Towards Text Processing System for Emergency Event Detection in the Arctic Zone

© Dmitriy Deviatkin © Artem Shelmanov

Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

devyatkin@isa.ru shelmanov@isa.ru

Abstract

We present the ongoing work on text processing system for detection and analysis of events related to emergencies in the Arctic zone. The peculiarity of the task consists in data sparseness and scarceness of tools / language resources for processing such specific texts. The system performs focused crawling of documents related to emergencies in the Arctic region, text parsing including named entity recognition and geotagging, and indexing texts with their metadata for faceted search. The system aims at processing both English and Russian text messages and documents. We report the preliminary results of the experimental evaluation of the system components on Twitter data.

Keywords: focused crawling, event detection, monitoring, named entity recognition, text processing, information search

1 Introduction

Due to ever-growing amounts of data available on the web, monitoring and searching in textual streams is still one of the most urgent problems today that has inspired researchers to develop many general-purpose information-retrieval methods and systems. However, the development of applications for specific domains often reveals lack of suitable techniques that could address challenging tasks arising in these domains, which require significant research.

This paper describes an ongoing development of a search and monitoring system for a specific domain and a task. It is oriented on detection and analysis of emergency events in the Arctic zone. Since a lot of textual information is generated during emergencies and crises, as during major events of other types, it is crucial to have automated tools for filtering and processing of unstructured textual data for support of search and rescue operations, as well as for helping people in affected areas. The Arctic zone is a hard but important and promising region that has a lot of potential for the development. The remarkable peculiarity of the chosen domain is data sparseness and scarceness of tools / language resources for processing such specific data, which poses a difficult problem.

The most significant features of the system are focused crawling and faceted search.

Since it is impossible to store all available data on the web, the developed system is designed to accumulate only data related to emergencies in the Arctic zone from multiple textual streams. The sources of such information include but are not limited to mass media, social networks, reports (e.g., official sources like national transportation safety boards[1,2]). The focused crawler is intended to narrow down the amount of indexed text and extract basic metadata of downloaded documents. At first sight, the problem of crawling messages about emergency events is very similar to topic crawling. The key difference lies in the fact that emergency related messages can be devoted to multiple topics and the composition of these topics can change over time. It means that using the ordinal topical approaches leads to inappropriate accuracy and laboriousness of the crawling process. To mitigate this problem, we have implemented the following ideas in the proposed framework:

- Multiple topic crawlers with narrow focuses outperform a single data collecting process in terms of recall.
- Geographical coordinate extracting and considering them for further filtering improve the accuracy of the crawling process. One could get topically irrelevant, but important messages from emergency zone.
- Topic models for crawled texts could be periodically built and verified for better tracking of topic shifts in text streams.
- Reposts and fuzzy duplicates can be effectively detected via inverted full-text indices [28].

The faceted search provides the abilities to retrieve and analyze texts in different perspectives: topic, time,

Proceedings of the XVIII International Conference «Data Analytics and Management in Data Intensive Domains» (DAMDID/RCDL'2016), Ershovo, Russia, October 11 - 14, 2016

[1]http://www.tsb.gc.ca/eng/rapports-reports/marine/index.asp

[2]http://www.ntsb.gov/investigations/AccidentReports/Pages/marine.aspx

location, relations with the given object, etc. The developed system performs deep natural language processing of texts (including syntax parsing and semantic role labeling), named entity recognition, as well as geotagging. The extracted metadata is indexed for the faceted search.

We evaluated the developed subsystems for geotagging, crawling, and faceted search on the data acquired from Twitter. Although this social network accumulates only short messages and is not designed for providing data for the considered tasks, many researchers, as shown in Section 2, demonstrated that tweets could be a useful source of information about emergencies. When common communication services are down, Twitter provides a channel, which is used by affected people and emergency response teams [22]. Therefore, we used messages crawled from Twitter for preliminary experiments, testing our approaches, and evaluation of the system components. However, we note that the developed system is designed to handle all sorts of textual information, not just short messages.

The rest of the paper is organized as follows. Section 2 reviews the related work about monitoring emergency events with help of social networks and focused crawling. Section 3 describes the details of the system in development; it presents the natural language processing pipeline, method for focused crawling, and faceted search techniques. In section 4, the results of the preliminary experiments are presented and discussed. Section 5 concludes and outlines the future work.

