

Global Media Monitoring

<http://eventregistry.org/>

Marko Grobelnik

Marko.Grobelnik@ijs.si

Artificial Intelligence Lab, Jozef Stefan Institute

Ljubljana, Slovenia

Contributions from Gregor Leban, Blaz Fortuna, Janez Brank, Jan Rupnik,
Andrej Muhic, Aljaz Kosmerlj, Evgenia Belyaeva, Pat Moore

Outline

- Introduction
- Collecting Media Data
- Document Enrichment
- Cross-linguality
- News Reporting Bias
- Event Representation
- Event Visualization
- Event API
- Future/Ongoing Projects

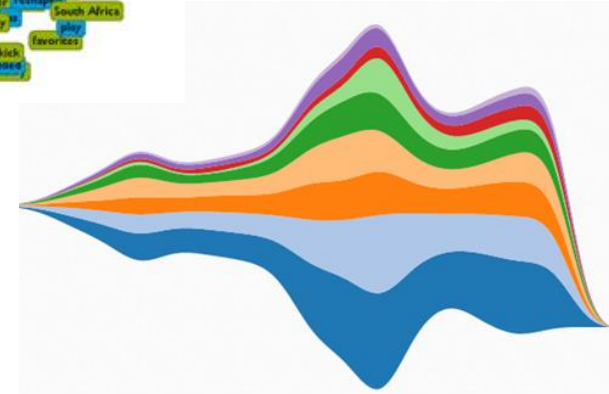
Introduction

What questions we'll try to answer?

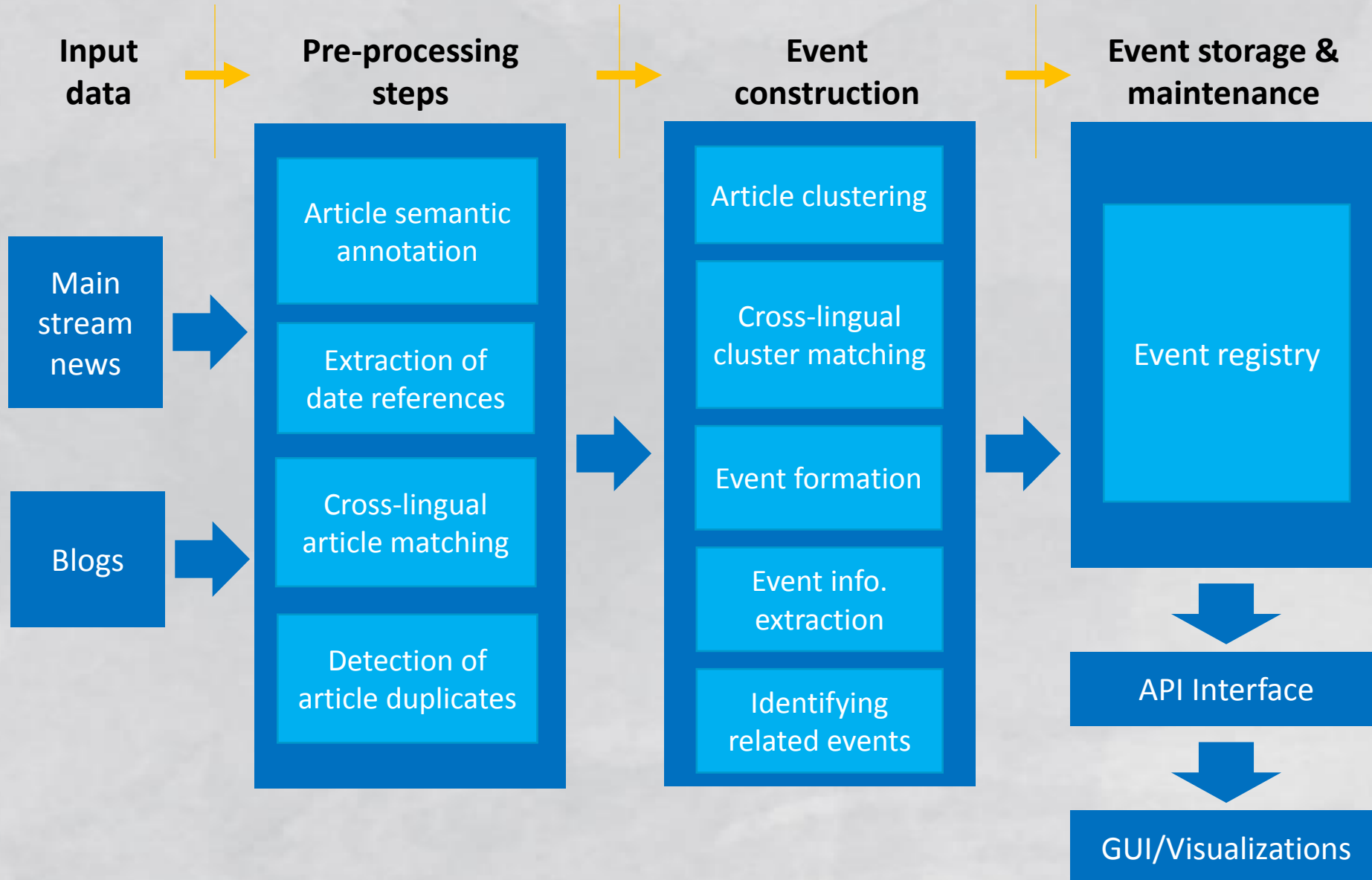
- Where to get global media data?
- What is extractable from media documents?
- How to connect information across languages?
- What is an event?
- How to approach diversity in news reporting?
- How to query & visualize global event dynamics?

