



HAL
open science

Using Geocoding and Topic Extraction to Make Sense of Comments on Social Network Pages of Local Government Agencies

Pedro Lima, Raissa Barcellos, Flavia Bernardini, Jose Viterbo

► To cite this version:

Pedro Lima, Raissa Barcellos, Flavia Bernardini, Jose Viterbo. Using Geocoding and Topic Extraction to Make Sense of Comments on Social Network Pages of Local Government Agencies. 17th International Conference on Electronic Government (EGOV), Sep 2018, Krems, Austria. pp.263-274, 10.1007/978-3-319-98690-6_22 . hal-01961527

HAL Id: hal-01961527

<https://inria.hal.science/hal-01961527v1>

Submitted on 20 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Using Geocoding and Topic Extraction to Make Sense of Comments on Social Network Pages of Local Government Agencies

Pedro C. R. Lima¹, Raissa Barcellos²[0000-0003-1710-2161],
Flavia Bernardini²[0000-0001-8801-827X] and Jose Viterbo²[0000-0002-0339-6624]

¹ Institute of Science and Technology – ICT, Fluminense Federal University – UFF
Rio das Ostras, RJ, Brazil

² Institute of Computing – IC, Fluminense Federal University – UFF
Niteroi, RJ, Brazil

pedrocrl@gmail.com, raissabarcellos@id.uff.br, {fcbernardini,viterbo}@ic.uff.br

Abstract. Social networks have become an important channel for exchanging information and communication among citizens. Text mining, crowdsourcing and data visualization are some approaches that allow the information and knowledge extraction from texts in comment formats, exchanged between citizens in social networks. This movement can be indirectly used as a bias for popular participation, gaining prominence in the construction of smart cities. The objective of this work is to present a method that geocodes citizens' comments made on posts in Social Network Pages of Local Government Agencies, and extracts the most frequent topics present in these comments. In order to validate our method, we implemented a web system that implements the steps of the proposed method, and conducted a case study. The tool, and consequently the steps of the presented method, was evaluated by four software developers, which indicated that the tool was easy to use, new knowledge could be extracted from it, and some interesting improvements were pointed out by them.

Keywords: Social Networks, Data Analysis, Information Visualization, Smart Cities, E-Participation

1 Introduction

The Smart City concept still draws much attention when it defines urban development policies and popular participation. In this context, Teran, Kaskina and Meier [1] propose a maturity model for Cognitive Cities, strongly based on a maturity model for electronic government, which values popular participation in government decision-making processes. Dameri and Rosenthal-Sabroux [2] discuss the issue of building public value by citizens, which can only be acquired by popular participation. Currently, social networks offer a range of information, provided by citizens, on pages of local governments. However, this source of information is still little explored to understand the quality of the provided services, and leverage new public services.

We could observe several posts by citizens on social networks, specifically in some local government agencies pages, suggesting or complaining about public services provided to citizens. Reddick, Chatfield and Ojo [3] present a conceptual framework for using social media text mining analytics and visualization e-participation. Their study case showed that using analytics techniques for mining Social Networks to gather information from citizens are quite useful. So, exploring techniques for detecting frequent mentioned topics in social network pages of local government agencies is an interesting approach. Also, data visualization has been explored in many works for leveraging open data exploration and interpretation [4][6], which can also be used in this scenario.

In this work, we propose a method to analyze user's comments, published on a social network page of a local government agencies. We premised that some comments on posts in such pages often contain comments about locations in a city that are not necessarily related to the related post. The method comprises the identification of (i) the locations (streets, avenues and others) mentioned in the comments, to display those comments in maps; and (ii) the most frequent topics. For both tasks, we used techniques from Text Mining area for automatic geocoding the comments and automatic extraction of the topics. Geocoding comments is an interesting approach even when the comments has metadata indicating the location it was posted, because not necessarily the comment is related to the place where the user is located when posting the message. For instance, someone could comment, when arrives at home, that on the street where he/she works there is a large hole. Tools like Google Maps API are able to show points in a map given the zip code. Hence, we propose to geocode comments by linking the comment to the zip code of the location mentioned in the comment. For this task, a dataset containing the name of the streets, its respective neighborhoods and its zip code is needed. It is important to observe that this method comprises the first step for acquiring new information and knowledge from social networks related to citizens claims. In order to evaluate the proposed method, we developed a web system that shows the map visualization, with points representing comments that refer to the respective location; and the most frequent identified topics mentioned in the processed page. In addition, we also conducted an experimental analysis of the method using the page of a governmental agency from a Brazilian city.

