Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A Spatial-Temporal Model for Event Detection in Social Media

Șerban Boghiu* , Daniela Gîfu<sup>a,b</sup>

*Faculty of Computer Science, “Alexandru Ioan Cuza” University, General Berthelot, 16, 700483, Iasi, Romania

<sup>a</sup>Institute of Computer Science, Romanian Academy - Iasi branch, Bulevard Carol I, 8, 700505, Romania

Abstract

Nowadays, the interest in data modelling from the spatial-temporal perspective is constantly increasing. Moreover, a wide variety of applications, such as social network data, need to be done to study spatiotemporal patterns. In general, however, these patterns are highly complex and challenging, so it is a demanding process to analyze or to classify them as the conventional context in various types of event data. In order to analyze the traffic viral within the text from the perspective of impressive negative effects, we should spatial-temporally localize the event and geographical regions and give a semantically interpreting of what happened. We propose a review of the best models and techniques applied for social media data processing to formalize a novel theory of action and time. This investigation intends to draw the basic knowledge level over which research intended to decipher in texts the occurrence of events, together with their involved characters, and their relationship with time and space.

© 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0).
Peer-review under responsibility of the scientific committee of the KES International.

Keywords: formalism; event analysis; spatiality; temporality;

1. Introduction

The concepts of time and space, from a pragmatic perspective, play a significant role in our life [1] that correspond to the ability to distinguish between alternatives. Most approaches must allow us to analyse and model
various processes with the use of interactions in spatial temporal space [2]. When data is collected across time as well as space and has at least one property (spatial or temporal), we can define a spatiotemporal model. Furthermore, to understand better what an event is, we must describe a spatiotemporal dataset, according to a spatial (certain place) and temporal (certain time) phenomenon. An example would be that of the patterns of COVID-19 mortality in China between December 2019 - March 2020, where the spatial property is the location – China with COVID-19 mortality rate information, and the temporal property is the time interval for which the event is valid – December 2019 - March 2020 COVID-19 mortality.

2. In recent years, more and more spatiotemporal models in social networks have been introduced to help with diverse features, being classified depending on how data is collected (across time or space) as follows [3]: (1) link-based - those models that apply link analysis (e.g. Page Rank) to recognize users with experience and locations with something special [4; 5]; (2) content-based - those models that use data from a user’s profile and the features of locations. [6; 7]; (3) collaborative filtering - those models that affect users’ preferences (e.g. location history). [8; 9]; (4) time-progressive - those models that consider the imminence of impact (i.e., the immediate, near, and far future) [10].

3. In order to be processed the text should be regarded as a set of chronologically and logically connected events or a set of words that are linked through relations. The relations can be extended from words to sentences or even paragraphs, but also from one event to another.

4. The legitimate research question of this paper intends to answer: How to efficiently implement a dynamic system using spatiotemporal event data on Social Media to recognize critical issues in real-time? Our research intends to draw the basic knowledge level over which research intended to decipher in texts the occurrence of events, together with their involved characters, and their relationship with time and space, using Social Media data. The final goal is to develop an integrated model for textual social media data with spatiotemporal dimensions.

5. The paper is structured as follow: Section 2 presents a short overview of studies of spatiotemporal models, considering the semantic level in order to clarify the role of them and what can we do to implement a new model with complex spatiotemporal features in Social Media data, while Section 3 refers to the data set, architecture and the design structure of our spatiotemporal system, named Real-Time Context of Virus Detection (RTCViD). Section 4 briefly discusses the evaluation of this platform before drawing some conclusions in the last section.

2. Previous Studies of Spatiotemporal Models

Previous studies of spatiotemporal models showed that even a short-term exposure to spread information proved can be hazardous to the ordinary people. Text, time and space are different domains with their own representation scales and methods [11]. In this study, we develop a spatiotemporal model with a semantically interpreting of what happened on social media.

2.1. Temporal Modelling

In general, for temporal processing the TimeML annotation scheme [12] is used, being a widely spread ISO standard that has great flexibility for expressing temporal events along with their relations with each other. This language model introduces three categories of time expressions that can be expressed are: durations (two hours), underspecified time events (Monday) and fully specified time events (December 2nd, 2019). Also, it allows notating correlations between events and time by using techniques such as anchoring one event to another, ordering the events from a temporal point of view. The available corpus annotated in TimeML standard are numerous (e.g. TimeBank 1.2 and TimeBank 1.1) [13; 14].