2 Related work

The problem of event detection in text streams has a lot of attention from the research community. Methods that were developed to address this problem were applied to many domains. One of them is monitoring emergencies. It was noticed that mass emergencies initiate the intensive exchange of information in social networks. This immense text stream contains cues about a situation in an affected area, infrastructure damage, human casualties, requests and proposals for help. It is a crucial information that can enhance the situation awareness [18] of both affected people and participants of rescue operations. However, it is mixed up with heavy noise: irrelevant or useless messages. Therefore, to put it to good use, new methods and technologies are required. The need of such technologies became apparent, which facilitated the development of many diverge systems for mining emergency related information in social networks. We review the most significant recent work on such systems.

Papers [20] and [17] present an information flow monitoring system Twittris designed for processing of short messages from mass and social media, as well as SMS-messages. Researchers tested the system on Twitter data. The system crawls messages from Twitter using a set of keywords, which is expanded over time by the most significant n-grams extracted from acquired messages. The system extracts the spatial and temporal information, as well as topics, which are used for message clustering. The clusters are considered as events found in an information stream. Researchers tested the system on the data acquired during hurricane Sandy. They showed that the system could be used for searching messages from affected people considering their location.

Another monitoring system SensePlace2, described in [12], specializes on analysis of the geographical data extracted from tweets. The system aims at improving the situational awareness during search and rescue operations. The main goal of the system is text stream filtering and searching of messages related to the given topic, place, and time. The system utilizes the geographical tags, as well as the information extracted from message texts. Besides text, Senseplace2 also indexes geographical and temporal information of messages. This enables the system to filter a message stream by place and time and build analytical reports for topic-time-location data. Senseplace2 can visualize results in different ways: as a common search result list, present them on a time scale as a histogram, and visualize results on a heat-map, which displays the intensity of messages about particular topic near the given location. Researchers tested the system using data related to the Haiti earthquake. They showed that SensePlace2 could be useful for finding refugee streams that are not represented in official sources.

In [24], researchers present a method for classification of messages acquired from a message stream. They demonstrate its capabilities of finding useful emergency related messages on Twitter data. The method can classify messages as useful and non-useful via standard supervised machine learning methods (Naïve Bayes and Maximum entropy). The most remarkable thing is a feature set used for training. Besides low-level features, they also conducted experiments with high-level features like message objectivity, whether it is personal or impersonal, whether it is formal or informal. The authors show that high-level features substantially improve the quality of classification. The out-of-domain evaluation showed accuracy from 30 to 80%. The experiments are conducted on the data acquired during Haiti earthquakes, USA wildfires, and floods.

The system EMERSE (Enhanced Messaging for the Emergency Response Sector) [4] collects messages from different sources, translates them, and classifies them into topics for better search and filtering. EMERSE consists of a smartphone application, a Twitter crawler, a translation subsystem, and a subsystem for classification. The smartphone application is considered to simplify a process of collecting messages and their metadata such as location, time, and associated media files (photo, video). Besides, the system crawls Twitter considering timestamps and eliminating duplicates (reposts). EMERSE classifies messages into multiple classes using support vector machine. In [4], authors experimented with different features and feature selection methods: bag of words, feature abstraction methods [21], Latent Dirichlet Allocation (LDA), and others. The system was tested on a collection of messages submitted to the

Ushahidi[3] web-service during the Haiti earthquake. In this example, the authors demonstrate that EMERSE can improve coordination of people during the emergencies.

In [25, 26], a system ESA (Emergency Situation Awareness) is presented. It can monitor social networks and blogs in real time and visualize information about different emergencies. The main task of the system is to enhance situational awareness of people in an affected area. The system is oriented on New Zealand and Australia regions. ESA gathers tweets and detects topical bursts in information streams. The retrospective data is used for building a language model, which is applied for the further burst detection. The algorithm searches lexis that has a very diverge distribution comparing to the language model. For convenient representation of bursts for end-users, ESA performs thematic clustering of messages. The system also selects informative messages that signal about emergencies, destructions, and requests for help. ESA has a component that extracts relevant spatial data using explicit geotags of messages (GPS-coordinates received from a smartphone) and implicit information, found in user profiles. The conversion from geographical names to coordinates is performed by Yahoo geo-service[4] (retired today). ESA also performs named entity recognition: it extracts names of organizations, names of people, geographical entities, dates, and timestamps. All these data can be visualized on a map, which could be useful for providing better representation of found events for end users. Visualization of data in ESA is also enhanced with media files (images, videos), extracted from messages. The authors tested ESA in Australian crisis center, which is responsible for monitoring of natural disasters and other national security threats.

