Visualisation of User-Generated Event Information: Towards Geospatial Situation Awareness Using Hierarchical Granularity Levels

Heidelinde Hobel[1], Lisa Madlberger[2], Andreas Thöni[3], and Stefan Fenz[4]

[1] Vienna University of Technology,
Doctoral College Environmental Informatics, Austria
[2] Vienna University of Technology,
Vienna Phd School of Informatics, Austria
[3] Vienna University of Technology,
Institute of Software Technology and Interactive Systems, Austria
[4] SBA Research, Vienna, Austria
hobel@geoinfo.tuwien.ac.at, lisa.madlberger@tuwien.ac.at,
andreas.thoeni@tuwien.ac.at, sfenz@sba-research.org

Abstract. In recent years, enterprises and emergency response teams have started to use user-generated content to monitor crises, events and trends. Especially in critical situations, decision makers must, above all, quickly assess huge amounts of data. Effective geographical visualization and aggregation of collected data is an important prerequisite to enable decision makers to infer the impact of a detected event on, for example, their supply chains and other physical establishments. However, in existing literature the aspect of geographical visualization of automatically analysed events is hardly addressed. In this paper, we propose to introduce hierarchical levels of detail, a concept from Geographic Information Systems, for the visualization of user-generated data describing a local event. We developed a tool which can improve the assessment of regional impacts by offering the possibility to browse and visualize results on layers aggregating data along individually defined hierarchical dimensions, e.g. geographical or political districts.

1 Introduction

Social networks, RSS feeds, and platforms for microblogging are becoming more and more important for users to share feelings, experiences and to report about recent events at any time. As a result, huge amounts of data are created, which are often publicly available. This raw information can be used by enterprises as well as governments to rapidly learn about the latest events and to monitor the public opinion about certain topics. However, processing these huge amounts of information is a challenging task and, in case of a crisis, the accessible information must be quickly assessed to be useful for decision-making. Globalization led to an increase in the multinational footprint of enterprises and their supply chains. Consequently, the critical infrastructure is spread over large and disconnected territories. Knowing geographical references is a key input for decision-making processes in such an environment.

Therefore, the assessment of the collected information needs to be linked with geospatial information about regions and points of interest. Microblogs and news feeds are often enriched with temporal and geospatial data, either explicitly provided by tags in the meta-data or implicitly in the messages' content itself. However, this information inherits the intrinsic properties of user-generated data and is therefore likely to be incomplete, incorrect, and imprecise. Furthermore, it is a non-trival task to monitor large regions of interest while still quickly assessing the impact of a detected event with globally dispersed points of interest. Therefore, knowing the geographical reference area of a feed can be an important starting point.

The goal of our research is, therefore, to improve geospatial assessment of events reported in microblogs by using hierarchical levels for the analysis of important events threatening the infrastructure of enterprises and visualization of results by browsing through these layers. While the detection of the geographic origin of an incident is often determined by the measurement of bursts, e.g., [7], we focus on monitoring regions of interest, assessing the impact of detected events, and providing better user-support for decision-making. To this end, we developed a tool for the assessment of collected microblogs at hierarchical geospatial levels while considering relevance factors that are assigned to individual microblog messages. The contributions of this paper are as follows:

- We developed a tool aimed at supporting enterprises and emergency response workers in geospatial assessment of incidents by using hierarchical levels.
- We discuss architectural decisions, implementation details, and our semantic model for the analysis of microblogs.

It is important to note that any monitoring-approach relying on user-generated web data is restricted to situations where both users as well as decision makers are still able to access communication infrastructure. Moreover, it is limited by the extent of data being augmented with geospatial information, either by explicit tags (e.g. GPS-tags) or implicitly in the content. Considering for example the microblog platform Twitter[5], around 2 percent of microblogs had been GPS-tagged in 2012. Given a baseline of around 400 million entries per day, the amount tagged was still significant. A fulltext geocoder including additional fields such as the location could even reference around 28% of the entries [6].

Throughout the paper we are considering exemplary User Generated Text Content (UGTC), which includes microblogs, RSS feeds and content from social media platforms. We focus on the generic aspects when using UGTCs in critical response systems - an implementation in a certain domain using a specific provider always requires the consideration and adherence to the specific applicable data protection laws, privacy terms, and terms of use of the provider.