In this model, four main XML data structures are used to temporarily mark a text: (1) EVENT; (2) TIMEX3; (3) SIGNAL, and (4) LINK, briefly described below.

(1) Event attribute describes actions that can occur or have already happened, usually being associated with words in a sentence, mainly verbs, and have the corresponding tag <EVENT>. The EVENT tag can classify events through a series of characteristics with precise values [15]. There are several attributes that can be specified for this tag:
a) Class can have the values as Reporting, Perception, Aspectual, I_action, I_state, State and Occurrence. These values can offer important information about the narration and its subjectivity. In order to build a prediction algorithm, different syntax patterns are considered.

b) POS can describe the main parts-of-speech used in a sentence.

c) Tense can have the value as Past, Present, Future, Infinitive, Present part (ending in -ing), Past part (as on verbs ending in -en) and None.

d) Aspect can refer to certain temporal properties of the action, having the values as Progressive, Perfective, Perfective-Progressive and None.

e) Polarity can be 0 or 1, indicating the negative or positive form of the event.

f) Modality can be seen like Polarity, meaning that it has a value 0 or 1 that indicates the presence of a modal verb (should, could, would, may, have to).

(2) TIMEX3 attribute is usually used when we want to annotate specific temporal expressions (moments and periods of time) like fully specified interval or dates (2nd February 2019), underspecified intervals (Monday) or duration (six hours).

(3) SIGNAL attribute is usually used for marking the temporal relationship between events. The words that are annotated with <SIGNAL> are mainly words that express a temporal order of the events, such as: on, before, after.

(4) LINK attribute is done through the ID of events. There are several types of links that are supported in TimeML: ALINK, TLINK, SLINK, allowing the annotation of events relationships. ALINK comes from Aspectual Link and marks an event that is an aspect of another event or it refers to it [16]. TLINK comes from Temporal Link and marks a temporal relationship between events or the positioning in time of the events. SLINK comes from Subordination Link and is usually used whenever an event instance subordinates another event instance type.

TimeML contains many libraries that offer support for temporal text interpretation. One of the libraries is SUTime, being developed by Stanford University to recognize and normalize time expressions. The system has a deterministic rule-based designed for extensibility and flexibility.

In order to identify the temporal expressions based on the predefined rules for each language, SUTime uses RegexTokens, a framework that helps mapping patterns found in text with semantic objects. The main types of algorithms used for recognizing and normalizing expressions are: (1) text regex rules - temporal expressions are mapped with simple regex over characters; (2) compositional rules - temporal expressions are mapped using regular expressions over chunks (tokens and temporal objects) or (3) filtering rules - ambiguous expressions are removed from the candidates list.

We decided to use SUTime because it has above average results and it is easy to integrate with our technology stack. In order to extract temporal events, the Time Yards Model is helpful for ordering events and assigning them to actors. A time segment can be described as a sequence of events that can evolve linearly or not at all and are narrated uninterruptedly in a span of text [17].

2.2. Spatial Modelling

The spatial processing part is done using the SpatialML standard that offers support for spatial text interpretation. The main reason for our decision to use it is the high number of libraries that use this format to express spatial expressions. In general, Spatial ML focuses on geographical annotation including relevant landmarks and provides spatial information about a domain [18]. Note that, only expressions that mark a specific place, such as a country, continent, state, city, river etc. will be annotated. The values used in the annotation are NAM and NOM.

One of the libraries that we will use is Named Entity Recognition (NER), which marks any named locations in a text. It will be very useful as well when the time tracks will be built because each entity can be located both by time and by space in a span of text, increasing the expressivity. We opted for the language independent NER of Stanford University. It suggests potential tags depending on the selected configuration such as: location, organization, facility, date, money, people, percent, time (Stanford Named Entity Recognizer - https://nlp.stanford.edu/software/CRF-NER.html).
2.3. Semantic Modelling

The semantic analysis intends to explore the nature of meaning in language. In this context, this level of analysis is opening new ways for exploring challenging questions about the most pressing contemporary social and political issues [19]. Additionally, as spatial and temporal events domains become larger, this analysis level is more diverse and has higher perplexity [20].