This paper is organized as follows: Section 2 presents the theoretical background on Smart Cities and E-Participation that conceptualize the importance of this work. Section 3 presents theoretical reference and related work regarding to the techniques used in our work. Section 4 presents our proposed method. Section 5 describes a case study, considering (i) the collected data; (ii) its analysis; and (iii) opinions collected from software developers regarding to the steps of the presented method, implemented in the tool. Finally, Section 6 draws our final conclusions and discuss future work.

2 Smart Cities and E-Participation

Gil-Garcia, Pardo and Nam [7] affirms that, when considering the smartness of a city, rather than holding the dichotomy in terms of “being or not being” smart, smart city

concept should hold a continuum in which local government, citizens and other stakeholders think about implementing initiatives that turns the city “smarter”. In this way, they present a comprehensive view of smart city components and its elements, after compiling many different definitions found in literature and tools for evaluating or assessing the smartness of cities. Data management and information processing are two elements of the data and information component. In addition, [8] states that the smart city-building initiative “seeks to improve urban performance by using data, information and information technologies to provide more efficient services to citizens, monitor and optimize existing infrastructure, the collaboration between different economic actors and encourage innovative business models in both the public and private sector”. In this scenario, it is important to consider the huge amount of data on social networks that is not commonly used by governors for firstly understand what issues in a city can emerge from this data source.

Also in this scenario, models raised in literature for e-participation, i.e., models for citizen interaction by e-government. Reddick, Chatfield and Ojo [3] choose and summarize three models, which can be represented on a continuum, being the managerial model the lowest form of e-participation, and the participatory model being the highest form of citizen interaction with government. In the managerial model, citizens are viewed as “customers”, and government provides information and services to satisfy the demand of these customers, i.e, governments merely respond to their demands. In the consultative model, instead of being focused only on providing more efficient service delivery, the role for government is creating better policy decisions considering citizens claims and other inputs in the decision-making process. Finally, in the participatory model, there is a complex flow of information between governments and citizens, designed to enhance and shape policy. Regarding to the managerial model, it is worth to observe that several cities around the world provide open data in forms of reports, so that citizens can follow the actions of the government. The consultative model is somewhere in between the managerial and the participatory in its level of active. However, Reddick, Chatfield and Ojo [3] state that, when considering social networks for extracting knowledge about citizen claims, these different stages do not occur in a linear fashion. So, they propose a framework considering the use of text mining, analytics and visualization for explain their loop for e-participation. Specifically, visualization is a very important instrument for helping humans making sense of data [9]. This work specializes two aspects of this framework, when considering specific techniques for linking data, constructing visualization and presenting frequently addressed topics in comments on Social Network pages from local government agencies.

It is important to differentiate the terms “information” and “knowledge”. Information is not a synonym for knowledge, which is an intellectual concept, referring to the condition of knowing or understanding something. Knowledge is organized information in people's heads. Selecting and analyzing data, information can be produced; by selecting and combining information, knowledge can be generated; from this decisions can be made and action taken [10]. Data and Text Mining has been applied both for extracting information and knowledge from data [11]. In this work, we focus on using techniques commonly used in text mining for chunking and POS tagging words, briefly described next, from texts and data visualization techniques, aiming to enrich the

comments and visualizing them (information extraction) to allow knowledge extraction by humans.

3 Theoretical Reference

3.1 Text mining (TM)

Inzalkar & Sharma [12] state that the amount of unstructured stored data has tremendously increased, mainly in social networks. TM is the process of extracting interesting information or patterns from unstructured text from different sources. Tools for the different steps TM have the ability to analyze large amounts of text in natural language, and detect lexical and linguistic usage patterns, in an attempt to extract useful information [12, 11]. Nowadays, researchers use these tools in real-world applications, mining social networks for health or financial information, for identifying emotions about products and services, for example [13]. Also, sentiment analysis on social networks around political dimensions has been explored [14, 15].

