Text Analysis Tool TWeet lOcator – TAT2

Adam Flizikowski[1,2]
[1]University of Technology and Life Sciences, Bydgoszcz, Poland
adamfli@utp.edu.pl

Marcin Przybyszewski
[2]ITTI Ltd., Poznań, Poland
mprzybysz@itti.com.pl

Anna Stachowicz
ITTI Ltd., Poznań, Poland
astachowicz@itti.com.pl

Tomasz Olejniczak
ITTI Ltd., Poznań, Poland
t_olejniczak@itti.com.pl

Rafał Renk[2,3]
[3]Adam Mickiewicz University, Poznań, Poland
rrenk@amu.edu.pl

ABSTRACT

Information about location and geographical coordinates in particular, may be very important during a crisis event, especially for search and rescue operations – but currently geo-tagged tweets are extremely rare. Improved capabilities of capturing additional location from Twitter (up to 4 times improvement) are crucial for response efforts given a vast amount of messages exchanged during a crisis event. That is why authors have designed a tool (Text Analysis TWeet lOcator – TAT2) that relies on existing open source text analysis tools with additional services to provide additional hints about people location. Validation process, complementing experimentation and test results, included involvement of end-users (i.e. Public Protection and Disaster Relief services and citizens during a realistic crisis exercise showcase. In addition, the integration of TAT2 with external tools has also been validated.

Keywords

Social media, crisis management, location of Twitter messages, iSAR+, AIDA

INTRODUCTION AND RATIONALE

The use of social media, including Twitter, during the crisis has been reported in many previous works (Palen, 2008; Vieweg, Hughes, Starbird and Palen, 2010; Wilensky, 2014; Yin, Lampert, Cameron, Robinson and Power, 2012). The iSAR+ project[1] (full name - Online and Mobile Communications for Crisis Response and Search and Rescue) in which proposed TAT2 (Text Analysis TWeet lOcator) has been developed, is focused on using social media, mobile and online technologies for the purpose of crisis response and search and rescue operations. Surveys performed recently within this project (Flizikowski, Hołubowicz, Stachowicz, Hokkanen, Kurki, Päivinen and Delavallade, 2014) have shown that PPDR (Public Protection and Disaster Relief) organizations are expecting citizens to support ongoing disaster relief by providing information about the current situation (messages, photos, video) and information about location of events using social media. Results of a similar survey are presented in (MacEachren, Jaiswal, Robinson, Pezanowski, Savelyev, Mitra, Zhang and Blanford, 2011). Crisis/emergency management professionals have been asked about the potential use of social media in crisis management. 75% of respondents indicated that the most desirable information for them are maps showing location of individuals sharing the information during the crisis events. Both presented works show how important and valuable for PPDRs are geographical origins of social media messages (leading to location of individuals and events) for the purpose of crisis response efforts. An interesting conclusion is that tweets with geo-location are more likely retweeted than tweets that do not contain that information (Vieweg et al., 2010), proving how powerful geo-localization is.

The potential of social media and Twitter in particular has been also recognized by humanitarian organizations.

[1] http://isar.i112.eu/

This recently motivated The United Nations Office for the Coordination of Humanitarian Affairs[2] to create and release a report entitled "Hashtag standards for emergencies" (Meier, 2014; UN OCHA report, 2014). This report aims at standardizing the approach to using social media during crisis events, especially focusing on Twitter and the policy of using hashtags. One of recommendations is that response agencies should encourage citizens to use geo-location in the tweets related to crisis events.

According to different sources currently only about 1% to 3% of tweets are geo-tagged (Herfort, de Albuquerque, Schelhorn and Zipf 2014; Schulz, Hadjakos, Paulheim, Nachtwey and Mühlhäuser, 2013). This information has been confirmed during the tests presented in this paper (performed with the TAT2 tool). When retrieving tweets to be used in tests, only 1% or even less had geo-tags included. Moreover, not all users providing relevant information during the crisis reside in the location they report about, which makes geo-tags useless. Citizens may avoid activating geo-tags because of high battery consumption when GPS (Global Positioning System) functionality is activated or low level of trust towards sharing their exact location. That is why Text Analysis Tool TWeet lOcator (acronym TAT2) has been developed and presented in this paper. For the messages lacking geo-tags information, this tool attempts to estimate localization of a given tweet origin based on its contents. The idea behind TAT2 is that it especially uses the information contained in the Twitter message body in order to extract possible location names (toponyms). However, concluding on geographical location from toponyms raises the following challenges:

- the same place can be named differently, e.g. using abbreviations,
- the same name may refer to different places,
- the same name can refer to a geographical location or to something different, like the name of people, things, organizations etc.

Thus, disambiguation of possible location is needed and this process is called toponym resolution. As a result, the tool gives the list of possible locations inferred from a Twitter message. It needs to be emphasized that TAT2 is focused on localization of the message origins, and not on the user location indicated by his/her profile. The proposed TAT2 architecture leads to the results which provide localization information for up to 4% (to be precise – about 4% in the radius of 200 km and about 2% in the radius of 10 km) of tweets instead of the current 1-3% of geo-tagged tweets. This objectively small increase becomes important when thousands (or millions) of tweets are considered in a real event.

RELATED WORKS

The concept of using Twitter content and its analysis in terms of geo-localization is presented and applied in many research works, for a variety of purposes. Different methods present different ways in which geographic information can be extracted from tweets, i.e.:

- using geo-tagged tweets,
- analyzing the content of the message,
- analyzing meta-data such as user profile, user location field, time zone,
- using toponym resolution,
- analyzing relationships with other users of social media,
- using geo-coding services.

