Making Sense of the Arab Revolution and Occupy: Visual Analytics to Understand Events

Thomas Ploeger[1], Bibiana Armenta[2], Lora Aroyo[1,⋆], Frank de Bakker[2], and Iina Hellsten[2]

[1] Computer Science Department VU University Amsterdam
[2] Organization Sciences Department VU University Amsterdam

Abstract. Knowledge on the Web comes in ever larger amounts and in a wider variety of structure and semantics that ever before. In order to exploit this knowledge in different applications, many researchers investigate techniques for making sense of Web data. Objects that the techniques try to identify and extract are, for example, people, organizations, and locations. Many applications though observe how events play an increasingly more important role. Capturing and extracting events for sense making analysis is what this research is aiming for, and in this paper we present the first results and contributions from our research. We consider how events get extracted, how they get conceptualized, and how visual analytics helps to make sense of the represented events. All of this is illustrated in a representative example where driven by questions from social scientists we apply our pipeline to the domain of activism, e.g. Occupy, Arab Revolution.

1 Introduction

Events play an increasingly important role in representing and organizing knowledge on the Web. We observe how more and more applications are centered around events, specially in the Social Web. That is in line with the important role events play in all our lives: social networks and applications for our personal communication include events as central elements. Events are not only important for ourselves for organizing personal information, but the way we use events is also of interest to third parties, such as commercial stakeholders (i.e. event organizers and providers) or social scientists (to model and explain social phenomena). A first step in the process of dealing with events is their representation, e.g. in terms of formats that allow further processing, application and sense making. This is evidenced by recent research projects on the modeling of events (cf. SEM [8], LODE [6] and EO [5]) and projects such as Agora [1] and Poseidon [2] that capture events from unstructured and semi-structured texts, respectively. Capturing and modeling events is the first step towards answering domain questions and sense making.

There are a number of reasons why capturing events from unstructured or even semi-structured text is a no trivial task. Primarily, this is because of the

⋆ Also affiliated with IBM Watson Research Center, NY.

inherent limitations of current natural language processing technology. Additionally, the source texts that are relevant for event information are often scattered and may present information that is incorrect, out of context, or biased. For a complete overview of a certain event, all the different perspectives on the given event would have to be found and captured.

Many of the aforementioned research initiatives concentrate on different aspects of these challenging problems. In this research we focus on the contribution of visual analytics. This research aims to find out if visualizing events based on their properties (e.g. location, type, involved actors, timestamps) would help overcome the aforementioned problems for making sense of events. In this way, we hope to facilitate better understanding of events and their properties, by both social scientists and the general public.

To validate our work and demonstrate the concrete contributions, we conduct our research within the social sciences domain of activist organizations. We identified this to be an interesting use case as activists have always had an impact on the present and have a significant role in shaping the future: *In 2011 activists occupied the West and revolutionized the Middle East.* From a social science point of view, if we look at this use case, we see the importance of the events that are involved - for both individual people to see what is happening in their own local environment, or for scientists to tell and explain what triggered and caused what effect.

It is important to differentiate this event-oriented research from research into issues (cf. [4]) and sentiments (cf. [3]). An event is more clearly delineated in terms of spatial aspects than an issue – an issue is more vague, it doesn't require a spatial location where it takes place. Sentiment analysis, in turn, does not capture one particular event, but rather reactions to that event.

At the same time, as computer scientists we can also be impressed by the challenges to identify and extract these events from the large amounts of textual (user-generated) information available, e.g. in newspapers, personal or organizational blogs, social networks, and other forms of social media. Next to the obvious technical challenges, we can also easily identify that the way in which people, individual citizens or social scientists, can visually overview and subsequently interpret the massive amounts of event information is limited. Not only is the capability to visually overview and interpret limited compared to the size and volume. Also, the fact that events are often perceived from different angles makes it even more difficult to account for the different perspectives. For visual analytics the support of different perspectives provides another interesting challenge. Of course, also the ambition to warrant objective (unbiased) presentation of events plays an important role in social science as well as in our daily lives.

