Detecting Emergent Conflicts through Web Mining and Visualization

Fredrik Johansson, Joel Brynielsson, Pontus Hörling,
Michael Malm, Christian Mårtenson
Swedish Defence Research Agency (FOI)
Stockholm, Sweden
Email: firstname.lastname@foi.se

Staffan Truvé, Magnus Rosell
Recorded Future AB
Göteborg, Sweden
Email: lastname@recordedfuture.com

Abstract—**An ocean of data is available on the web. From this ocean of data, information can in theory be extracted and used by analysts for detecting emergent trends (trend spotting). However, to do this manually is a daunting and nearly impossible task. We describe a semi-automatic system in which data is automatically collected from selected sources, and to which linguistic analysis is applied to extract e.g., entities and events. After combining the extracted information with human intelligence reports, the results are visualized to the user of the system who can interact with it in order to obtain a better awareness of historic as well as emergent trends. A prototype of the proposed system has been implemented and some initial results are presented in the paper.**

Index Terms—**Information fusion; trend analysis; trend spotting; web harvesting; web mining**

I. Introduction

Open source intelligence (OSINT) has become a very important type of information for intelligence analysts all around the world. In a number of recent FOI (Swedish Defence Research Agency) research projects, the need for tools that can support analysts and decision makers with mining web pages, social media, etc. has become obvious.

The European Union has a number of battle groups, of which one is the Nordic Battle Group (in which Sweden and other countries are participating with officers and soldiers). Each of these battle groups consists of a battalion sized force, reinforced with combat support elements. At all times, two such battle groups are active and ready for deployments in conflicts where there is a critical need for humanitarian, peacekeeping, or peace enforcing operations. One of the recent conflicts in which the involvement of European Union battle groups has been discussed is the civil war in Libya, where the troops potentially can be used for securing the delivery of aid supplies. Despite the preparedness of the battle groups to act in different environments and circumstances, a lot of planning is required before an actual deployment. Therefore, it is important that escalating or upcoming conflicts can be discovered as early as possible, in order to allow for sufficient planning and training before military units are deployed [1]. Moreover, it is easier to influence global events in their earlier stages than later on. Our belief is that there are many signs of new conflicts or other types of "trends" before they are reported in traditional media and start to become common knowledge. We think that early signs of new conflict areas and trends such as suicide bombings, piracy, or systematic political, religious, and ethnic persecution in many cases are to be found in reports from non-governmental organizations (NGOs) such as Human Rights Watch[1], Institute for War & Peace Reporting[2], and International Crisis Group[3], working in this kind of areas. In addition to detecting new trends of man-made disasters and events such as those already mentioned, the same sources may also be used for obtaining a better preparedness for humanitarian operations due to natural disasters such as drought. Besides NGOs and news agencies, it is at least for developed countries also very likely that signs of emergent trends are to be found in social media of various kind, e.g., Twitter or Facebook.

In this paper, we present an idea of how to identify emergent trends by letting an intelligence analyst select a number of data sources (such as web pages and RSS feeds from NGOs), automatically harvest these sources and extract information such as events, places, dates, and actors from the unstructured harvested data, apply statistical analysis on the extracted information, and to present the analyzed information to the user in a graphical user interface which also allows the user to dig deeper into the underlying data and reports.

The rest of this paper is organized as follows. In Section II, we present an overview of the proposed tool for semi-automatic trend analysis. In Section III, it is described how a tool called Recorded Future can be used to automatically harvest data and extract useful information such as event and entity types by the use of linguistic analysis. Section IV presents the Ushahidi platform and how this can be used to visualize OSINT extracted from the web, as well as manually created human intelligence (HUMINT) reports. The suggested idea has been implemented into a prototype tool which is described in Section V, in which we also present some initial findings from experiments where we have extracted information regarding protest events from the web. Section VI provides a discussion together with some ideas for future work. Finally, a number of conclusions are presented in Section VII.