Systems/Demos used within the presentation

- NewsFeed (<http://newsfeed.ijs.si/>)
 - News and social media crawler
- Wikifier (<http://wikifier.org/>)
 - Language and Semantic annotation
- XLing (<http://xling.ijs.si/>)
 - Cross-lingual document linking and categorization
- Event Registry (<http://eventregistry.org/>)
 - Event detection and topic tracking



Global Media Monitoring pipeline



<http://EventRegistry.org>

Collecting Media Data

<http://newsfeed.ijs.si/>

Where to get references to news publishers?

- Good start is Wikipedia list of newspapers:
 - http://en.wikipedia.org/wiki/Lists_of_newspapers

Lists of newspapers

From Wikipedia, the free encyclopedia

Below are lists of **newspapers** organized by continent.

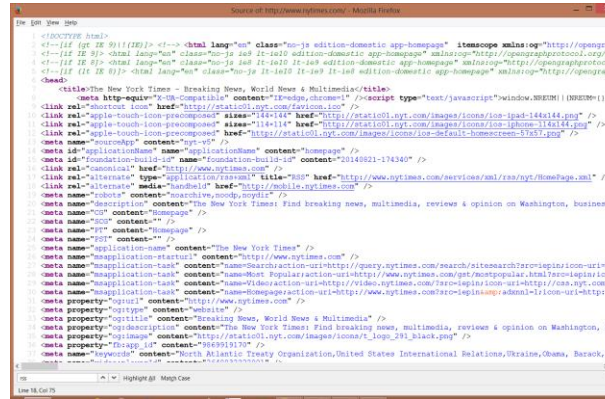
Contents [hide]
1 Africa
2 Asia
3 Europe
4 North America
5 Oceania
6 South America

Europe [edit]
<div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div><div><div><div></div></div><div><div></div></div></div></div></div> <div><div><div><div><div><</div></div></div></div></div>

From a newspaper home-page to an article

<http://www.nytimes.com/>

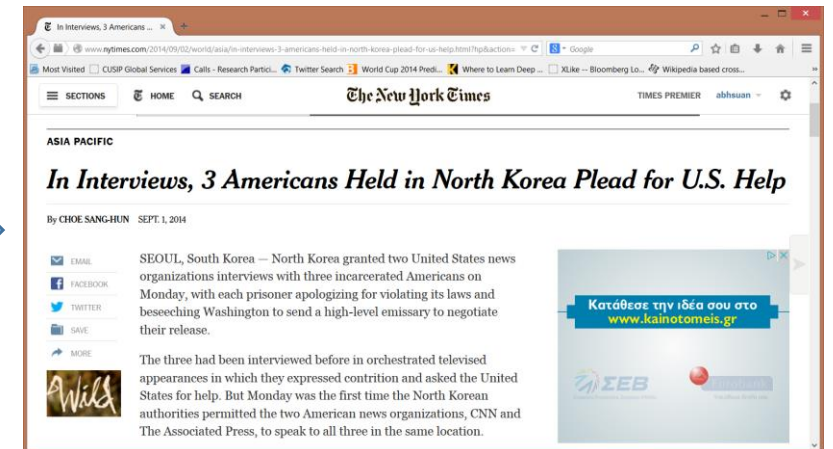
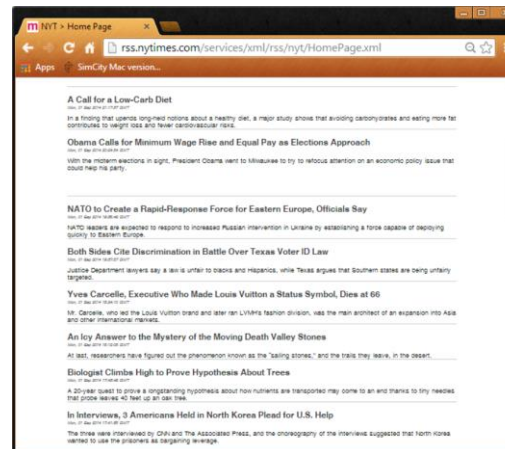
HTML



```
<link rel="alternate" type="application/rss+xml" title="RSS" href="http://www.nytimes.com/services/xml/rss/nyt/HomePage.xml" />
```

