Automated Production of Volunteered Geographic Information from Social Media

Maxwell Guimarães de Oliveira, Cláudio de Souza Baptista, Cláudio E. C. Campelo, José Amilton Moura Acioli Filho and Ana Gabrielle Ramos Falcão

Information Systems Laboratory – Federal University of Campina Grande (UFCG)
Av. Aprígio Veloso, 882, Bloco CN, Bairro Universitário – 58.429-140
Campina Grande – PB – Brazil

maxwell@ufcg.edu.br, {baptista, campelo}@dsc.ufcg.edu.br, {joseamilton, anagabrielle}@copin.ufcg.edu.br

Abstract. *The easy production of data with geographic context has enabled a deeper engagement with people and has led to the emergence of Location-Based Social Networks - LBSNs. Such environments have proved to be very useful in the context of smart cities, however, one of the main challenges has been how to keep users willing to contribute and keep the LBSNs in a continuous operation. Concerning this problem, we propose an automated production of Volunteered Geographic Information - VGI - based on Geographic Information Retrieval techniques with the aim of providing valuable and up-to-date information for LBSN environments taking advantage of social media messages around the web. A prototype software was developed and evaluated through a case study using microtexts.*

1. Introduction

Volunteered Geographic Information (VGI) has emerged in the last years as an alternative spatial data source on the Web. It consists of data pooled with the geographic context, which is produced and disseminated by individuals spread throughout the world, forming an environment known as Crowdsourcing [Surowiecki 2005]. These individuals are called volunteers and most of these volunteers are not experts in Geography or Geographic Information Sciences, but ordinary people interested in sharing their viewpoints and knowledge about geographic locations [Goodchild 2007].

The easy production of data with geographic context has enabled a deeper engagement with people in everything involving location. It can be explained by the technological evolution of the last years and novel tendencies of the Web 2.0, including the emergence of devices featuring GPS and the spread of Internet connectivity around the world. Users have been increasingly consuming this type of information by means of location-based applications, and also sharing this information in several domains.

This scenario has led to the emergence of applications such as the Location-Based Social Networks (LBSN). The LBSNs provide context-aware services which allow assigning users to content [Vicente et al. 2011], and provide many different types of services, from entertainment to public utilities. In such a social network, much information is voluntarily created, be it textual, multimedia or geographic. Falcão et al. (2012) developed Crowd4City, a LBSN that can be applied to the domain of smart cities, which supports participatory human sensors, aiming to create an environment for identification and discussion of matters concerning the government of the cities, a common interest of the population.

Despite initiatives like Crowd4City, one of the main challenges in the use of human sensors has been keeping them willing to contribute and consequently maintain the LBSNs in a continuous operation. Only a few users are in charge of providing a significant volume of information. This phenomenon is visible in terms of geographic location, where many areas around the world are mapped by just one user [Haklay and Weber 2008]. One of the factors regarding users' motivation can be associated with the existence of costs for these volunteers. These costs can be inherent to the learning curve for correct operationalization of a LBSN, or related to the contribution routines. These costs can also be associated with the volunteers' available time and demands persistence from them. Therefore, it becomes necessary to find alternatives that will allow keeping the LBSNs up-to-date even when the volume of contributions of the volunteers is below the expected.

One of the purposes of researches in Geographic Information Retrieval (GIR) is the development of techniques for inference of geographic locations associated to text documents. This is an area full of challenges, involving Natural Language Processing (NLP), handling of uncertainty, disambiguation, context identification, among other tasks [Bordogna et al. 2012]. With GIR techniques, it is possible to process and assign geographic locations to text from websites, blogs and social networks, such as the task known as geoparsing [Purves and Jones 2011]. Hence, we raised the hypothesis that, after the identification of their referenced geographic location, texts from the web, such as messages publicly exchanged in social networks, could automatically turn into useful information in applications such as the LBSNs. Thus, their authors can become non-intentional volunteers in the production of the VGI.