Thus, in this research, the overall goal is to explore the social sciences domain of activist events and their properties and do so with suitable event visualization techniques. This implies that we are contributing with relevant modeling and analysis techniques for event knowledge. The rest of this paper is structured as follows. In Section 2, we describe the activist use case in more detail. In Section 3,

we present a pipeline for extracting, modeling and visualizing events. In Section 4, we present our concluding remarks and expectations for future work.

2 Activist Events Use Case

Our research focuses on activist events. As stated in the introduction, activists have an influence on the present and play a role in shaping the future. Because even a single activist event might have consequences, we think they are worthy of study, by both the general public and social scientists. In this paper we will use events related to the Arab Spring and the Occupy Movement as examples because of their recency and social relevance.

This can be seen in many real-world examples of activist events. One such event was a particular confrontation between police and 'occupy' protesters in New York, where a police officer sprayed 4 protesters with pepper spray. A video recording of the event was uploaded to YouTube [3] and several news outlets picked up the story as it went viral[4,5]. Protesters argued that the use of pepper spray was uncalled for. Initially, the police department defended the officer, saying that the use of pepper spray was appropriate. The officer in question stated that the event was taken out of context. Later, a more detailed investigation concluded that the officer was at fault and he was reprimanded.

In order to make fair and unbiased judgments during such an investigation, it is important to represent different perspectives on an event and to take into account the larger context. Another example event that demonstrates this is the self-immolation of Mohamed Bouazizi, a Tunisian street vendor who set himself on fire to protest after officials confiscated his wares. This event is seen by many as the 'spark' that ignited the Tunisian Revolution and the Arab Spring. Many people see Bouazizi as a martyr, standing up to a dictatorial regime. Nevertheless, whether Bouazizi's exact motivations were personal, political or economic is the subject of some debate[6]. Placing this event in context could help investigators make sense of Bouazizi's motivations. How was he treated by officials when they confiscated his wares? How did officials respond to his complaints? Are there any earlier encounters between him and officials that might have played a role in the process? Detailed information about the event in question and the relations to earlier events are important when investigating Bouazizi's motivation.

As stated in the introduction, while the relevance and need are obvious, from a computer science perspective, it is challenging to identify, extract and aggregate mentions of these events from large amounts of textual information, such as newspapers, blogs, social networks and other forms of social media. In the next section we describe our pipeline for doing so.

[3] http://www.youtube.com/watch?v=TZ05rWx1pig
[4] http://online.wsj.com/article/SB10000872396390443866404577565341948999820.html
[5] http://cityroom.blogs.nytimes.com/2011/09/28/police-department-to-examine-pepper-spray-incident/
[6] http://www.frumforum.com/how-an-entrepreneur-sparked-the-arab-spring/

3 Main Process & Pipeline

To be able to explore the social sciences domain of activist events and their properties, as illustrated in Section 2, we turn to suitable event modeling and visualization techniques. Before we can consider the modeling, we must overcome the challenge of building a corpus of source material (such as newspapers, blogs, social networks and other forms of social media) from which to extract events. The process of constructing a corpus and extracting events is described in Section 3.1. It involves using natural language processing technology to extract enough information from unstructured text to 'build' event descriptions from. In Section 3.2 we describe the type of information we are looking for by describing our event model. This is a necessary step in order to be able to prepare for our main contribution, which is in visual analytics (Section 3.3), where we report on exploratory investigations into available instruments for visual event analysis.

3.1 Extracting Activist Events

In our use case, the event extraction process starts with creating a corpus from a (Web) source of news articles. Currently, we gather articles from The Guardian[7] via a simple search using keywords related to activism. This is a representative example, also since the extraction itself is not the main focus here in this paper but a necessary prior step in the process, as we have argued before. In future work, multiple sources of news articles will be used to strengthen validity.

Using keyword extraction and concept tagging, we attempt to determine the type of event that is described in the article. Actors, locations and timestamps of the event are identified using a named entity recognizer. Relation extraction is used to determine the type of involvement an actor had in an event. The details of this process are beyond the scope of this paper and will be described in a forthcoming paper.

