



Multi Agent-Based Addresses Geocoding for More Efficient Home Delivery Service in Developing Countries

Al Mansour Kebe^{1(✉)}, Roger M. Faye², and Claude Lishou¹

¹ Ecole Supérieure Polytechnique, Université Cheikh Anta Diop, Dakar, Senegal
manskebe@gmail.com

² Université Amadou Mahtar MBOW, Diamniadio, Senegal

Abstract. In this study, we present an original method that enhance geocoding system in poorly mapped areas thanks to multi-agent system. In contrast with industrialized countries, many developing countries lack formal postal address systems assignments and usage, making the operation of translating text-based addresses to absolute spatial coordinates, known as geocoding, a big challenge. We recreated a standard of address as it is perceived and used by local people, a kind of non-official national address standard since there is no official one in these areas. Then, we designed a multi agent system in which agents are assigned different tasks of geocoding process and can perform negotiation to achieve global objective: find the best possible match or approximation of a location based on current knowledge. A verification of the usefulness of the proposed approach is made in comparison with Google geocoding API which shows that the proposed approach has great potential to geocode addresses considering local context semantic issues.

Keywords: Geocoding · Multi agent · Text mining · Knowledge discovery · Address standard

1 Introduction

From standard service delivery to emergency system dispatching, addresses are the most common and convenient way to locate people. Addresses are easily comprehensible to people, but not directly suitable for use in an IT environment. Translating text based addresses to absolute spatial coordinates is known as geocoding. However geocoding technology is very common on commercial Geographical Information Systems (GIS) products, it's usually limited to fully standardized structures with a respect of elements order as well as writing style.

In contrast with industrialized countries, many address systems in developing countries lack standards, making the addressing system ambiguous, incomplete or imprecise. The prompt provision of a precise location from unstructured or even vague addressing data provided by people is critically important, especially in emergency situations, and have a socio-economic impact in day-to-day life.

Furthermore, with the rapid development of e-commerce and internet, products and services delivery to customers is a challenge for companies, therefore increasing the

need for appropriate geocoding tools in poorly mapped area. In fact, even if online mapping system is widely used, many developing countries suffer from good addressing and geocoding system, even internet connectivity.

Due to the essential life services they provide, water, energy and garbage collection utility companies have mastered addressing and geocoding systems for years now. Through their day-to-day interaction with customers to quickly troubleshoot water supply breakdowns or electricity distribution failures, these utility companies need to have a good knowledge of the field they operate in. To deal with this permanent challenge, these companies have to create their own geolocation systems that become over the years an important source of geographical and customers' addresses data.

Hence, we investigate ten years of addresses data collected by Senegal power company through customer's trouble calls. On any power outage, verbal description of their location is given by customer to callcenter officers without any account references. This collected data is characterized by its vagueness and full of misspelling or unknown places from common mapping facilities.

To express these challenges, we first needed to produce an address standard by text mining historical data and creating a reference dataset before being able to achieve the matching operations.

Afterwards, we designed a multi agent system in which agents are assigned different tasks of geocoding process and can perform negotiation to achieve global objective: find the best possible match or approximation of a location based on current knowledge.

The rest of the paper is organised as follow: Sect. 2 presents the background of the study including geocoding process and literature review. In Sect. 3, the model of the proposed system is set out with a definition of a standard address format from data and description of the proposed multi agent system with its implementation. In Sect. 4, experiments to measure the accurateness of the proposed approach is carried out in comparison to the results of Google Maps Geocoding API against a set of local addresses for which we have verified latitude-longitude and we discuss results.

2 Background

2.1 The Process of Geocoding and Its Challenge

As well documented in the literature [1–3], within classical approach, the geocoding process is divided in three main phases: address normalization, address matching and address locating (Fig. 1).

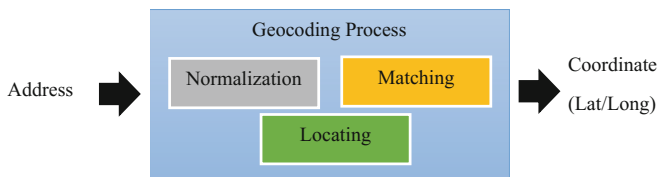


Fig. 1. The process of geocoding

The normalization phase transforms the address to a standardized form which allows to accomplish the comparison with reference data. Because addresses are not free from errors, they must be cleaned and standardized to get the same format as the reference data. The cleaning process involves solving issues such as case variation, abbreviation and punctuation. Standardization is accomplished by tokenizing the address string and assigning meaning to each token from different address element types.