Several researches have addressed the assignment of geographic locations on web documents (Georeferencing) [Rupp et al. 2013] [Watanabe et al. 2011], including social network messages. However, the percentage of the information concerning this geographic context is still very low. Furthermore, approaches based on matching the users' locations and their messages have proved to be very inaccurate, since users can freely disseminate information about the most diversified geographic contexts, which, in most times, mismatch their geographic position at the very moment the information is shared.

On that account, this paper presents an approach for the automated production of VGI based on the application of geoparsing and georeferencing techniques to texts published on the web, especially on social media. We have kept our focus in the body of messages and therefore do not considered previously geotagged texts due to such occurrence is still low. Furthermore, we cannot ensure that an embedded geolocation is the same location that message refers to. The VGI produced by the authors of the processed texts will become available for the users of a LBSN. These users will be the main consumers and also validators of this information, being capable of pointing incoherences as well as stressing the relevance of that information for other users of the network, enriching the crowdsourcing environment.

The main contributions of this paper are: the development of an artifact for automatic production of Volunteered Geographic Information, based on the content of social media texts; and a discussion about geoparsing in informal texts published in microblogs and the value that such information may reveal whether the geographic context is explored. The remainder of this paper is structured as follows. Section 2 discusses related work. Section 3 describes our approach. Section 4 addresses a case study carried out to evaluate the proposed ideas. Finally, section 5 concludes the paper and highlights further work to be undertaken.

2. Related Work

Research on VGI has prevailed in several parts of the world. Besides Computer Science, many correlated disciplines, such as Geography [Goodchild 2007], Geographic Information Science [Du et al. 2011] [Haklay et al. 2010] [Jackson et al. 2010] and Human Factors [Parker et al. 2011] have investigated issues concerning this kind of volunteered information.

One of the most representative VGI project is the OpenStreetMap (OSM) [Haklay and Weber 2008] [Koukoletsos et al. 2012]. The OSM database consists of a significant collection of volunteered spatial data based on the Wikipedia collaborative model [Mooney and Corcoran 2012]. The OSM project has received many contributions from the community. Haklay (2010), for instance, has focused on assessing VGI quality and how VGI can be reliable and usable. Ballatore and Bertolotto (2011) focused on semantic relationships within OSM data. They highlight how OSM is spatially rich but semantically poor and investigate ways of linking OSM to other distributed repositories.

Besides the OSM project, several works have revealed VGI as a promising research field. Horita et al. (2013) made a thorough literature review on VGI with the objective of verifying its applicability for aiding in disaster management. In that study, it was possible to observe that the VGI has been more frequently used in fires and floods. Havlik et al. (2013) discussed VGI mobile applications concerning several aspects, such as functionalities and user experience. Ballatore et al. (2013) explored the semantic side of VGI and presented a technique for computing the semantic similarity of geographic terms in VGI based on their lexical definitions and using WordNet. The authors based themselves on the intuition that similar terms tend to be recursively defined by similar terms.

While the research on VGI is still relatively novel, the research on GIR has many studies focused on the identification and indexing of geographic locations through the application of Natural Language Processing (NLP) techniques. Some related works on GIR will be described as follows. Rupp et al. (2013) discussed the customization of geoparsing and georeferencing tools to be applied in collections of historical texts. The authors made an analogy between the storage/indexing of files about the medieval era and the storage/indexing of Twitter feeds, and discussed questions involving standardization and use of gazetteers. There is no discussion about the spatial precision of the geoparsing, but this could be a motivational factor for such customization.

Liu et al. (2013) proposed QGIR, Qualitative Geographic Information Retrieval, as a better option to deal with geographic information described in natural language in web documents. The authors argue the replacement of GIR by QGIR for cases where the place name and thematic representations are necessary, considering the use of semantic spatial relations and domain-specific ontologies. An experiment was carried out in order to compare QGIR with the standard GIR, and the results proved the superiority of QGIR for queries like “precious metals in the Hebei province”. Freire et al. (2011) described an approach for recognition of place names expressed in metadata of digital libraries. That approach should be better at capturing features of the non-structured text found in metadata records and at the exploration of the relevant information in the structured data of those records.