Thus, for each news article in the corpus, we obtain several properties of the event described in the article: The type of the event, the actors involved in the event, the locations at which the event has taken place, and the time at which it has taken place. Additionally, we attempt to identify the type of involvement or 'role' of an actor in an event. We represent this event information using the Simple Event Model (SEM) as described in the next section. Additionally, we try to identify the authors of event information to represent different perspectives or viewpoints on the same event.

3.2 Conceptualizing the Activist Domain

When modeling events, there are a number of interesting additional challenges that come with the nature of events. In particular, we focus on the implications of different viewpoints: events are perceived from different perspectives, are thus being reported in text from different perspectives, and are also consumed (interpreted) from different perspectives.

[7] http://www.guardian.co.uk/

Modelling: ACTEVE In this section we report on the explorations we performed in the use case to make the challenges concrete and explore the possible solutions. As we followed an exploratory and evolutionary way to capture and model the events, we describe the different revisions of the model. By doing this we illustrate and motivate our modeling decisions and define and explain concepts important to this research. We call our model "ACTEVE" for "ACTivist EVEnts".

Initial Model The first version of the model was based on the Social Science concepts for describing networks of activist organizations and their activities. Essentially, the model represented *Organizations* targeting *Companies* through *Campaigns* that consist of *Tactics* with the aim of reaching a certain *Solution* (to a problem). *Campaigns* and *Solutions* were associated with a *State*, e.g. 'resolved' or 'partial'. *Organizations* were modeled with a Type, e.g. 'radical' or 'reformative'. *Campaigns* and *Tactics* were also associated with *Time* and *Location*, e.g. '12-06-12' or 'Amsterdam'. This version of the model can be seen in Figure 1A.

Revised Model After reviewing the initial model, we observed that the focus was much more on the campaigns and that the events were not yet considered as first class citizens in our model, like the event of the self-immolation of Bouazizi in the Arab Spring. The second version of our model was therefore made more event-centered. We introduced the concepts of an *Event* and an *Issue*. *Organization* was changed into an *Actor*. Essentially, the new model consisted of *Actors* organizing *Events* (as part of *Campaigns*) using a *Tactic* and targeting an *Issue* (related to *Company* responsible or causing the *Issue*). Each *Issue* has a *State* associated and each *Event* had a *Location* and *Time* associated. What follows are the definitions of each concept. This version of the model can be seen in Figure 1B.

1. *Event:* An action undertaken by an actor as part of a campaign with the aim of influencing the state of an issue.
2. *Tactic:* Defines the type of an event.
3. *Actor:* May be a person, group, or organization who performs tactics.
4. *Company:* An organization that triggers an issue.
5. *Issue:* Is a topic or problem important to actors and companies.
6. *Campaign:* consists of a set of events undertaken by an actor aiming to influence the state of an issue.

ACTEVE-SEM Model We then considered how the revised ACTEVE model could be expressed with the Simple Event Model (SEM) (Figure 2, [8]), to profit from the way SEM allows for a minimal modeling of events to facilitate interoperability (similar to Lode [6] and the Event Ontology [5]) and SEM's compatibility with external vocabularies. We observed how SEM allows us to express all of the constructs of the ACTEVE model. As can be seen in Figure 2, SEM models events in terms of who did what with what to whom where and when, modeled as *Actors*, *Events*, *Objects*, *Roles* and *Places*, each of which has a *Timestamp*.

