

Using OSM for LBS – an analysis of changes to attributes of spatial objects

Peter Mooney and Pdraig Corcoran

Abstract. The quality of volunteered and crowd-sourced spatial data is not in most cases audited prior to being made accessible to end-users. Studies have shown that this spatial data varies significantly in terms of its geometric quality, its semantic consistency, in terms of its comprehensiveness of coverage and in terms of its currency. Subsequently it often compares poorly with the authoritative data capture and mapping undertaken by national mapping agencies and commercial companies. In this paper we highlight a specific type of problem encountered with volunteered geographic information (VGI) – the naming of real-world features. Many Location-based Services (LBS) applications are using VGI as a spatial data source. The volatility in VGI, as shown by the results in this paper, will require LBS developers to carefully consider how they manage and use the spatial data sources generated by VGI and crowd-source paradigms.

Keywords. VGI, OpenStreetMap, Crowd-sourcing, Spatial Data

1. Introduction

The explosive growth of location-aware devices, wireless communications, and mobile databases has resulted in the realization of location-based services as commercial products and research prototypes. (Mokbel and Levandoski, 2009). These technologies now allow citizens to capture information about their position in space. This includes the capture of both the geometry and attribution of the landscape or environment that they are: currently living in, passing through, or otherwise interested in. OpenStreetMap (OSM) is an example of a collaborative initiative where such spatial data and information can be collated to build-up comprehensive databases operating at all levels from the local and specialist to the global and general. OSM is probably the most famous example of a collaborative, crowd-driven, social-network project for spatial data and information (Over et al 2010, Mooney et al 2010b). However the quality of volunteered geographic information (VGI) and crowd-sourced data is not in most cases audited prior to being made accessible to the public. The data varies significantly in terms of its geometric quality, its semantic consistency, in terms of its comprehensiveness of coverage and in terms of

its currency. Subsequently it compares poorly with the authoritative spatial data capture and mapping regimes undertaken by national mapping agencies (NMA). However, despite the lack of resources compared to NMA, crowd-sourced and volunteered data is in many cases more up-to-date than the authoritative source of information, can incorporate features of interest not covered by mapping agencies and is often broader and richer in the meta-data captured. Crowd-sourced data is, almost by definition, data linked to where people are carrying out their everyday lives and related to things that are of interest or importance to them. In this paper we analyse the naming of spatial objects in OSM in this collaborative environment. The data in OSM is most commonly gathered by OSM contributors surveying areas with GPS then uploaded to OSM using one of the many editors (Potlatch, JOSM, Merkaartor, etc) or by using an editor to trace the outline of geographic features from aerial imagery. In both cases tags or attributes (key-value pairs) can be optionally associated with the geographic object. The OSM community maintain a community endorsed ontology of key-value pairs on the “Map Features” page (OSM, 2011)

1.1. Tagging objects in OSM

The name attribute of objects in spatial databases is arguably one of the most important for Location-based Services (LBS) applications. Accessing information by its geographic reference is natural and useful in several contexts: for example when looking for information resources on a city, town, or area for tourism activities from mobile devices it is common practice to attempt to access information by specifying the geographic place of interest. Edwardes (2009) states that “one of the quintessential ways in which people make sense of space is by naming it” and that identification of places is “perhaps the most important perspective in LBS”. More generally, it has been estimated that about 15% of the queries submitted to general purpose search engines contain geographic names (Bordogna et al, 2011). Searches for geographic information on the Internet often fail due to people referring to locations, such as local neighbourhoods, with commonplace or vernacular names that are not recognised by conventional administrative gazetteers (Twaroch et al, 2009). de Longueville et al (2010) comment that in many real-world applications, vernacular names, which do not correspond to those official place names in gazetteers, are often used to name places by people and those who contribute to VGI projects. In fact the authors use the term “open gazetteer” to designate the concept of gazetteer enriched with vernacular place names

Joshi et al (2010) remarks that one of the most “potent forms of noise in collaborative environments is the incorrect or alternate spellings associated with place-names”. In the English language this form of noise is compounded by the inherent polysemy and synonymy of words in the English language.

According to the TagInfo service in May 2011 (TagInfo, 2011) there are just over 26 million objects in the OSM global database with a name tag assigned. In OSM the name tag is a free text field. When a contributor adds or edits an object in OSM they can optionally create, edit, or delete a name tag. The highway tag is used to indicate: roads, streets, paths, lanes, etc. This tag is specified formally in the pseudo ontology provided on the OSM Wiki pages called “Map Features”. This specifies values for the highway tag from (motorways, Interstates, Autobahnen) to (laneways, paths, tracks, trails). Generally the highway-tag represents the importance of the highway in the street/road network grid. Consequently contributors should understand where in the street/road network hierarchy the feature they are working on fits. All tagging in OSM should be performed in accordance with the community agreed ontology on the “Map Features” page. The OSM editor software packages all provide functions in the user interfaces to make selection of keys, and where appropriate corresponding values, easy and partially automated. However contributors have the ability to ignore these suggestions and provide their own tags.