Many of the NLP tasks involve searching for patterns in text that can be arduous if applying basic string operations. Regular expressions (**Regexes**) are primarily used in string search and substitution tasks in texts search and editing. **Regexes** are strings that define a search pattern [16]. To define a grammar, along with the use of regular expressions, it is necessary to use a technique called chunking [17], a fundamental mechanism of language: words can combine with other words, forming chunks. These can be combined with other chunks, to form even larger chunks, until a sentence is established. One way to construct the referred grammar is using POS (part-of-speech) taggers, which aims to assign a tag to each word in a text, or equivalently classify each word in a text to some specified classes such as noun, verb, adjective, etc. Both POS tagging and chunking are used in this work for extracting frequent topics from the comments.

3.2 Geocoding Data and Visualization

Up to our knowledge, there is not any work that links a comment/post in social networks to the location it mentions in its content. So, in this section we present similar works that reinforces the importance of our proposal. Cammarano et al [18] considered the problem of visualizing heterogeneous data sets, describing a system capable of automatically finding specific information in the set, necessary to create a visualization. The researchers introduced a mechanism capable of describing views, regardless of the data, and a data recovery algorithm appropriate for a given view. Initial experiments demonstrated that the created system had the ability to find appropriate data through visualization. So, exploring the linkage and processing of different types of data to generate different views can lead to enrich knowledge acquisition from unstructured data. This fact reinforces the importance of our method, as each comment must be geocoded by the zip code of some mentioned location, for generating the map visualization.

MacEachren, Brewer & Pickle [19] present a web analytics approach based on geovisualization using the social network Twitter, for supporting crisis management. Crisis management is the process by which an organization handles an unexpected event that threatens to undermine the organization and its audience. The proposed approach is implemented as a web application on a geographic map, which allows the user to search for information using indexing and tweet viewing, based on specific place and time characteristics. Our work is similar to this one, but different techniques for linking the data to be visualized had to be used. This is due to our interest in geocoding comments by its content, and not by the place it was posted.

Azevedo et al [5] present an approach that enables the integration of unstructured data located in different public organizations. They present concepts and technologies that provide information visualizations from open data using Geographic Information Systems (GIS). From the proposed framework, the application is able to identify vulnerable communities and provide effective preventive and emergency actions. The main contribution of the work is to include the use of tools and methods for data publication. To validate their approach, the authors used flood data from the Rio Doce basin. Our work is similar to this one. However, we worked with the reality that comments on Facebook cannot be used as linked data, and neither the comments are geocoded by its content. So, we had to propose and implement another ways for recovering the geospatial data, as well as linking the comments to geospatial information.

Li et al [6] present a survey about spatial technology and social media in remote sensing. They observe that there is a massive amount of data originated from remote sensing, social media, and GIS systems that are completely different sources of data. According to them, although important progress has been made in mining spatial and temporal data from social media, there is a need for investigating how these data can be used for decision making, particularly in the context of its integration with Geographic Information Systems (GIS). On the other hand, in our point of view, although there are many evolving solutions for GIS and possibilities for linking data, there are many data sources that do not follow this approach, as in our case. Also, [6] shows many applications, including in Brazil, that georeferenced image data and Twitter data, or similar approaches, for finding new knowledge in many specific domains. However, they do not present any work that explores generating new knowledge from different topics for understanding problems in a city, as we propose in this work.

4 Our Method

Considering all works described before, we understand that, up to our knowledge, there are gaps in literature that we explore in our method: (i) there is not any work that geocode comments based on some place mentioned in its content, which can help governments and citizens to visualize the distribution of comments over a city area; (ii) there is not any work that extract frequent topics from comments using POS tagging and chunking, which can help governments and citizens to visualize what are the most frequent topics mentioned in some social network page.

Figure 1 shows a schematic representation of our proposed method in BPMN (Business Process Modeling Notation). Each of the activities are described in what follows:

1. Collecting Social Data: Extract all posts and comments in a social network page of a local government agency, in a given period.
2. Collecting Postal Service Data: Extract data from some postal service of a given city, containing the names of the streets, avenues and others; and, for each one, the respective neighborhood and the zip code.
3. Processing Comments: This is the core of our method, better explained later, responsible for identifying (i) the locations mentioned in comments; and (ii) their topics. We explain better this task later.
4. Calculating frequent topics: Detect the most frequent topics in the set of topics, counting the frequency of all the detected topics.
5. Geocoding locations (streets in general): Geocode each comment by the mentioned place, if it is founded.
6. Generating data visualization: Generate a data visualization in map format with the geocoded comments, and a list of the most frequent topics (in descending order of the number of mentions). The view should display, beyond the map, the city data (name and number of neighborhoods, sites, pages, posts and comments) and the most frequent topics, followed by the number of mentions.