In (Cheng, Caverlee and Lee, 2010) authors focus on user's location within a city and the content is searched for geo-scoped words. The experiment results show that this system placed 51% of Twitter users within 100 miles from their actual location. Geocoding services are the alternative to retrieval of coordinates from the content of the message (e.g. street addresses or postal codes) when the tweet is not geo-tagged. In (Yin et al., 2012) Yahoo geocoding service[3] has been used. An analysis of content of the message in order to extract location has been performed in (Wilensky, 2014). The downside of approaches based solely on the message text is that they do not disambiguate location mentions, which leads to inconsistent and misleading results. In order to provide more accurate results and increase efficiency of geo-localization, many of the researchers use more than one method to assess tweet or user location. They combine different sources of information about location such as text of tweets, location provided by geography coordinates, city or country name, administrative division information meta-data related to this message, information contained in the user's profile (location and time zone) and place-related hashtags or automated entity extraction methods (MacEachren et al., 2011; McClendon and Robinson, 2012; Schulz et al., 2013; Wilensky, 2014; Yin et al., 2012). The problem with disambiguation of location and

[2] http://www.unocha.org/,
[3] http://developer.yahoo.com/geo/placefinder

location homonyms highlighted e.g. by (McClendon et al., 2012; Wilensky, 2014) is addressed in (Schulz et al., 2013), making the work particularly interesting from the iSAR+ point of view.

The results presented there are very promising and the approach combines various spatial indicators (Schulz et al., 2013):

- tweet message with the addition of identification of found entities coordinates,
- location field using Geonames[4] for toponym resolution process,
- website entries that might provide information about the home country of the user,
- Time Zone – initially set by Twitter or can manually be changed by the user.

The overall results from c.a. 1 million of extracted tweets show the efficiency of the estimation to be of 54% of tweets within a 50 km radius. Unfortunately, there is no clear information about how this estimation was calculated. TAT2 presented in this paper takes advantage of the previous research results and utilizes different approaches combined together. It uses spatial indicator as defined in (Schulz et al., 2013), as well as geocoding services as in (Yin et al., 2012).

The geo-tagged tweets have been used for different kinds of analysis of tweets distribution, their content and correlations with geographical location during crisis events (Banford, Bernhardt, Savelyev, Wong-Parodi, Carleton Titley and MacEachren, 2014, Caragea, Squicciarini, Stehle, Neppalli, and Tapia, 2014). There are also platforms and tools which use geo-tagged tweets to perform the analysis of social media posts during crisis events. An interesting example is Geofidia (Geofidia) enabling the user to display and analyze all geo-tagged social media posts from such platforms as e.g. Twitter, Facebook, Instagram, YouTube. "Artificial Intelligence for Disaster Response" (AIDR) platform also uses geo-tagged tweets and aims at identifying automatically informative content on Twitter during disasters (Meier, 2013). However, very low percentage of geo-tagged tweets (1-3%) is a major limitation for pursuing research analyzing social media according to users and/or messages location. This is the reason why alternatives of social media messages locations have been investigated. In the next steps, TAT2 could be integrated with such systems, which use only geo-tagged tweets as additional source of input data.

METHODS AND TOOLS

The general approach used inside TAT2 is based on localization using disambiguation tools. It has been decided to use existing open source tool for entity detection and disambiguation for TAT2. From a short list of the existing tools that detect and disambiguate various entities from the text, currently DBpedia Spotlight[5] and AIDA[6] (Accurate Online Disambiguation of Named Entities in Text and Tables) are distinguishing themselves to be the most reliable frameworks. In order to examine their capabilities, experiments including entering five different examples of tricky sentences were carried out on the web demo versions of both tools. The experiment results show that AIDA seems to be a vastly better framework by taking into consideration the whole sentence rather than individual entities. For the purpose of integration with TAT2, full version of AIDA has been used. The code of AIDA is open and thus easy to integrate with TAT2. Moreover, TAT2 includes additionally geo-coding services, which makes the results more accurate, as well as time zone and area filtering.

TAT2 ARCHITECTURE

An input data for TAT2 is a tweet message consisting of content and metadata. As output information the tool returns twitter message together with a list of possible locations associated with that tweet. The location list provides information on e.g. the coordinates (or bounding box for area), associated toponym and possibly confidence value for each of the locations provided for the message. Messages coming from social media (Twitter) are first analyzed in AIDA module (see TAT2 architecture in **Error! Reference source not found.**). In the iSAR+ system, those tweets are collected and crawled by OsintLab (Open Source Intelligence Lab) tool developed by Thales Communications and Security[7] (Mabiala, De Maupeou, Gouttas, Huyot and Delavallade, 2013) and adapted to crisis management issues within the iSAR+ project. OsintLab provides TAT2 with tweets gathered, pre-analyzed and associated with metadata and relations between texts, authors and sources of tweet.

[4] Geonames – geographical database that covers all countries and contains over eight million placenames,http://www.geonames.org/