The remainder of this paper is structured as follows: In Section 2, we present related work on the use of microblogs for event-detection and in Section 3, we

[5] http://www.twitter.com

discuss the opportunities to retrieve geospatial information from microblogs and RSS feeds. We present the visualization approach in Section 4, and describe our architecture and implementation in Section 5. In Section 6, we conclude our work and offer an outlook on future work.

2 Related Work

The huge amounts of publicly available user-generated data motivated research in various areas. Several studies focus on detecting events in microblog data - one of the earliest of this kind was developed by Sakaki et al. [11], who developed a system to reliably detect earthquakes in Japan exclusively based on Twitter data. They do not particularly address the aspect of visualization in their study, however in a provided screenshot they use coloured pins to depict individual tweets on related to earthquakes on a map. The problem with pins is that multiple instances at the same location cannot be visually distinguished from single occurrences. The same applies for Sadilek et al. [10] who use microblog data to predict disease transmission and used pins to visualize geographic locations of users, but again, visualization was not a core aspect of their work. However in both cases, the possibility to aggregate and view results on higher hierarchical levels, e.g. for each district might enable a better overview and lead to additional insights. Chunara et al. [3] use news and microblog data to visualize disease outbreaks on a health map. They present alerts, derived from Tweets on a heatmap, which indicates high and low-density of relevant messages both on a detailed level and aggregated on up to two hierarchical levels according to administrative districts. While this provides a good example for the use of geospatial hierarchies for data-visualization, our approach aims for a general solution allowing for hierarchical aggregations along multiple dimensions, e.g. administrative but also according to geographical or political attributes.

Additionally, several efforts have been made to detect events independently of a specified domain [1, 2, 5, 7]. These methods typically extract events based on the detection of high occurrences of words. While in most of these studies temporal and geospatial properties of the detected events are extracted, only little attention has been paid yet to the geographic representation of events.

One of the few studies explicitly devoting attention to visualization aspects was conducted in [7], where the authors validated their map-based visualization approach in a user study. As one of their results they found, that for an intuitive user experience the additional possibility to zoom in and out of visualized data as well as the aggregation of mapped results would be required. Rosi et al. [9] also point out the need for better visualization techniques and tools to view and understand data at multiple levels of granularity. Pouliquen et al. [8] geocoded news items and experimented with different visualization options. They suggested representing news stories as points on a map leveraging WorldKit[6] or used placeholders in GoogleEarth[7] with icons representing the frequency of news

[6] `brainoff.com/worldkit/`
[7] `earth.google.com`

items found referencing a specific place. In the later, they relied on the zooming features naturally provided by GoogleEarth. Furthermore, they experimented with Scalable Vector Graphics (SVGs), but relied on only one country level.

In our study, we want to address this gap by proposing a method to implement hierarchical geospatial layers, which allow for different aggregation levels during event-detection, visualization and assessment. We use UGTCs such as microblogs and RSS feeds to illustrate our approach.

3 Inferring Geospatial Attributes from UGTC

Multiple ways to infer geospatial information are applicable to Microblogs as well as to RSS feeds. By using microblogging platforms users can often decide whether the exact location (identified by GPS-information), the place (such as the city or neighbourhood) or no location information is attached to a message. Additionally to these location-tagged microblogs, geographic information could be obtained from the user's profile if available and accessible at a platform. The user's profile may include a location field, the time zone and may include further location information in the profile description or on a linked personal website. Eventually, information can also be extracted from the message text, which might relate to a certain event or directly to an area, territory, or jurisdiction. Using profile, user location and geo tags as input and again taking Twitter as an example for a microblogging platform, Leetaru et al. [6] reported a share of 34% of microblogs mappable at a correlation level of 72% against a baseline.

Considering standard RSS feeds, location information can be obtained analogously from the content or the author's information. Since RSS feeds are linked to more comprehensive blogs or news articles, more information about where an event occurred could be provided. Moreover, the W3C GeoRSS standard is designed to explicitly provide information about the geographic location a post relates to in form of geographical points, lines and polygons, which can be automatically processed by geographic software. In this case, the location information is certainly more accurate since it is explicitly annotated and aims at describing the geospatial features of a report. However, even when having only full-text with geographic information available, accuracy rates of 77% have been achieved [8].