Activities in Processing Comments: For this task, a tagger must be used for classifying a word according to its class. The word classes are noun, proper noun, personal pronoun, adjective, adverb, verb, numeral, preposition, subordinating and coordinating conjunction, interjection, and others. It is worth to mention that the tagger is dependent on the language of the text in analysis.

For identifying the topics, the following activities are executed for each comment: (i) Cleaning the comments: Remove abbreviation, links and words with more than one repeated letter; replace a repetition of more than one of the same punctuation symbol by only one symbol; and remove blank spaces before a punctuation symbol; (ii) Processing the comments (POS tagger): Divide each comment into sentences, and tag (each word of) each sentence using the given tagger; and (iii) Applying the chunking process: Process each tagged sentence by a syntactical analyzer, using a grammar with one regular expression (RegEx grammar) of the form:

$$\langle N.* \rangle + \langle PREP.* \rangle * \langle KS \rangle * \langle ADJ \rangle * \langle N.* \rangle +$$

This analyzer looks only for subjects that form a sequence of a noun (N, which may be a regular or a proper noun), followed by a preposition, a subordinating conjunction, adjective or another (regular or proper) noun. The symbol “*” means that sequences that attends this grammar has no limit of size. The output is a collection of subjects found in all sentences of the comment. The output of this task is a set with all topics found in all the set of considered comments.

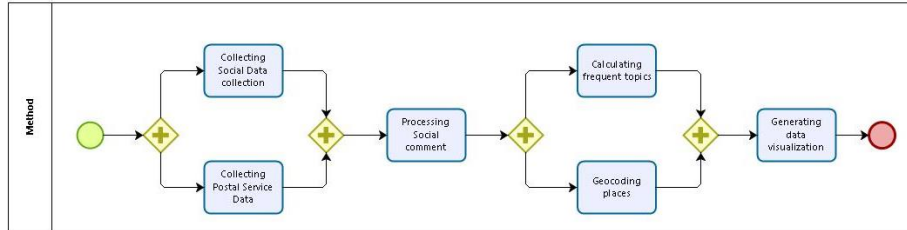


Fig. 1. Our method for processing comments in local government pages in social networks for knowledge acquisition.

For identifying the locations mentioned in comments, the following steps are executed for each comment: (i) Cleaning the comments: The same in the previous task, added to another activity: Remove stopwords (prepositions, pronouns and others); (ii) Processing the comments: Look for words that represent the type of the location (street, avenue, plaza and others); and (iii) Retrieving the names of possible locations: Retrieve from the database all possible locations of the type found in the previous activity, and find which one occurs in the comment. Return the related zip code of the location found.

5 A Case Study

In order to validate our method, we implemented a web system¹. It was logically divided into three modules: extraction (responsible for Activities 1 and 2), processing (responsible for Activities 3, 4 and 5) and visualization (responsible for Activity 6). All of them were implemented using .Net Core, specifically ASP .NET for constructing web applications, and Angular and Bootstrap frameworks for the graphical interface. We also used NLTK² (Natural Language Toolkit) for facilitating the POS Tagging and Chunking activities. We also constructed a tagger using a corpus for Brazilian Portuguese called Mac-Morpho³ [20].

5.1 Collecting and Processing the data

For the analysis of the operation of the tool, several information from different sources were collected, which are (i) the city's streets and neighborhoods; (ii) address geocoding; and (iii) comments of interest in the social network. Our case study addressed the municipality of Rio das Ostras, in the state of Rio de Janeiro, where a part of our research group is located. In what follows, we detail the data and information collection for each of the three items:

¹ Available at github.com/pedrocrl/tcc

² Available at <https://www.nltk.org/>

³ Available at <http://nilc.icmc.usp.br/macmorpho/>

(i) Rio das Ostras streets and neighborhoods: We extracted the information from the streets and districts of the city of Rio das Ostras on the website of the Brazilian Post Office⁴. It was difficult to obtain this information, because it is necessary to specify at least the neighborhood to retrieve the information of a municipality's public place. We obtained this information (neighborhoods of the city) with an employee of the Information Technology Department of Rio das Ostras. Also, to increase the efficiency of the application and to be able to be used in other contexts, we developed a robot to go through the Brazilian post office and obtain the data required. This collected data was inserted into the database for each neighborhood. A total of 1505 names of locations (streets, avenues, and so on) were obtained by the robot.