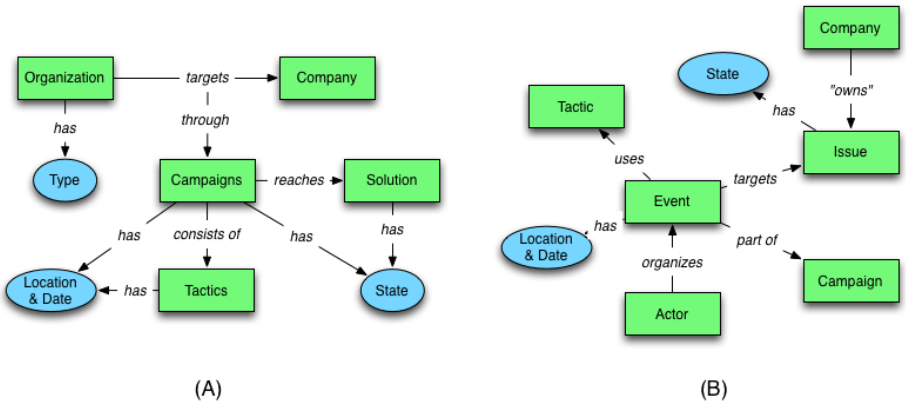

Fig. 1. (A): Initial model: Campaign-centered. (B): Revised model: Event-centered

SEM also allows us to specify certain 'views' on an event. This important concept for ACTEVE is explained in more detail in the next section and is one of the main reasons for choosing SEM over Lode, EO or a custom model.

Modelling Different Viewpoints As stated in the previous section, SEM also allows us to specify certain 'views' on an event, which hold according to a certain authority. This makes it possible to model different perspectives on the same event, which is an important notion as illustrated by the examples in Section 2.

Specifically, SEM allows us to specify three aspects of viewpoints [8]:

1. Event-bounded roles.
2. Time-bounded validity of facts.
3. Attribution of the authoritative source of a statement.

Being able to specify event-bounded roles is an important notion, specifically in our domain, because it allows us to differentiate between role types that hold according to different authoritative sources. In the Occupy event example of Section 2, the police officer might be seen by the protesters as having the role of an 'aggressor', while the officer might be seen by the NYPD as a 'peacekeeper'. This can be expressed in SEM using the 'sem:View' according to some 'sem:Authority' construct.

Similarly, it is useful to be able to specify the time-bounded validity of a certain fact. For example, the role of Mr. Bouazizi in the second example from Section 2 was initially 'street vendor' but changed after a certain point in time to 'martyr', at least according to some people. This can be expressed in SEM using the 'sem:Temporary' construct.

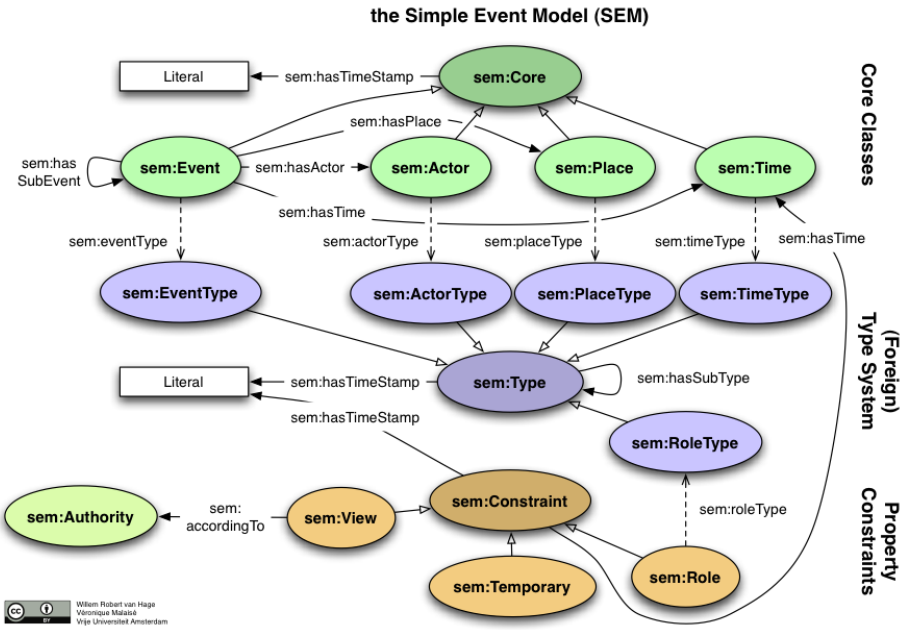

Fig. 2. Simple Event Model constructs

3.3 Visual Analytics for Activist Events

In the previous section we showed how our investigations from the use case learned us how the event information can be represented, including the different viewpoints. In our further use case exploration, we have seen how the representation of event information is not only geared towards a sound and complete storage of the events inside a system: the actual usage of the event information is done as part of the analysis that end-users do. The end-users are the ones that need to make sense of the information. Usually, they are accessing the information with some questions in mind: these questions are the ones that drive them in their explorations of the information. Through interviews with potential end-users (e.g. social scientists and interested lay people) we have identified 3 kinds of questions that could be posed.