AIDR[5] (Artificial Intelligence for Disaster Response) is an open-source platform for classification of messages related to emergencies [9]. The system detects messages about different topics: infrastructure damage, casualties, required or available donations. The authors point out that classifiers trained on the data collected during one disaster perform badly on the data acquired from new disasters. They address this problem by introducing human annotation into the process of adapting the system to new tasks. When a new emergency happens, the system should be retrained. The training dataset for supervised machine learning is composed from the old labeled data and the data urgently annotated via crowdsourcing services. The systems have elements of active learning; it chooses for human annotation the most informative samples that can significantly leverage classification performance. The authors tested the system on the collection of messages related to Pakistan earthquake in 2013.

TEDAS [10] is the system for emergency detection via focused crawling of Twitter messages. TEDAS collects topic relevant messages using the Twitter search API. The system uses the original crawling strategy that consists in dynamic shifting of crawler focus.

Another system for vertical search of information about emergencies is described in [27]. The system includes a focused ontology-based crawler. An extensive ontology describing various emergencies is designed for the crawler.

It is also worth mentioning Tweedr [2] – an open-source system that can find informative messages from Twitter for information support of people involved in rescue operations. It can distinguish general messages from the ones that have particular information about infrastructure damage and human casualties. Another recent effort in constructing tweet classification system is described in [5]. The authors use deep natural language processing techniques and rich set of features to determine whether a message contains information about damage dealt during natural disasters. In [14], an approach for construction of crisis-related terms is proposed. Authors used pseudo-relevance feedback mechanisms to expand a number of seeding terms during crawling, which results in recall improvement of retrieving of messages related to mass emergencies. Another lexicon called EMTerms is described in [23]. The authors claim that it is the biggest crisis-related lexicon for Twitter analysis so far.

Solutions for monitoring events in text streams heavily depend on focused crawling techniques. We review some of the state-of-the art approaches below.

ICrawl system [7] is a framework for focused crawling of social networks. It adopts ontology based crawling strategy. The novel feature of this system is a usage of Internet search engines for generation of bootstrap crawling points. In [3], researchers propose a distributed crawler for continuous message-gathering from particular user communities, which can circumvent limits of Twitter API. In [11], automatic Topic-focused Monitor is presented. It samples tweets from the message stream and selects keywords to track target topics based on the samples.

The review shows that there are a plenty of systems for monitoring emergency related events in textual streams intended to improve situational awareness of affected people and rescue teams. In our work, we consider a particular geographical region – the Arctic zone, which complicates focused crawling and filtering of data. Many aforementioned systems specialize on narrow problems like message classification, whereas our research is oriented on the development of a full-stack system that solves many tasks: from focused crawling and information extraction, to faceted search leveraged with spatial and temporal metadata. Unlike the aforementioned systems, the framework proposed in this paper is oriented on processing messages in both English and Russian languages. This is significant because of the large area of the Arctic territories of Russia. We note that many systems use Twitter data for evaluation, and we also use this approach in our work.

[3]https://www.ushahidi.com/
[4]https://developer.yahoo.com/boss/geo/
[5]http://aidr.qcri.org/