1.2. Research contributions of this work

The research focus of this paper is as follows. Current research on OSM data quality and user behaviour is predominantly focussed on the „current“ snapshot of the OSM database for a given region or country. Fresh updates to these databases can be downloaded as frequently as every 3 hours. However this approach makes the assumption that the spatial data (geometries and attributes) within the OSM database have been collected and managed in a process similar to NMA of: survey/collect data, assimilate, combine, correct, check, and release. Mooney and Corcoran (2011c) argues that is not the case and that edits and updates are made in often unpredictable ways. In LBS some of the most popular applications involve queries of the form “How do I get to location X?” or “Plan me the shortest route to street/place X”. In this paper we provide an analysis of the entire version editing history of over 24,000 objects in 4 OSM databases in Europe. We show that significant changes can occur over time as a contributor or contributors edit/update object's geometry or spatial attributes. Specifically we analyse changes to objects with a “name” attribute and objects with a “highway” attribute. We highlight, and attempt to quantify, the uncertainty over the names assigned to objects and/or the

assignment of designation status to the highway attribute. We are not aware of any other work of this type on VGI/OSM which specifically addresses these problems through analysis of the full version history. The key outcomes are as follows: (1) uncertainty in name or highway attribute value is not strongly correlated to increasing number of contributors editing a given object, (2) approximately 10% of objects in the OSM databases analysed exhibited changes in the name or highway attribute, and (3) this provides a warning to LBS developers that OSM may need to be supported, in parallel, by gazetteers etc to ensure accuracy of spatial queries.

The remainder of the paper is organised as follows. In section 2 we give a brief overview of the current literature on this topic. Section 3 outlines the experimental setup. The results of the analysis performed on the four OSM databases are provided in Section 4. The paper closes with section 5 which provides a summary of the main conclusions from this work and offers some possibilities for future work.

2. Overview of related work

OpenStreetMap (OSM) provides a highly dynamic source of spatial data for use in Location-based Services (Jacob et al, 2010). The purpose of this paper is to provide information on how geometries and attributes change in the OSM database and to comment on the effect this might have on LBS using OSM as the primary source of spatial data. Changes occur in OSM due to the upload of new spatial data and/or the updating of existing contributions (geometry and spatial attributes) by the original creator or another contributor. Some of the reasons for these changes could be: contributor disagreement, changes to geometry (shortening of roads, resizing, etc), actual real world changes reflected in the data, combination of polylines/polygons into multipolygon relations, mistakes being made by contributors through not understanding spatial data handling or incorrect use of the editor software.

Some studies have highlighted problems with OSM in terms of spatial data quality. In Mooney et al (2010a, 2010b) the authors provide analysis of how the representation (points used to represent features, tagging, etc) of features in OSM can vary greatly from country to country and within object classes. This work also highlighted problems with the quality of the shape representation of polygon features representing natural features in OSM when compared with NMA data for the same features. However other studies such as Over et al (2010), Haklay (2010), Girres and Touya (2010) compares OSM very favourably, in terms of geometric accuracy, against road and street network databases from National Mapping Agencies and

commercial sources. Fritz et al (2009) even suggest at using OSM and other crowd-sourced VGI as an alternative approach for “validating and calibrating global land cover”.

OSM and Wikipedia are similar in that they are both crowd-source Wikis allowing collaboration on knowledge and information generation. It is therefore obvious to look to research work on Wikipedia to see if there are similar issues occurring in terms of: contributor behaviour, editing, etc. Wikipedia has grown to be the world largest and busiest free encyclopedia, in which articles are collaboratively written and maintained by volunteers online (Hu et al, 2007). They suggest for Wikipedia that given its collaborative nature the “user interaction data” must be analysed with article quality metrics to obtain a good overall understanding of the quality of an article.

Neis (2011) reports during the first two weeks of May 2011 17000 new “Unconnected Roads” errors were introduced into the OSM Europe database. This is of concern for LBS applications using OSM databases as the street/road network data source. However this problem of unconnected roads is complimented by Neis stating that in OSM Europe for the months (February, March, April 2011) there has been an increase of “about 2850000 new OSM way segments” (polylines) for routing. So while there are errors in the OSM Europe database (and others) this is somewhat offset by continued contributions. Overall literature is beginning to appear on analysis of the quality and usability of OSM data. However there is little literature providing an analysis of OSM through investigation of the history of edits and contributions to the OSM database.