Article to be retrieved

RSS Feed
(list of articles)



Example: North Rhine Westphalia Local Newspapers

Aachener Nachrichten
Aachener Zeitung
Aegidienberger Bote
Ahlener Zeitung
Sonntags Rundblick
Altenaer Kreisblatt
Sonntags Rundblick
Werbepost
Bergisches Handelsblatt
Bergisches Sonntagsblatt
Neue Westfälische Zeitung
Bocholt Borkener Volksblatt
Die Tageszeitung
Bonner Rundschau
General Anzeiger
Rheinischer Merkur
Blickpunkt Schaufenster
Borkener Zeitung
Briloner Anzeiger
Bruhler Schlossbote

Hellerthaler Zeitung
Allgemeine Zeitung
Dattelner Morgenpost
Lippische Landes Zeitung
Ruhr Nachrichten
Dulmener Zeitung
Dusseldorfer Nachrichten
Handelsblatt
Rheinische Post
Terz
Westdeutsche Zeitung
Emsdettener Volkszeitung
Neue Ruhr Zeitung
Blickpunkt am Sonntag
Wochenende
Neue Regionale
Anzeigen Echo
Westfalen Post
Haller Kreisblatt
Allgemeiner Anzeiger

Sonntags Rundblick
Westfälischer Anzeiger
Hertener Allgemeine
Ibbenburener Volkszeitung
Echo Nord
Iserlohner Kreisanzeiger
Kvelaerer Blatt
Aktiv
Express
Kolner Stadt Anzeiger
Kölnische Rundschau
Krefelder Zeitung
Der Neue Postillon
Lippische Wochenschau
Siegerland Kurier
Lokale Informationen
Der Patriot
Ludenscheider Nachrichten
Marler Zeitung
Blickpunkt Schaufenster

Mendener Zeitung
Mindener Tageblatt
Münster am Sonntag
Münstersche Zeitung
Westfälische Nachrichten
Hellerthaler Zeitung
Neuss Grevenbroicher Zeitung
Die Glocke
Stimberg Zeitung
Sauerland Kurier
Sonntags Post
Recklinghauser Zeitung
Remscheider General Anzeiger
Münsterlandische Volkszeitung
Extra Blatt
Rhein Sieg Rundschau
Sonntagsanzeiger
Solinger Tageblatt
Steinheimer Blickpunkt
Hellweger Anzeiger
Lokalanzeiger
Waltroper Zeitung
Suderlander Volksfreund
Werbekurier
Meerbuscher Nachrichten
General Anzeiger

Example: news publishers from Rio de Janeiro

General News Media

- Agencia O Globo - agency
- Correio do Brasil - newspaper
- Destak - newspaper
- Extra Rio de Janeiro - newspaper
- Globo News - TV
- Ilha Noticias - newspaper
- Jornal Cidadania - newspaper
- Jornal da Barra - newspaper
- Jornal do Brasil - newspaper
- O Dia Rio de Janeiro - newspaper
- O Globo - newspaper
- Posto Seis - newspaper
- Povo do Rio - newspaper
- Tribuna da Imprensa - newspaper

Business Media

- Jornal do Commercio - newspaper
- Monitor Mercantil - newspaper

Entertainment Media

- Canal Brasil - TV
- Canal USA - TV
- Fox Life South America - TV
- Multishow - TV
- Piaui Magazine - magazine
- Rede Globo - TV
- TV Brasil - TV

Sports Media

- Diario Lance - newspaper
- Jornal dos Sports - newspaper
- Sportv - TV
- TV Esporte Interativo - TV

Society Media

- Folha Universal - newspaper

Collecting global media data

- Data collection service News-Feed
 - <http://newsfeed.ijs.si/>
 - ...crawling global main-stream and social media
- Monitoring
 - ~150k main-stream publishers (RSS feeds+special feeds)
 - ~250k most influential blogs (RSS feeds)
 - free Twitter feed
- Data volume: ~350k articles & blogs per day (+5M tweets)
- Languages: eng (50%), ger (10%), spa (8%), fra (5%), ...