[5] https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki

[6] http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/aida/

[7] www.thalesgroup.com

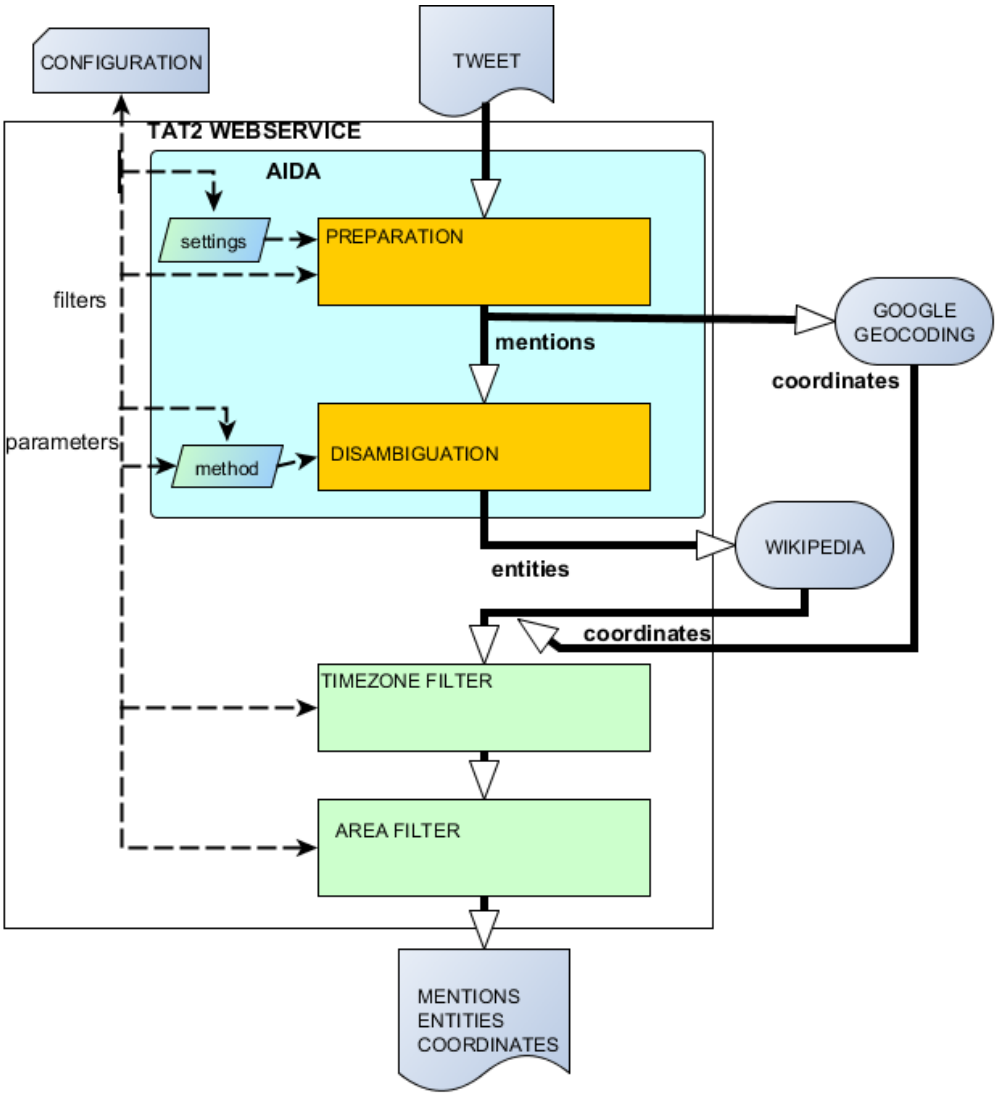

Figure 1 TAT2 architecture

Using AIDA tweets are then processed in the following two functional blocks:

1. The preparation module – which detects mentions[8] in provided input text. It performs information extraction (IE) using built-in methods (StanfordHybrid and StanfordManual). Additionally, the preparation module allows defining filters for detected mentions and thus defining what types of mentions should be detected (e.g. geographical places or events).
2. The disambiguation module – mentions found in the preparation module are resolved by identifying the entity that each name means. An entity is the unambiguous definition for a given mention supplemented with Wikipedia URL (Uniform Resource Locator). For example, for the mention "Seine", the disambiguation block can return two entities: "Seine, a major river in France" and "Seine (department), a former administrative subdivision of France". This process is called toponym resolution (TR).

AIDA is a framework for entity detection and disambiguation. Given a natural-language text, it maps mentions of ambiguous names onto canonical entities (e.g., individual people or places) registered in its own knowledge base called YAGO[9]. This database is supplied mainly with information from Wikipedia[10] (also from WordNet[11]

[8] Term "mention" used in this document comes from AIDA documentation and refers to places mentioned in text, and it has nothing in common with the meaning used in Twitter where mentions are started with the At sign '@' and are used for replying to, citing or indicating other posts.
[9] http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/
[10] http://en.wikipedia.org/wiki/Main_Page

and GeoNames[12]) and each provided entity can be supplemented with URL to Wikipedia article, which can be used for further processing. Six disambiguation methods (algorithms) available in AIDA[13] are presented in Table 1.

Method	Description
PriorOnlyDisambiguationSettings (PRIO)	Annotate each mentions with the most prominent entity
LocalDisambiguationSettings (LO)	Use the entity prominence and the key phrase-context similarity to disambiguate
FastLocalDisambiguationSettings (LO_FAST)	Same as above but sacrificing a bit of accuracy for roughly 5 times quicker disambiguation by dropping low weight key phrases
CocktailPartyDisambiguationSettings (COCTAIL)	Use a graph algorithm on the entity coherence graph to disambiguate.
FastCocktailPartyDisambiguationSettings (COCTAIL_FAST)	Same as above but sacrificing a bit of accuracy for roughly 5 times quicker disambiguation by dropping low weight key phrases
CocktailPartyKOREDisambiguationSettings (COCTAIL_KORE)	Use a graph algorithm on the entity coherence graph to disambiguate

Table 1 Six disambiguation methods (algorithms) available in AIDA

"Graph" algorithms are characterized by the fact that they use approach of building graph of relations between found mentions. "Local" algorithms on the other hand consider mentions without other mentions in the text. Each algorithm is defined by a set of configuration parameters, which can be used to tune the particular method. Entities returned by disambiguation block do not contain geographical coordinates but they contain the URL to article on Wikipedia. In turn, TAT2 downloads the article, parses it and extracts geographical coordinates which are usually available inside the article. Mentions returned by the preparation block can be also passed to the Google Geocoding web service by the publically available interface. The Google Geocoding can be considered another disambiguation method with one essential advantage: it immediately returns the correct coordinates. Coordinates got either by parsing Wikipedia page or from Google Geocode web service can next be verified using two filters:

- Timezone filter – which checks whether coordinates, especially longitude, belong to the time zone provided with tweet text,
- Area filter – which checks whether coordinates belong to predefined area.

In this way, resulting locations are narrowed to the most suitable ones. The final output from the TAT2 web service contains mentions found in the tweet text and associated entities with geographical coordinates. The TAT2 web service interface enables also applying configuration changes i.e.: preparation method, filters for preparation block, disambiguation method and predefined area coordinates. Additionally, for the reason of validation during the project showcase in Paris, France, an online translator has been incorporated. It has been used to translate tweets and alerts in both ways, English-French. The graphical user interface (GUI) of TAT2 was developed for the purpose of tests and to make it a stand-alone tool. It has been equipped with functionalities such as downloading tweets, displaying tweets on the map, translating messages and sending them with coordinates to the information distribution module. Finally, the tool can be integrated with systems which use only geo-tagged tweets, such as OsintLab, AIDR, Geofidia, as a provider of additional input data.