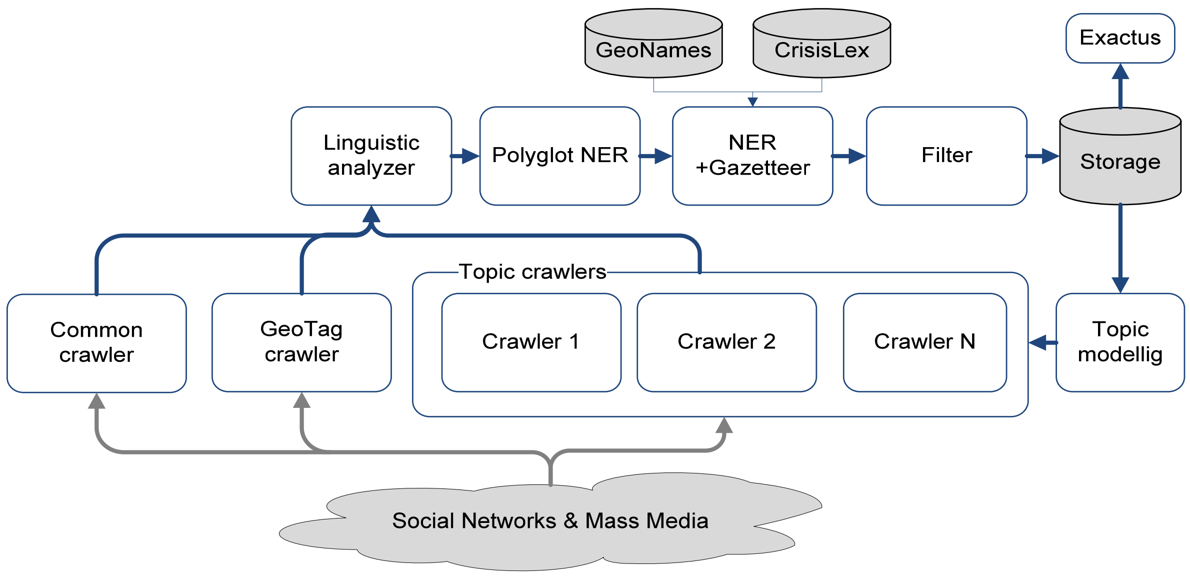

Figure 1 Framework for crawling of emergency messages

3 System components

3.1 Natural language processing pipeline

The system performs deep natural language processing of Russian and English texts. Besides basic processing tools, the pipeline also includes syntax parsing, semantic role labelling, and named entity recognition.

The basic analysis for Russian texts is performed by AOT.ru[6]. This framework is used for tokenization, sentence boundary detection, POS tagging, and lemmatization, including morphological disambiguation. We use MaltParser[7] trained on SynTagRus [13] for dependency parsing of Russian texts and our semantic parser for semantic role labelling [19]. The same types of linguistic analysis of English texts are performed via Freeling [16]. Note that the syntax and semantic annotations are used for information search (see section 3.3).

For the basic named entity recognition, we used Polyglot NER framework [1]. It implements language agnostic approach and due to this provides named entity recognition for many languages including English and Russian. It produces annotations for locations, organizations, and person names. However, we found that the basic NER processor is not suitable for extracting toponyms related to a particular region (e.g., Arctic zone); it yields low recall in this task. Therefore, we complemented Polyglot with a gazetteer.

The gazetteer was created on the basis of Geonames[8] database. It contains more than 11 million geographical locations of different types around the world with their names (in many languages including Russian and English), geographical coordinates, and other metadata. From Geonames, we extracted location names that are situated on the north of the 60th latitude. The gazetteer uses these data to mark spatial information in texts. It also implements rather simple rules to filter out common false positives that take into account parts of speech and capitalization of words.

We also tag crisis-related lexis in texts; it enhances and simplifies filtering and search. The data for this purpose is taken from CrisisLex lexicon, proposed in [14].

3.2 Focused crawling framework

We deal with several social networks, such as Twitter, Facebook, and VKontakte, and with some news feeds (ArcticInfo, BarentsObserver, BBC, etc.) These sources provide different kinds of content. The Twitter provides API for crawling of recent messages by keywords. However, the limitations of the API make the topical crawling process challenging. Since results commonly contain much irrelevant noise, additional filtering is necessary. We access Facebook and VKontakte primarily via links in twitter messages that are considered topically relevant. The news feeds have a static structure, therefore, they can be processed by a common crawler with a preliminary created static task. The data acquired from news feeds do not need topical filtering, because the crawling task can be restricted to process only relevant sections. Since we deal with a number of heterogeneous sources, we use several kinds of crawlers (see Fig. 1).

The first type is a GeoTag crawler. It is used for collecting messages from Twitter with specified coordinates. Tweets may include geographical coordinates or geo-tags, which could be used for localization of their authors. We filter all messages, whose geo-tag latitude is less than 60 degrees.