Downloading the news stream (1/2)

- The stream is accessible at <http://newsfeed.ijs.si/stream/>
- To download the whole stream continuously, you can use the python script (<http://newsfeed.ijs.si/http2fs.py>)
- The script does the following:

```
timestamp = [when you want to start downloading, e.g. now() minus 1 hour]
while True:
    fetch http://newsfeed.ijs.si/stream/?after=timestamp
    if [404 error]:
        # there is no new data
        pause 1 minute
    else:
        save data
        timestamp = [extract it from the Content-Disposition HTTP header]
```



Downloading the news stream (2/2)

- News Stream Contents and Format
 - The root element, <article-set>, contains zero or more articles in the following XML format:
- ...more details:
 - Mitja Trampus, Blaz Novak: *The Internals Of An Aggregated Web News Feed*. Proceedings of 15th Multiconference on Information Society 2012 (IS-2012). [\[PDF\]](#)

```
<article id="internal article ID; consistent across streams">
  <source>
    <hostname> Publisher hostname </hostname>
    <title> Name of the publisher; failing that, title of the RSS feed </title>
    <location?>
      <longitude?> publisher longitude in degrees </longitude>
      <latitude?> publisher latitude in degrees </latitude>
      <city?> publisher city </city>
      <country?> publisher country </country>
    </location>
    <tags?>
      <tag+> a tag for the publisher; the vocabulary is not controlled </tag>
    </tags>
  </source>
  <feed?>
    <uri> URL from which the article was discovered; typically the RSS feed </uri>
  </feed>
  <uri> URL from which the article was downloaded </uri>
  <publish-date?> The publication time and date. </publish-date>
  <retrieve-date> The retrieval time and date. </retrieve-date>
  <lang> 3-letter ISO 639-2 language code </lang>
  <story_id?> story cluster this article was grouped into (at download time) </lang>
  <location?+>
    <longitude?> story content longitude in degrees </longitude>
    <latitude?> story content latitude in degrees </latitude>
    <city?> story city </city>
    <country?> story country </country>
  </location>
  <tags?>
    <tag+> a tag for the article; the vocabulary is not controlled </tag>
  </tags>
  <img?> The URL of a related image, usually a thumbnail. </img>
  <title> Title. Can be empty if we fail to identify it. </title>
  <body-cleartext>
    Clear text body of the article, formatted only with <p> tags
  </body-cleartext>
  <body-rych?; only English, Slovene>
    Enriched article body; an XML subtree as returned by Enrycher.
  </body-rych>
  <body-xlike?; only English, Spanish, Catalan>
    Enriched article body; an XML subtree as returned by iSOCO; experimental.
  </body-xlike>
</article>
```