TEST SCENARIOS AND RESULTS

In order to verify correctness and effectiveness of TAT2 tool, three different test scenarios have been defined and performed. Tests aimed also at identifying the methods and parameters that provide the most accurate results. All of the tests were performed on various datasets, which are described together with the particular scenario. For verifying results only geo-tagged tweets were used and geolocation data was assumed as ground truth. Geographical coordinates returned by TAT2 were compared to the original ones and the distance was calculated. This distance is considered as a main metric for test cases. It need to be emphasized that since only 1-3% of tweets are geo-tagged, obtaining tweets for some special event and special location having both geo-tags and mentions of location in the text is a challenging task in itself. It has been noticed especially during the test scenarios 1. That is why for the scenarios 2 and 3 we have used data sets collected from different locations

[11] http://wordnet.princeton.edu/
[12] http://www.geonames.org/
[13] https://github.com/yago-naga/aida

and not during the particular crisis event. Generally, scenarios have been divided into two cases where:

- data sets are preprocessed with regards to information extraction,
- data sets are not preprocessed.

The preprocessed datasets contain only the properly separated tweets that have mentions of locations in the text. In such case TAT2 performs only the process of toponym resolution (TR). In the other case, before the process of toponym resolution, the operation of information extraction is performed (IE+TR). Test scenarios presented in this section are summarized in Table 2.

Sceanrio	Description
Scenario 1	Comprehensive analysis and comparison of main methods in both IE and TR cases. Analysis of combining methods to improve the performance (IE + TR)
Scenario 2	Verification how TAT2 works with external toponym resolution mechanism and comparing results to the standard AIDA resolution methods (TR and IE+TR)
Scenario 3	Verification whether information provided by external toponym resolution system can be used by TAT2 for non-English tweets how usage of external language translation tools improve effectiveness of mention detection and disambiguation (IE+TR)

Table 2 Comparison of effectiveness of algorithms

Scenario 1 – Overall comparison of all methods

The objective of this scenario is to estimate the overall performance of all 6 algorithms used in TAT2 (i.e. provided by AIDA) with default settings (algorithms: LO, LO_FAST, COCTAIL, PRIO, COCTAIL_FAST, COCTAIL_KORE). The test was performed on the dataset that has not been preprocessed. Therefore, this scenario was carried out by two operations. First, the information extraction process was performed and then each method executed process of toponym resolution to connect those locations with their coordinates.

Data set

Data set for this scenario has been retrieved using the website[14], where identifiers of tweets collected during hurricane Sandy were available. It consists of tweets posted on Twitter during Hurricane Sandy and collected from October 25, 2012, to November 4, 2012, using the keywords 'hurricane' and 'sandy' (and also variations like #hurricane and #sandy). The whole data set contains nearly 15 million tweets, for TAT2 purposes the first 290.000 tweets have been downloaded and filtered out regarding the geo-tagging. Finally, only1855 tweets were geo-tagged and it was possible to use them for testing.

Results

Each algorithm has been tested according to information extraction (IE) and toponym resolution (TR). Summary results of these tests are presented in Table 3. Out of the 1875 tweets, each algorithm found mentions related to locations indicated in percentage value according to whole data set as IE (column 2). From locations that were found, toponym resolution identified tweets located within a radius of 200 km (column 3). Percentage of tweets located within a radius of 200 km according to whole data set is indicated as TR-overall (column 4). Radius of 200 km includes all the tweets for which location have been found within radius 0-1 km (exact location), 1-10 km, 10-50 km and 50-200 km. Test results show that "local" algorithms PRIO, LO and LO_FAST have found significantly more locations, i.e. in about 33% of all tweets compared to "graph" algorithms with only about 14,5% of tweets. However, number of toponyms found within 200 km radius for detected locations is considerably greater for "graph' methods (28%) than for local methods (12%). Since the absolute number of locations found within a radius of 200 km was approximately similar in all methods (4%), it may be concluded that "local" algorithms found much more incorrect locations.

Algorithm	IE performance (IE detections/dataset size)	TR+IE performance (TR within 200 km radious from	TR-overall (TR within 200 km radious from ground truth

[14] http://www.zubiaga.org/datasets/hurricane-sandy-tweets/

		ground truth divided by number of IE detections)	divided by size of the dataset)
	(2)	(3)	(4)
PRIORONLY	33,12%	12,1%	4%
LOCAL	32,64%	12,7%	4,1%
FASTLOCAL	32,64%	12,9%	4,2%
COCKTAILPARTY	15,04%	28,3%	4,3%
FASTCOCKTAILPARTY	13,92%	28%	3,9%
COCKTAILPARTYKORE	12,96%	26,3%	3,4%

Table 3. Comparison of effectiveness of algorithms

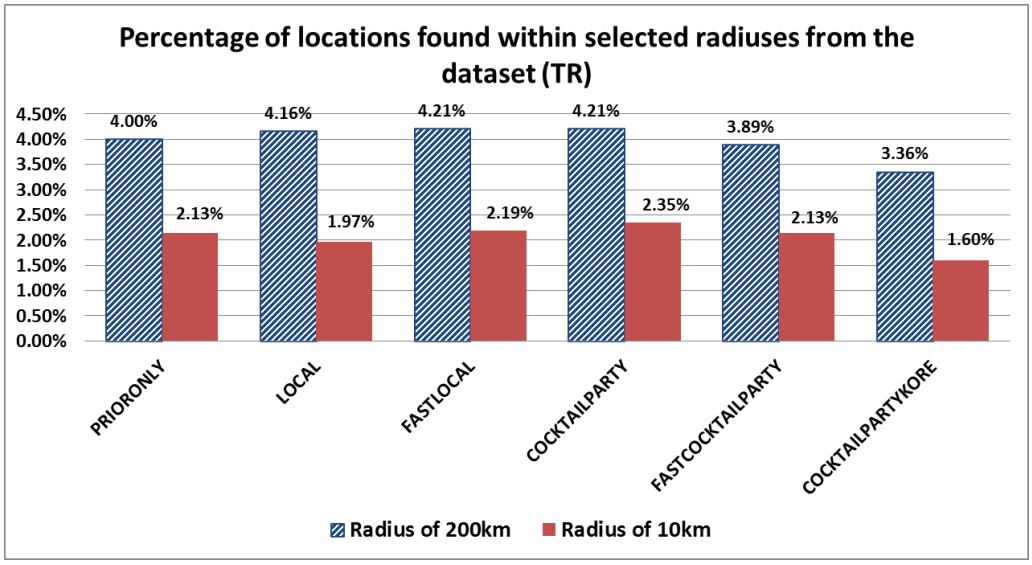

Figure 2 Results of toponym resolution process (TR-overall)