[1] http://www.hrw.org/
[2] http://www.iwpr.net/
[3] http://www.crisisgroup.org/

978-0-7695-4406-9/11 $26.00 © 2011 IEEE
DOI 10.1109/EISIC.2011.21

FOI-S--3801--SE

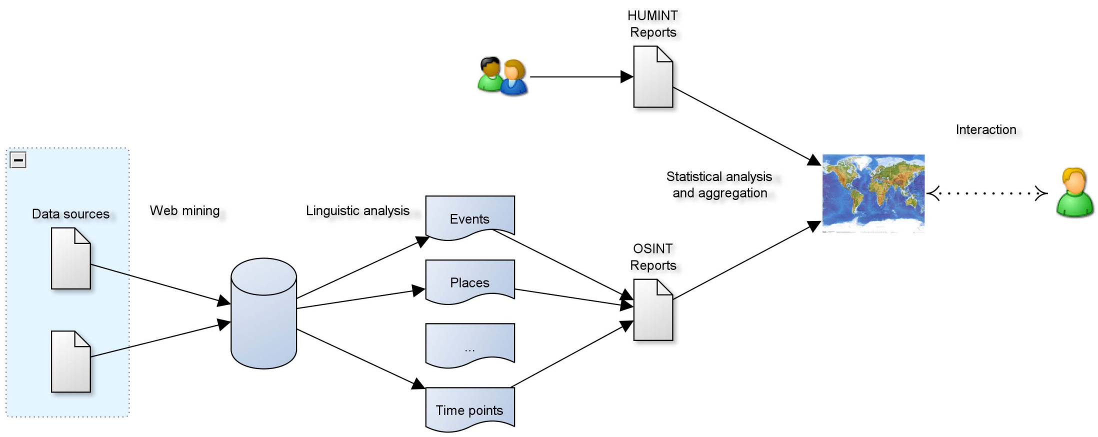

Figure 1. Overview of the proposed tool for semi-automatic trend analysis.

II. System overview of the proposed tool

We envision a system for detecting emergent trends in which an intelligence analyst (or some other kind of user) sets up a project by:

- selecting a set of data sources (e.g., a number of web sites and RSS feeds from NGOs, news agencies, or other kinds of organizations that may contain useful information concerning the development in potential conflict areas).
- determining the rate for how often data is to be updated.
- choosing one or several existing extractors for finding entities and events (or by constructing new ones)
- selecting a taxonomy for how to group the extracted event types.

Once a project has been set up and the analyst chooses to activate it, data is automatically harvested from the selected data sources. The content of the data sources is collected on a regular basis determined by the chosen update rate parameter. New data (or all data in the case of a new project where no previous collection has been made) is after harvesting fed to a linguistic analyzer, in which useful information such as event types, names, dates, and places is extracted. The extracted information is used to construct structured OSINT reports, to which the underlying unstructured source data is attached. In addition to these automatically created reports, the idea is that users can add manually created reports from other kinds of sources, e.g., human intelligence (HUMINT). All the constructed reports are then aggregated into groups, whereupon statistical analysis is applied. The clustered reports are finally visualized geographically on a map, on which the user can choose to zoom in and out, watch the development over time, and to dig down to the actual data content on which the visualized reports are based. An illustration of this suggested process is shown in Figure 1.

III. Harvesting the web and extracting information using Recorded Future

In recent years there has been a growing interest in automatic and semi-automatic harvesting and mining of data from web pages, social media, RSS feeds, etc. (see e.g., [2], [3], [4]). Web harvesting can according to Chen et al. [5] be defined as:

> "the process of gathering and organizing unstructured information from pages and data on the Web,"

while the closely related term web mining often is defined as [6], [7]:

> "the use of data mining techniques to automatically discover and extract information from web documents and services."

Many systems have been proposed for various web mining and web harvesting tasks, but here we will focus on the tool Recorded Future (see e.g., [8]), since it offers advanced linguistic analysis capabilities such as entity and event extraction using natural language processing (NLP). Another strength of Recorded Future is the way it handles time. Rather than using the publication date of a text, semantic information from sentences such as:

> *About two weeks ago, a hurricane destroyed five houses in the southern parts of Massachusetts.*

can be used to analyze when an event has taken place, or even *will* take place if the text refers to an event that is expected to take place in the future.

Overall, Recorded Future has a database-centric view, as illustrated in Figure 2. Around this database, five main components can be identified [8]:

- **Harvesting**: a process in which text documents are retrieved from various sources and stored in the database. The documents are stored for long-term if permitted by terms of use and IPR legislation, otherwise they are only stored temporarily for the needed analysis.
- **Linguistic analysis**: the process in which the retrieved texts are analyzed in order to extract entities, events, time and location, etc. In contrast to the other components, the linguistic analysis is language dependent.
- **Refinement**: additional information can be obtained in this process by synonym detection, ontology analysis, and sentiment analysis.

- **Data analysis**: application of statistical and AI-based models such as Hidden Markov Models (HMMs) and Artificial Neural Networks (ANNs) in order to generate predictions about the future and to detect anomalies in the data.
- **User experience**: a web interface for ordinary users to interact with, and an API for interfacing to other systems.