The second type is a Topic crawler. These crawlers download topically relevant messages from Twitter with unspecified coordinates. Each topic crawler has lists of "permissive" and "restrictive" terms that are fed to

[6]http://aot.ru/
[7]http://maltparser.org/
[8]http://www.geonames.org/

Table 1 Examples of topics for crawled data

No	Keywords	Relevant
1	Bay, charity, Amazon, Antarctica, cdnpoli.	False
2	Starling, Tuktoyaktuk, community, visit, bird, southern, blackbird.	True
3	Ice, national, ship, circle, arctics, photography, day, june, pewenvironment.	True
4	Rescue, buntings, air, guardsmen, squadron, cranes, divers, spot.	True
5	Haha, dart, Trump, meepismurder, white, sales, gauges, street.	False
6	Icebreaker, Nunavut, hardy, apithanny, piece, fascinating, blue, warming, bear.	True
7	Home, conservation, thebigbidtheory, may, island, science, hydrazine.	False
8	Spring, noaa, climatechange, water, super, sail, challenge, Mediterranean.	False
9	Arctic, Alaska, skuas, Greenland, road, amb, melt, Anchorage, Bering.	True
10	Life, natgeomag, trip, journey, remote, team, chukchi, collaborating.	True

Twitter search API. In the initial steps, several bootstrap terms are used for defining a target topic. The challenge lies in limitations of topic search API, provided by Twitter. It restricts a size of a query and a response, which leads to insufficient recall of the crawling process. The simplicity of the query language causes the low precision and recall of the collected data. We use multiple topic crawlers with different keyword subsets to solve the insufficient recall problem. NER and filtering are used to improve the precision.

The last type of crawlers is a common crawler. It collects data from topically related sections of news feeds. The crawlers of this type can also download pages from VKontakte and Facebook referenced by relevant Twitter posts.

The whole schema of data processing in our framework is the following. In the first step, messages are collected by GeoTag and Topic crawlers. In the second step, we apply linguistic analyser, NER, and gazetteer to the collected texts. Then, we filter all messages that do not contain any crisis lexis, toponyms, and geotags. URLs from the remaining messages are fed to the common crawler that also processes topically related news feeds. The selected useful messages and documents are indexed by the Exactus search engine [15].

For Topic crawler, we build a topic model [8] of the crawled messages every several days. It helps to track topic shifts in the message stream. We summarize topic content with a keyword cloud and a set of the most significant messages from the cluster. Then each topic is marked as relevant or irrelevant by several assessors (see Table 1). We define the following types of posts as relevant:

1. Posts about arbitrary events (past, current and planned) and locations in the Arctic.
2. Arbitrary posts from users, who currently are in the Arctic zone.

The most significant terms from the relevant topics are sent to "permissive" keyword collections of topic crawlers, and terms from irrelevant topics are sent to "restrictive" ones. Thus, the crawling process becomes responsible to trend shifts.

3.3 Faceted search

The faceted search became a backbone for professional search applications [6]. In this type of search, users can iteratively specify queries using metadata and keywords extracted from search results of previous iterations. Additionally, search results could be filtered using different sets of meta fields that can be static or dynamic.

In the developed system, the faceted search is powered by the Exactus technology [15]. Its main advantage lies in ability to efficiently index rich linguistic information including syntax relation, semantic roles, or other types of semantic annotations extracted from natural language text (e.g. named entities). This enables phrase search (results have to contain given syntactically connected phrases) semantic search (results are ranked taking into account semantic similarity of the query and indexed documents). We take advantage of this technology by introducing indexing by geographical tags, timestamps, and emergency-related tags. This provides the ability to filter results efficiently by semantic information like location, time, organizations, persons, and topics. It also provides the ability to retrieve information with certain tags filtered by other metadata producing the results that can be sifted with consequent queries.

4 Evaluation of system components

We have conducted a series of experiments to assess the quality of the created components for focused crawling, named entity recognition, and faceted search. The source of the data for evaluation is Twitter social network. The experimental dataset contains approximately 100 thousand messages in English and Russian. In the first experiment, we assessed accuracy of the proposed focused crawling framework. More specifically, we evaluated the quality of filtering. We labelled several subsets of posts devoted to accidents in Alaska and Bering Sea. Each post from the subsets was labelled by three assessors to reach sufficient coherence of the test data. We have not applied a cross-validation approach here because the labelling was not used for the crawler training, just for testing. The standard measures for supervised learning: precision, recall and F_1-score, were used for each subset. Macro-averaging was used to evaluate the result assessments. Table 2 refers to the results of the crawling without and with filtering as "Impure data" and "Filtered data" correspondingly.