The chart presented in Figure 2 relates to the toponym resolution process (TR-overall) and shows that assuming that the locations found within 10 km can be classified as good results, the COCKTAIL algorithm seems to be the most reliable method. Hence, this scenario showed that "graph" algorithms are more reliable methods in terms of toponym resolution. As shown in Figure 2, the presented process leads to the results which provide localization information for up to 4% of tweets, i.e. about 4% in the radius of 200 km and about 2% in the radius of 10 km. The overall conclusion is that the information extraction process has a significant impact on the overall performance of TAT2. "Local" methods that have implemented poor information extraction feature, exhibit significantly worse results. Moreover, characteristics such as the quality of input data, the precision and level of details of information about location in the message have great influence on results. As the next step, an analysis of combination of the two considered as most effective "graph" algorithms COCKTAIL and COCTAIL_FAST with comparison to Google Geocoding – has been performed in scenarios 2 and 3.

Scenario 2 – Testing the effectiveness of the toponym resolution

The aim of this scenario was to verify how TAT2 works with external toponym resolution mechanism and comparing results to the standard AIDA resolution methods. Two methods have been tested: COCTAIL and Google Geocoding. The same dataset was processed three times:

1. First the method StanfordManual was used within preparation module which means that the tool (TAT2) will not try to find mentions itself, but instead only use mentions provided by external system (OsintLab), and then TAT2 using AIDA will use them for disambiguation.

2. Next the method StanfordHybrid was used for detecting mentions. It means that the tool uses mentions provided by external system OsintLab, but it also will try to find other mentions in the tweet text (e.g. not found by the method).
3. The last time information about mentions detected by external system will be removed from the dataset, and the method StanfordHybrid will be used for detecting mentions. This will verify whether TAT2 using AIDA detects the same set of mentions as the external system.

Mentions detected during all three cycles were processed by all available disambiguation methods, and the geo-location information will be used as a proper target value.

Data set

In this scenario the dataset used consisted of 2177 tweets written in English language, from around the world, not related to any particular crisis. The rationale for this approach was to obtain as much as geo-tagged tweets, which have also mentions of location in the text. All tweets have geo-location information and mentions already detected and identified by an external system (OsintLab). Data set has been collected during the period around two weeks between June and July 2014.

Results

Results of this test scenario have been divided into effectiveness of the toponym resolution and results of disambiguation process.

1. Effectiveness of the toponym resolution

For the 2177 tweets the external system found 2797 mentions in total. That means that statistically each tweet had ca. 1.28 mentions detected. All in all TAT2 has found (Figure 3):

- 2588 mentions (StanfordManual, external system),
- 3205 mentions (StanfordHybrid, external system),
- 2691 mentions (StanfordHybrid, external system mentions removed).

The most effective is using both external and internal toponynm resolution methods. TAT2 empovered with AIDA is able to use information about mentions provided from external system and additionally it finds even more mentions than it was primarily found by external system (with StanfordHybrid method). The StanfordManual method was not able to recognize all mentions provided by external system. It can happen when mentions are incorrectly defined (incorrect offset and length), or when particular mention has incorrect text – i.e. local name used instead of English name, hyphens or underscores, typo in the name. After removing mention information, the StanfordHybrid alghoritm was not able to recognize all removed mentions, but it was anyway more effective than StanfordManual alghoritm working only on pre-recognized mentions.

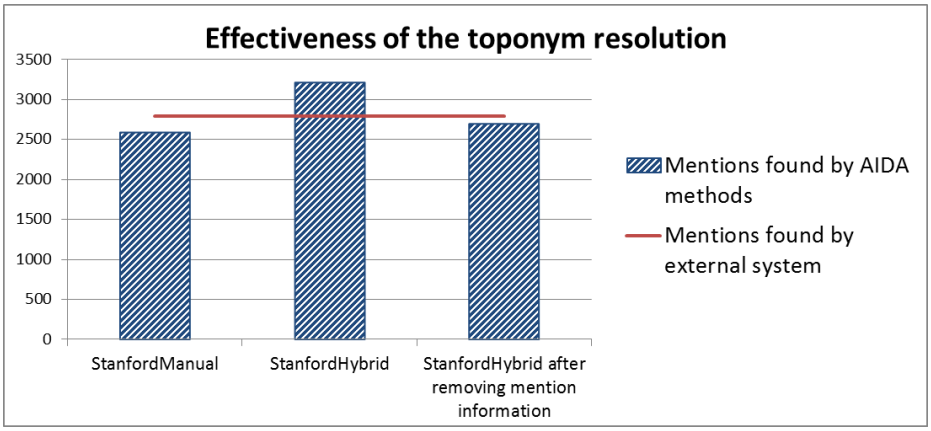

Figure 3. Effectiveness of the toponym resolution

2. Disambiguation results

Mentions produced by all three toponym resolution procedures described above, were processed by all available disambiguation alghoritms. In figures Figure 4 - Figure 6 results for two methods: COCTAIL and GoogleGeocoding for all three procedures are presented.