3. Experimental Setup

Four OSM databases were accessed for this research: Ireland, United Kingdom, Austria, and Germany. The raw data was downloaded from the Geofabrik web service during April 2011. The raw data is made available in OSM-XML format. The raw data contains the most up-to-date version of all features in the OSM global database for that country. All tags (annotations) are also included in the XML. For this paper we focused on „high edit“ features: that is features (polylines and polygons) that have been edited by OSM contributors 15 times or more. These “high edit” features were chosen in an attempt to select a subset of features from OSM which have undergone collaborative editing from multiple contributors or editors. From the four databases selected this provided our study with over 24,000 spatial objects and these objects are summarised in Table 1. Objects with feature types highway, land-use, natural, amenity, and waterway were

chosen. According to the Tag-Info web-service (taginfo.openstreetmap.de) these are the 5 most popular feature types in OSM.

Processing the OSM-XML is made difficult by the size of the downloaded files (Mooney and Corcoran, 2011). For example the uncompressed OSM-XML file for Germany is approximately 14Gb in size. Conventional desktop tools for XML data handling cannot process this volume of data. Using Linux command-line scripting (grep and sed) combined with PHP scripts the following steps are performed. Firstly all “ways” (polygons and polylines) in the OSM-XML are extracted into smaller, more manageable sized, files (~ 100Mb per file). These files are then processed by the PHP scripts to extract the “high edit” objects. The most time-resource consuming part of the data gathering process is the download of the entire version history, in OSM-XML format, for each of the 24,000 spatial objects. The OSM API (Application Programming Interface) is used for this purpose. When the version history for an object is downloaded the OSM-XML is processed by a PHP script and: the tags, contributor history, geometry history, and other information for the object is stored in the PostGIS database for further analysis.

4. Experimental Analysis

In this section we provide results of the experimental analysis of the four OSM databases described above. For the purposes of analysis we grouped the UK and Ireland databases together.

Attribute	Germany	Austria	UK + Ireland
<i>Total Objects</i>	10,603	3,367	10,693
<i>Total Highways</i>	902	2,359	8,999
<i>Objects with „name“ tag</i>	5,062	2,571	7,642
<i>„name“ changes</i>	1110	767	1815
<i>Highway change</i>	408	1248	3703

Table 1: Summary of three OSM databases used in this case-study

In Table 1 the summary of the contents of the three OSM databases is provided. Each database contains objects representing: highways, land-use, natural features, amenities, and waterways. The UK-Ireland database is dominated by highway features. The Germany and Austria database provides a better distribution of object type. The fourth row of the table provides a count of the number of objects where a “name” tag is assigned to the object. The “name” changes row shows the total number of objects where the name of the object changed one or more times over the lifetime of the object. The “highway-change” row indicates the number of highway objects where the designation (highway status ie motorway, secondary, path, lane, etc) was changed one or more times.

4.1. Contributor behaviour

In the PostGIS database we store each version of all objects as an atom of a three-dimensional (3-D) unit geographic information ‘location (geometry), time, attributes’. The term “place name” has several synonyms, including “toponym”, “geographical place name”, and “geographic place name”. Confusion, uncertainty, and misunderstanding may occur when the name for an entity is spelled in different ways, when different names are used for the same place, when the same name is used for different places, or when a name is applied to a feature in an unexpected or different way from the general understanding of how it should apply. The term “allonym” is occasionally used to refer to two or more names for the same place; it essentially means place–name synonym (Beall, 2010). Tables 2, 3, and 4 show examples of collaborative editing to three different features in Austria, UK, and Germany. In Table 2 a street in the city of Atmunt, Austria, is assigned a slightly different name tag by 4 different users (see *User_ID* column). The final row of the table shows the current value assigned to the name tag and the version number of the object in the OSM database (indicated in brackets). Table 3 shows an example of a street in Birmingham, UK. Two contributors are involved in the tagging of this object. There are several variations on spelling of the streetname. The example in Table 4 is from Hamburg, Germany. In this example five contributors are involved. The interesting aspect of this example is the number of changes in designation to the “highway” object. This introduces uncertainty related to the actual physical designation of this object.