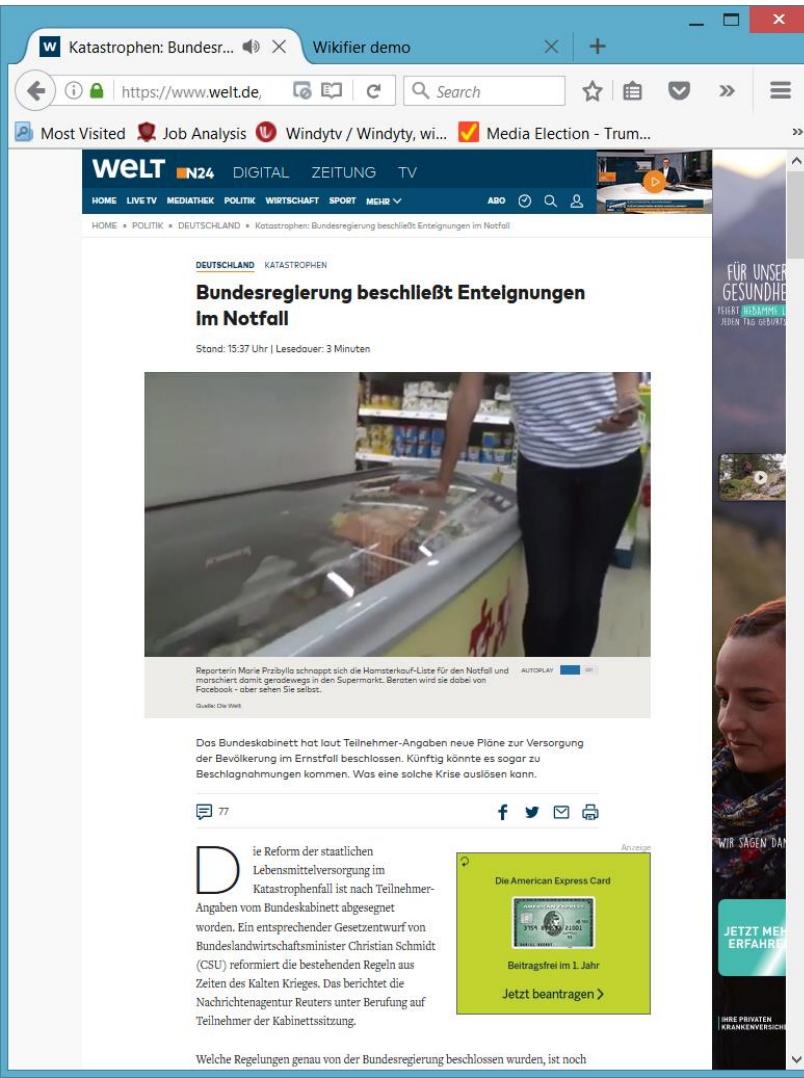

What can be extracted from a document?

- Lexical level
 - **Tokenization** – extracting tokens from a document (words, separators, ...)
 - **Sentence splitting** – set of sentences to be further processed
- Linguistic level
 - **Part-of-Speech** – assigning word types (nouns, verbs, adjectives, ...)
 - **Deep Parsing** – constructing parse trees from sentences
 - **Triple extraction** – subject-predicate-object triple extraction
 - **Name entity extraction** – identifying names of people, places, organizations
- Semantic level
 - **Co-reference resolution** – replacing pronouns with corresponding names; merging different surface forms of names into single entity
 - **Semantic labeling** – assigning semantic identifiers to names (e.g. LOD/DBpedia/Freebase) including disambiguation
 - **Topic classification** – assigning topic categories to a document (e.g. DMoz)
 - **Summarization** – assigning importance to parts of a document
 - **Fact extraction** – extracting relevant facts from a document

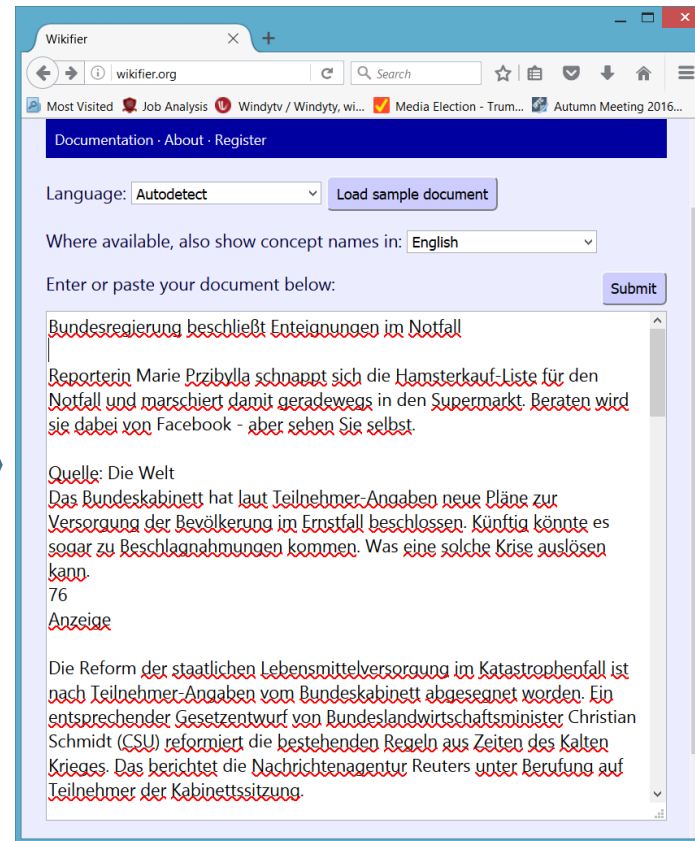
Semantic annotation with Wikifier.Org

(operates for 100 languages)