Matching phase is the attempts to link the normalized address to a corresponding record from the reference data set. The performance of this stage depends on the completeness of the reference data. If exact matching cannot be obtained, approximation is made to get the best available adjacent area in the data set.

Locating process allow to return a geocode, thanks to geographic coordinate assigned to matched address in reference data. The geographic coordinate goes from polygons representing locality (which have a coordinate assigned to his centroid) through line segment representing street (coordinate assigned to his mid-point) to point representing single address.

From this brief description of geocoding process, we can notice that we face two big challenges when attempting to geocode address in poorly mapped area:

- Input address must be converted to a standardized format but there is no standard address in many developing countries (as reported in [4]).
- Standardized address should be compared to a reference data set but since these areas are poorly mapped, incomplete database with unknown places are very common from all mapping facilities.

2.2 Related Work

Geocoding is a well-studied question with plenty of contributions that have been proposed to expand the process specially in the domain of classical geocoding [1] but less in what we can call intelligent geocoding which deal with geocoding difficult address by using control and knowledge improvements [3, 5] to depart from simply matching and table lookups approaches.

Since addressing systems vary largely between countries, as addresses have a strong cultural bias, one main part of studies revolve around some issues introduced by the natural language in the process of geocoding such as in Chinese [6–8] or Croatian [9] and more specifically on twitter [10].

Another part of the current geocoding literature deals with geocoding application and issues solving in an explicit geographic area such as in Croatia [11], Turkish [12], Brazil [13], China [6–8], Australia [14], South Africa [15], Morocco [16], Cuba [17] and India [18]. Most of these countries are developing ones and have to deal with rapid urban growth which introduces problems such as ambiguous region boundaries and lack of convention in spellings of toponyms. Concerning developed area, the studies deal mainly with the verification of suitability of online geocoding tools [2] for a specific region as we can find in [11, 19, 20] for Austria, Quebec and Germany. Address Standard issues are also discussed in [1, 4, 21–23].

This literature survey shows a strong dynamic in geocoding process from regional and cultural point of view. Thus, to improve the geocoding course, knowledge need to be added in the process to take into consideration lack of standard, semantic issues and complicated logic existing in many countries. Therefore, Multi Agents System (MAS) which utilizes theories and concepts from many areas such as computer science, artificial intelligence, distributed systems, social sciences, economics, organization and else is a good candidate for this purpose.

In this tendency, in [5] Wei et al. present a knowledge-based agent prototype for Chinese address geocoding. Toward the statement that Chinese address geocoding is a difficult problem to deal with due to intrinsic complexities in Chinese address systems including Chinese language and civil history and a lack of standards in address assignments and usages, the authors propose a spatial knowledge-based agent prototype to improve existing address geocoding algorithm. To construct this agent, they first introduce a knowledge base consisting of a basic ontology for Chinese address validation domain and an internal fact database. An inference rule set is integrated into this agent to deduce the spatial accuracy of these potential matches. However, this approach is limited to some inferential ability to help matching process and had to be improved to solve issue like geocoding addresses containing two or more geographical classes of the same kind.

In [24] Hutchinson and Veenendaal present an agent-based framework for intelligent geocoding. Having noted that despite progress in the field of geocoding, there remain a sizable proportion of addresses that are difficult to geocode due to missing information and wrong addresses, they explore how agent-based processing, which utilizes the belief, desire, intention (BDI) model, can add intelligence to the geocoding process. The goal of the system is to correct address element allowing to find them in reference data set and every agent pursues its own intention to this goal. Nevertheless, this study is confined to geocoding matching process assuming that, it will take longer but complex sites and rural sites will have reliable geocode data.

Our work extends those of Hutchinson and al which is conducted in context of a developed country with a comprehensive geocode data updated every three months (the Australia G-NAF files). In contrast with Hutchinson we are in a poorly mapped area [25] which conduct to include more dynamics in all geocoding process (data normalisation, data set acquisition and cleaning).

3 Proposed Approach

3.1 Definition of an Address Standard

To define an address standard, an ontology for this domain is constructed as the basic vocabulary to represent spatial knowledge in the address geocoding domain. Hence we analysed ten years of addresses data collected by Senegal public energy utility through customer's trouble calls. This collected data is characterized by its sketchiness and full of misspelling and unknown places from common online mapping. This basic ontology includes geographical classes such as city, county, road, house number, community, building number, points of interest (POI) and other concepts needed for comparing and evaluating address matches.