(ii) Address Geocoding. We used Google Maps Geocoding API service for geocoding the addresses (names of the streets), associating coordinates with each zip code. All the 1505 locations were geocoded through this process.

(iii) Frequent Topics in Comments. Facebook Graph API was used for collecting Facebook data. Each request in this API returns a document. For this, an Access Token, or authentication key, is needed. For this, we created an application in the Facebook Graph page. We collected posts and comments from the Facebook page “Cidadão Riostrense” (in Portuguese, which could be freely translated to Citizens of Rio das Ostras). We collected a total of 881 posts and 32157 comments between January 2016 and November 2017.

5.2 Visualizing the data

Processing the collected data, we identified the mention of locations in 2054 comments, while in the georeferenced sites, we found 184 comments. Figure 2 shows the map present on the screen of the constructed web system, with the extracted information from the collected data. In the original web system page, the left column displays the city data (name and number of neighborhoods – 66, locations – 1505, pages – 1, posts – 881, and comments – 32157), the middle one displays the refereed map, and the right column displays the most frequent topics, followed by the number of mentions. From the list of the hundred most frequent topics, we filtered the most twelve important ones, shown in Table 1. We observed that the main topic in the collected comments was 'street lighting', and this is not an information that could be obtained in any other place. There are some other curious observations in this list. The mayor's name (“Carlos Augusto”) and “Public Power” appeared many times as topics, and typically was related to some negative sentiments, when we observed the entire comments. Also interesting is the topic “God Comfort”. This indicates that sentiment analysis in comments of local government agencies should also be interesting. A deeper investigation of the remaining topics should be interesting with the local government, but it should not be executed up to the publication of this work.

⁴ We used the search tool available at <http://www.buscapep.correios.com.br/sistemas/buscapep/resultadoBuscaLogBairro.cfm>

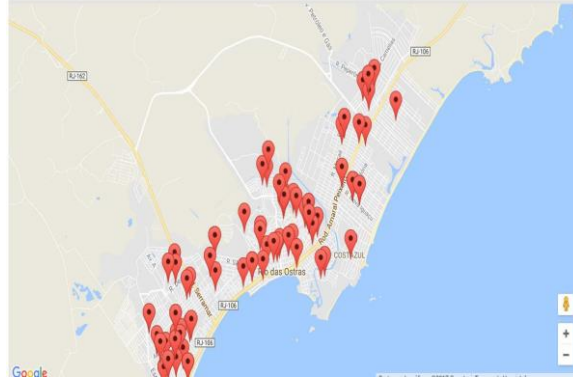


Fig. 2. Map of the constructed website for Rio das Ostras city.

Table 1. Frequent topics and their frequencies. “Carlos Augusto” is the mayor's name.

Topic	Frequency	Topic	Frequency
Public lighting rate	181	Municipal guard	41
Carlos Augusto	76	Lack of education	40
Public Power	75	God comfort	39
Fireworks	51	Public money	36
Basic sanitation	46	Garbage collection	34

It is important to observe that manually processing more than 30 thousand comments for extracting comments is not an easy task. When the task is complete by humans, probably the information will be obsolete. A more difficult task would be manually identifying the streets (locations), present in the content of the comments, and also geocoding them. This task was only possible to be executed by the computer using APIs and services available in the Web, as discussed before. So, our method allowed us to observe which citizens' claims emerge in a particular city, helping us to better understand people's real feelings, and which locations have the most problems.