<i>Version</i>	<i>Name Tag</i>	<i>Creation Time</i>	<i>User_ID</i>
1	Ennstal Bundeststrasse,	2008-05-05 15:10:06	16170

	Ennsradweg		
7	Ennstal Bundeststraße, Ennsradweg	2008-09-13 20:20:55	45347
8	Ennstal Bundeststraße	2008-10-25 09:05:33	12408
10 (15)	Hauptstraße	2009-07-19 21:24:51	6470

Table 2: osm_id 24228123 - street in Atmont City, Austria

Version	Name Tag	Creation Time	User_ID
2	Oakthorp Drive	2008-05-08 19:39:45	35691
6	Over Green Drive	2008-05-09 08:50:30	35691
9	Oak Thorp Cr	2008-05-09 08:52:52	35691
10	Oak Thorp Dr	2008-05-09 08:53:10	35691
15	Oak Thorpe Dr	2008-05-11 13:54:37	35691
18	Oak Thorp Drive	2010-02-07 14:38:14	9065
19 (current)	Oak Thorpe Drive	2010-08-24 11:32:25	35691

Table 3: OSM_ID 24276789 in North-east Birmingham UK

Version	Name Tag	Creation Time	User_ID
2	Unclassified	2007-10-18 11:10:53	4902
3	Secondary	2008-01-11 15:15:07	21021
4	Unclassified	2008-01-11 15:25:52	21021
13	Construction	2009-10-22 12:47:15	124032
16	Secondary	2010-02-17 11:36:30	211280
17	Unclassified	2010-02-18 09:48:43	211280

18	Pedestrian	2010-02-22 15:21:24	211280
19 (current 23)	Tertiary	2010-02-25 16:09:54	44838

Table 4: OSM_ID 9782645 in Hamburg, Germany

4.2. Changes to “name” tags and highway designation

In an attempt to quantify the types of changes that the value of the name tag for objects in the OSM databases undergo we applied two well known and robust string matching metrics from the domain of text similarity matchin. The Jaro–Winkler distance (Jaro, 1989) is a measure of similarity between two strings. The higher the Jaro–Winkler distance for two strings is, the more similar the strings are. The literature indicates that the Jaro–Winkler distance metric is most suitable for short strings such as person names or placenames. The Jaro-Winkler score is normalized such that 0 equates to no similarity and 1 is an exact match (Top et al, 2007). The Levenshtein distance between two strings is defined as the minimum number of character edits required to transform one string into the other (Yujian and Liu, 2007). The allowable edit operations are: deletion, insertion, or substitution of a single character at one time. The Levenshtein distance has a lower bound of zero if and only if the strings are identical whereas the upper bound is either at maximum the length of the longer string or at minimum the difference of the lengths of the two strings (Ackroyd, 1980). Both metrics are used for similarity searching in databases. For example for the strings “A171” and “Scalby Road” the Jaro-Winkler distance is 0 (no similiarity) while the Levenshtein distance is 11 indicating the number of operations required. Another example is „Western Road“ and “Western Avenue”. The Jaro-Winkler distance shows good similarity at 0.74 while the Levenshtein distance is 6. These two metrics are suitable to take into account spelling errors introduced by contributors when naming objects (or renaming). For example in Table 3 the values assigned to the street over the editing period are very similar.

For each object in the OSM databases we extracted those objects where the name tag value was changed at least once as described in Table 1. We calculated the Levenshtein and Jaro-Winkler distance metrics for each name transformation pair, in order – for example “Western Road” changing to “Western Avenue”. For each object when calculate the mean Levenshtein and Jaro-Winkler distance for each object. Figure 1 and 2 below show mean Levenshtein and Jaro-Winkler distances for each object plotted in scatter plot format for the UK and Germany databases. The pattern of distribution is very similiar. There are instances where there are clusters of very similar name transformations - those with mean Jaro-Winkler close to 1 having

very small (<5) mean Levenshtein distances. On the other hand there are examples of some completely dissimilar name transformations – those with mean Jaro-Winkler distance at 0 and large values for mean Levenshtein distances.

As shown in Table 2, 3, and 4 the changes to names of objects or the designation values for highways can happen in the presence of a single contributor or multiple contributors. We calculated the a number of statistics to investigate if the number of changes to names or objects or highway designation increased as the number of contributors increased. We calculated the correlation and also the Spearman correlation (rho ρ and p-value). The results are summarised in Table 5. Unfortunately there does not appear to be any significant statistical relationship between the number of contributors and the number of changes to name attributes. We calculated the two-sided p-value for a hypothesis test where the null hypothesis is that two sets of data (number of contributors, number of changes performed by contributors) are uncorrelated. In all cases the p-value exceeds 0.05 so we must accept the null hypothesis. While this is somewhat disappointing it is not unexpected. The examples in Table 2 and 3 show the effects of multiple contributors and then a single contributors.