Locations inferred from user-generated data can be obtained by different methods. However, it must be distinguished between reports about events or incidents from the place where the event or incident occurred and reports in which the event or incident are discussed. While certainly both of categories are important, we require schemas to reflect these geospatial dependencies. Furthermore, location information is often ambiguous or may be incorrect. As a consequence, we have to account for the possibility to allow users to correct errors and infer the geographic dependencies between analysed topics.

4 Geospatial Visualization of UGTCs

The goal of our tool is to improve geospatial assessment of events reported in UGTCs by using hierarchical levels for the analysis of important events threatening the infrastructure of enterprises and the visualization of results by browsing through these layers. These layers are diversely designed according to the needs of the domain, which is encompasses the enterprises' requirements and the (global) dependencies of its supply chain infrastructure. For example, comparing the states of the US with the countries of the European Economic Area requires to set these areas on the same hierarchical level. This might be interesting when enterprises are planning or evaluating establishments of their infrastructure in these regions of interest. Furthermore, the user has to decide which regions and layers are of importance in the analysis. For example, in an industrial use case, an enterprise might concentrate on geographic regions where critical infrastructure or suppliers are located. Measuring the influence of earthquakes might motivate the user to define the center of the earthquake as central point and then to define concentric circles as hierarchical levels around the earthquake's epicentre. For our implementation, we created a hierarchical structure according to the United Nations Statistics Division[8]. According to this website, the geographical regions and compositions are structured in the following hierarchy: World $\mapsto$ continental regions $\mapsto$ geographic sub-regions (e.g., Eastern Africa) $\mapsto$ Countries. We extended the structure with Countries $\mapsto$ country-specific Regions.

In the following two scenarios, we will motivate the usage of hierarchical layers which facilitate the browsing of data in two conceptually contrary models: top-down and bottom-up.

Top-down assessment. In the top-down approach, users are monitoring the highest level of interest. This visualization approach is aimed at providing a holistic overview of gathered information and allowing to zoom into the defined levels of interest, where each area reveals its own and from sub-areas inherited UGTCs. By using the top-down approach, the user can compare regions of interest globally (i.e., at the highest abstraction level) with other areas at this level. Our tool then provides the possibility to seamlessly start zooming in to explore more specific areas in more detail. For instance, this visualization scenario is suitable when an enterprise is planning new establishments for critical infrastructure or to monitor geographically large areas. In case of an emergency or crisis this approach allows to zoom into the relevant local areas, determine the geospatial impact and to plan appropriate measures.

Bottom-up assessment. The bottom-up approach is useful for monitoring specific geographic locations and to access the impact as as soon as an event is detected. Users can also zoom out of the region in order to browse the UGTCs in the history of upper-levels. The predefined hierarchical levels allow to systematically analyze the global impact of an event. For instance, in case of a detected earthquake an approximation of the epicenter is calculated and the user's definition

[8] `https://unstats.un.org/unsd/methods/m49/m49regin.htm`

of the radii of the concentric circles around the epicenter are used to infer the impact of the earthquake.

The following sections explain the features of our tool based on the Visual Analytics Mantra "Analyse First – Show the Important – Zoom, Filter and Analyse Further – Details on Demand" [4]. We shortly describe how the visualization tool can be used to browse through the analyzed data (Show the Important – Zoom Filter and Analyse Further), and how details of analyzed UGTCs could be accessed. The setup of the system and the processing of UGTCs is explained in Section 5.

4.1 Interactive Maps for Visualization

We use Leaflet[9], which is an open source javascript library for interactive maps, for the visualization of results using the defined hierarchical layers. Figure 1 shows the main interface for the analysis of collected UGTCs.