5.3 Evaluating the usefulness of the tool

In order to try to identify in which level people graduated in computing undergraduate courses can use the tool for obtaining new knowledge, and how the tool can be evolved, we conducted an experiment with four participants, aiming to obtain a qualitative insight in the usefulness of the tool. For this, we constructed a question and answer form, containing five questions. Questions Q.1 to Q.4 were of the Likert scale type, and Q.5 asked the user to indicate how the tool could be evolved. The first four questions are: [Q.1] What level of difficulty did you feel in using the tool to identify which regions of Rio das Ostras are most commented? [Q.2] How much difficulty did you feel in using the tool to identify the most frequent topics? [Q.3] What level of difficulty did you feel in acquiring knowledge about regions with the most problems in the municipality of

Rio das Ostras? [Q.4] What level of difficulty did you feel in acquiring knowledge about the most frequent topics in the comments? Q.1 and Q.2 scales range from 1 to 5, where 1 means “Very difficult” and 5 means “Very easy”. Q.3 and Q.4 scales range from 1 to 10, where 1 means “I could not extract new knowledge” and 10 means “I discovered many new and interesting things”. The difference is due to the understanding that the question of extracting new knowledge can be much more sensitive when the user changes. Tables 2 and 3 show the number of responses obtained by scaling (first column) for each question (remaining columns). Table 4 shows the answers of the participants in Q.5. It is worth to notice that we chose software developers for this first analysis due to their ability to understand how is difficult to analyze unstructured data. They suggested interesting ways to evaluate our tool.

Table 2. Answers from participants to Q.1 to Q.4.

Q.	1	2	3	4	5
1	0	0	0	2	2
2	0	0	1	2	1

Table 3. Answers from participants to Q.1 to Q.4.

Q.	1-6	7	8	9	10
3	0	1	1	0	2
4	0	0	2	0	2

Table 4. Answers to question Q.5

Q.5: How do you believe that the tool could be evolved?	
1.	“Very good idea and tool! On evolution: to make comparative analyzes in the future on the evolution of topics/comments, to identify if there was any progress in the problems and what new ones arose, allowing analyzes on the performance public agencies and perception of the population in the region.”
2.	“[The system] could have a space to register the responsible public agencies concerned [to the related problems], and a notification for submission and possibility of response of the type “complaint here”.”
3.	“Twitter API can be integrated, which allows users to get more information about the problems (so other channels can take advantage of this feature).”
4.	“In the technical part, I think that an implementation using Artificial Intelligence over the topics found is required. Determining the really important topics is worth – unnecessary topics are appearing”.

6 Conclusions and Future Work

This work presents a method that allows extracting new information in comments in a social network page of a local government agency, through identifying the most frequent topics and geocoding these comments based on the location it refers to. Our initial experiments allowed us to observe that visualizing the map with the geocoded comments, joined to the frequent topics, allowed (advanced) users extracting knowledge from the collected comments.

Based on literature review in e-participation and smart cities, citizens' claims are information of great value to the government. In this work, we used two forms of analyzing and extracting knowledge about their claims posted in social networks, which are frequent topics and geovisualization. We understand that this is the first step for exploring this huge amount of data. In this way, sentiment analysis and sarcasm identification others are interesting methods for being investigated in this scenario. Specifically, this is a challenge in Portuguese language. In the future, we intend to investigate the use of these approaches to measure and interpret citizen claims in social network.

We also consider the following limitations of our work, for improvement in the future: (i) We did not evaluate the precision of our method from Information Retrieval perspective. For this, we need to collect a sample of the comments, label the comments with the name of the location (street and others), use the technique for detecting the name of the location, and verify the precision of the technique. This type of information, if precise, can be added as a metadata of the comments; (ii) Other ways of geocoding comments, as using, for instance, geolocation of the user that created the comment, could be interesting, in the case of this information is available; (iii) Our qualitative experiment allowed to observe future improvements of the method and the tool, although the number of participants is too low. We desire to evaluate our tool using citizens of different profiles, as well as investigate the usefulness of the toll with people working with local government agencies.

Acknowledgements

We would like to thank the precious comments of the anonymous referees that helped us to improve our work. We also thank CAPES, connected to the Brazilian Ministry of Education, FAPERJ (Foundation of Support to the Research of State of Rio de Janeiro) and CNPq (Brazilian National Council for Research) for partially funding this work.

References

1. Teran, L., Kaskina, A., Meier, A.: Maturity model for cognitive cities. In: Portmann, E., Finger, M. (eds.) *Towards Cognitive Cities: Advances in Cognitive Computing and its Application to the Governance of Large Urban Systems*, pp. 37-59. Springer (2016).
2. Dameri, R.P., Rosenthal-Sabroux, C.: *Smart city: How to create public and economic value with high technology in urban space*. Springer (2014).