In order to decipher positive and negative emotions that are present in our dataset we create a neural network. The main purpose of this approach is to correlate syntax with semantics within dependency grammar. This can be done using a tool named Treeops that converts an XML format to another following a set of rules [21].

Previous works include the UAIC Ro-Dia Dependency Treebank, where the annotated data was divided into 14 types of circumstantial modifiers: concession, condition, consecutive, cumulative, causal, exception, instrumental, local, modal, opposition, relative (or referential), purpose, associative and temporal. The corresponding semantic tags for these are CNCS, COND, CSQ, CUMUL, CAUS, EXCP, INSTR, LOC, MOD, OPPOS, REFR, PURP, ASSOC, and TEMP. For the moment, only unambiguously data can be annotated automatically. The training data consists of a collection of long sentences that have the structure of a tree where the leaves (annotated parts of the sentence) are linked between them through operators (that resemble logical operators, such as conjunction, implication etc.). A few modifiers can specify purpose, location or temporal meanings so they were annotated as an expression.

In this study, we focused on the COVID-19 topic, being classified as a high-scale event. We consider that detecting high-scale events poses general interest since they are usually shown or inherent and there is a significant prior knowledge of events that may happen (Lasswell's 5 W's: Who, What, Where, Whom, Why) [22]. For the semantic approach, based on Krippendorf's interpretative approach [23], a survey should suggest, in an objective and systematic way, encoded meanings in the analysed dataset. Of course, according to the context in which it is created and transmitted.

Our system, RTCoViD, aims to provide a reliable way to interpret tweets and textual information and correlate them to global events as well as predicting the evolution of the event based on existing data set from previous global pandemics.

3. Data Set and Method

This section involves collecting and compiling social media dataset and inspecting its metadata. In order to decipher in texts, the occurrence of events, together with their involved characters, and their relationship with time and space, we describe a machine learning approach to build a new system. In general, small documents like tweets do not have enough information to make their semantic context clear. In contrast, novels or other larger documents have too much variation. The big challenge is to find a good segmentation. In order to elucidate this issue, we preferred to combine both types of texts.

3.1. Data Set

Due to the relevance of the COVID-19 global pandemic, we use a dataset acquired from John Hopkins University (2019-nCoV Data Repository - https://github.com/CSSEGISandData/COVID-19), which operates a dashboard for tracking the COVID-19 and shares a data repository. Their model resembles our previously defined model: every record provides a Province, Country, Timestamp, Number of confirmed cases, Number of deaths, Number of recovered cases, Latitude and Longitude. This model can easily be mapped to our above described STM model. We are also going to use a dataset of tweets acquired from the Twitter Stream. From the start day to the end day, we can analyse the dramatic increase as the awareness for the virus is spreading. From March 11th to 1st of April we gathered about 4 million tweets a day regarding COVID-19 related topics only. All the gathered data are going to be parsed and imported into our SQL databases. This way we can easily access, manipulate and analyse all the information. Every existing database entry will be further analysed through our STM Pipeline.

In order to deeply analyze a record, it can be split into sub records, and it can be displayed as a collection of them.
3.2. System Architecture

The web application will be hosted in the cloud in order to provide high performance and reliability. As described above, to persist our data we are going to use Microsoft SQL Server. In order to assure the cross-platform development we are going to use the ASP.NET Core framework.

![System Architecture Diagram](image-url)

**Fig. 1. The System Architecture and the components of the basic NLP pipeline components**

Figure 1 describes the components that constitute the most useful NLP pipeline and the corresponding dependencies (https://gokulchittaranjan.wordpress.com/2015/09/08/nlp-pipelines-1/). A description of each of the phases used in this system and state-of-art performances will be presented below. Note that all significant components have been listed, the NLP pipeline is constructed by customizing, adding, and/or removing all these.

Therefore, each text record will go through the following steps:

- Sentence Boundary-Detection
- Part-of-speech Tagging (POS-tagger)
- Named Entity Recognition (NER)
- Coreference Resolution
- Temporal Tagging
- Spatial Tagging
- Sentimental Analysis (SA) using Neural Network

Most of these NLP phases are especially challenging when applied to noisy social media texts [24]. Each step is going to alter the existing text record where metadata will be added - through predefined XML tags.

