MONA
Interdisciplinary research into activist events

Organizational Sciences
Frank de Bakker
Inna Hellsten
Maxine Kruijt
Bibiana Armenta

Computer Science
Lora Aroyo
Thomas Ploeger
Chun Fei Lung

Computational Linguistics
Antske Fokkens
Why?

Activists shape the future by influencing opinions and policies
How?

Read news articles, identify the events activists participate in

Simple Questions
Narrative Questions
Interpretative Questions

Monday, October 21, 13
Problems

Answering these types of questions is difficult for various reasons.

Processing and aggregation is time consuming.

Individual news articles may lack context and present an incomplete and biased story.
Ideally

Automatically extract event mentions from multiple news sources

Store them in a structured format to enable aggregation

Present statistics and visual overviews such as timelines
Research Goals

Explore semantic models for storing and interpreting events

Build (semi)-automated tools to support question answering

NLP Tools for extraction

Visual Analytics for browsing
An ontology to describe events in terms of “Who did what, where and when?” (Van Hage 2011)
Extraction

Are there any existing methods we can re-use?

Gene and protein interactions from molecular biology papers

Illegal migration events from news articles for border security

Football events for creating match summaries
Our method for extracting activist events

Article Text

Sentence Splitting + Word Tokenization

Tuesday  Greenpeace  protested  against  Shell  in  London
Extraction

Our method for extracting activist events

Part-of-Speech Tagging

Tuesday  Greenpeace  protested  against  Shell  in  London
NNP      NNP      VBD      IN       NNP     IN     NNP

EVENT
Extraction: Our method for extracting activist events

Named Entity Recognition

Tuesday DATE
Greenpeace ORG protested against Shell ORG in London LOC
TIME ACTOR
Extraction Our method for extracting activist events

Dependency Parsing

Tuesday Greenpeace protested against Shell in London

NN NSUBJ PREP_AGAINST PREP_IN
Extraction
Our method for extracting activist events

Named Entity Disambiguation

Date Normalization

SEM Event
“Tuesday Greenpeace protested against Shell in London.”
45 articles concerning arctic oil exploration

A variety of sources, including NYT, Guardian, and Amnesty

Annotated by domain experts to create a gold standard
<table>
<thead>
<tr>
<th>Annotators</th>
<th>Event</th>
<th>Actor</th>
<th>Place</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46%</td>
<td>35%</td>
<td>36%</td>
<td>28%</td>
</tr>
<tr>
<td>2</td>
<td>34%</td>
<td>32%</td>
<td>28%</td>
<td>43%</td>
</tr>
<tr>
<td>3</td>
<td>20%</td>
<td>33%</td>
<td>36%</td>
<td>20%</td>
</tr>
</tbody>
</table>
We extracted 1829 events, 1609 actors, 772 places, and 456 times

<table>
<thead>
<tr>
<th></th>
<th>Event</th>
<th>Actor</th>
<th>Place</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>71.00%</td>
<td>46.00%</td>
<td>36.00%</td>
<td>65.00%</td>
</tr>
<tr>
<td>Recall</td>
<td>58.00%</td>
<td>35.00%</td>
<td>54.00%</td>
<td>60.0%</td>
</tr>
</tbody>
</table>
Errors

Low Place Precision
Low Actor Recall

“Shell is working with wary United States regulators.”

“The European Union has long shown leadership on the issue of climate change”
Errors  

Low Actor Recall

“The Kulluk Oil Rig was used for test drilling last summer.”

“The Coast Guard flew over the rig for a visual inspection.”

“Scientists say the drilling may influence...”
Errors | Gold Standard Annotation Instructions

Task was too loosely defined, resulting in different interpretations

<table>
<thead>
<tr>
<th>Annotator</th>
<th>Event F1</th>
<th>Actor F1</th>
<th>Place F1</th>
<th>Time F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63%</td>
<td>41%</td>
<td>49%</td>
<td>57%</td>
</tr>
<tr>
<td>2</td>
<td>54%</td>
<td>44%</td>
<td>19%</td>
<td>35%</td>
</tr>
<tr>
<td>3</td>
<td>60%</td>
<td>43%</td>
<td>48%</td>
<td>61%</td>
</tr>
</tbody>
</table>

Monday, October 21, 13
Visual Analytics
Visual Analytics
Lessons Learned

Off-the-shelf NLP tools yield useful, but imperfect results with little domain-specific tuning

Minimally constrained notion of event makes it difficult to instruct annotators

...and to fine-tune extraction process.
Future Work

Implement and evaluate solutions to known errors, re-run evaluation

New gold standard with more rigorous annotator instructions and majority voting

Methods for identifying complementary events
Recap

Study activists, they shape the future by influencing opinions and policies.

Used a pipeline of off-the-shelf NLP tools to extract events.

Extracted events used in visual analytics to answer questions.

Loose event definition is challenging to annotate and fine-tune.
Photo Attributions

1 http://www.flickr.com/photos/greenpeacesuomi/
2 http://www.flickr.com/photos/greenpeace_switzerland/
3 http://www.flickr.com/photos/greenpeace_nederland/
4 http://www.flickr.com/photos/x1klima/
5 http://www.flickr.com/photos/oxfam/
6 http://www.flickr.com/photos/remonrijper/
7 http://www.flickr.com/photos/davepattern/
8 http://wrvh.home.xs4all.nl/wrvhage/
9 http://www.flickr.com/photos/uccelino/
10 http://www.flickr.com/photos/cactusmelba/
11 Own work
12 http://www.flickr.com/photos/liloh
13 http://www.flickr.com/photos/liloh
14 http://www.flickr.com/photos/dgoodphoto/
15 http://www.flickr.com/photos/kaiban/
16 http://www.flickr.com/photos/kaiban/
17 http://www.flickr.com/photos/kaiban/
18 *
19 *
20 http://www.flickr.com/photos/rich_w/
21 http://www.flickr.com/photos/9948354@N08/
22 http://www.flickr.com/photos/christinawelsh/
23 *
24 http://www.flickr.com/photos/39955793@N07/
Shell oil rig ran aground in the Gulf of Alaska on the 31st of December 2012

18 days later, United States government expresses concerns about Shell’s operational accidents and impact on the environment

A few months later, Shell is granted a permit to begin drilling oil wells in Alaska, despite his earlier concerns