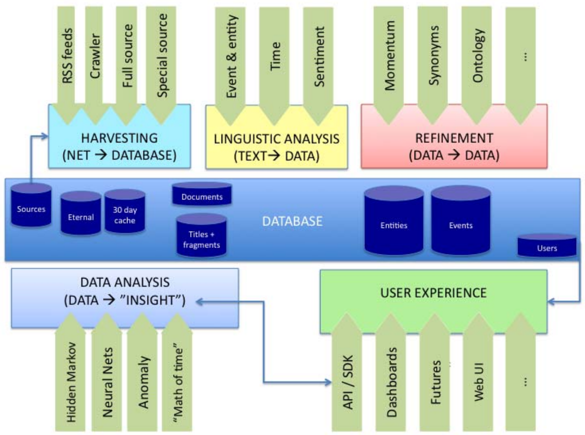

Figure 2. A database-centric view on Recorded Future (from [8]).

In this work, it is the two first components, i.e., the harvesting and the linguistic analysis, that are of largest interest.

IV. Constructing Reports and Visualizing the Results Using Ushahidi

Ushahidi is Swahili for "testimony" or "witness," and is also the name of an open source platform that has been used in several recent crises, such as Haiti, Libya, and Japan [9]. Originally the platform was launched during the post-election violence in Kenya in the beginning of 2008. Ushahidi makes it possible for local observers to submit eyewitness reports using their mobile phones or the Internet (e.g., by text messages or web forms) and increases the situational awareness of what the crisis situation looks like for both ordinary people and decision-makers by presenting the reported events in a graphical interface using Google Maps, Yahoo Maps, or OpenStreetMap. An example of the Ushahidi interface is shown in Figure 3, taken from the recent crisis in Libya.

Ushahidi's possibility to cluster the submitted eyewitness reports and present them geographically is very appealing and close to what we would like to accomplish in this project, except for that we would like to present the information extracted from harvested data sources rather than observations submitted by web forms or SMS. Due to this, we have decided to use Ushahidi as a basis for the proposed system.

In addition to the automatically created OSINT reports, we want the ability to combine those with manually created reports, concerning HUMINT or other kinds of intelligence that does not originate from harvested web data. For this kind of reports, we will use Ushahidi's possibility to send in reports by the use of web forms, as illustrated in Figure 4. Hence,

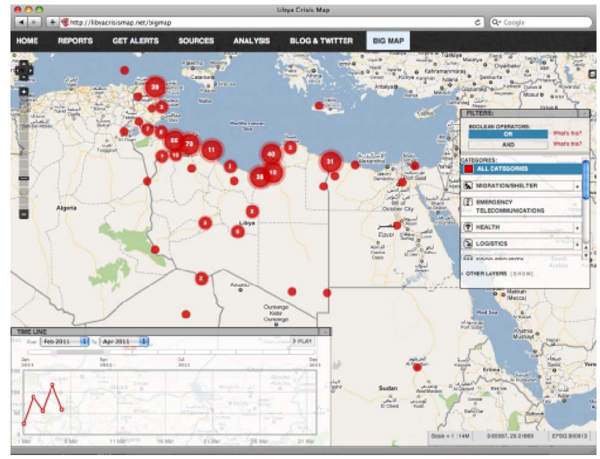

Figure 3. An example of the GUI in Ushahidi (from the Libya crisis).

Figure 4. An example of a HUMINT report created using Ushahidi's web form for submitting reports.

we envision an analyst-centric tool where human intelligence and information from web sensors are fused and visualized in order to obtain situational awareness [10] and facilitate sensemaking [11] with regard to trends concerning upcoming conflicts.

To put this idea into context, we envision a system design based on the mixed initiative paradigm [12], [13], i.e., a system where the analyst uses a number of different "tools" to both interact with and influence the system in order to continuously refine his/her understanding of both the situation itself as well as the reasons behind it. Here, we are motivated by the increase of publically available on-line news information in combination with the ease for first responders to contribute with own reports via, e.g., mobile devices. Hence, the analyst has got hold of many types of information fragments stemming from his own subjective ideas, HUMINT reports, news feeds, governmental/organizational information, classified intelligence databases, etc. A natural way for an organization to implement a trend spotting tool according to these ideas would then be to maintain its own Ushahidi server installation connected to databases of interest and a graphical user interface letting the analyst combine different kinds of information pieces that are temporally and geographically coded. Running one's own server makes it possible to combine public sources with own, possibly secret, information. We think primarily of three types of information: 1) structured, continuously updated, and relevant web information such as the Recorded Future database, 2) information gathered in real-time by observers, 3) classified intelligence products. When combined, a system comprising all these three aspects would 1) provide enough statistic and linguistic power to detect emerging trends, 2) provide the means to gather reports from, e.g., soldiers present at an international mission or first responders reporting directly from a crisis situation, 3) make it possible to let the trend analysis take a, possibly large, classified data set into account. In particular, events can be visualized both spatially and temporally over time which is helpful for understanding eventual trends. In practice, the actual design of a secret system would require the use of a unidirectional security gateway that lets the open information flow into the secret system but not the other way around.