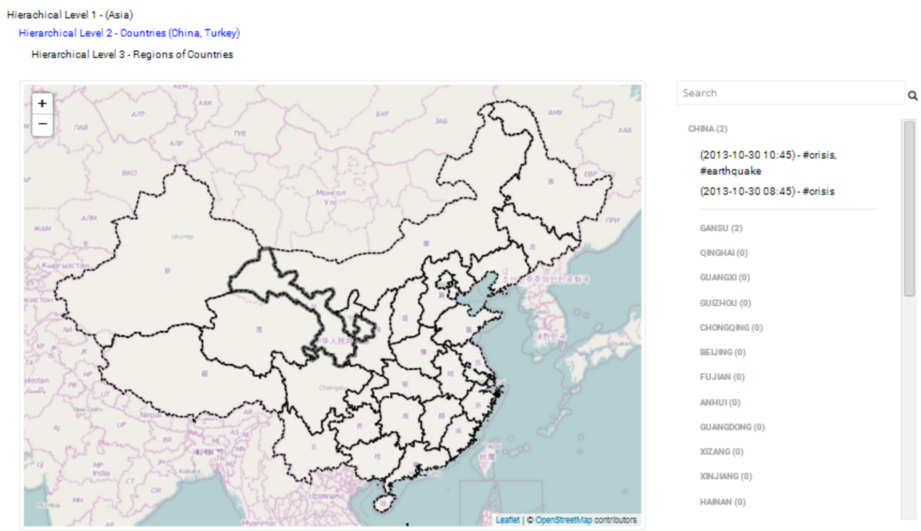

Fig. 1. Using Leaflet and a hierarchical tree to visualize the geographic correlation of UGTCs on interactive maps. (See for tools and data: Leaflet, OpenStreetMap, and NaturalEarth)

On the top of the tool, the user can choose the appropriate level for the analysis. In the example presented in Figure 1, the user is analyzing the secondary level, which comprises the countries China and Turkey. Furthermore, the user has opened the category China in the tree on the right side, which centers China's geometry on the map. China's category shows two identified

[9] `http://leafletjs.com`

UGTCs, which where classified into one of China's provinces, i.e., Gansu. How many UGTCs where detected in an area is displayed in brackets beneath after the name of the region of interest. Zooming into one of the provinces of China could be done by either opening the respective category or by double clicking onto the specific area on the map. In the case that a high number of UGTCs regarding defined topics, e.g. crisis, bomb, etc., for an area is detected, the specific area and every superordinate area is colored red. To specify "high", users may define a threshold value of number of UGTCs detected. Moreover, in order to minimize the effort to assess single UGTCs, only the most relevant UGTCs (see Section 5) are presented to the user. The time-frame, which could be used to retrieve relevant UGTCs can be manually adjusted.

When zooming into an area of the lowest level of the defined layers, the vector overlay will be transparent and UGTCs that have a location tag are shown on the map as markers, as shown in Figure 2.

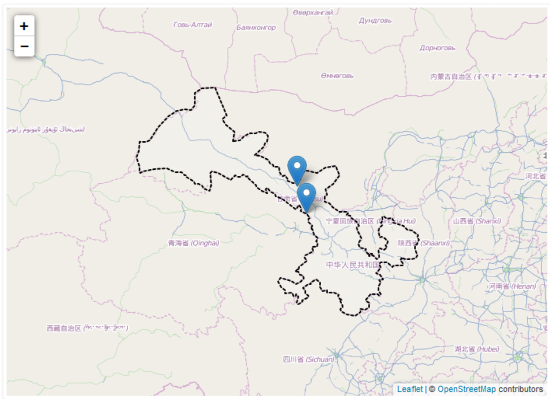

Fig. 2. Highlighting of detected UGTCs with location tags in the lowest hierarchical level. (See for tools and data: Leaflet, OpenStreetMap, and NaturalEarth)

4.2 Details on Demand

In our first prototype, we use timestamps of retrieval and the actual timestamp of exemplary messages, the content of exemplary messages, topical and geospatial tags, as well as information about the preprocessed information of the UGTC. Our system pre-classifies UGTCs based on a keyword analysis and assigns an indicator which is reflecting the relevance of each UGTC, as explained in Section 5. However, the classification of text content is a non-trivial task and a complete accuracy is almost not possible, therefore users should be able to manually assess detected incidents and correct possible misclassification. Hence, our interactive interface allows to open detail-sites when clicking onto a link of a incident that is displayed in the tree structure on the right side, in which the user can quickly assess the relevance of a message, edit its tags, and link further online resources. Semantic annotations allow to explore further details by following the links.