**Sentence Boundary-Detection**
Here, each sentence will be included in a <SENTENCE> element, having its own attributes like: ID - a GUID that will uniquely identify the sentence, NUMBER.

Lorem Ipsum is simply a dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularized in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

All described NLP steps do not modify the existing content of the text. They only add deeper levels of understanding for the textual data.

**POS-tagger**

Based on these annotations, the rendering engine will display relevant information for any attribute and tag that are available on the content. For POS tagging, we are going to use Stanford Log-linear Part-Of-Speech Tagger for .NET. The tagger is licensed under the GNU General Public License (https://github.com/sergey-tihon/Stanford.NLP.NET/blob/master/LICENSE.txt) and based on text records, it assigns part of speech to each word (and other tokens) - such as noun, verb and adjective.

Example:

```csharp
// Text for tagging
let text = """A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'."

tagTextFromReader <| new StringReader(text)
```

As we can see, the tagger format† does not respect our tagging convention. In order to use it we must parse the provided output and map it to our existing XML tag model. This way, all the libraries and tools that we use provides us a unified output.

**NER & Coreference Resolution**

Coreference Resolution is dependent on NER (Fig. 2). It assigns predefined categories to them such as: person names, organizations and locations. For this task, we chose the Stanford Deterministic Coreference Resolution System.

> “I voted for Nader because he was most aligned with my values,” she said.

† [https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html](https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html)
Temporal Tagging

The text content of each record is temporally tagged using the TimeML standard that is described above. Each tweet or text content will be analysed, and any temporal expression will be marked with an TIMEX3 tag as illustrated below. If there is no identified expression, the date of the tweet will be annotated and attached to the resource. Possible LINK for each tweet can be made using its’ comments.
Example:

```
<TIMEX3 tid="t1" type="DURATION">Thursday</TIMEX3>, I <EVENT eid="e1" class="I_ACTION">took</EVENT> the plane to Copenhagen.
```

SUTime is a library developed by Stanford University to recognize and normalize time expressions. It detects temporal expressions and computes a timespan between that moment and the current time reference.

For example, the expression “next Wednesday at 3 p.m.” will be translated as “2019-06-25T15:00” depending on the current date and time.

Below we can see few examples of temporal expressions:

| Expression       | Type   | Value         |
|------------------|--------|---------------|
| October          | DATE   | 2011-10       |
| last Monday      | DATE   | 2019-07-01    |
| January of 1994  | DATE   | 1994-01       |
| the nineties     | DATE   | 199X          |
| winter of 2000   | DATE   | 2000-W1       |
| 6 p.m. Tuesday   | TIME   | 2019-09-20T18:00 |
| 6th century B.C. | DATE   | -06XX         |
| now              | DATE   | PRESENT_REF   |

Spatial Tagging

Spatial ML is a standard that focuses on geographical annotation including relevant landmarks and provides spatial information about a domain. The tagging is done using the Named Entity Recognition task which also recognizes the locations in the given text content. If there is no location discovered, the location of the tweet will be added to its content as described in the example below.

```
I visited many trattorias in Rome, Italy.
```

Sentiment Analysis

SA using Neural Network is done through the CNTK.GPU. This is a library developed by Microsoft trains that runs deep neural networks. This network is using a single LSTM layer to process the tweets and text content, and a single dense layer to classify the results into a positive or negative prediction.

To increase the accuracy of prediction I will use the IMDB Movie Dataset. This is a dataset with 25,000 positive movie reviews and 25,000 negative movie reviews.
Mining the abundant information on Social Media can reflect responsive plans for ongoing events questioned by users and situational awareness companies [25].

RTCoViD offers the option of uploading own records that will be processed by our pipeline. Above, we have described a few use cases for the platform and what workflows are enabled by it.