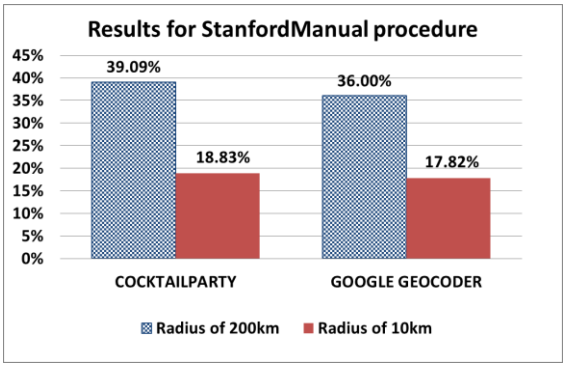

Figure 4. Results for StanfordManual procedure

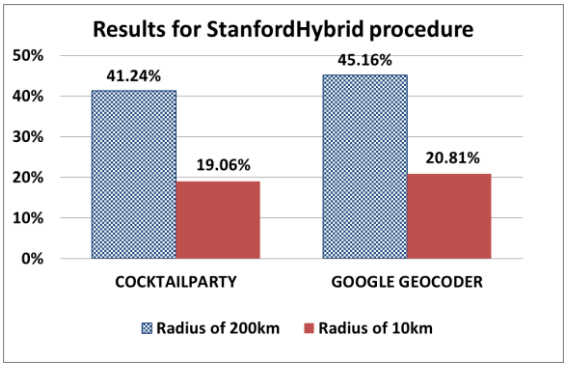

Figure 5. Results for StanfordHybrid procedure

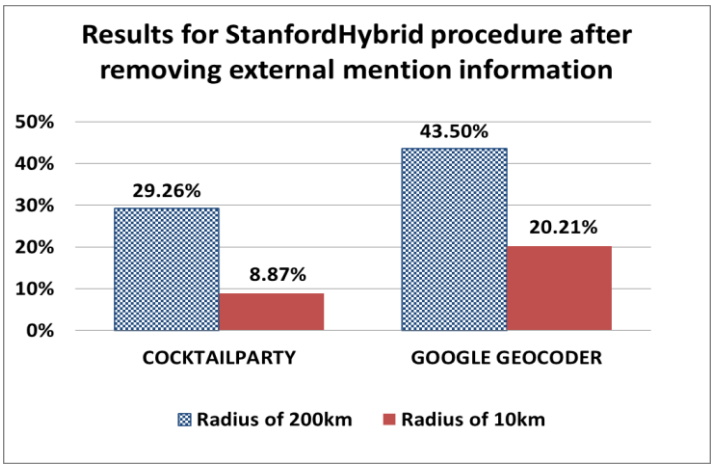

Figure 6. Results for StanfordHybrid procedure after removing external mention information

Similarly as for effectivess of the toponym resolution - from the disambiguation perspective the best results are achieved when using both external and internal method of toponym recognitions, and the best results are achieved by the StanfordHybrid alghoritm runing on the already pre-recognized mentions. When the externally recognized mentions have good quality (i.e are valid names of places, without distortion, abbreviations, slang names, and declinations) the best results are returned by the GoogleGeocoding method. Additionally it has been verified how TAT2 handles data set in which an external toponym resolution mechanism did not find any localization mentions. As it is shown in Figure 7 TAT2 was able to find geographical mentions in the dataset that was considered not having any geo-location information. Coordinates for ca. 14% of tweets used in the scenario have been found.

The latter example clearly shows that it is worth to use TAT2 for processing tweets that are claimed not having any geo-location information by external system.

Scenario 3 – Using external translation tools

Objectives of this scenario are as follows:

- To verify how TAT2 processes tweets written not in the English.
- To verify whether information provided by external toponym resolution system can be used by TAT2 for non-English tweets.
- To verify whether using external language translation tools for translating tweets to English can improve effectiveness of mention detection and disambiguation.

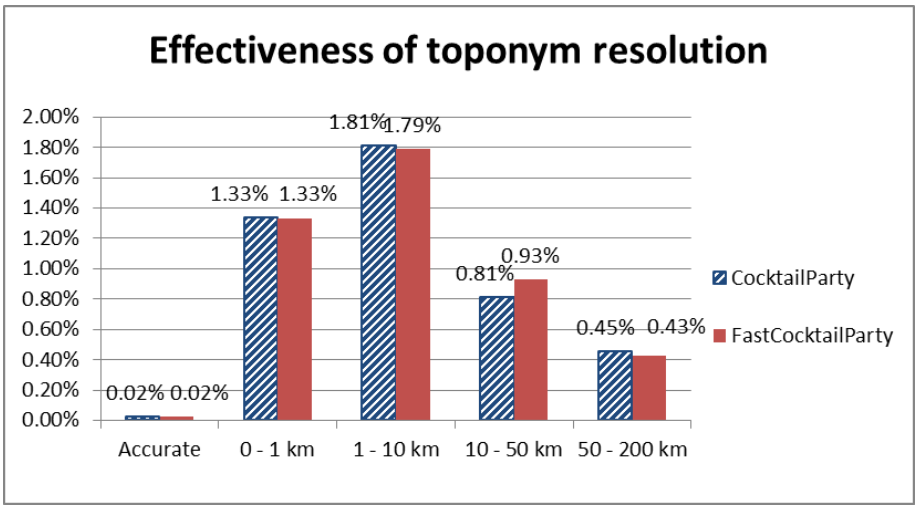

Figure 7. Effectiveness of the toponym resolution

The data set was processed three times:

1. First time the dataset will be processed as it is – French tweets with mention recognized by external system will be processed by TAT2 using AIDA with StanfordHybrid method. This will check whether TAT2 using AIDA is able to use mentions detected for non-English language.
2. For second time information about mentions detected by external system will be removed from the dataset, and tweets will be processed by StanfordHybrid method. This will check how the tool recognizes mentions for other languages.
3. For the third time the whole data set will be automatically translated into English language with using online translator. This will check how automatic translation impacts on the quality of mention detection and disambiguation made by TAT2 using AIDA.

Two methods have been tested: COCTAIL and Google Geocoding.

Data set

Data set used in this scenario contains tweets, which have geo-location information and mentions already detected and identified by an external system (OsintLab). The tweets have been collected during in the three events:

- tweets collected in the first three months of the French intervention in Mali (January- March 2013),
- tweets about the elections in India collected in the last 3 months of 2013,
- tweets collected from around the world using keywords related to fires, e.g. fire, smoke or flames.

The overall number of tweets used in this scenario is 275 tweets written in French language.