Watanabe et al. (2011) proposed an automatic method for identification of geographic location in non-geotagged tweets. Such method is based on the clustering of messages according to the type of event, considering short time intervals, small geographic areas and geotagged tweets. Thus, geotagged tweets are used to allocate geotags in tweets which do not have the geographic tag yet. The authors do not consider the possibility of the geotagged tweets having a different geographic reference than the location discussed in the messages. Also, it is possible that users are not necessarily talking about their current locations. Therefore, there is a possibility of errors in the geographic precision and this must be considered. In a similar way, Jung (2011) presented a method for analyzing sets of microtexts, aiming at identifying contextual clusters of tweets. By establishing a contextual relation between the messages, a set of microtexts can be considered as a single document and make the process easier for the geoparsers. This task, however, can be very costly, depending on the volume of related tweets. In addition, there is also a possibility of errors in the geographic precision.

Campelo and Baptista (2009) proposed a model for extraction of geographic knowledge from web documents. They developed GeoSEn, a search engine with geographic focus, which enables the geographic indexing of documents extracted from the web. Thus, it is possible to infer geographic locations cited in a text written in Portuguese in a political-division hierarchy, going from the least precise levels (Brazilian regions) to the most precise ones (cities).

As we can notice, there are several researches on identification of geographic location in social media messages focusing exclusively on the text. However, the majority does not address specific issues of Portuguese language. Furthermore, they do not also address LBSN domain and the aim of providing valuable information for such environments in an automated way. In this sense our proposal comes as a solution to cover this gap.

3. Automated Production of VGI based on Crowdsourced Social Media

This section presents our approach for automated production of VGI based on crowdsourced Social Media.

The main objective of this approach is the automated VGI production based on information published on social networks and focusing solely on microtexts. Thus, the expected result is the production of spatiotemporal markers with the content of these messages, which can be widely viewed and handled by the users of a LBSN. This spatiotemporal information may help users interested in learning more about specific geographical locations, for instance, through people who freely share information in social media. An illustration of the proposed approach is presented in Figure 1.

Figure 1. The main idea of our proposal: turning social network messages into spatiotemporal markers in a LBSN

Figure 1 (left side) illustrates the social networks as information sources for the production of the VGI visualized in a LBSN such as Crowd4City (right side), for example. In this context, each message posted by the users of these networks can be turned into a spatiotemporal marker, which can then be used by the users of the LBSN through recommendation and feedback actions, or just getting information. It is important to highlight that the VGI term in our work is related to the spatiotemporal markers that will be produced automatically by social media users who become volunteers even without necessarily access a LBSN.

In order to achieve such goal, it is necessary to have a computational processing involving capture and treatment of information and application of GIR techniques. This processing is illustrated in Figure 2.

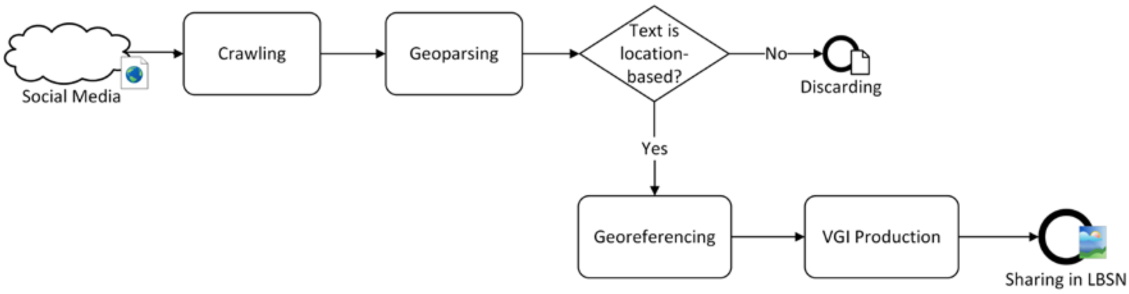

Figure 2. Computational processing flow for automated VGI production

As shown in the flow presented in Figure 2, the computational processing of this approach involves, basically, four distinct stages: Crawling, Geoparsing, Georeferencing and VGI Production. The initial stage is Crawling, in which occurs the capture of the messages posted on the social networks. We developed a real-time algorithm to capture microtexts (tweets) posted on Twitter[1]. This algorithm focuses on the original text of the messages posted on the network, discarding the other metadata of the tweets, except the timestamp containing the time the message was published.