4. Evaluation

In order to improve RTCoViD, several tests were conducted. Accuracy (see formula 1) is commonly used in order to evaluate the performances of the proposed models. Therefore, we used the Multinomial Naïve Bayes classifier for this study. In fact, this formula is described as the total number of correct classifications over the total number of classifications made at a given point in time.

\[
\text{Accuracy} = \frac{\text{(correct classifications)}}{\text{total classifications made}}
\] (1)

In addition, we also created the confusion matrix, known as the error matrix, that is extracted based on the predictions made by the classifier. This matrix helps us to clearly visualize the classifier confusion.

Some quantifiers will be retrieved from the daily reports such as number of deaths per day or the growth of active cases per day. These quantifiers are mapped to tweets based on their date time information. In order to increase our accuracy, we must further research each parameter that is used in the prediction algorithm. Based on this research, we may add or subtract parameters to provide a better accuracy.

Confusion matrix

Table 2 shows the results of Multinomial Naïve Bayes Multi-Class classification model (Prediction result - horizontally, Actual result - vertically) on a set of test data which the true values are known. We can observe a remarkable symmetry on the diagonal meaning that over 65% of tweets have been correctly classified. The prediction rate of diametral opposed classification class (I – H, H – I) is very good (2 – 70, 70-2 for insignificant tweets).

| Insignificant (I) | Low (L) | Medium (M) | High (H) |
|-------------------|---------|------------|----------|
| Insignificant     | 70      | 20         | 8        | 2        |
| Low               | 8       | 65         | 12       | 15       |
| Medium            | 10      | 10         | 68       | 12       |
| High              | 2       | 10         | 16       | 72       |

This matrix can also be visualized as a chart (Fig. 3):

\[
\begin{array}{cccc}
70 & 20 & 8 & 2 \\
8 & 65 & 12 & 15 \\
10 & 10 & 68 & 12 \\
2 & 10 & 16 & 72 \\
\end{array}
\]

I L M H

As presented, the community has responded in the coronavirus outbreak by generating datasets that can be used for epidemic modeling. However, as we show, there is still a need for more rigorous approaches and tools that will be able to handle the variety and complexity of the social media data.
Fig. 3. Confusion matrix

Basically, our detection algorithm has a good prediction accuracy: from 100 randomly picked insignificantly tweets, 70 are classified as correctly, and only 20 are classified with a low chance because of the features of similarity, 8 with a medium chance, 2 with a low chance for pandemic occurrences.

5. Conclusions

As presented, the community has responded in the coronavirus outbreak by generating datasets that can accelerate research for a new treatment or forecasting models that can better predict the behavior and even warn us about future disasters.

Previous literature has shown that spatio-temporal-semantic models have a great potential to provide reliable predictions of infectious diseases in time and space. This study could be beneficial for authorities, epidemiologists, physicians, etc. to keep track of dynamics of virologic indicators in order to help build smart vaccine tech, by providing new ways of detecting potential viruses and their effects on human health. Based on our research, governments can be informed so they can easily act on time imposing specific restrictions that can help stop the spread of pandemic diseases.

Data Availability

All datasets are available through our git repository at https://github.com/inextricable/SS_Annotator.NET. Our data collection method complies with the terms of service of Twitter. Also, the datasets are anonymized to protect the identity of users.

Acknowledgements

This work was supported by a grant of the Ministry of Research and Innovation, Program PN-III-P1-1.2.-PCCDI, nr. 73/2018, as part of the ReTeRom project.