Example of address data from the trouble call data corpus:

- Guediawaye Cite des Enseignants Villa N° D/10
- Yeumbeul Ben Barack Darou Salam 4/C Qrt Elhadj Ablaye Diop
- Ngor Diongoran Qrt Pape Moustapha Ba
- Yeumbeul Cite Comico 4 Villa N°129/D

Thanks to the determination of most often used terms we have identified the terminology used to describe an address.

Words like “villa” (house), “qrt” (abbreviation of French word quartier which mean neighbourhood), ‘rue’ (street) are in focus.

Many mistakes and spellings are encountered in the data due to typos and abbreviations knowing that they are collected verbally by phone from call center.

In order to reunite the various spelling of frequent words, language processing algorithms have been used such as Levenshtein distance and Jaro-Winkler distance combined with TF-IDF (Term Frequency-Inverse Document Frequency) [26].

We classify these address concepts in family and settle a hierarchy between them. Considering the current national administrative division of the country including Administrative district and Municipality we get the Fig. 2 with the main families in focus: property, subdivision, administrative division, space for circulation, municipality and administrative district.

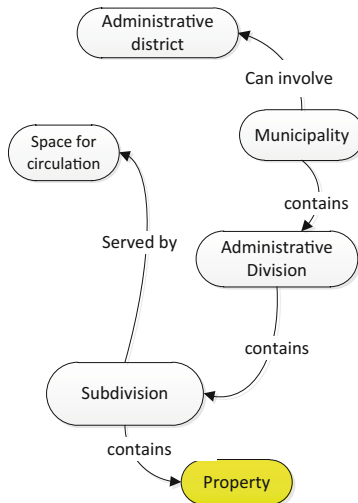


Fig. 2. Simplified addressing concept hierarchy

The developed version of the addressing concept is presented in Fig. 3 with a detail of words used frequently in each family of concept giving an ontology of how address is nowadays expressed by Senegalese People.

This ontology is an input used by the multi agent system to accomplish the goal of the LOGEMAS system.