[1] Twitter: http://www.twitter.com/

Once captured in the crawling stage, the microtexts are submitted to the geoparsing stage. In order to accomplish this stage, we used the GeoSEn Geoparser [Campelo and Baptista 2009], which is responsible for the detection of geographic terms in the process of parsing the analyzed texts written in Portuguese. At this stage, all the candidate locations are identified and then sent to the next stage, in which the text will be georeferenced. Figure 3 illustrates a microtext after the geoparsing stage, where the candidate locations are detected and highlighted.

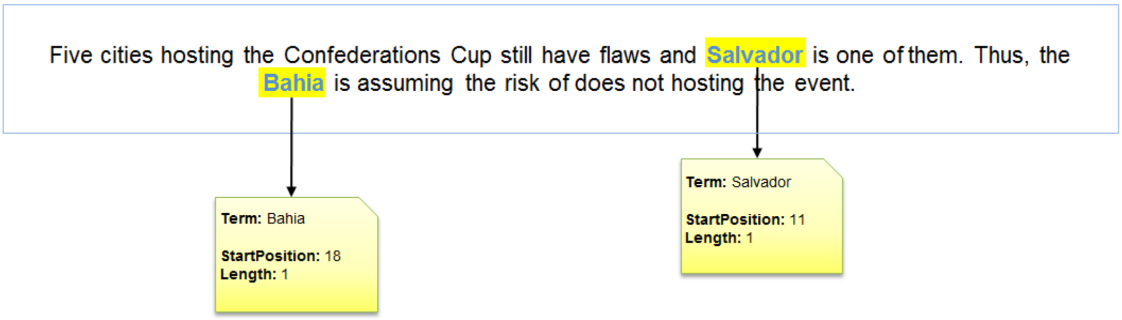

Figure 3. Result of the Geoparsing process applied to microtext (translated from Portuguese)

In Figure 3, it is possible to view all the candidate locations identified in the sample microtext. The Geoparser considers information such as the position of the term in the text and its length, that is, the number of words that form the term. Such position of the term can be used to correlate spatial terms which may appear closely in the messages. In the case where the geoparsing of a microtext returns an empty set of candidate locations, this microtext is discarded and its VGI production process is interrupted.

In the georeferencing stage, the candidate locations pass through a relevance evaluation in order to define the geographic scope of the microtext. In this stage, we used the Geo Scope Modeler featured by GeoSEn. The process of modeling the geographic scope explores the geographic hierarchy

city → micro-region → mesoregion → State → region

in order to generate the scope and compute the relevance for its highest levels, based on references found in lower levels. Therefore, the most precise geographic level which can be employed in the production of the VGI from a microtext is the City level. In order to georeference a microtext, the local gazetteer of GeoSEn containing all of the spatial data, structured according to this geographic hierarchy, was also used. The result of the georeferencing stage, applied to the sample microtext of Figure 3, is shown in Figure 4.

In Figure 4 we can notice that only one of the two candidate locations highlighted was considered for the georeferencing of the microtext. Since one of these locations (the city of Salvador) is inside the other one (the State of Bahia), the geographic scope modeling algorithm returned just the most geographically precise.