Database	N	corr	rho ρ	p-value
Germany	1110	0.022	0.045	0.105
UK Irl	1815	-0.171	-0.013	0.561
Austria	767	0.088	0.078	0.051

Table 5: Analysis of the effects of the number of contributors on the changes to name attributes on objects in the case study databases

In Figure 3 we show a simple plot of the numbe of unique contributors for the $N = 1815$ objects used for the analysis in Table 5 plotted against the number of unique “name” tag values for those objects. It is immediately obvious why the correlation tests return such inconclusive results. Small numbers of contributors per object (single, two or three) appear just as likely to introduce uncertainty into the accuracy of the value of a “name” tag value as the arrival or introduction of additional contributors to the edit history of an object.

Table 6 provides the same analysis for objects with the highway tag and the effects of the number of contributors on the number of changes to highway designation. Again we must accept the null hypothesis.

Database	N	corr	rho ρ	p-value
Germany	408	0.04	0.025	0.604
UK Irl	3703	0.03	0.033	0.019
Austria	1248	0.041	0.045	0.105

Table 6: Analysis of the effects of the number of contributors on the changes to highway designation on objects in the case study databases

Some specific examples will help to highlight the highway designation problem. In the Austria database there are 93 highway objects who have their designation changed from “primary” to “secondary” (or vice-versa), in the German database this is 46, while in the UK and Ireland this is 350. We feel that this aspect of the collaborative contributions to OSM requires further investigation to explain the rationale behind these decisions.

4. Conclusions

In this paper we have analysed “high edit” geographical features from four OSM databases from Europe. “High-edit” features, those with at least 15 versions in their history, were chosen to provide us with features which exhibit the edit patterns of a collaborative crowd-sourced environment.

The results of the experimental analysis on the number of contributors examined against the corresponding number of changes to objects do not reveal any significant patterns. In studies on Wikipedia it was found that most Wikipedians contribute to a relatively small set of articles each. Their contribution was biased towards one or very few article(s) (Zhang et al, 2010). At the same time, each article's contributions are often championed by very few active contributors including the article's creator. Stein and Hess (2007) found in an analysis of the German Wikipedia that the number of contributors and who those contributors were was very strongly correlated with article quality. In the data for this work we found that there were many different “patterns” of contribution: single contributors making multiple changes, multiple contributors making multiple changes, and a mixture of these. The analysis here shows significant changes to the names of objects. There is also significant changes to the designation of highways. Both of these issues should be considered very carefully before OSM is used as the primary source of data for LBS applications. Codescu et al (2011) provides a recent and valuable reference for our work. In their paper they develop a web service focusing on finding locations not only by their

address, but by systematically relating the places to activities that a person could perform there. This is helpful if a person wants to explore a new city, or plans leisure activities. The authors remark that "OpenStreetMap provides a rich set of tags that can be used for activity-oriented search". However, they acknowledge the evolving nature of tags in social media and collaborative web projects by integrating several ontologies that are related to each other and develop "matching tools to cope with the evolving nature of the tags".

A potential drawback of the OpenStreetMap crowdsourced model is the lack of a specific set of "administrators" or "moderators" for the spatial content uploaded and edited to the OSM database. It could be easily argued, based on the statistics Neis (2011) provides for short periods of time (3 months) that a very large number of administrators or moderators would be required to check and verify uploaded data. This might not be feasible or may as Girres and Touya (2010) warn "kill the joy" of contribution to VGI projects such as OpenStreetMap. While Allen (2010) outlines the key advantages Web 2.0 and crowdsourcing has brought to knowledge sharing he states the open question if lack of coordination and quality issues actually provide barriers to knowledge sharing and usage.

Acknowledgements

Dr. Peter Mooney is a research fellow at the Department of Computer Science NUI Maynooth and he is funded by the Irish Environmental Protection Agency STRIVE programme (grant 2008-FS-DM-14-S4). Dr. Pdraig Corcoran is a lecturer and post-doctoral researcher also at the Department of Computer Science NUI Maynooth. Dr. Corcoran is part of STRAT-AG which is a Strategic Research Cluster grant (07/SRC/I1168) funded by Science Foundation Ireland under the National DevelopmentPlan. The authors gratefully acknowledge this support.