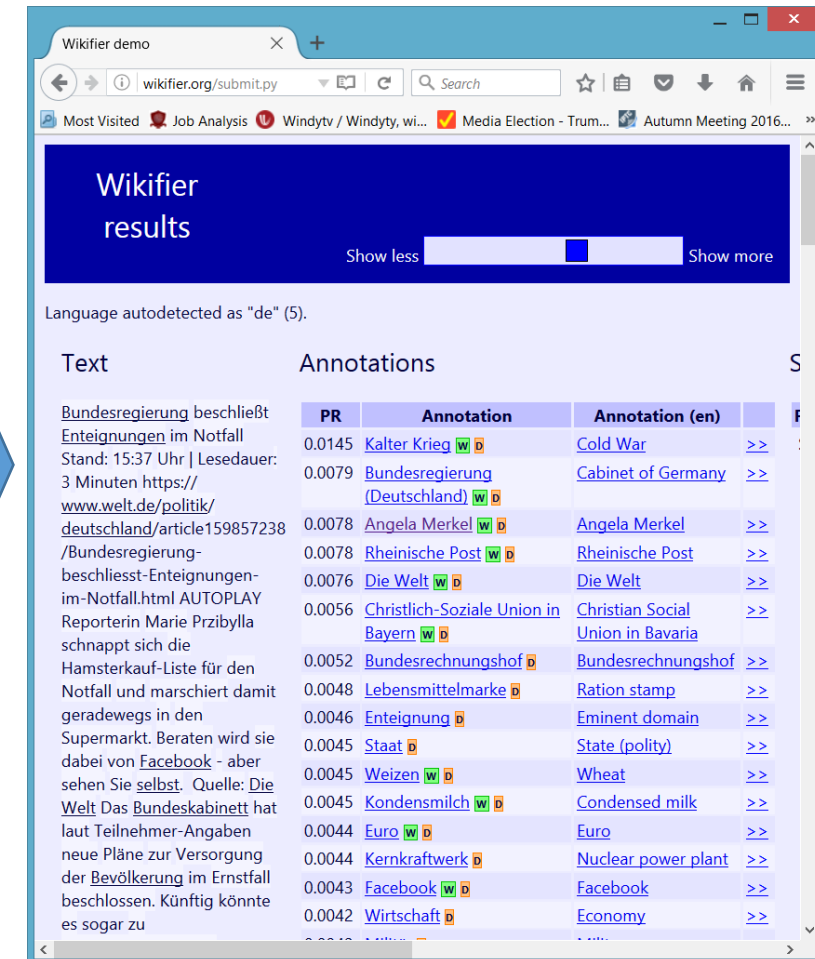
Original text



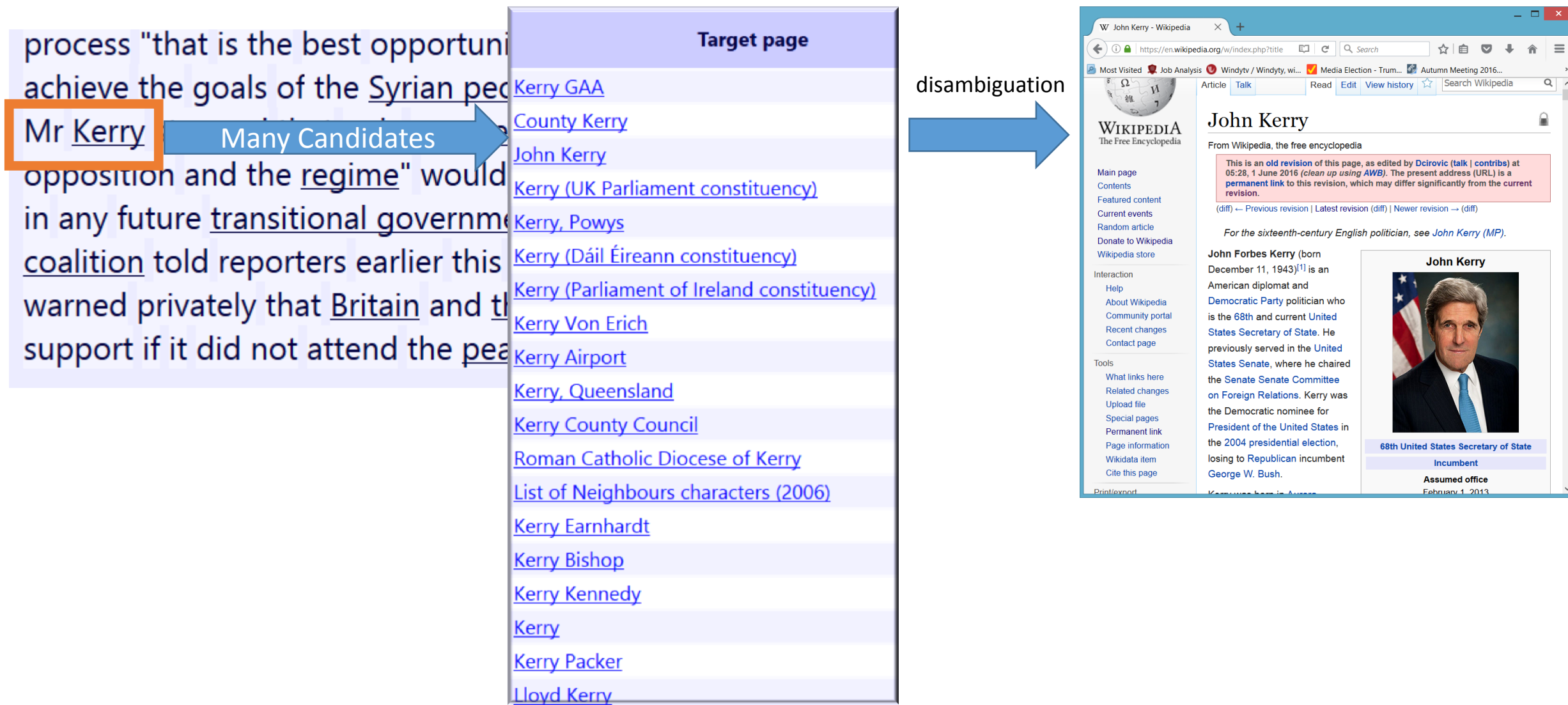
Plain clean text



Annotated text in original language and English



Key problem in semantic annotation (wikification) is concept disambiguation

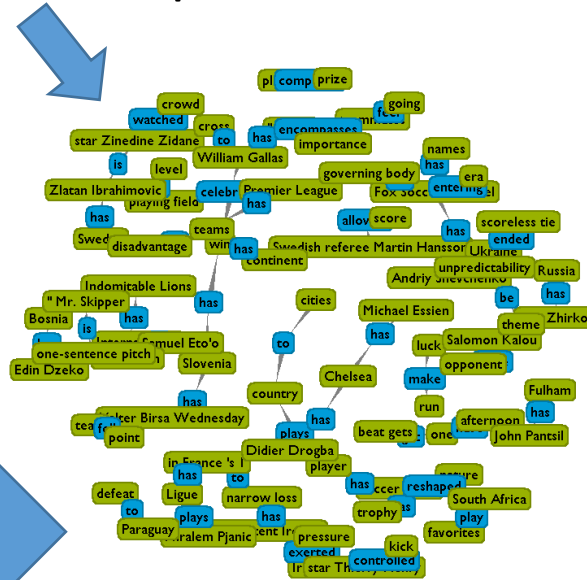


Enrycher (<http://enrycher.ijs.si/>)

Plain text



Extracted graph of triples from text



Text Enrichment

Slovenia's dramatic win over Russia Wednesday, and to a lesser extent Ireland's narrow loss to France, capped off a grueling two-year qualifying period that saw some of the smallest countries in the world kick some of soccer's biggest names in the teeth. After a century of near domination from the likes of Brazil, Italy and Germany, international soccer is entering the era of the Cinderella. It may not happen this time around, but given the increasing flow of talent, training and information across borders, it's almost certain that a small upstart nation blessed with good athletes and better luck will make a legitimate run at the world's most coveted trophy.

Russia's Yuri Zhirkov, right, fights for the ball with Slovenia's Wednesday.

entities

- [Brazil](#)
- [Italy](#)