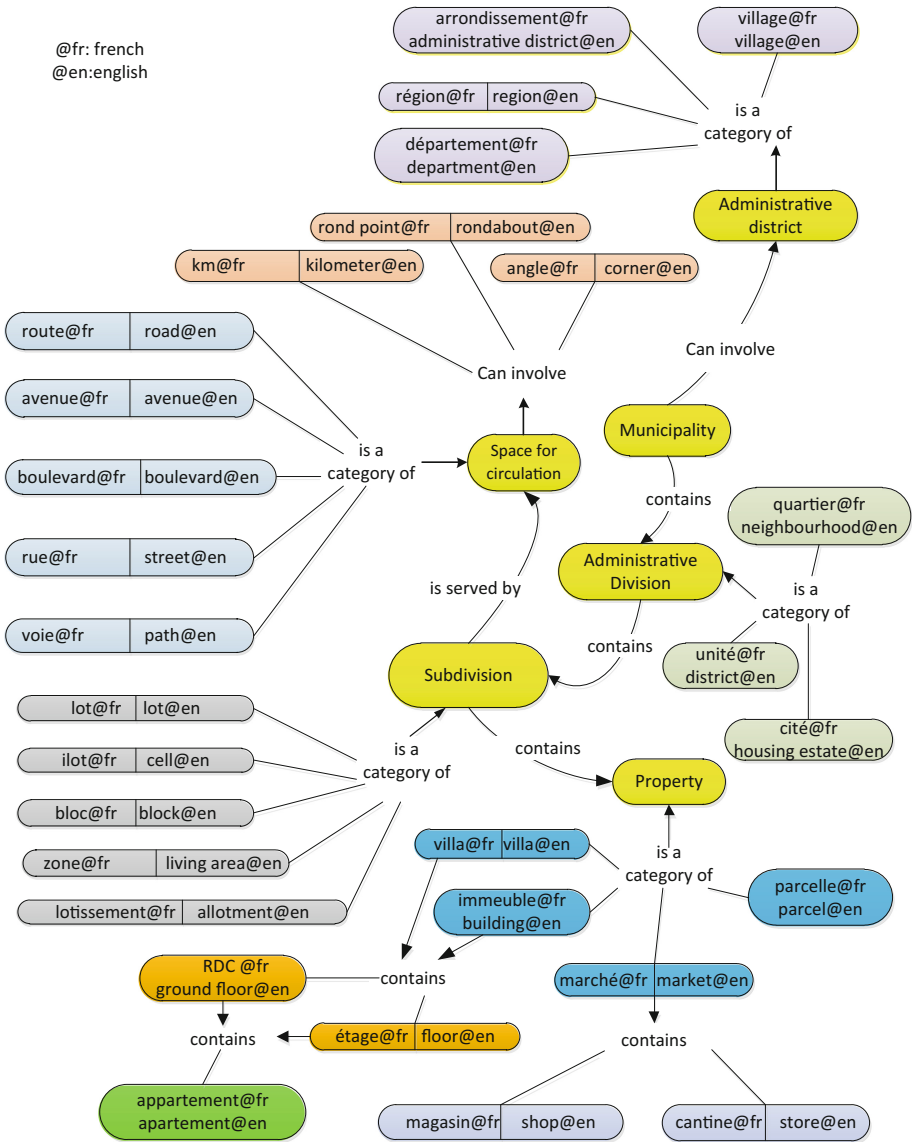


Fig. 3. Ontology of urban geography in Senegal

3.2 LOGEMAS-Location Geocoding with Multi Agent System

In this section, we present the LoGeMAS (Location Geocoding with Multi Agent System) architecture in which the various geocoding tasks are delegated to different agents.

Multi-agent Architecture. To achieve the task of geocoding, agents are assigned different tasks of geocoding process and can perform negotiation through FIPA protocol to achieve global objective: text mining and pattern recognition to find the best possible match or approximation of a location based on current knowledge.

Figure 4 depicts the interaction in the model composed of five type of agents.

- **Supervisor Agent (SA):** his role is to manage the different processes through a well-defined sequence of events. As a coordinator agent, SA dispatches the different tasks to other agents.
- **Cleaning Agent (CLA):** his role is to clean the provided address, as in the normalization phase in geocoding process (see Sect. 2). Algorithm used by CLA include transforming text to lower case and deleting all special characters.
- **Keyword Finder Agent (KFA):** his objective is to find the different entity of the ontology, called keyword, present in the provided address and to classify them as described by the knowledge provided by ontology (see Fig. 4).

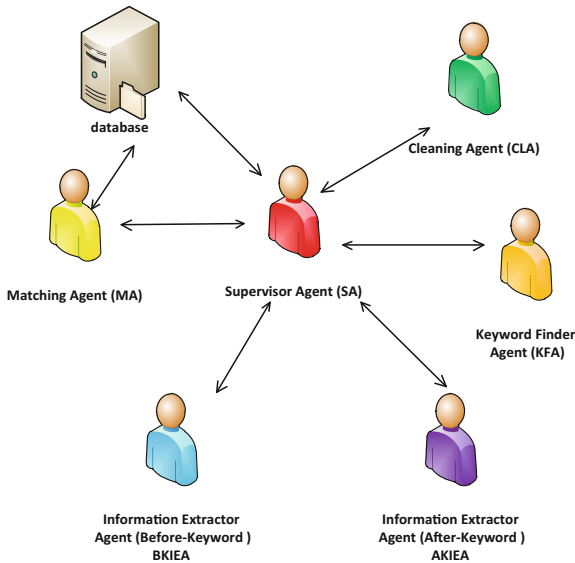


Fig. 4. Multi agent LOGEMAS model

- **Before Keyword Information Extractor Agent (BKIEA):** as his name shows, this is an information extractor agent specialised on information appearing before a keyword. Example in the address:
 “Guediawaye **Cite** des Enseignants **Villa** N° D/10”
 BKIEA will retrieve the couple [‘Guediawaye’, ‘des Enseignants’]
- **After Keyword Information Extractor Agent (AKIEA):** as his name show, this is an information extractor agent specialised on information appearing after a keyword. Example in the address:

“Guediawaye **Cite** des Enseignants **Villa** N° D/10”

AKIEA will retrieve the couple [‘des Enseignants’, ‘N° D/10’]

- **Matching Agent (MA):** The Matching Agent role is to do the matching action by finding geocode corresponding to address as described in geocoding process in Sect. 2. From each found keyword completed with his proprieties, information are looked in the database and the process is completed with heuristics rules that allow to define the accuracy.

3.3 Data Set Acquisition

To track crew performance during outage recovery, Senelec utility has introduced a vehicle tracking system in 2015. The data collected during breakdown repairation combined with these captured tracking data gives a potential of learning new regions and landmarks by reverse geocoding concerned customer address.