V. The implemented tool

In order to demonstrate the usefulness of the proposed system, we have implemented a prototype tool which combines the strengths of Recorded Future and Ushahidi, as will be explained below. We have in Recored Future chosen a number of data sources, which e.g., consist of RSS feeds from the organization Human Rights Watch and web sites from some news agencies. A number of event extractors have been constructed which are used to extract events. Once the events have been extracted from the harvested data, the event type is together with location, date and time, etc. used to construct reports which are saved in a .csv (comma-separated values) file, following the format:

1) Id (e.g., 2022)
2) Title (e.g., Wrenching Photos Of The Inflation And Unemployment Protests Tearing Apart Tunisia)
3) Date and time (e.g., 2010-12-27 00:00:00)
4) Location (e.g., *Tunisia*)
5) Message (e.g., Frustration over high unemployment and rising food prices has erupted in deadly riots in Tunisia where clashes between youth protesters and security forces killed at least 23 people over the past week according to UPI.)
6) Event type (e.g., *Protest*)
7) Latitude (e.g., 35.418736)
8) Longitude (e.g., 9.987493)
9) Approved (e.g., *YES*)
10) Verified (e.g., *NO*)

We have set up an Apache web server on which we have created a MySQL database and installed the Ushahidi platform after downloading the source code from https://github.com/ushahidi.

Currently we are importing the OSINT reports created with Recorded Future by logging into the admin-interface of Ushahidi and manually uploading the generated .csv-files using its web interface. In the future this will most likely be replaced by a process in which the Ushahidi API will be used to automatically load OSINT reports generated with Recorded Future, but the current simplified solution is enough to demonstrate the idea. For HUMINT reports, web forms are used to create and submit the reports into Ushahidi, as illustrated in Figure 4. The reports are after that clustered based on their geographical position (latitude and longitude) and shown on a map, as shown in Figure 5. The event type(s) to which a report belongs can be used for filtering which reports should be displayed on the map, either by selecting individual event types (such as *Protest* or *Ethnic/religious violence*), or by selecting a category in the used event type taxonomy. The user of the tool can choose a start date and

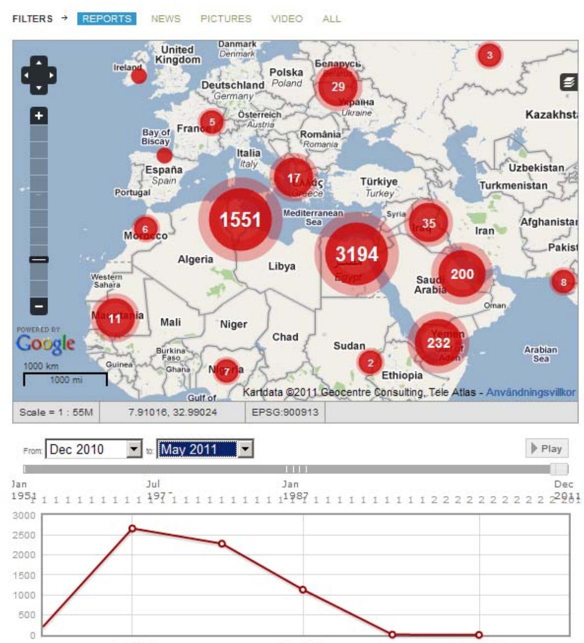

Figure 5. Clustered reports of event type Protest.

an end date (as shown in the middle of Figure 5), and start a visualization of how the number of reports with the selected event or category type evolves over time geographically. In this way the user can become aware of the development of trends in retrospect, and hopefully also as new trends emerge.

A. Findings

We have from our selected data sources collected data during the first quarter of 2011. The harvested data also include mentions of some historic as well as future events, since Recorded Future uses the extracted and estimated time for the described event rather than the publication time. However, the major portion of the data concern events during Q1, 2011. From this data set containing texts on various themes, 7251 *Protest* events have been extracted, out of which 5861 are believed to concern the first three months of 2011 according to the linguistic analysis performed by Recorded Future. The protest events have been identified using a trained protest extractor, which was constructed using the Recorded Future event extractor framework.