References

- Ackroyd, M.; , "Isolated word recognition using the weighted Levenshtein distance," *Acoustics, Speech and Signal Processing, IEEE Transactions on* , vol.28, no.2, pp. 243-244, Apr 1980
- Allen, J.P.; (2010), "Knowledge-Sharing Successes in Web 2.0 Communities," *Technology and Society Magazine, IEEE* , vol.29, no.1, pp.58-64, Spring 2010

- Beall, J (2010) Geographical research and the problem of variant place names in digitized books and other full-text resources, *Library Collections, Acquisitions, and Technical Services*, Volume 34, Issues 2-3, Pages 74-82
- Bordogna, G., Ghisalberti G., and Psaila, G., Geographic information retrieval: Modeling uncertainty of user's context, *Fuzzy Sets and Systems* (2011), doi:10.1016/j.fss.2011.04.005
- Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T. and Rau, R. 2010. DO-ROAM: Activity-Oriented Search and Navigation with OpenStreetMap, Fourth International Conference on Geospatial Semantics. Brest, France.
- De Longueville, B., Ostländer, N. and Kesitalo, C. (2010) Addressing vagueness in Volunteered Geographic Information (VGI) – A case study. *International Journal of Spatial Data Infrastructures Research*, Vol 5 ISSN 1725-0463
- Edwardes, A.J. 2009. Geographical perspectives on location for location based services. In *Proceedings of the 2nd International Workshop on Location and the Web (LOCWEB '09)*. ACM, New York, NY, USA, , Article 5 , 4 pages.
- Fritz S., McCallum I., Schill C., Perger C., Grillmayer R., Achard F., Kraxner F., Obersteiner M. Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sensing*. 2009; 1(3):345-354.
- Girres, J.-F., and Touya, G. Quality assessment of the french openstreetmap dataset. *Transactions in GIS* 14, 4 (2010), 435-459.
- Haklay, M. How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets. *Environment and Planning B: Planning and Design* 37, 4 (2010), 682–703.
- Meiqun Hu, Ee-Peng Lim, Aixin Sun, Hady Wirawan Lauw, and Ba-Quy Vuong. (2007). Measuring article quality in wikipedia: models and evaluation. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (CIKM '07)*. ACM, New York, NY, USA, 243-252.
- Jacob, R., Mooney, P., Corcoran, P., and Winstanley, A. C. (2010). Haptic-GIS: exploring the possibilities. *SIGSPATIAL Special* 2, 3 (November 2010), 13-18.
- Jaro, M. A. (1989). "Advances in record linkage methodology as applied to the 1985 census of Tampa Florida". *Journal of the American Statistical Society* 84 (406): 414–20.
- Joshi, D., and Luo, J., and Yu, J., and Lei, P., and Gallagher, A. (2010). Rich location-driven tag cloud suggestions based on public, community, and personal sources. In *Proceedings of the 1st ACM international workshop on Connected multimedia (CMM '10)*. ACM, New York, NY, USA, 21-26.
- Mokbel, M and Levandoski, J. (2009). Toward context and preference-aware location-based services. In *Proceedings of the Eighth ACM International Workshop on Data*

- Engineering for Wireless and Mobile Access* (MobiDE '09). ACM, New York, NY, USA, 25-32.
- Mooney, P., Corcoran, P., and Winstanley, A. C. A study of data representation of natural features in openstreetmap. In Proceedings of the 6th GIScience International Conference on Geographic Information Science (2010a), GIScience 2010, University of Zurich, p. p150.
- Mooney, P., Corcoran, P., and Winstanley, A. C. Towards quality metrics for openstreetmap. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (New York, NY, USA, 2010b), GIS '10, ACM, pp. 514–517.
- Mooney, P and Corcoran, P., Accessing the history of objects in OpenStreetMap. Proceedings of the 14th AGILE International Conference on Geographic Information Science, Utrecht, The Netherlands, Eds: Stan Geertman, Wolfgang Reinhardt and Fred Toppen p 141 April 2011.
- Neis, P. (2011) OpenStreetMap Routing View Europe (May 2011) Online Blog <http://neis-one.org/2011/05/routing-view-2011-05/> (Last checked May 2011)
- M. Over, A. Schilling, S. Neubauer, A. Zipf, Generating web-based 3D City Models from OpenStreetMap: The current situation in Germany, Computers, Environment and Urban Systems, Volume 34, Issue 6, GeoVisualization and the Digital City - Special issue of the International Cartographic Association Commission on GeoVisualization, November 2010, Pages 496-507
- OSM (2010) The Map Features Page on OpenStreetMap Wiki http://wiki.openstreetmap.org/wiki/Map_Features (Last Checked May 2011)
- Klaus Stein and Claudia Hess. 2007. Does it matter who contributes: a study on featured articles in the german wikipedia. In *Proceedings of the eighteenth conference on Hypertext and hypermedia* (HT '07). ACM, New York, NY, USA, 171-174.
- TagInfo The TagInfo Application (2011) <http://taginfo.openstreetmap.de> (Last Checked May 2011)
- Top, P.; Dowla, F.; Gansemer, J.; , "A Dynamic Programming Algorithm for Name Matching," Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on , vol., no., pp.547-551, March 1 2007-April 5 2007
- Twaroch, F.A and Purves, R.S. and Jones, C.B (2009) Stability of Qualitative Spatial Relations between Vernacular Regions Mined from Web Data, International Workshop on Geographic Information on the Internet, Toulouse , France
- Twaroch F.A., Jones, C.B. and Abdelmoty, A.I.: (2008) Acquisition of a Vernacular Gazetteer from Web Sources. In: Workshop on Location on the Web 2008, Locweb 2008, Beijing, China, pp. 61--64.