3.4 The Multi-agent System Design and Implementation

To implement a multi-agent system, there are several open-source agent platforms available in the literature that helps developers to build a complex agent system in a simplified manner. The GAMA platform [27, 28] was chosen to implement our system. This platform which is a new trend in multi agent developing, enables multi agent development through an intuitive interface and programing language GAML. GAMA has the advantage to implement many features like FIPA compliance, Geographical Information System (SIG) and the ability to build spatially explicit multi-agent simulation which is one characteristic of geocoding system.

4 Experiment Result and Discussion

To challenge our approach and demonstrate its capability, a case study in Dakar City Senegal is carried out with text mining and pattern recognition.

First we measure the ability of the system to geocode up to different levels of precision from the proposed address standard including property, subdivision, administrative division, space for circulation, municipality and administrative district. To achieve this test, we run the LOGEMAS process through one month of trouble call data containing 6321 addresses (June 2015 data). The test is compliant since only 23 addresses were not recognised. The Fig. 5 shows an example of the results provided by LOGEMAS system.

```

Address : 'HLM 5 VILLA 2209 angle 2'
hlm5 is : ['quartier', 'unite', 'cite']**Familly: administrative district
2209 is villa **Familly: property
2 is angle **Familly: space for circulation

```

Fig. 5. Sample of address recognition through LOGEMAS process

When analysing the address that was not well recognized we found that they had in common the fact of containing natural language part of speech like in ‘Yarakh Hann Pecheur ne connait pas le nom du Chef de qrt’ (meaning Yarakh Hann Pecheur doesn’t know the name of the head of district!!!).

Finally, a comparison of LOGEMAS results with those of Google’s geocoding service against a set of addresses verified latitude-longitude coordinates was conducted and we arrive at these conclusion:

- Google Maps was not able to consider semantic issue while LOGEMAS well recognised standard address as expressed by local people;
- Google geocoding service is limited to street name, municipality and completed Point of Interest (POI) and LOGEMAS try to give a level of precision from its known database.

We can conclude that the architecture we are developing has great potential to geocode addresses considering local context issues with integration of semantic question (Fig. 6).

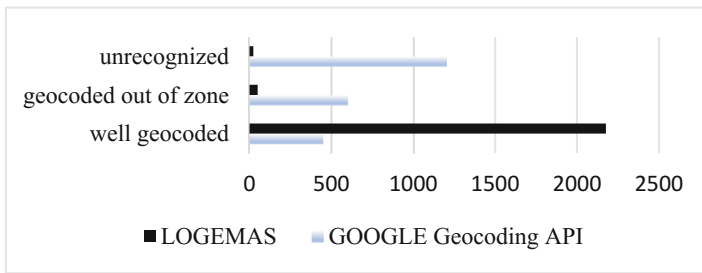


Fig. 6. Comparison between LOGEMAS and Google Maps Api geocoding process results

5 Conclusion and Future Work

In this paper, we presented a system called LOGEMAS - Location Geocoding with Multi Agent System. We investigated ten years of addresses data collected by Senegal public energy utility through customer’s trouble calls. We recreated a standard of address as it is perceived and used by people. Afterwards, we designed a multi agent system in which agents are assigned different tasks of geocoding process and can perform negotiation to achieve global objective: find the best possible match or approximation of a location based on current knowledge. A verification of the usefulness of the proposed approach is made in comparison with Google geocoding API which shows that the proposed approach has great potential to geocode addresses considering local context semantic issues and can contribute to more efficient home delivery service in developing countries. In future work the tool will be enhanced with more semantic skills and propose an application to outage management in power companies.