Finally, the VGI production stage is responsible for producing the spatiotemporal marker that will be shared on the LBSN. The marker is basically formed by the original microtext captured from the social network, the spatial data obtained in the georeferencing stage and the timestamp of the moment that the message was first published on the social network. For the generation of spatial markers, we compute the centroid points of the geometries georeferenced in the texts. Moreover, these markers produced automatically are assigned an exclusive type defined in the Crowd4City so that they can be easily distinguished from the types originally managed by the LBSN users, like education, transportation, security, etc. Thus, this exclusive type can highlight that the marker was not produced by a LBSN user.

Figure 4. Result of the georeferencing process applied to a microtext (translated from Portuguese)

A software application called *text2vgi* was implemented taking into account the whole flow illustrated in Figure 2, which was detailed throughout this section. The purpose of this application is to validate the proposed approach, confirm our raised hypothesis discussed previously in the introductory section, and identify points which may possibly need further improvements in order to ensure the most spatially-accurate VGI production.

4. Non-intentional VGI from non-geotagged microtexts – A Case Study

In this section, we present a case study using the *text2vgi* software application with microtexts from a social network.

4.1. Methodology

Our study used a dataset formed by 329,732 microtexts written in Portuguese, published on Twitter during the FIFA's Confederations Cup, which took place in Brazil in 2013. We chose to use this dataset because it is related to an event in which people normally use terms that can be associated to geographic location, such as the name of the host cities. The methodology used for conduction of this study is illustrated in Figure 5.

The Crawler implemented in *text2vgi* was responsible for capturing the messages and storing them in a local database. As the messages were received by the application, the geoparser was activated to identify the candidate locations. Then, the georeferencing module modeled the geographic scope of the microtexts that presented at least one candidate location. Finally, the VGI production module concluded the work creating the spatiotemporal marker.

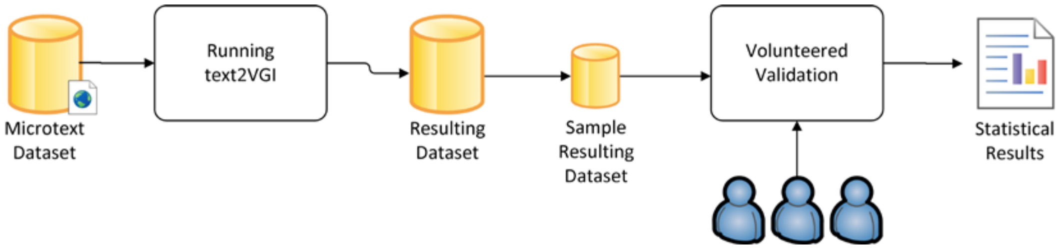

Figure 5. The process flow for the case study

4.2. Volunteered Validation

The whole set of microtexts processed by *text2vgi* needed to be validated concerning the identified geographic locations and the spatiotemporal markers created. Thus, we could measure the performance of the automated production of VGI. For such, we needed to recruit some volunteers and instruct them in the validation process. It was also necessary to develop a web application in order to assist these volunteers on validating the processing performed.

The application for volunteers' assistance presented a random list of processed microtexts, which were to be analyzed individually. For each one of these validated microtexts the volunteers answered the following questions: geoprocessing accuracy (boolean, star ☆ [1→5]), if it refers to more than one place (boolean) and if it can be more precise (boolean).

The geoprocessing accuracy question could receive the combination (TRUE, ☆☆☆☆☆) as answer in the cases in which the georeferencing was totally accurate according to the georeferencing strategy used, or in the cases where the VGI was not produced because the microtext did not express any geographic location. It could also receive the tuples (FALSE, [0 | ☆ | ☆☆ | ☆☆☆ | ☆☆☆☆]) as answer, depending on the geographic and semantic distances between the georeferencing location and the location identified manually by the volunteers on reading the microtext. One example of geographic distance is a microtext expressing the city of "Campina Grande" however it was georeferenced as "Paraíba" (the State) or "Nordeste" (the Region). The semantic distance is related to misunderstanding of the georeferencing such as a microtext expressing the "Bahia" (Football Team) in which was georeferenced as "Bahia" (the State) instead of non-location.