5 Implementation and Architecture

In this section, we present our framework for geospatial assessment of UGTCs. For our first prototype, we implemented a Web application for the processing of collected microblogs and visualization of results, its architecture is illustrated in Figure 3.

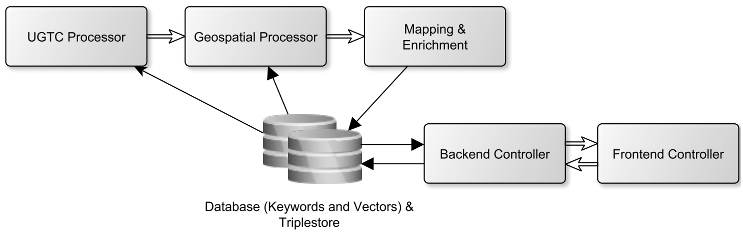

Fig. 3. Architecture

In the following, we will showcase the process using an exemplary fictive microblog message: "The earth is shaking - earthquake in Gansu".

UGTC Processor Our system is designed that it can be connected to various data sources, including microblogs and RSS feeds. The UGTC processor is designed to collect UGTCs from online sources or import datasets to search based on topical and geospatial keywords for relevant UGTCs. At the time this paper was written, we have tested our system based on an imported historic dataset. The keywords should be initially defined according to the domain of the system in order to restrict the input of UGTCs to the system. Enterprises with globally dispersed critical infrastructure could define keywords for the names of establishments and critical facilities (i.e. topical keywords) as well as all related names of areas corresponding to the facilities (i.e. geospatial keywords). In more generic use cases, the users should define keywords such as crisis, earthquake and thunderstorm. In our example message, we detected the following keywords: `earthquake` and `Gansu`.

Geospatial Processor. Geospatial keywords and hierarchical dependencies can be inferred from the GeoNames API[10]. GeoNames provides a huge amount of geospatial features, however, the location of many areas is often just provided as a single point, since the boundaries are not yet available for every record.

[10] `http://www.geonames.org/export/ws-overview.html`

Furthermore, the hierarchies of queried areas are not customizable and must be mapped to self defined hierarchies to allow customized comparisons.

The hierarchical geospatial structure for the prototype is based on continental and administrative boundaries, where we used the following taxonomy as mentioned in Section 4: World $\mapsto$ continental regions $\mapsto$ geographic sub-regions (e.g., Eastern Africa) $\mapsto$ Countries $\mapsto$ country-specific regions. We identified the datasets provided by NaturalEarth[11] as the most appropriate dataset for the visualization, since the data layers are preprocessed and provide consistent geographic shapes which lineworks are independent from other shapes, e.g. countries that share one line as a boundary. For countries and their regions we used the large scale data (1:10m) and directly exported the vector data in GeoJSON format by using the open source tool Quantum GIS[12]. For the world-wide and continental layers, and geographic sub-regions, we started from the dataset for administrative level one and merged the country-specific features into the appropriate structure. For each of the extracted areas, we added GeoJSON properties to designate the hierarchical type of the layer and the part-of relation of the respective area. Each layer and feature is enriched with the geospatial features, names and alternative names retrieved from GeoNames. The hierarchical geospatial database for our first prototype implementation is based upon a NoSQL database to store vectorial features in GeoJSON[13] format.

Each UGTC that is passed on from the UGTC Processor is processed according to the hierarchical information stored in the databases. Our first prototype supports geospatial keyword analysis for the content of a UGTC as well as geospatial queries such as "is this point in this polygon" for possible location metadata information of the UGTC. If the UGTC is classified based on the location information, then it is assigned to the feature of the lowest level of the hierarchical layers. If the UGTC contains a geographical keyword of the layers, then it is assigned to the specific feature. Since the fictional microblog message encompasses no explicit location tag, it is assigned to the feature that relates to "Gansu".