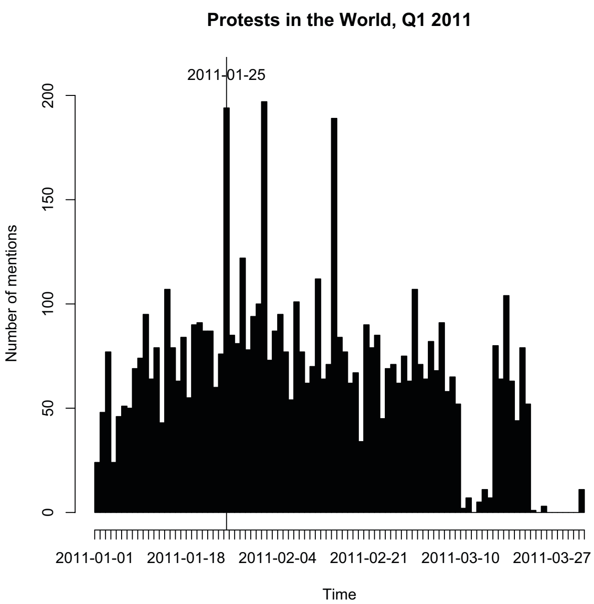

Figure 6. Number of *Protest* events in the world.

Once the *Protest* events have been extracted, they have been used for constructing OSINT reports on the format described earlier (it should be noted that many of these reports in a sense are "duplicates," since they contain the same information but come from various sources). In Figure 6 it is shown how the number of mentions of protests taken place anywhere in the world has varied over the first quarter of 2011 (based on the dataset described above). As seen in Figure 5, most of the *Protest* events in the dataset originate from countries such as Tunisia and Egypt where the so called Arabian spring has taken place. Therefore, we also made histograms of the number of extracted protest events in Tunisia and Egypt, see Figures 7 and 8. To find the events in Tunisia and Egypt we used a list of major cities in both countries to group events together. Although this simple approach misses some of the events it still gives a good picture of the trend. From the histograms, it is quite obvious how the Tunisia protests precede those in Egypt.

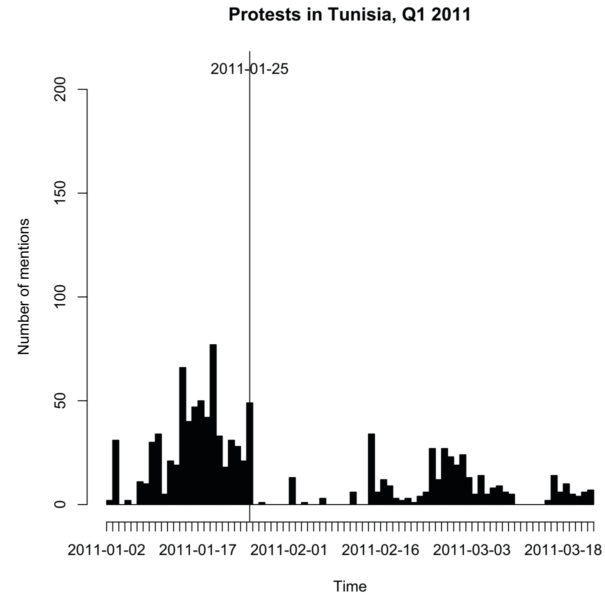

Figure 7. Number of *Protest* events in Tunisia.

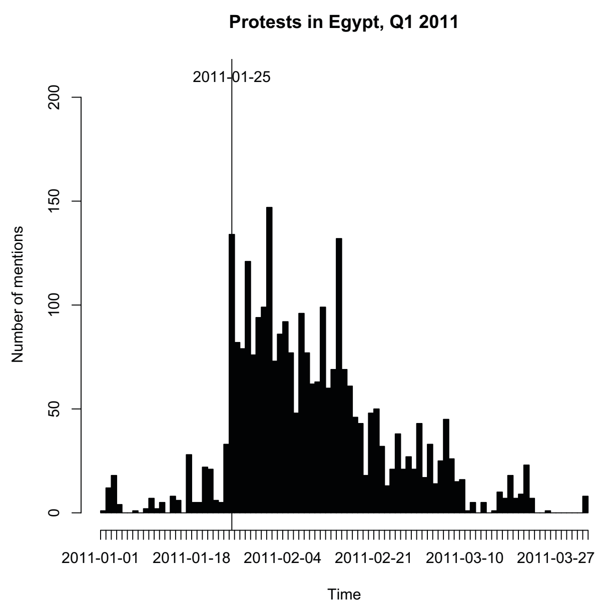

Figure 8. Number of *Protest* events in Egypt.