The question about whether a microtext refers to more than one place could receive TRUE when the microtext refers to more than one geographic location and, therefore, would allow the production of more than one spatiotemporal marker for the same microtext; and receives FALSE, otherwise. Finally, the question about whether a microtext can be more precise could receive TRUE when the microtext presents evidence that might make the modeling of the geographic scope more precise at city level, such as neighborhood names, streets, or specific buildings, such as parks, squares, stadiums and tourist spots.

From the whole set of processed microtexts, 2.3% (about 7,500) had at least one geographic location automatically assigned by *text2vgi* and could then produce spatiotemporal markers. It is important to highlight that there might be several microtexts which did not express a geographic location and such fact can explain this rate. The volunteered validation may help to understand this aspect.

Considering the huge volume of microtexts of the dataset used in this study, a random sample of these microtexts needed to be defined so it could then be validated by the volunteers. With a trust level of 99% and a sampling error of 0.65%, the sample validated by the volunteers consisted of 35,120 microtexts. In this sample, 975 microtexts (2.7%) had a geographic location automatically assigned by *text2vgi*, nearly the same proportion presented by the whole set of processed microtexts. Since the validation was performed by humans, we still consider a margin of error of 2.0%.

4.3. Results

The mean processing time of each microtext in *text2vgi*, from the moment of the capture of the message to the production of the spatiotemporal marker, was of 0.25 seconds. It took nearly 23 hours to process the whole dataset in only one computer, with an Intel Core i7 processor, 8 GB of RAM and 1 single thread.

Considering the sample validated by the volunteers, Figure 6 presents the results for true positives, when the geographic location was identified correctly; false positives, when the geographic location was not identified correctly; true negatives, when there was no geographic scope assigned due to the lack of evidence in the text; and false negatives, when no geographic scope was assigned, but there was evidence for it.

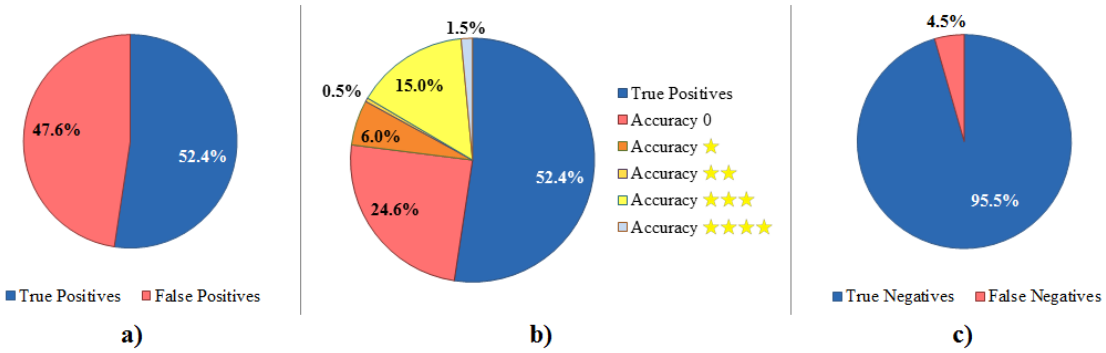

Figure 6. Pie charts representing the percentages of each result: a) True/False Positives Relation, b) True/False Positives Relation considering the False Positives in five subdivisions, and c) True/False Negatives relation

Figure 6a show that there was a balance between true and false positives, if we consider as true positives only the 100% precise location detections. In Figure 6b, it is possible to see the false positives in five classifications levels. Each classification level represents how geographically close the false positive was to a true positive. We can notice the false positives that are very far from the location expressed in the microtext (which received no stars in the accuracy question), represent only 24.6% - about half the total number of false positives. Finally, in Figure 6c, it is possible to observe a good result for true negatives. It confirms the lower rate of the processed microtexts which had at least one geographic location automatically assigned by *text2vgi*: in fact there were several microtexts that did not express one location at least.