Results

Results of the presented test scenario for the Google Geocoding and COCTAIL disambiguation methods are presented in figures Figure 8 - Figure 10. The COCTAIL method was able to find more mentions that it was pre-recognized by external system regardless of the language used. The Google Geocoding method was able to find accurate positions only for good quality mentions. The number of recognized mentions and found coordinates for all cases is presented in Figure 8. The input data set consisted of 275 tweets and each had one mention recognized by external system. The best results are for original French tweets with mention information provided by external system. When mention information were removed, results decreased considerably for both methods Google Geocoding and COCTAIL. Translating tweets from French to English using automatic translator increased the effectiveness of the COCTAIL method by a certain amount but not to the original level. For non-English tweets best results are achieved for data sets for which mention information were provided by an external system. The Google Geocoding method work well also with mentions in different languages. Translating tweets to English involves removing information about mentions detected by external system. It is because mentions found by external system are identified by the position in the tweet text, and when the text is changed by translation, positions of mentions are also changed, and should be updated or removed.

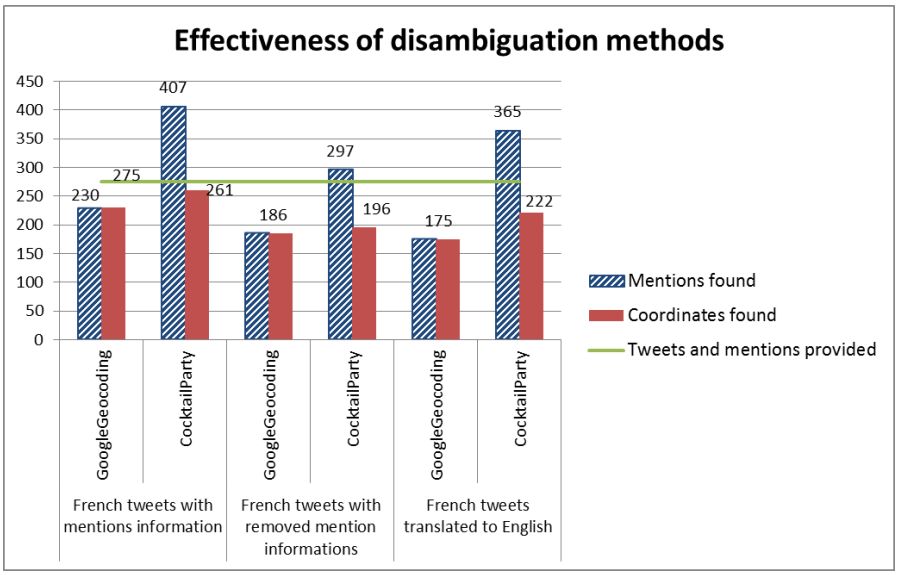

Figure 8. Effectiveness of disambiguation methods

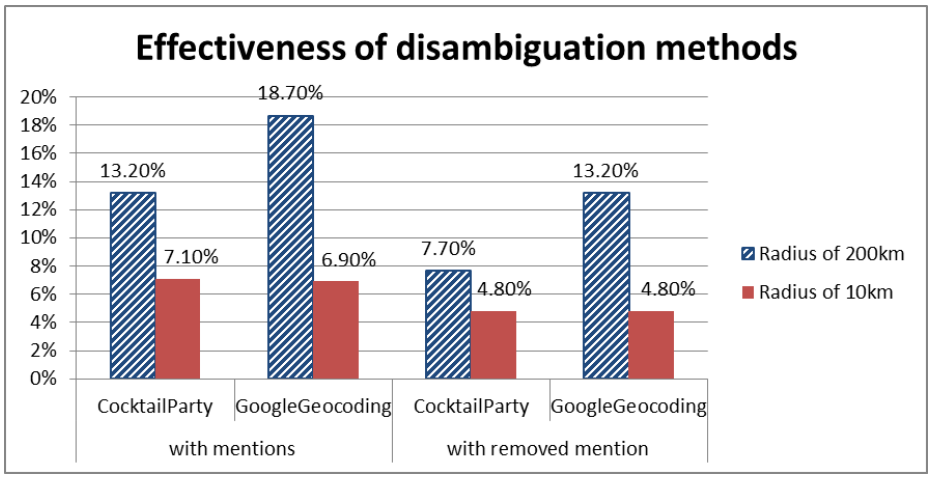

Figure 9. Effectiveness of disambiguation methods

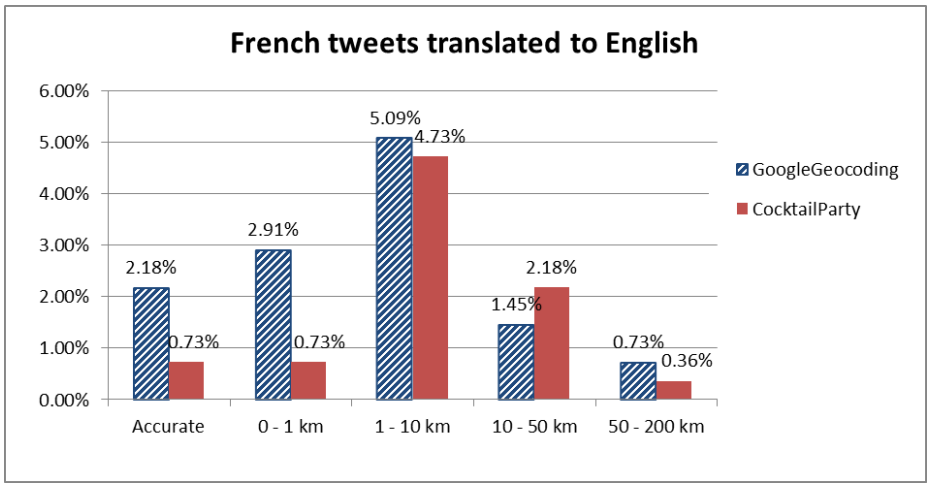

Figure 10. Effectiveness for French tweets translated to English

VALIDATION – DURING THE ISAR+ SHOWCASE

The TAT2 together with other iSAR+ tools have been validated during the 2nd project showcase on 24.09.2014

in Paris, France. The showcase included a real exercise of crisis operation to demonstrate, validate and use the functionalities of the of iSAR+ solutions. The TAT2 application was used during the showcase by OsintLab analysts to complement OsintLab by providing an assessment of the geo-location of the relevant messages and by translating those tweets which were written in any language other than French. The showcase results were discussed during dedicated workshops with its participants organized just after the showcase. Workshop aimed at collecting end-users feedback (on the usefulness of these tools and the level of trust towards them) in order to validate solutions and make recommendations for further developments. PPDRs and citizens participants had the opportunity to express their opinion on the use of iSAR+ tools orally and through questionnaires. PPDRs have been asked directly about the functionality provided by TAT2. Results showed that PPDRs wouldn't trust location indicated neither by unregistered citizens nor paid citizens (not professional). But they would trust the location if this would be confirmed by registered Citizens or paid professionals. Opinions about the level of trust towards the tool that automatically locates messages by reading message contents were divided. Some respondents would trust such information and some not. Interesting (as a side effect) is the fact that PPDRs asked about their trust towards automatic translations of messages, answered that they wouldn't trust such translation. However during the showcase, online translator was used, but it remained unnoticed due to good quality of the translation. This proves that online translator was sufficient to provide understandable and proper level of language.