Since we have not collected data long enough to see the trend development over a long time horizon, we currently have to focus on trend development on a short basis for the harvested data. It is rather troublesome to illustrate the interactive aspects of the tool in paper format, but Figures 9 and 10 give an idea of how the number of protests have developed in Northern Africa from January 2011 to February 2011 (notice that the numbers in the figures indicate the total number of *Protest* events, i.e., the number of reported events for Tunisia was $990 - 721 = 169$ in February). By watching the development in the region, it would in February 2011

Figure 9. Number of *Protest* events until the end of January 2011.

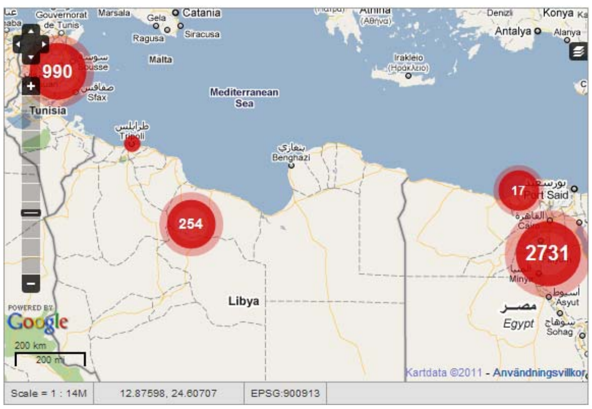

Figure 10. Number of *Protest* events until the end of February 2011.

have been possible for a user of the tool to see a sudden rise in the number of protests in Egypt, but also that there were indications that the Arabian Spring was about to spread to Libya. In the same manner, it can be seen that the situation in Tunisia had started to calm down at that time. However, it might also be the case that the decline in media attention in Tunisia is due to that all attention was brought over to Egypt on January 25 when the 2011 Egyptian revolution began. A preliminary hypothesis that we have is that a rise of media attention covering a specific area tends to dominate other attention covering the same domain area, e.g., that the area "protests in north Africa," came to be dominated by the 2011 Egypt revolution although there were still frictions in Tunisia that would have been reported on if the Egypt crisis would not have taken place. This trend can be seen in Figures 7 and 8 where the media reports covering Tunisia suddenly drops to almost zero on January 25 when the Egyptian revolution began ("the day of rage"). Also, it is interesting to note that the number of protest events in Egypt almost dominates the amount of protest reports in the whole world at the time of the onset of the revolution, which might provide empirical evidence for a second hypothesis with regard to the total amount of news information in the media. That is, if a protest event is of significant interest, it might be the case that this protest event totally dominates the media coverage for some time.

VI. Discussion and future work

A related idea to discover emerging trends through harvesting and analyzing data from the web is to analyze what kind of web searches that are made and how this varies over time. A tool that can be used for such trend analysis is Google Trends [14], in which the number of searches on specific terms can be compared relative to the total number of Google searches done over time. In a number of recent papers it has been suggested that the use of Google Trends can be useful to detect present [15], [16] as well as future trends [17], [18] in diverse areas. However, there is also recent research suggesting that Google Trends are not always a good predictor, e.g., for some congressional elections in the United States [19], [20]. Whether web searches alone are suitable for trend spotting or not is probably very context dependent. For the situation in which we would like to identify potential upcoming conflict areas, we do not think that the frequency of web searches are enough, since most people will not make searches on e.g., "Aden" and "piracy" before the reasons for making such searches are already common knowledge. However, it is our hypothesis that an increasing number of texts about such things will be published on web pages from NGOs (or other types of similar organizations) working in the areas before most of the rest of the world learns about it, making it possible to do trend spotting based on such material.

Although we think that many trends can be identified through web mining and text analysis, this does not mean that specific events can be predicted. If a trend of suicide bombings is started in some region, this can be discovered by harvesting the news media, but we will not be able to predict individual suicide bombings using this method since it is very unlikely that people will make postings of this on the web before the actual attack. Furthermore, it should be noted that we currently only deal with texts in English, since the linguistic analysis depends on natural language processing. Support for analyzing Swedish as well as other languages can therefore be of interest for the future.

