country having broad internet access. The methods for geo-coding would need adaptation to a country's administrative data sources. Valuable feedback received during the pilot phase has been considered in the revision of the incrementally growing platform. We are positive that the suggested framework will be able to create the much needed evaluation data set of vernacular place names for geographic information retrieval in the near future.

6. ACKNOWLEDGMENTS

7. REFERENCES
[1] García Adeva, J. J. 2008. Translating Vernacular Terms into Geographical Locations, in Sample, J.; Shaw, K.; Tu, S. & Abdelguerfi, M. (ed.) Geospatial Services and Applications for the Internet Springer, 2008, 135-153

[2] Jones, C. B., Purves, R.S., Clough, P.D. And Joho, H. 2008. Modelling vague places with knowledge from the Web. IJGIS 2008, 22(10): 1045-1065

[3] Evans, A. J. and Waters, T. 2007. Mapping Vernacular Geography: Web-based GIS Tools for Capturing 'Fuzzy' or 'Vague' Entities *International Journal of Technology, Policy and Management,* 2007*, 7*, 134-150

[4] Robinson, V. B. 1990. Interactive machine acquisition of a fuzzy spatial relation. *Comput. Geosci., Pergamon Press, Inc.,* 1990*, 16*, 857-872