3. Reddick, C.G., Chatfield, A.T., Ojo, A.: A social media text analytics framework for double-loop learning for citizen-centric public services: A case study of a local government Facebook use. *Gov. Inf. Quarterly* 34(1), 110-125 (2017).
4. Barcellos, R., Viterbo, J., Miranda, L., Bernardini, F., Maciel, C., Trevisan, D.: Transparency in practice: using visualization to enhance the interpretability of open data. In: *Proceedings of 18th International Annual Conference on Digital Government Research*, pp. 139-148. ACM, New York (2017).
5. Azevedo, P.C.N., Pinto, V.A., Bastos, G.S., Parreiras, F.S.: Using linked open data in geographical information systems. In: Grueau, C. Gustavo Rocha, J. (eds.) *Geographical Information Systems Theory, Applications and Management*, pp. 152-166. Springer (2016).
6. Li, J., Benediktsson, J.A., Zhang, B., Yang, T., Plaza, A.: Spatial technology and social media in remote sensing: A survey. *Proceedings of the IEEE* 105(10), 1855-1864 (2017).
7. Gil-Garcia, J.R., Pardo, T., Nam, T.: What makes a city smart? Identifying core components and proposing an integrative and comprehensive conceptualization. *Information Polity* 20(1), 61-87 (2015).
8. Marsal-Llacuna, M.L., Lopez-Ibañez, M.B.: Smart urban planning: Designing urban land use from urban time use. *Journal of urban technology* 21(1), 39-56 (2014).
9. Baker, J., Jones, D.R., Burkman, J.: Using visual representations of data to enhance sense-making in data exploration tasks. *Journal of the Association for Information Systems* 10(7), 533-559 (2009).
10. Webb, S.P.: *Knowledge management: Linchpin of change*. Routledge, London (2017)
11. Aggarwal, C.C., Zhai, C.X. (eds.): *Mining Text Data*. Springer Science+Business Media, New York (2012).
12. Inzalkar, S., Sharma, J.: A survey on text mining-techniques and application. *International Journal of Research in Science & Engineering* 24, 1-14 (2015).
13. Hirschberg, J., Manning, C.D.: Advances in natural language processing. *Science* 349(6245), 261-266 (2015).
14. Weitzel, L., Bernardini, F., Quaresma, P., Alves, C., Zacharski, W., Figueiredo, L.: Brazilian social mood: The political dimension of emotion. In: Fuhr N. et al. (eds) *Experimental IR Meets Multilinguality, Multimodality, and Interaction, CLEF 2016, LNCS, vol. 9822*, pp. 247-252. Springer, Cham (2016).
15. Paula Filho, W., Rosseti, I., Viterbo, J.: On tweets, retweets, hashtags and user profiles in the 2016 american presidential election scene. In: *Proceedings of 18th International Annual Conference on Digital Government Research*, pp. 120-128. ACM, New York (2017).
16. Chapman, C., Wang, P., Stolee, K.T.: Exploring regular expression comprehension. In: *Proceedings of the 32nd IEEE/ACM International Conference on Automated Software Engineering*, pp. 405-416. IEEE Press, Piscataway (2017).
17. Lu, Q., Xu, C., Liu, H.: Can chunking reduce syntactic complexity of natural languages? *Complexity* 21(S2), 33-41 (2016).
18. Cammarano, M., Dong, X., Chan, B., Klingner, J., Talbot, J., Halevey, A., Hanrahan, P.: Visualization of heterogeneous data. *IEEE Transactions on Visualization and Computer Graphics* 13(6), 1200-1207 (2007).
19. MacEachren, A.M., Robinson, A.C., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J., Mitra, P.: Geo-twitter analytics: Applications in crisis management. In: *Proceedings of the 25th International Cartographic Conference*, pp. 3-8. GeoVISTA Publications, State College (2011).
20. Fonseca, E., Aluísio, S.M., Rosa, J.: Evaluating word embeddings and a revised corpus for part-of-speech tagging in Portuguese. *Journal of the Brazilian Computer Society* 21(2), 1-14 (2015).