References

[1] Gif, Daniela, Mirela Teodorescu, and Dan Ionescu (2014). “Pragmatical Rules for Success”. International Letters of Social and Humanistic Sciences, 26(2014), 18-28.
[2] Klimešová, Dana and Eva Ocelíková (2012). “Spatial-temporal modeling and visualization”. International Journal of Mathematical Models and Methods in Applied Sciences 6(1) - https://www.researchgate.net/publication/265193407_Spatial-temporal_modeling_and_visualisation
[3] Zarezade, Ali, Sina Jafarzadeh, and Hamid R. Rabiee (2018). “Recent spatio-temporal modeling of check-ins in location-based social networks”. PLoS One. 2018; 13(5): e0197683, doi: 10.1371/journal.pone.0197683
[4] Liu B, Fu Y, Yao Z, Xiong H. (2013). “Learning geographical preferences for point-of-interest recommendation”. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 1043–1051.
[5] Scellato S, Noulas A, Mascolo C. (2011). “Exploiting place features in link prediction on location-based social networks”. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 1046–1054.
[6] Xiao X, Zheng Y, Luo Q, Xie X. (2014). “Inferring social ties between users with human location history”. Journal of Ambient Intelligence and Humanized Computing, 5(1), 3-19
[7] Liu B, Xiong H. (2013). “Point-of-Interest Recommendation in Location Based Social Networks with Topic and Location Awareness”. Proceedings of the 2013 SIAM International Conference on Data Mining. SIAM, 396-404.
[8] Yuan Q, Cong G, Ma Z, Sun A, Thalmann NM. (2013). "Time-aware point-of-interest recommendation". Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, 363-372.
[9] Huayu, R H, Li YG, Zhu H. (2016). "Point-of-interest recommendations: Learning potential check-ins from friends". In: Proceedings of the 22th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.
[10] Appel, G., Grewal, L., Hadi, R. et al. (2020). The future of social media in marketing. Journal of the Acad. Mark. Sci. 48, 79-95 - https://doi.org/10.1007/s11747-019-00695-1
[11] Diaz, Juglar, Barbara Poblete, Felipe Bravo-Marquez (2020). “An integrated model for textual social media data with spatio-temporal
dimensions.” Information Processing & Management - https://doi.org/10.1016/j.ipm.2020.102219

[12] Pustejovsky, J., Castaño, J.M., Ingria, R., Sauri, R., Gaizauskas, R.J., Setzer, A., Katz, G., & Radev, D.R. (2003). “TimeML: Robust Specification of Event and Temporal Expressions in Text”. New Directions in Question Answering.

[13] Verhagen, M. and Moszkowicz, J. (2008). “AQUAINT TimeML 1.0 Corpus Documentation”, Brandeis University.

[14] Pustejovsky, J., Verhagen, M., Sauri, R., Littman, J., Gaizauskas, R., Katz, G, Mani, I., Knippen, R., Setzer, A. (2006). “TimeBank 1.2”. Philadelphia: Linguistic Data Consortium.

[15] Sauri, R., Goldberg, L., Verhagen, M., Pustejovsky, J. (2009). “Annotating Events in English TimeML Annotation Guidelines”, Version TempEval-2010.

[16] Sauri, R., Littman, J., Knippen, B., Gaizauskas, R., Setzer, A., Pustejovsky, J. (2006). “TimeML Annotation Guidelines”, Version 1.2.1

[17] Cristea, D. (2017). “The Time Yards Model: a Way to Decipher the Evolution of Actors over Time in Texts”. The 6th International Conference on Mathematics and Informatics, Tg. Mureș.

[18] Mani, I., Hitzeman, J., Richer, J., Harris, D., Quimby, R., Wellner, B. (2008). “SpatialML: Annotation Scheme, Corpora, and Tools”. The 6th LREC 2008.

[19] Nguyen, D., Liakata, M., DeDeo, S., Eisenstein, J., Mimno, D., Tromble, R., Winters, J. (2019). “How we do things with words: Analyzing text as social and cultural data” - https://arxiv.org/abs/1907.01468

[20] Yuqian Huang, Y., Li, Y, Shan, J. (2018). “Spatial-Temporal Event Detection from Geo-Tagged Tweets”. International Journal Geo-Information, 7, 150 - doi:10.3390/ijgi7040150.

[21] Mărânduc, C, Mititelu, C., Bobicev, V. (2018). “Syntactic Semantic Correspondence in Dependency Grammar”. Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories (TLT16), Prague, Czech Republic, 167-180.

[22] Signori, P. & Rigon, S. & Flint, D. (2015). “Sentiment and Content Analysis to cluster neutral messages online”. Annual International Marketing Trends 2015, Paris.

[23] Krippendorf, K. (2004). “Content Analysis: An introduction to its Methodology”, II ed. Thousand Oaks, Calif., Sage.

[24] Eisenstein, J. (2013). “What to do about bad language on the internet”. Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, Georgia, 359–369

[25] Xu, S., Li, S, and Huanh, W. (2019). " A spatial ‐ temporal ‐ semantic approach for detecting local events using geo ‐ social media data". Transactions in GIS, 24:1, https://doi.org/10.1111/tgis.12589.