Mapping and Enrichment of Microblogs. Once a UGTC is detected from the geospatial processor and preclassified according to the identified feature, it is mapped to our RDF schema and stored in a triple store. To allow queries and aggregation functions on the stored set of UGTCs, we map each UGTC into a semantic model. We link the UGTC to the geospatial feature of interest by using a feature tag. The term location tag is used to refer to explicit GPS information in the meta data of the UGTC if available. Identified topical and geospatial keywords are annotated as tags, as well as location and temporal tags of UGTC are added as annotations. For the annotation we used the geonames[14] and dcterms[15] vocabularies. The following triples show an excerpt of the tags

[11] http://www.naturalearthdata.com/
[12] http://www.qgis.org/en/site/
[13] http://geojson.org/
[14] http://www.geonames.org/ontology/documentation.html
[15] http://dublincore.org/documents/dcmi-terms

used to annotate our exemplary microblog (we simplified it for presentation by linking geonames' RDF resource for Gansu and one entry for dcterms).

```
@prefix geovis: <http://environmental.tuwien.ac.at/geovis/1.0/> .
<http://environmental.tuwien.ac.at/geovis/1.0/id12938572552353>
    <http://purl.org/dc/terms/created> "2014-04-23 2:11:02" ;
    geovis:relatesToGeoNames <http://sws.geonames.org/1810676/about.rdf ;
    geovis:feature "feature8984762834456" ;
    geovis:tag "earthquake" ;
    geovis:geo-tag "Gansu" ;
    geovis:relevance 0.4 .
```

Moreover, we added a relevance tag which is calculated as follows: Each topical tag is assigned a relevance factor between 0 and 1. Assume that a UGTC has the topical tags: $tkw_1, \ldots, tkw_n$. Each topical tag can be mapped to a relevance factor ($rel(tkw_i)$, defined according of the domain of the enterprise. The maximum of these relevance factors is then multiplied by a location factor (f_{loc}), as shown in equation (1). We explain this process below.

$$UGTC_{relevance} = f_{loc} \times \max\{rel(tkw_0), \ldots, rel(tkw_n)\} \tag{1}$$

First we calculate the maximum of the relevance factors of the respective tags. In addition, χ_{locTag} denotes the boolean decision whether there is a location tag in this specific UGTC. In the second step, it is analyzed where the UGTC originates. If the UGTC was annotated with a location tag that is within one of the the lowest level of the regions of interest, then its values obtained by the previous maximum calculation is multiplied by 1. When the UGTC contains geospatial keywords referring to the region of interest, however the location tag indicates that it originates from another location, it is multiplied by 0.3. If the UGTC does not contain a location tag, it is multiplied by 0.5, since it is not clear whether the UGTC originates from one of the regions of interest. In our example microblog, the relevance would be calculated as follows: $max(rel(\text{``}earthquake\text{''})) = 0.8$ and $f_{loc} = 0.5$, which results in the relevance indicator 0.4. The described semantic model is still in an early stage, however it provides the basic links to the fundamental information for the assessment. Future extensions and links between the UGTCs could provide useful information in the assessment of the UGTC and aggregated results.

We believe that the usage of the maximum function is justified, because it allows the direct translation of the modeled relevance factors as the most important contributing element, as we assume that the topical keywords with the highest relevance has the highest impact. This is in accordance with the informal understanding, that the most severe element is the major contributor to its impact. We would like to note that the assessment of these values is out of the scope of this paper, since we are focussing on presenting a coherent visualization tool.

Controller. The controller is responsible for the processing of queries. The system allows to request the defined hierarchical layers and their assigned features. The request for a specific layer of interest is accomplished by the following steps:

First, the features of the layers are loaded that are displayed as overlays on the map. Second, for each feature of the layer all UGTCs are retrieved that were assigned as relevant to the feature. Subsequently, all UGTCs of all lower levels are retrieved. To limit the amount of results, the user has the option to specify the timeframe, the minimum of the calculated relevance, and specific topical keywords. The user can further explore the details of a UGTC by requesting the stored information and the enriched tags as well as he has the option to update the information and link the UGTC to further online resources.

