

Streaming geocoding of incidents based on information from social networks

*Yana Bekeneva**, and *Alexandr Lopushanskiy*

Department of Computer Science and Engineering, Saint Petersburg Electrotechnical University “LETI”, ul. Professora Popova, 5, Saint Petersburg, 197022, Russia

Abstract. Users daily publish and discuss information about various kinds of situations occurring in their places of residence (district, city). Sometimes it is necessary to determine the extent to which a situation is affecting the quality of life of citizens, and sometimes published information can be used to identify problems. The work proposes a method for processing text information from messages on social networks and applying geotags, including in real time. The proposed tool can be useful for visualizing certain problems and can be used to make decisions about how to fix those problems.

1 Introduction

The most common geocoding task is to convert one or more full addresses, which usually contain the name of a city, and sometimes an administrative unit or country. However, many geocoding systems are capable of handling other types of location information. You can enter city names, postal codes, as well as the names of regions or countries into the system. Apart from directly entering immediate coordinates, the most accurate result can be achieved using the full address. However, users of social networks rarely indicate full addresses in a form convenient for geocoding: they indicate house numbers in free form, indicate landmarks along with street names. Quite often, a situation arises when it is necessary to assess the scale of a situation, for example, to identify victims of a natural disaster and send assistance to the areas from which reports are received.

Social networks can be used for cases where there is no direct threat to human life, but users of the social network exchange among themselves about the impact of the situation on objects at certain points. For example, after a utility accident, some houses may be left without water supply or receive water of inadequate quality. After explosions or earthquakes, cracks and other damage to buildings may appear.

This paper presents an approach to streaming geocoding to indicate the occurrence of emergency situations or their mention in real time.

* Corresponding author: yabekeneva@etu.ru

2 Related works

Modern geoservices and navigators show information about road accidents. As a rule, information is entered by users, who can be both road users and representatives of government services.

Let's consider existing solutions for processing streaming information and applying geotags to the map.

The task of processing geolocation data to place marks on a map is relevant for research in various subject areas. For example, using data on crimes committed, safe and dangerous ones are identified in a city or an entire region [1]. Using medical data, the spread of diseases is analyzed [2]. The construction of heat maps for natural disasters shows both the source and the affected areas [3]. The most common problem of interest is the analysis of traffic accidents. Some studies are aimed at identifying particularly dangerous areas where accidents most often occur [4], some are aimed at predicting accidents taking into account traffic and weather conditions [5]. Most of the research related to the processing of messages from social networks considers Twitter as a data source [6]. This is a platform where users can post short text messages, their length is strictly limited. As a rule, such messages are not accompanied by other unstructured information, while other social networks do not limit the length of the message and suggest the possibility of attaching photos or video material. In addition, a number of studies involve processing geotags accompanying a message, rather than extracting information from the text.

The goal of this development is to automatically map events based on the analysis of information from social networks in real time. In this case, the type of event is determined by researchers and users, i.e. is not limited to road accidents only.

3 Proposed approach

Different events have different priorities and different speed of response to these events. Therefore, for some events, batch processing and analysis at specific time intervals can be considered. For example, reports of pavement conditions may accumulate and be reviewed for weeks before a regularly cited street is listed for repair. Mentions by users of social networks of tap water of inadequate quality require a timely response, but do not need immediate processing. At the same time, situations such as traffic accidents entail consequences in the form of traffic jams within a few minutes, and emergency situations such as explosions or fires require immediate response.

In this work, road traffic accidents are subject to research, since they are most often mentioned by users of social networks, and for them the task of geocoding in real time is the most relevant.

Let's consider what information can help the user identify a post that potentially describes an emergency situation. One possible indicator could be the frequency of sending messages, which increases when an abnormal event occurs. In this case, you have access not only to the time the post was created, but also to its characteristics - people's reactions to it. Based on this, it was decided to attempt to display to the user only the characteristics of the records and not display their frequency. This is also due to the fact that some communities publish posts not as news emerges, but evenly throughout the day.

Taking into account all the requirements, the application interface was developed. The PyQt6 library was used for development. The interface consists of several windows, among which the main one is a window with information about all posts. The program also creates a window for selecting a group whose posts will be analyzed, and a window for viewing a map with marked coordinates.

The window with information about posts consists of a set of widgets, each widget performs a separate function in the application. The main widget contains the post text and highlighted keywords. Another widget allows you to mark on the map in real time. And finally, a widget that displays the statistics of posts in the group.