Yujian, Li; Bo, Liu; , "A Normalized Levenshtein Distance Metric," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.29, no.6, pp.1091-1095, June 2007

Yi Zhang, Aixin Sun, Anwitaman Datta, Kuiyu Chang, and Ee-Peng Lim. (2010). Do Wikipedians follow domain experts?: a domain-specific study on Wikipedia knowledge building. In *Proceedings of the 10th annual joint conference on Digital libraries (JCDL '10)*. ACM, New York, NY, USA, 119-128.

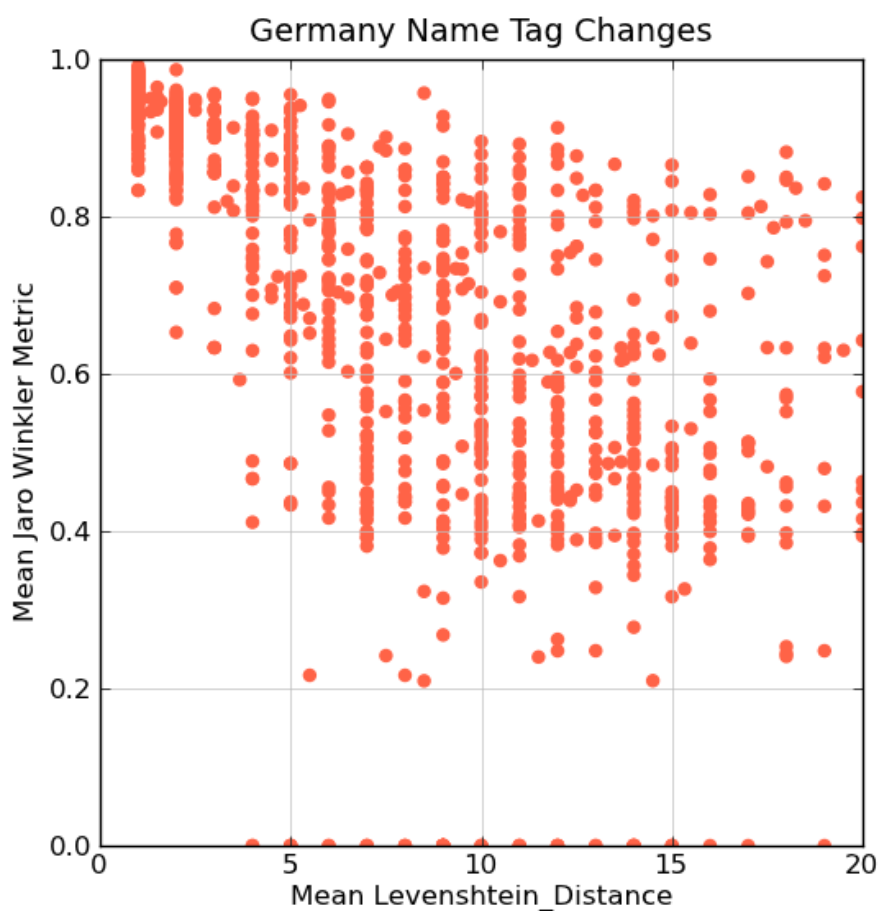


Figure 1: Mean Levenshtein Distance plotted against the Mean Jaro-Winkler distance for the changes to "name" tags for objects in the Germany OSM Database