6 Conclusion and Future Work

Enterprises and emergency response teams can make use of user-generated data to monitor events impacting critical infrastructures and, in case of crises, quickly assess resulting impacts to be able to decide about appropriate measures. In existing research, little focus has been onto the visualization and user-interaction in such systems for event-detection and the assessment of impact. We developed a tool for the visualization of geospatial impacts, which allows to analyze and assess impacts on different abstraction layers using geospatial hierarchical layers and enables therefore more intuitive and more user-specific representations.

This work presents a first approach of mapping UGTCs onto different hierarchical visualization layers. For future work we want to extend our prototype to a completely generic framework with comprehensive vocabulary for the hierarchical layer for visualization as well as the annotation of collected data. This would allow to set up such a system in a generic way and share and reuse information processed from the system. Moreover, we identified the following major aspects that allow further improvements. First, during the manual assessment of UGTCs supported by the hierarchical visualization layers, possible links between identified topics and contents could be identified and integrated into the data set. This semantically enriched data could be shared between domains and significantly improve the analysis and future decision-making in critical situations. Second, the recognition of geographical references could be improved by applying more sophisticated approaches in information retrieval. This could, among others, include the exploration of machine learning or the usage of further geographical attributes provided. Third, the presented visualization could be advanced by including additional attributes of analyzed feeds or by seamlessly adjusting polygons to different projections. Finally, the relevance calculation could be derived in a more sophisticated way including spatial features. All these adjustments could enhance the approach presented to a more sophisticated tool that would be applicable in multiple domains.

Acknowledgements This research was partially funded by the Vienna University of Technology through the Doctoral College Environmental Informatics and the Vienna PhD School of Informatics as well as by COMET K1, FFG – Austrian Research Promotion Agency.

References

1. L. Aiello, G. Petkos, C. Martin, D. Corney, S. Papadopoulos, R. Skraba, A. Goker, I. Kompatsiaris, and A. Jaimes. Sensing trending topics in twitter. *Multimedia, IEEE Transactions on*, 15(6):1268–1282, Oct 2013.
2. A. Boettcher and D. Lee. Eventradar: A real-time local event detection scheme using twitter stream. In *GreenCom*, pages 358–367, 2012.
3. R. Chunara, J. R. Andrews, and J. S. Brownstein. Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak. *American Journal of Tropical Medicine and Hygiene*, 86(1):39–45, 2012.
4. D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. Visual analytics: Scope and challenges. In *Visual Data Mining*, pages 76–90. 2008.
5. M. Krstajic, C. Rohrdantz, M. Hund, and A. Weiler. Getting There First: Real-Time Detection of Real-World Incidents on Twitter. In *Published at the 2nd IEEE Workshop on Interactive Visual Text Analytics "Task-Driven Analysis of Social Media" as part of the IEEE VisWeek 2012, October 15th, 2012, Seattle, Washington, USA*, 2012.
6. K. Leetaru, S. Wang, G. Cao, A. Padmanabhan, and E. Shook. Mapping the global twitter heartbeat: The geography of twitter. *First Monday*, 18(5), 2013.
7. A. Marcus, M. Bernstein, O. Badar, D. Karger, S. Madden, R. Miller, and O. Arenz. Twitinfo: Aggregating and visualizing microblogs for event exploration. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 227–236. ACM, 2011.
8. B. Pouliquen, M. Kimler, R. Steinberger, C. Ignat, T. Oellinger, K. Blackler, F. Fuart, W. Zaghouani, A. Widiger, and A.-C. Forslund. Geocoding multilingual texts: Recognition, disambiguation and visualisation. `http://publications.jrc.ec.europa.eu/repository/handle/111111111/12654`, 2006.
9. A. Rosi, M. Mamei, F. Zambonelli, S. Dobson, G. Stevenson, and J. Ye. Social sensors and pervasive services: Approaches and perspectives. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 525–530, March 2011.
10. A. Sadilek, H. A. Kautz, and V. Silenzio. Modeling Spread of Disease from Social Interactions. In *Sixth AAAI International Conference on Weblogs and Social Media (ICWSM)*, pages 322–329, 2012.
11. T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *In Proceedings of the Nineteenth International WWW Conference (WWW2010). ACM*, 2010.