The developed tool allows you to select communities in which to search for information on a given topic. The community “Road Accidents and Emergency | St. Petersburg | Peter Online | SPb”, where users offer and administrators post information about various types of incidents. The search parameter was traffic accidents. Start of the experiment: October 9, Monday, 8:00. At this time there is a large flow of cars, as many people go to work at this time.

As a result of launching the program, several incidents that occurred over the last hour were marked on the map (Figure 1).

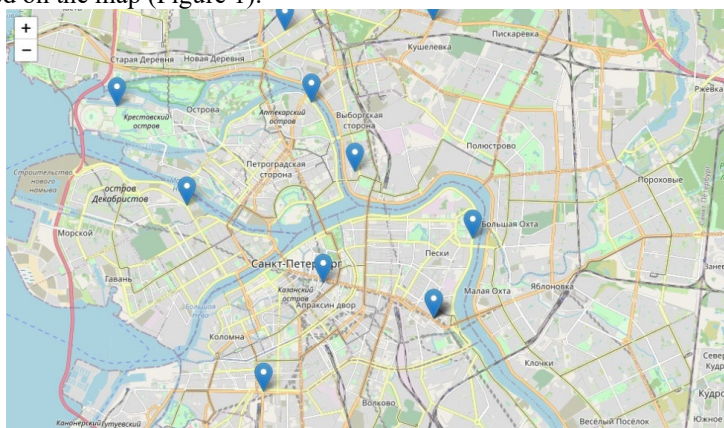


Fig. 1. Geotags about the traffic situation at the time the program was launched.

Next, the program automatically processed posts published in the community, highlighted geoinformation and added a mark to the map. Please note that information about one incident is published once, therefore, repetitions are excluded. Also, comments that may contain references to events that happened in the past, which could affect the reliability of the noted data, were not subject to consideration. In addition, the discussions do not provide any useful information for geotagging.

Figures 2 and 3 show geotags placed 1 hour and 3 hours after starting the program. As you can see, new geotags were added during this time.

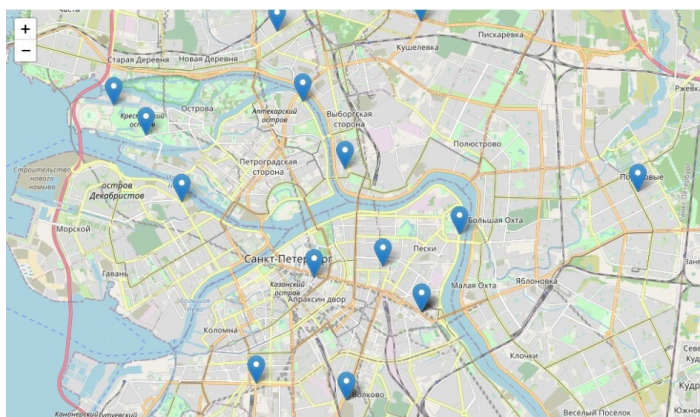


Fig. 2. Geotags about the traffic situation after 1 hour the program was launched.

Further developments may involve increasing the number of settings for more convenient display. Also important is the task of developing a tool for direct extraction of a toponym without converting it to a named entity and back.

This work has been supported by Russian Federation President Grant MK-918.2022.1.6.

References

1. Z. Wang, Y. Yao, Evaluating the quality of public geocoding services for crime analysis in China Preprint rs-1989875/v1 (2022)
2. M.E. Quinteros et al., Chile *Cadernos de Saude Publica* **38**, e00288920 (2022)
3. K. Lai, J.R. Porter, M. Amodeo, D. Miller, M. Marston, S. Armal, *Inf. Proc. & Manag.* **59(1)** 102735 (2022)
4. B. Kilic, M. Hacar, F. Gulgen, *Transactions in GIS* **27(5)**, 1599-1613 (2023)
5. A. Azhar, S. Rubab, M.M. Khan, Y.A. Bangash, M.D. Alshehri, F. Illahi, A.K. Bashir, *Cluster Computing* **26(1)**, 477-493 (2023)
6. G. Razis, I. Maroufidis, I. Anagnostopoulos, *EventMapping: Geoparsing and Geocoding of Twitter Messages in the Greek Language*, In IFIP International Conference on Artificial Intelligence Applications and Innovations, 14-17 June 2023, Leon, Spain (2023)