References

1. Goldberg, D.W.: A Geocoding Best Practices Guide. North American Association of Central Cancer Registries (2008)
2. Karimi, H.A., Sharker, M.H., Roongpiboonsopit, D.: Geocoding recommender: an algorithm to recommend optimal online geocoding services for applications. *Trans. GIS* **15**(6), 869 (2011)
3. Hutchinson, M.J.: Developing an agent-based framework for intelligent geocoding. Curtin University of Technology (2010)
4. Coetzee, S., et al.: Towards an international address standard. In: 10th International Conference for Spatial Data Infrastructure (2008)
5. Wei, R., Zhang, X., Ding, L., Ma, H., Li, Q.: A knowledge-based agent prototype for Chinese address geocoding. In: *Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Advanced Spatial Data Models and Analyses*, p. 71,461D. International Society for Optics and Photonics (2009)
6. Tian, Q., Ren, F., Hu, T., Liu, J., Li, R., Du, Q.: Using an optimized Chinese address matching method to develop a geocoding service: a case study of Shenzhen. *ISPRS Int. J. Geo Inf.* **5**(5), 65 (2016)
7. Gao, D., Li, Q.: Preliminary results of a Chinese address segmentation algorithm based on self-organizing neural network. In: *Geoinformatics 2007*, p. 67,530U. International Society for Optics and Photonics (2007)
8. Fang, L., Yu, Z.Y., Zhao, X.: The design of a unified addressing schema and the matching mode of China. In: 2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3987–3990. IEEE (2010)
9. Ugrina, I., Zigo, M.: Searching for semantically correct postal addresses on the Croatian web. In: *Central European Conference on Information and Intelligent Systems*, p. 276. Faculty of Organization and Informatics Varazdin (2014)
10. Zhang, W., Gelernter, J.: Geocoding location expressions in twitter messages: a preference learning method. *J. Spat. Inf. Sci.* **2014**(9), 37 (2014)
11. Cetl, V., Kliment, T., Jogun, T.: A comparison of address geocoding techniques—case study of the city of Zagreb. *Surv. Rev.*, 1–10 (2016)
12. Yildirim, V., Yomralioglu, T., Nisanci, R., Inan, H.: Turkish street addressing system and geocoding challenges. *Proc. Inst. Civ. Eng.* **167**(2), 99 (2014)
13. Davis Jr., C.A., de Alencar, R.O.: Evaluation of the quality of an online geocoding resource in the context of a large Brazilian city. *Trans. GIS* **15**(6), 851 (2011)
14. Christen, P., Churches, T., Willmore, A., et al.: A probabilistic geocoding system based on a national address file. In: *Proceedings of the 3rd Australasian Data Mining Conference*, Cairns (2004)
15. Coetzee, S., Cooper, A.K.: What is an address in South Africa? *S. Afr. J. Sci.* **103**(11–12), 449 (2007)
16. Malaainine, M.E.I., Rhinane, H., Baidder, L., Lechgar, H.: Omt-g modeling and cloud implementation of a reference database of addressing in Morocco. *J. Geogr. Inf. Syst.* (2013)
17. de Armas Garcia, C.J., Gutierrez, A.A.C.: Deployment of a national geocoding service: Cuban experience. *URISA J.* **25**(1), 53 (2013)
18. Chatterjee, A., Anjaria, J., Roy, S., Ganguli, A., Seal, K.: SAGEL: smart address geocoding engine for supply-chain logistics. In: *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, p. 42. ACM (2016)

19. Burns, S., Miranda-Moreno, L., Stipancic, J., Saunier, N., Ismail, K.: Accessible and practical geocoding method for traffic collision record mapping: Quebec, Canada, case study. *Transp. Res. Rec. J. Transp. Res. Board* **2460**, 39 (2014)
20. Ahlers, D., Boll, S.: On the accuracy of online geocoders. *Geoinformatik 2009* (2009)
21. Davis, C.A., Fonseca, F.T.: Assessing the certainty of locations produced by an address geocoding system. *Geoinformatica* **11**(1), 103 (2007)
22. Karimipour, F., Alinaghi, N., Weiser, P., Frank, A.U.: *Cognitive engineering for spatial information processes: from user interfaces to model-driven design* (2015)
23. Coetzee, S., Rademeyer, M.: Testing the spatial adjacency match of the Intiendo address matching tool for geocoding of addresses with misleading suburb or place names. In: *Proceedings International Cartography Conference*, pp. 15–21 (2009)
24. Hutchinson, M.J., Veenendaal, B.: An agent-based framework for intelligent geocoding. *Appl. Geomat.* **5**(1), 33 (2013)
25. Irwin, M.: Mapping missing maps (2014). <https://www.mapbox.com/blog/mapping-missing-maps/>, Accessed 20 Jan 2017
26. Graham, S.A.: *String Searching Algorithms*. World Scientific Publishing Co. Pte Ltd. (1994)
27. Kravari, K., Bassiliades, N.: A survey of agent platforms. *J. Artif. Soc. Soc. Simul.* **18**(1), 11 (2015)
28. Documentation sur la plateforme GAMA (2016). <https://github.com/gama-platform/gama/wiki>, Accessed 4 Dec 2016