Applying the proposed filtering technique results in the substantial growth of the precision without the significant decrease of the recall. This means that during the crawling process we do not lose much topically relevant data but substantially decrease the stored noise. We decided to choose a fairly soft filtering because, although a stricter procedure would improve the precision, it would also imply a more significant recall drop, which contradicts the purpose of the monitoring system.

Table 2 Focused crawling evaluation

	P	R	F_1
Impure data	0.26	1.00	0.41
Filtered data	0.57	0.94	0.70

In the second experiment, we estimated the performance of named entity recognition performed by Polyglot and gazetteer. We labelled all location mentions in 300 tweets that were downloaded by the Topic crawler and measured precision, recall, and F_1-score for extraction of spatial entities (Table 3).

Table 3 NER evaluation (on locations)

	P	R	F_1
Polyglot	0.78	0.57	0.66
Gazetteer	0.78	0.74	0.76
Polyg.+gazetter	0.76	0.82	0.79

Results show that proposed Gazetteer significantly outperforms Polyglot on location extraction in terms of recall. The knowledge source of Polyglot is Wikipedia that does not have the full coverage of locations. We conclude that it is reasonable to use the gazetteer and Polyglot together for the maximum performance.

In the last experiment, we assessed the performance gain of the information search achieved by using the proposed emergency faceted search method in comparison to the baseline algorithm. We deployed the Exactus full-text search algorithm without filtering by tag locations as the baseline. For the evaluation, we applied the NDCG score and peer reviewing approach. The results are presented in Table 4.

Table 4 Faceted search evaluation

	3-DCG	5-DCG	10-DCG
Faceted	0.76	0.76	0.70
Baseline	0.61	0.55	0.53

It was revealed that use of location and crisis tags for faceted search significantly improves the quality of ranking when searching posts about emergencies.

5 Conclusion

We presented an automated framework for crawling and processing textual documents about emergency events in the Arctic zone. The main functions of the proposed framework are focused crawling and faceted search that takes into account information about geographical locations and timestamps of messages. With the data crawled from Twitter, we experimentally demonstrated that the framework provides the basic abilities for analysis of message streams about emergencies in the restricted area.

In the future work, we are going to incorporate into the natural language processing pipeline components that extract information about ships and planes in the Arctic zone. Bulk information is available openly on the web (e.g., MarineTraffic service[9]). Tagging ship names and their coordinates in document and message streams potentially can improve the quality of emergency event detection and enhance the situation awareness.

We are going to accumulate more retrospective data from social networks and other sources to increase the recall of the crawling process. Among many other types of information sources, collections of reports from rescue services are the most prospect supplement for the crawling. Another way to improve topic crawling is detection of users and groups in social networks that constantly post topically relevant messages. This could be done semi-automatically by building topical models on users and groups. We are also going to create visualization tools for geotagged messages that can present events on the map.

Acknowledgments

The project is supported by the Russian Foundation for Basic Research, project number: 15-29-06045 “ofi_m”.

References

[1] Rami Al-Rfou, Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. Polyglot-NER: Massive multilingual named entity recognition. In Proceedings of the 2015 SIAM International Conference on Data Mining. SIAM, 2015.

[2] Zahra Ashktorab, Christopher Brown, Manojit Nandi, and Aron Culotta. Tweedr: Mining Twitter to inform disaster response. Proceedings of ISCRAM, pages 354–358, 2014.

[3] Matko Boanjak, Eduardo Oliveira, José Martins, Eduarda Mendes Rodrigues, and Luís Sarmento. Twitterecho: a distributed focused crawler to support open research with twitter data. In Proceedings of the 21st international conference companion on World Wide Web, pages 1233–1240. ACM, 2012.

[4] Cornelia Caragea, Nathan McNeese, Anuj Jaiswal, Greg Traylor, Hyun-Woo Kim, Prasenjit Mitra, Dinghao Wu, Andrea H. Tapia, Lee Giles, Bernard J. Jansen, et al. Classifying text messages for the Haiti earthquake. In Proceedings of ISCRAM, 2011.

[5] Stefano Cresci, Maurizio Tesconi, Andrea Cimino, and Felice Dell’Orletta. A linguistically-driven approach to cross-event damage assessment of

[9] http://www.marinetraffic.com/

natural disasters from social media messages. In Proceedings of the 24th International Conference on World Wide Web Companion, pages 1195–1200. International World Wide Web Conferences Steering Committee, 2015.