CONCLUSIONS AND FUTURE WORK

The TAT2 tool presented in this paper, aims at enriching the available information about localization of given tweet based on the tweet contents. TAT2 relies on integration with existing tools such as AIDA or Google geocoding service and their algorithms. One of additional functionality is translation capability, applicable during all crisis events, where the affected communities are using more than one languages. In order to verify TAT2 functionalities and effectiveness the test scenarios have been performed as presented in this paper. Overall results are promising since TAT2 is able to find location of around 2-4% of tweets, while available source say that only 1-3% of tweets are geo-tagged. TAT2 is a complete tool, available to be integrated into other systems analyzing social media, with inputs which currently use only geo-tagged tweets, e.g. OsintLab, AIDR, Geofidia. As a result, using TAT2 would increase the number of processed tweets leading to more comprehensive analysis of social media content for PPDRs during crisis. The tool has been developed, tested and validated within the research project iSAR+. Because of close cooperation with end-users during the project, tool has been tested and validated by them during project showcase. In the future, additional tests will be performed, e.g. relaed to scalability of the tool. One of the challenge identfied in the course of the reseach was to find large enough data set which contains geo-tagged tweets and mentions in the tweet content. This effected in data set not related to the particular crisis event needed to be used for the purposes of scenario 2 and 3. Another crutial challenge in using such a tool in connection with social media messages is consideration of data privacy and ethical and leagl issues. The tool must conform to European Union ethics standards (legal and ethical constraints) and social media platform Terms of Service. Current TAT2 state of development is in line in all those aspects. In the future works - the possibility of including and fusing information from other sources (metadata, referenced information from other social media platforms etc.) will be subject to feasibility and ethical and legal analyses.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 312850 (iSAR+ Project).

REFERENCES

1. Banford, J. I., Bernhardt, J., Savelyev, A., Wong-Parodi, G., Carleton, A. M., Titley, D. W. and MacEachren, A. M. (2014) Tweeting and Tornadoes, *Proceedings of the 11th International ISCRAM Conference* – University Park, Pennsylvania, USA
2. Caragea, C., Squicciarini, A., Stehle, S., Neppalli, K. and Tapia, A. (2014) Mapping Moods: Geo-Mapped Sentiment Analysis During Hurricane Sandy, *Proceedings of the 11th International ISCRAM Conference* – University Park, Pennsylvania, USA
3. Cheng, Z., Caverlee, J. and Lee, K. (2010) You are where you tweet: a content-based approach to geo-locating twitter users, *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 759-768), ACM

4. Flizikowski A., Hołubowicz W., Stachowicz A., Hokkanen L., Kurki T., Päivinen N. and Delavallade T. (2014) Social Media in Crisis Management – the iSAR+ Project Survey, *Proceedings of the 11th International ISCRAM Conference* – University Park, Pennsylvania, USA
5. Geofidia http://geofeedia.com/
6. Herfort, B., de Albuquerque, J. P., Schelhorn, S. J. and Zipf, A. (2014). Does the spatiotemporal distribution of tweets match the spatiotemporal distribution of flood phenomena? A study about the River Elbe Flood in June 2013, *Proceedings of the 11th International ISCRAM Conference* – University Park, Pennsylvania, USA
7. Mabiala Y., De Maupeou S., Gouttas C., Huyot B. and Delavallade T. (2013) Visual Analytics Innovative Services for Cyber-Security, Symposium on Visual Analytics (NATO IST-116)
8. MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., Zhang X. and Blanford, J. (2011) Senseplace2: Geotwitter analytics support for situational awareness, In *Visual Analytics Science and Technology (VAST)*, 2011 IEEE Conference (pp. 181-190), IEEE
9. McClendon, S. and Robinson, A. C. (2012) Leveraging geospatially-oriented social media communications in disaster response, *Proceedings of the 9th International ISCRAM Conference* – Vancouver, Canada
10. Meier, P. (2013) AIDR: Artificial Intelligence for Disaster Response (Blog), October 1, http://irevolution.net/2013/10/01/aidr-artificial-intelligence-for-disaster-response/
11. Meier, P. (2014) Establishing Social Media Hashtag Standards for Disaster Response (Blog), November 5, 2014, http://irevolution.net/2014/11/05/social-media-hashtag-standards-disaster-response/
12. Palen L. (2008) Online Social Media in Crisis Events, In *Educause Quarterly*, Vol.31, Number 3 https://net.educause.edu/ir/library/pdf/EQM08313.pdf
13. Schulz, A., Hadjakos, A., Paulheim, H., Nachtwey, J. and Mühlhäuser, M. (2013) A Multi-Indicator Approach for Geolocalization of Tweets, *Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM*.
14. UN OCHA report (2014) The United Nations Office for the Coordination of Humanitarian Affairs report: Hashtag Standards for Emergencies, October 2014, https://app.box.com/s/yvobt4n9wptqa8sd0887
15. Vieweg, S., Hughes, A. L., Starbird, K. and Palen, L. (2010) Microblogging during two natural hazards events: what twitter may contribute to situational awareness. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1079-1088), ACM
16. Wilensky, H. (2014) Twitter as a Navigator for Stranded Commuters during the Great East Japan Earthquake, *Proceedings of the 11th International ISCRAM Conference* – University Park, Pennsylvania, USA
17. Yin, J., Lampert, A., Cameron, M., Robinson, B. and Power, R. (2012) Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, 27(6), 52-59.