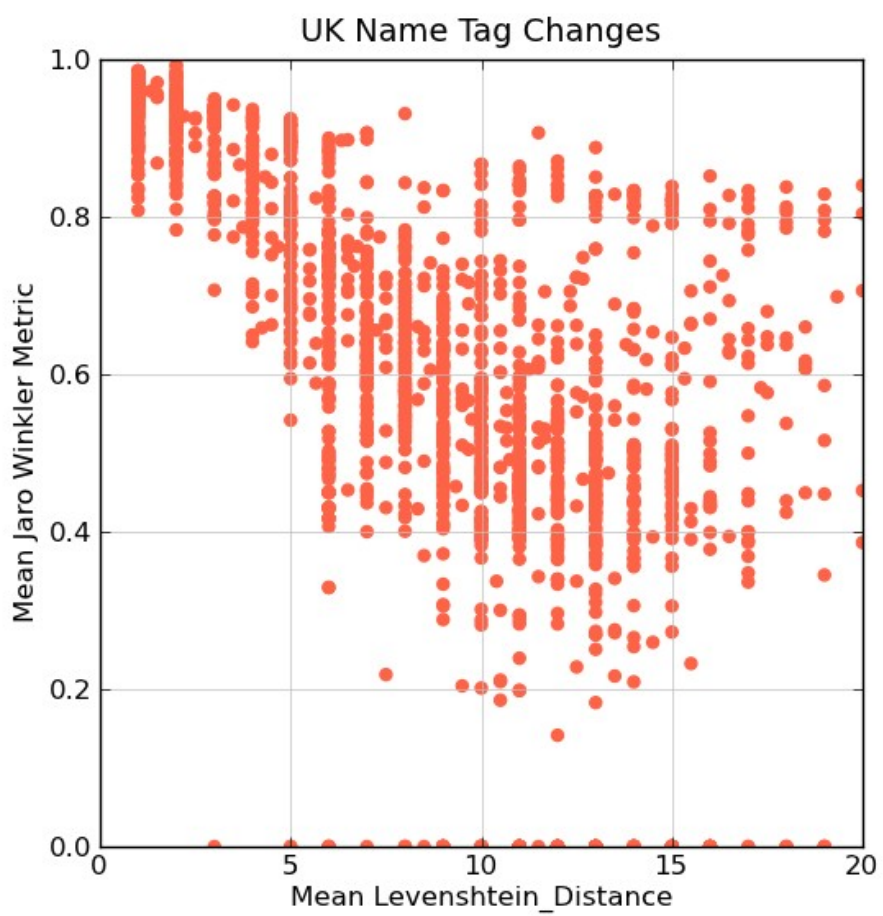


Figure 2: Mean Levenshtein Distance plotted against the Mean Jaro-Winkler distance for the changes to "name" tags for objects in the UK Ireland Database

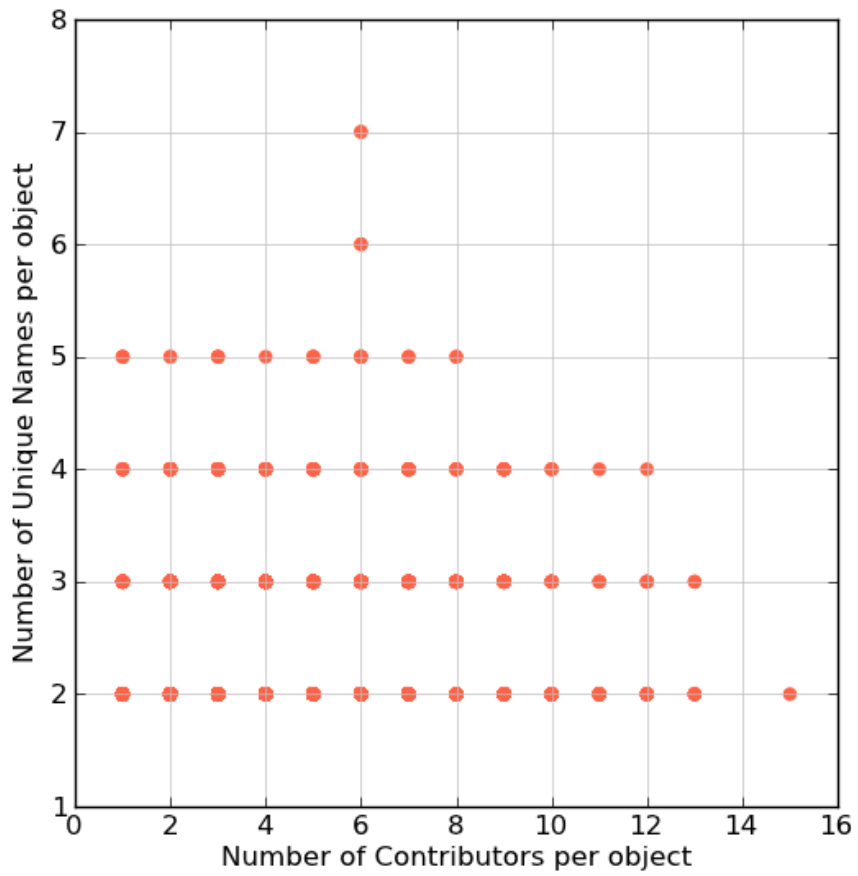


Figure 3: The number of unique contributors plotted against the number of unique names for objects in the UK Ireland database