The validation performed by the volunteers on the microtexts also resulted in the following data:

- 16.6 % of the microtexts have evidence for georeferencing of more detailed geographic locations. A georefececing strategy which takes this aspect into account may improve the overall accuracy;
- 3.2 % of the microtexts have evidence for the inference of more than one geographic location, thus producing more than one spatiotemporal marker.

Table 1. Statistical results of volunteered validation over VGI produced automatically

Overall Accuracy	Precision	Recall	F-Measure
74.1 %	92.3 %	52.6 %	0.67

Table 1 presents four metrics for evaluating the overall performance of VGI produced automatically by *text2vgi* and validated by volunteers during this case study. Among the analyzed metrics, we can notice a low recall rate, that is, 47.4% of the microtexts with geographic location evidence were not correctly identified by *text2vgi*.

Nevertheless, this result was already expected, since the geographic scope considered in the georeferencing strategy used considers only locations related to the Brazilian political territorial division. The geographic references that may be expressed in the set of microtexts used such as soccer stadiums and airports ended up not being properly interpreted. However, it is important to highlight the good precision rate resulted, which is justified by the number of true negatives.

5. Conclusion and Further Work

In this paper, we presented an approach for automated VGI production based on geoparsing and georeferencing of texts published on the web. Such approach was conceived with the objective of turning web authors into volunteers in the VGI context, contributing to the indirect production of information in a Location-based Social Network.

A prototype, called *text2vgi*, was implemented with the goal of validating the ideas proposed by our approach. In order to evaluate the prototype in a real context, we carried out a case study using a set of microtexts in the Portuguese Language concerning a sporting event of large impact on media, the 2013 FIFA’s Confederations Cup, held in Brazil.

Overall, the achieved results were considered satisfactory. However, we have confirmed the need for improving the georeferencing strategy in order to increase the amount of VGI produced from microtexts, to improve the spatial accuracy of the spatiotemporal markers created and to achieve better results for the recall and F-Measure. It is important to consider points of interest such as soccer stadiums and airports, and other buildings and well known places in a city context. Thus, the automatically produced VGI will become more spatially precise and the user's experience in the LBSN will be improved.

As future work, we consider the implementation of georeferencing strategies to address the specific treatment of microtexts like informal language. Besides, we will seek the development of heuristics that increase the precision of the locations detected, and consequently improve the F-Measure. Other future direction of our work is to improve our approach for production of VGI based on microtexts in other languages such as English, Spanish and French.

References

Ballatore, A., Wilson, D. C. and Bertolotto, M. (2013) “Computing the semantic similarity of geographic terms using volunteered lexical definitions”, International Journal of Geographical Information Science, Taylor & Francis, vol. 27, no. 10, p. 2099-2118.

Ballatore, A. and Betolotto, M. (2011) “Semantically enriching VGI in support of implicit feedback analysis”, Web and Wireless Geographical Information Systems, LNCS, vol. 6574, p. 78-93.

Bordogna, G., Ghisalberti, G. and Psaila, G. (2012) “Geographic information retrieval: Modeling uncertainty of user's context”, Fuzzy Sets and Systems, vol. 196, p. 105-124.

Campelo, C. E. C. and Baptista, C. S. (2009) “A Model for Geographic Knowledge Extraction on Web Documents”, In: Advances in Conceptual Modeling - Challenging Perspectives, LNCS 5833, Edited by Carlos Alberto Heuser and Günther Pernul, Springer Berlin Heidelberg, p. 317-326.

Du, H., Anand, S., Morley, J., Leibovici, D., Hart, G. and Jackson, M. J. (2011) “Developing open source based tools for geospatial integration”, In: Proceedings of the 3rd Open Source GIS Conference, Anonymous Nottingham University, UK.

Falcão, A. G. R., Baptista, C. de S. and Menezes, L. C. de (2012) “Crowd4City: Utilizando Sensores Humanos como Fonte de Dados em Cidades Inteligentes” (in Portuguese), In: Proceedings of the 8th Brazilian Symposium on Information Systems, São Paulo, Brazil.