1. Simple questions: basic 1-dimensional questions targeting simple statistical analysis, e.g "What is the most common type of event?"
2. Advanced questions: multidimensional questions targeting comparative analysis between different events, e.g. "Which Dutch actors have been involved in Occupy events?" This example covers two dimensions, i.e. actor and event type.
3. Interpretive questions: questions that cannot be answered only through reasoning with our event dataset, but require some interpretation and interaction by the users. For example, for the question "What are the most effective

tactics used by Tunisian actors during the Arab Spring?" we could retrieve statistical information from our dataset, with respect to tactics of Tunisian actors within a certain timespan, as well as issues and their states. However, the answer is in the user's interaction and final interpretation of this information. These questions are the ideal example to motivate the need for visual analytics.

In the context of these three types of questions, we envision the incorporation of different viewpoints to form an essential part of the visual analytics and end-users' exploration of the event information. In the study of our use case examples and domain conceptualization we have already seen several examples that demonstrate this.

To understand the best practices for supporting visual analysis of event information, the first step is to elicit visual analysis requirements for answering the above questions and then map existing tools to them. The ultimate goal that is behind this requirements elicitation of course is to understand what visual analysis support is needed in real-life use cases like the one we are considering here, and how such support could be realized

Visualization Requirements In this section, we present basic requirements for event visualizations. On the basis of these requirements, we will be able to select (or construct) different types of visualization techniques and evaluate their suitability. We distinguish between basic and advanced visualizations, which map to the question types of Section 3.3. For both types of visualizations, we report on how they answer the questions of Section 3.3, what is necessary to represent different viewpoints and the sources according to whom these viewpoints hold (Section 3.2 - Modeling Different Viewpoints).

Basic Visualizations Basic visualizations should present statistics about events and their properties. They use a simple numerical representation (e.g. percentage or ratio) and are typically used for representing statistical information on (parts of) large collections of events. Sorting, ranking and filtering by different criteria should be the most advanced features these types of visualizations have.

Even basic visualizations will have to facilitate representing different versions of the same statistic. This means that there should be a way to show according to whom a certain value is true, to represent different perspectives. Additionally, it should be possible to see to which point in time a certain value belongs, to be able to represent the temporal evolution of a statistic.

Unsurprisingly, these visualizations are best suited for answering the 'simple' questions as defined in Section 3.3. One example of a basic visualization would be a table, which lends itself well to showing one-dimensional information, but is poor for comparing multiple dimensions at the same time. This is where more advanced visualizations are necessary.

Advanced Visualizations Advanced visualizations should allow for links between events and/or links between event properties. They involve comparing multiple

dimensions at the same time, such as both location and time. Advanced visualizations typically map data to visual properties of geometric shapes to reveal trends and patterns in the data. Positioning data points in comparison to each other should be possible for categorical and temporal comparison.

As with basic visualizations, advanced visualizations will have to be able to represent different versions of the same data. Again, this means that there should be functionality to represent and switch between different versions of the same data over time or according to a certain authoritative figure.

Advanced visualizations are intended to answer the multidimensional 'advanced' and 'interpretive' questions of Section 3.3. An example is a scatterplot, where the x- and y-axis are mapped to two dimensions, while a third dimension could be represented by varying size/color of displayed symbols.