[6] Pavlos Fafalios and Yannis Tzitzikas. Exploratory professional search through semantic post-analysis of search results. In Professional Search in the Modern World, pages 166–192. Springer, 2014.

[7] Gerhard Gossen, Elena Demidova, and Thomas Risse. The iCrawl Wizard – supporting interactive focused crawl specification. In Advances in Information Retrieval, pages 797–800. Springer, 2015.

[8] Thomas Hofmann. Probabilistic latent semantic indexing. In Proceedings of the 22nd annual international ACM SIGIR Conference on Research and Development in Information Retrieval, pages 50–57. ACM, 1999.

[9] Muhamma-d Imran, Carlos Castillo, Ji Lucas, Patrick Meier, and Sarah Vieweg. AIDR: Artificial intelligence for disaster response. In Proceedings of the companion publication of the 23rd International Conference on World Wide Web Companion, pages 159–162, 2014.

[10] Rui Li, Kin Hou Lei, Ravi Khadiwala, and Kevin Chen-Chuan Chang. Tedas: A twitter-based event detection and analysis system. In Data engineering (ICDE), 2012 IEEE 28th international conference, pages 1273–1276. IEEE, 2012.

[11] Rui Li, Shengjie Wang, and Kevin Chen-Chuan Chang. Towards social data platform: Automatic topic-focused monitor for twitter stream. Proceedings of the VLDB Endowment, 6(14):1966–1977, 2013.

[12] Alan M. MacEachren, Anuj Jaiswal, Anthony C. Robinson, Scott Pezanowski, Alexander Savelyev, Prasenjit Mitra, Xiao Zhang, and Justine Blanford. Senseplace2: Geotwitter analytics support for situational awareness. In Proceedings of Visual Analytics Science and Technology (VAST) on IEEE Conference, pages 181–190, 2011.

[13] Joakim Nivre, Igor M. Boguslavsky, and Leonid L. Iomdin. Parsing the SynTagRus treebank of Russian. In Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), pages 641–648, 2008.

[14] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. CrisisLex: A lexicon for collecting and filtering microblogged communications in crises. In Proceedings of ICWSM, 2014.

[15] Gennady Osipov, Ivan Smirnov, Ilya Tikhomirov, Ilya Sochenkov, and Artem Shelmanov. Exactus expert – search and analytical engine for research and development support. In Novel Applications of Intelligent Systems, pages 269–285. Springer, 2016.

[16] Lluís Padró and Evgeny Stanilovsky. Freeling 3.0: Towards wider multilinguality. In Proceedings of the Language Resources and Evaluation Conference (LREC 2012). ELRA, 2012.

[17] Hemant Purohit and Amit P Sheth. Twitris v3: From citizen sensing to analysis, coordination and action. In Proceedings of ICWSM, pages 746–747, 2013.

[18] Nadine B Sarter and David D Woods. Situation awareness: A critical but ill-defined phenomenon. The International Journal of Aviation Psychology, 1(1):45–57, 1991.

[19] A. O. Shelmanov and I. V. Smirnov. Methods for semantic role labeling of Russian texts. In Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2014), number 13, pages 607–620, 2014.

[20] Amit P. Sheth, Hemant Purohit, Ashutosh Sopan Jadhav, Pavan Kapanipathi, and Lu Chen. Understanding events through analysis of social media. Kno.e.sis Center, Wright State University, Tech. Rep., 2010.

[21] Adrian Silvescu, Cornelia Caragea, and Vasant Honavar. Combining super-structuring and abstraction on sequence classification. In Proceedings of ICDM, pages 986–991. IEEE, 2009.

[22] Juan Sixto, Oscar Pena, Bernhard Klein, and Diego López-de Ipina. Enable tweet-geolocation and don't drive ERTs crazy! Improving situational awareness using Twitter. Proceedings of SMERST, pages 27–31, 2013.

[23] Irina Temnikova, Carlos Castillo, and Sarah Vieweg. Emterms 1.0: a terminological resource for crisis tweets. In ISCRAM 2015 Proceedings of the 12th International Conference on Information Systems for Crisis Response and Management, 2015.

[24] Sudha Verma, Sarah Vieweg, William J Corvey, Leysia Palen, James H Martin, Martha Palmer, Aaron Schram, and Kenneth Mark Anderson. Natural language processing to the rescue? Extracting "situational awareness" tweets during mass emergency. In Proceedings of ICWSM, pages 385–392, 2011.