Freire, N., Borbinha, J., Calado, P. and Martins, B. (2011) “A Metadata Geoparsing System for Place Name Recognition and Resolution in Metadata Records”, In: Proceedings of the 11th Annual International ACM/IEEE JCDL, Ottawa, Canada, p. 339-348.

Goodchild, M. F. (2007) “Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0”, International Journal of Spatial Data Infrastructures Research, vol. 2, p. 24-32.

Haklay, M. (2010) “How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets”, Environ Plan B Plan Des, vol. 37, no. 4, p. 682-703.

Haklay, M., Ather, A. and Basiouka, S. (2010) “How many volunteers does it take to map an area well?”, In: Proceedings of the GIS Research UK 18th Annual Conference, University College London, UK, p. 193-196.

Haklay, M. and Weber, P. (2008) “OpenStreetMap: user-generated street maps”, IEEE Pervasive Computing, vol. 7, no. 4, p. 12-18.

Havlik, D., Soriano, J., Granell, C., Middleton, S. E., van der Schaaf, H., Berre, A. J. and Pielorz, J. (2013) “Future Internet enablers for VGI applications”, In: EnviroInfo, Edited by B. Page, A. G. Fleischer, J. Göbel and V. Wohlgemuth, Shaker, p. 622-630.

Horita, F. E. A., Degrossi, L. C., Assis, L. F. G. de, Zipf, A., Albuquerque, J. P. (2013) "The use of Volunteered Geographic Information (VGI) and Crowdsourcing in Disaster Management: a Systematic Literature Review", In: Proceedings of 19th AMCIS, Chicago, USA, p. 1-10.

Jackson, M. J., Rahemtulla, H. and Morley, J. (2010) “The Synergistic use of authenticated and crowd-sourced data for Emergency response”, In: Proceedings of the 2nd Int. Workshop on Validation of Geo-Information Products for Crisis Management (VALGEO), European Commission Joint Research Centre, Ispra, Italy, p. 91-99.

Jung, J. J. (2011) “Towards named entity recognition method for microtexts in online social networks: a case study of Twitter”, In: Proceedings of the International Conference on Advances in Social Network Analysis and Mining, p. 563-564.

Koukoletsos, T., Haklay, M. and Ellul, C. (2012) “Assessing Data Completeness of VGI through an Automated Matching Procedure for Linear Data”, Transactions in GIS, vol. 16, no. 4, p. 477-498.

Liu, L., Gao, Y., Lin, X., Guo, X. and Li, H. (2013) “A framework and implementation for qualitative geographic information retrieval”, In: Proceedings of 21st International Conference GeoInformatics, p. 1-4.

Mooney, P. and Corcoran, P. (2012) “The Annotation Process in OpenStreetMap”, Transactions in GIS, vol. 16, no. 4, p. 561-579.

Parker, C. J., May, A. J. and Mitchell, V. (2011) “Relevance of volunteered geographic information in a real world context”, In: Proceedings of GIS Research UK 19th Annual Conference, University of Portsmouth, UK.

Purves, R. and Jones, C. B. (2011) “Geographic Information Retrieval”, SIGSPATIAL Special, vol. 3, no. 2, p. 2-4.

Rupp, C. J., Rayson, P., Baron, A., Donaldson, C., Gregory, I., Hardie, A. and Murrieta-Flores, P. (2013) “Customising Geoparsing and Georeferencing for Historical Texts”, In: Proceedings of the International Conference on Big Data, IEEE, p. 59-62.

Surowiecki, J. (2005), The Wisdom of Crowds, Anchor.

Vicente, C. R., Freni, D., Bettini, C. and Jensen, C. S. (2011) “Location-Related Privacy in Geo-Social Networks”, IEEE Internet Computing, vol. 15, no. 3, p. 20-27.

Watanabe, K., Ochi, M., Okabe, M. and Onai, R. (2011) “Jasmine: a real-time local-event detection system based on geolocation information propagated to microblogs”, In: Proceedings of the CIKM’11, Glasgow, Scotland, p. 2541-2544.