Visualization Tools Because of the highly specific nature of the requirements defined in the previous section, it is unlikely that we will be able to find a ready-made visualization technique or tool that meets these requirements. Therefore, it will be necessary to construct new visualization tools or modify existing visualization tools to incorporate the specific functionality that is necessary for event visualizations. In this section, we present a non-exhaustive overview of existing visualization tools that we believe are flexible and extensible enough to be usable in this process. In essence, we report on tools that might be useful when we start constructing visualizations per the requirements described above.

R R[8] is a free software environment for statistical computing and graphics. R allows for both basic one-dimensional statistical analysis as well as more advanced, multi-dimensional visual analytics (cf. [7]). R has a comprehensive library of plugins that extend its base functionality with additional functionality, such as the ability to generate different types of graphs and charts.

Google Chart Tools Google's Chart Tools[9] allow for the creation of various data visualizations, varying in complexity from pie, line or bar charts to maps, timelines or motion charts. The created visualizations have options for interactivity and are easily created on the fly for embedding in websites.

D3.js D3.js[10] is a JavaScript library for creating data visualizations using HTML, SVG and CSS. Like the Google Chart Tools, many types of data visualizations can be created, but D3.js is slightly more flexible as users can create new types of visualizations from scratch instead of having to select from a predefined library.

The study of the requirements and techniques has learned us what is needed to meet the research goal. We have seen how the elicitation of the questions that drive end-users in their analysis have determined the concrete targets for the visualization techniques in terms of features and functionality.

[8] http://www.r-project.org/
[9] http://developers.google.com/chart/
[10] http://d3js.org/

4 Conclusion and Future Work

In this paper we have reported the first results from our work concerning the modeling and analysis of events in the domain of activism. Many applications observe how events play an increasingly more important role. Capturing and extracting events for sense making analysis is what this research is aiming for. Reporting from the concrete context of our activism use case, e.g. Occupy, Arab Revolution, we show how events first get extracted, then how they get conceptualized, and then how visual analytics helps to make sense of the represented events. We emphasized the need to be able to represent different perspectives on events, as well as event properties. We have contributed the first SEM-based model for event modeling in the activism domain and we have identified the objectives and requirements for the visual analysis of these events. In future work we continue mapping the requirements for the visual analysis to the available techniques and tools, to design visual analysis support that can be evaluated with social scientists and lay people in the context of the activism domain.

Acknowledgements

This research is partially funded by the Royal Netherlands Academy of Arts and Sciences in the context of the Network Institute research collaboration between Computer Science and Social Sciences at the VU University Amsterdam. We would like to thank Willem R. van Hage and Jesper Hoeksema from VU University Amsterdam for their help in the NLP tools exploration.

References

1. van den Akker, C., Legêne, S., Van Erp, M., Aroyo, L., Segers, R., Van der Meij, L., van Ossenbruggen, J., Schreiber, G., Wielinga, B., Oomen, J., Jacobs, G.: Digital hermeneutics: Agora and the online understanding of cultural heritage. ACM Web Science Conference (Koblenz, Germany, June 14–17 2011)
2. van Hage, W., Malaisé, V., van Erp, M., Schreiber, G.: Linked open piracy. In: K-CAP. pp. 167–168. ACM (2011)
3. Leetaru, K.H.: Culturomics 2.0: Forecasting large-scale human behavior using global news media tone in time and space. First Monday 16(9), 2 (2011)
4. Marres, N., Rogers, R.: Recipe for tracing the fate of issues and their publics on the web. In: Latour, B., Weibel, P. (eds.) Making Things Public: Atmospheres of Democracy, pp. 922–935. MIT Press, Cambridge (Mass) (2005)
5. Raimond, Y., Abdallah, S.: The event ontology. Tech. rep., Queen Mary University of London (2007)
6. Shaw, R., Troncy, R., Hardman, L.: Lode: Linking open descriptions of events. The Semantic Web 5926(Section 2), 153–167 (2009)
7. Van Hage, W.R.: Sparql package for r - linked open piracy tutorial (2012)
8. Van Hage, W.R., Malaisé, V., Segers, R., Hollink, L., Schreiber, G.: Design and use of the simple event model (sem). Web Semantics Science Services and Agents on the World Wide Web 9(2), 128–136 (2011)