In the work presented in this paper we have used Ushahidi as a tool for mapping the reports geographically and visualizing them to the user. There are of course other options as well. One such option is to create .kml (keyhole markup language) files instead of .csv files, and to display the content of these files using Google Earth. An example of this is shown in Figure 11, in which we have plotted the extracted *Protest* events using Google Earth. Our impressions so far is that it as a user is more convenient to work with Ushahidi than Google Earth for the tasks of interest for this paper. As an example, the clustering in Ushahidi allows for a gradual zooming in on interesting regions, while reports from one country are either clustered together into one cluster or "exploded" into the individual reports in Google Earth. Much of this can be

Figure 11. *Protest* events plotted using Google Earth.

extended by making use of the Google Earth API, but Ushahidi is currently better suited for our needs out-of-the-box.

To extract locations, dates, names, etc. from unstructured text is a challenging active research problem often referred to as entity extraction or named entity recognition. Many systems for entity extraction have been developed, such as ANNIE (part of the GATE distribution [21]) and TREX [22]. Most approaches use either handcrafted rule-based algorithms or statistical machine learning models. Although much research has been devoted to entity extraction, it is to the best of our knowledge more uncommon to also use the extracted entities for extraction of events as done in the Recorded Future tool. An exception is the event data extracted in the Defense Advanced Research Projects Agency (DARPA) funded ICEWS project [1] and the previous National Science Foundation (NSF) funded KEDS project [23], [24]. However, their focus seems to be solely on political events, while our approach is more general and can take more types of intelligence into account.

A. Future work

Currently, the Verified attribute is automatically set to *NO* for all OSINT reports. The analyst can manually change the value of the attribute to *YES* in the Ushahidi interface for reports that in some way have been confirmed, but due to the amount of automatically created OSINT reports, this can be a labor intensive process. For the future it would be interesting to rank the source reliability and report credibility, and to use these measures in order to try to automatically verify some reports (e.g., reports that come from trusted news sources). One potential way of making such rankings is to use link analysis techniques such as HITS [25] or PageRank [26]. Such techniques can also be used to automatically detect data sources that may be useful for the trend analysis at hand.

As has been described in Section V, the way users of the tool can detect trends is by looking for how the number of reports with a certain characteristic (event type or category) changes over time. A possible continuation of this thought is to let the computer detect this kind of changes automatically, and offer a “relative” mode in which the user is shown places with a large positive derivative in the number of reports with a certain characteristic, in addition to the “absolute” mode which is offered in the tool already today. Since the number of reports from different sources varies over time, it would also be of interest to use the *fraction* of events of a specific type (cf. the approach used in [27]). Another plan for future work is to add support for analysis of more languages, as has been indicated earlier.

As has been argued in this paper, the proposed system is expected to (at least to some extent) be able to forecast escalating conflicts and “hot spots” of various types. But how well will it perform? One potential check could be to let it harvest old archived, since then unchanged, flows of information up to some specified date and from this try to forecast what was then the future. Today we know what actually happened then, and this can be compared with these forecasts. However, the concept “know what actually happened” is rather subjective, and some kind of objectively coded information on the relevant events, their type, actors, start dates, etc. is needed. One way could be to use the datasets maintained in The Uppsala Conflict Data Program (UCDP) at Uppsala University (see [28]) for future, more systematic, benchmarking. These datasets contain, coded with a given precision, the start and end dates for armed conflicts of various kinds where at least one side is a government of a state, and where the conflict results in at least 25 battle-related deaths during one or more calendar years [29].

VII. Conclusions

We have in this paper proposed a semi-automatic system for trend spotting to be used by e.g., intelligence analysts when trying to figure out where or what the next conflict will be that may need military intervention of some sort, be it humanitarian, peacekeeping or peace enforcing. An overview of the proposed system has been given, in which automated web harvesting is used to collect data from open sources such as web pages and RSS feeds. Once data have been harvested, next step is to extract entities and events from the collected texts using natural language processing, in order to use them for automatic report construction. By combining the extracted OSINT with manually constructed HUMINT reports, a good foundation for trend analysis is offered. The reports are clustered based on their geographical position (latitude and longitude), whereupon they are visualized on a map, where the analyst can zoom in and out, look at the development over time, and access the content of the individual OSINT and HUMINT reports when needed.

A prototype tool has been developed to demonstrate the suggested concept. This tool relies on Recorded Future for the web harvesting and the extraction of e.g., location, dates, and events, while the creation of HUMINT reports and the visualization is accomplished using the Ushahidi platform. An example of how the tool can be used to find information concerning protest events such as demonstrations has been given, in which we e.g., have analyzed how the protests have spread during the so called Arabian spring.

Acknowledgements

This work was supported by the R&D programme of the Swedish Armed Forces and by Recorded Future AB, and is also a supporting work for the Alert4All research project[4], which is funded under the European Union Seventh Framework Programme under contract no 261732.

References

[1] S. P. O'Brien, "Crisis early warning and decision support: Contemporary approaches and thoughts on future research," *International Studies Review*, vol. 12, pp. 87–104, 2010.
[2] M. A. Russell, *Mining the Social Web*. O'Reilly, 2011.
[3] N. Memon, D. Hicks, and H. Larsen, "Harvesting terrorists information from web," in *Proceedings of the 11th International Conference on Information Visualization*, 2007, pp. 664–671.
[4] M. L. Nelson, J. A. Smith, and I. G. del Campo, "Efficient, automatic web resource harvesting," in *Proceedings of the 8th annual ACM international workshop on Web information and data management*. New York, NY, USA: ACM, 2006, pp. 43–50. [Online]. Available: http://doi.acm.org/10.1145/1183550.1183560
[5] H. Chen, J. Qin, E. Reid, Y. Zhou, and M. Sageman, *Terrorism Informatics – Knowledge Management and Data Mining for Homeland Security*. Springer, 2008, ch. Case Study of Jihad on the Web, pp. 221–235.
[6] R. Kosala and H. Blockeel, "Web mining research: a survey," *ACM SIGKDD*, vol. 2, pp. 1–15, 2000.
[7] O. Etzioni, "The world wide web: quagmire or gold mine?" *Communications of the ACM*, vol. 39, pp. 65–68, 1996.
[8] Recorded Future, "Recorded future – a white paper on temporal analytics," 2010. [Online]. Available: http://blog.recordedfuture.com/tag/white-paper/
[9] Ushahidi. [Online]. Available: http://www.ushahidi.com/
[10] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, no. 1, pp. 32–64, Mar. 1995.
[11] K. E. Weick, K. M. Sutcliffe, and D. Obstfeld, "Organizing and the process of sensemaking," *Organization Science*, vol. 16, no. 4, pp. 409–421, Jul.–Aug. 2005.
[12] M. A. Hearst, J. F. Allen, C. I. Guinn, and E. Horvitz, "Trends & controversies: Mixed-initiative interaction," *IEEE Intelligent Systems*, vol. 14, no. 5, pp. 14–23, Sep./Oct. 1999.
[13] G. Tecuci, M. Boicu, C. Ayers, and D. Cammons, "Personal cognitive assistants for military intelligence analysis: Mixed-initiative learning, tutoring, and problem solving," in *Proceedings of the First International Conference on Intelligence Analysis*, McLean, Virginia, May 2005.
[14] Google. [Online]. Available: http://www.google.com/trends
[15] H. Choi and H. Varian, "Predicting the present with Google trends," Google Inc., Tech. Rep., 2009.
[16] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, pp. 1012–1015, 2009.
[17] S. Asur and B. A. Huberman, "Predicting the future with social media," HP Laboratories, Tech. Rep. HPL-2010-53, 2010.
[18] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with Twitter: What 140 characters reveal about political sentiment," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, 2010.
[19] C. Lui, P. T. Metaxas, and E. Mustafaraj, "On the predictability of the U.S. elections through search volume activity," in *Proceedings of the IADIS International e-Society 2011*, 2011.
[20] D. Gayo-Avello, P. T. Metaxas, and E. Mustafaraj, "Limits of electoral predictions using Twitter," in *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, 2011.
[21] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "Gate: A framework and graphical development environment for robust nlp tools and applications," in *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*, 2002.
[22] J. Iria, "T-rex: A flexible relation extraction framework," in *In Proceedings of the 8th Annual Colloquium for the UK Special Interest Group for Computational Linguistics (CLUK'05*, 2005.
[23] P. A. Schrodt, "Twenty years of the kansas event data system project," *The Political Methodologist*, vol. 14, pp. 2–8, 2006.
[24] P. A. Schrodt and D. J. Gerner, "An event data analysis of third-party mediation," *Journal of Conflict Resolution*, vol. 48, pp. 310–330, 2004.
[25] J. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the Association for Computing Machinery*, vol. 46, no. 5, pp. 604–632, 1999.
[26] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," Stanford University, Tech. Rep., 1998.
[27] R. Colbaugh and K. Glass, "Early warning analysis for social diffusion events," in *Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI 2010)*, 2010.
[28] Uppsala Conflict Data Program, "UCDP database." [Online]. Available: http://www.pcr.uu.se/research/ucdp/database/
[29] N. P. Gleditsch, P. Wallensteen, M. Eriksson, and M. Sollenberg, "Armed conflict 1946-2001: A new dataset," *Journal of Peace Research*, vol. 39, no. 5, pp. 615–637, 2002.

[4]http://www.alert4all.eu/