- [Germany](#)
- [Cinderella](#)
- [Paris](#)
- [John O'Shea](#)
- [Manchester United](#)
- [Robbie Keane](#)
- [Shay Given](#)
- [Greece](#)
- [Portugal](#)
- [Bosnia-Herzegovina](#)
- [Cristiano Ronaldo](#)
- [Uruguay](#)

keywords

Sports, Soccer, CONCACAF, Competitions, United States, Sports and Hobbies, Kids and Teens, World Cup, Women,

categories

- [Top/Kids_and_Teens](#)
[/Sports_and_Hobbies](#)
[/Sports/Soccer](#)
- [Top/Sports/Soccer](#)
[/Competitions](#)
- [Top/Sports/Soccer](#)
[/Competitions/World_Cup](#)
- [Top/Sports/Soccer](#)
[/CONCACAF](#)

“Enrycher” is available as
as a web-service generating
Semantic Graph, LOD links,
Entities, Keywords, Categories,
Text Summarization, Sentiment

Diego Maradona Semantics:

owl:sameAs: http://dbpedia.org/resource/Diego_Maradonaowl:sameAs: <http://sw.opencyc.org/concept/Mx4rvofERZwpEbGdrcN5Y29ycA>

rdf:type: <http://dbpedia.org/class/yago/ArgentinaInternationalFootballers>

rdf:type: <http://dbpedia.org/class/yago/ArgentineExpatriatesInItaly>

rdf:type: <http://dbpedia.org/class/yago/ArgentineFootballManagers>

rdf:type: <http://dbpedia.org/class/yago/ArgentineFootballers>

Robbie Keane Semantics:

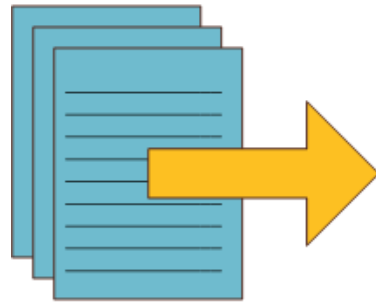
owl:sameAs: [http://dbpedia.org/resource/Robbie Keane](http://dbpedia.org/resource/Robbie_Keane)

rdf:type: <http://dbpedia.org/class/yago/CoventryCityF.C.Players>

rdf:type: <http://dbpedia.org/class/yago/ExpatriateFootballPlayersInItaly>

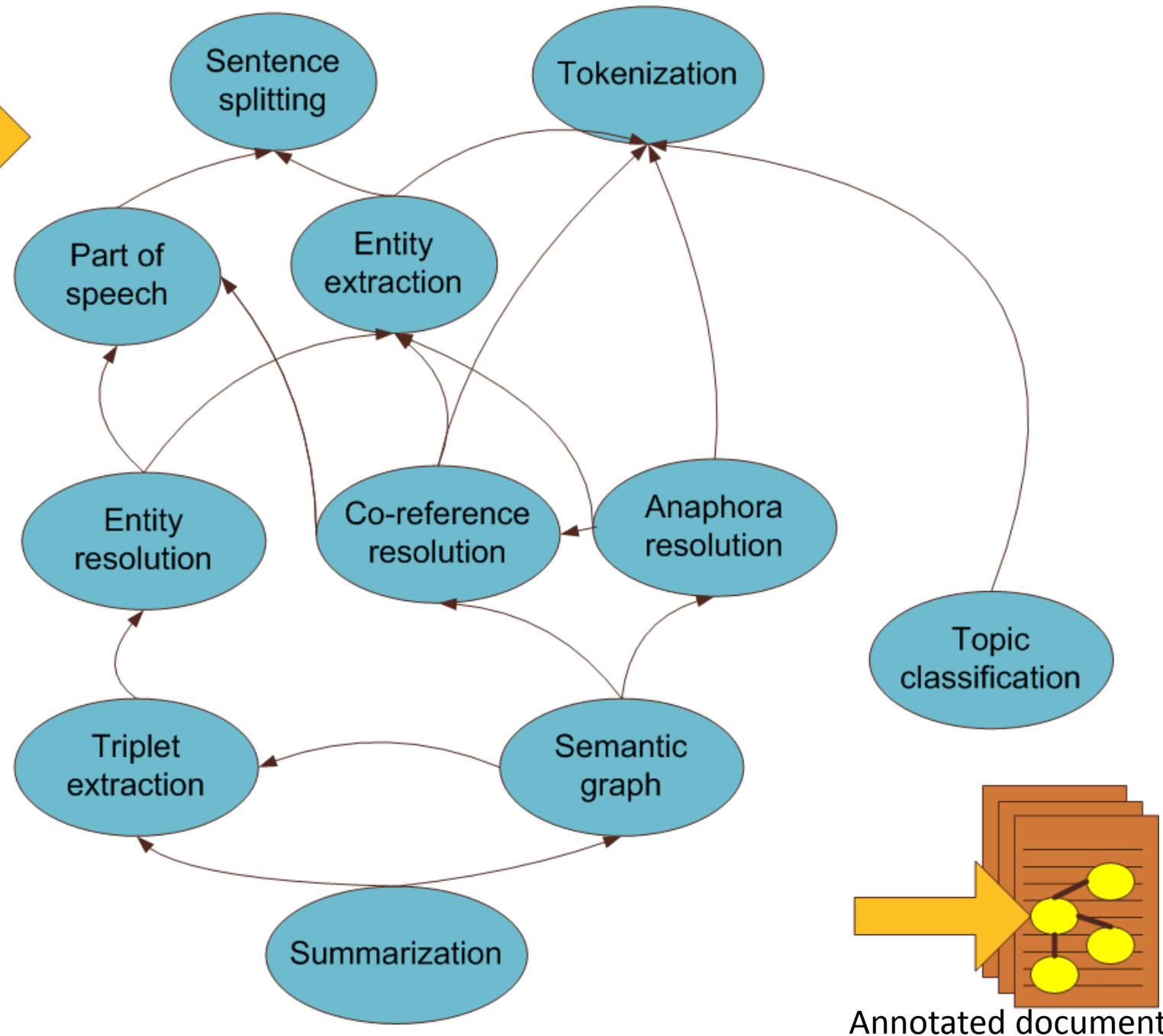
rdf:type: <http://dbpedia.org/class/yago/F.C.InternazionaleMilanoPlayers>

Enrycher Architecture



Plain text

- **Enrycher** is a web service consisting of a set of interlinked modules...
- ...covering lexical, linguistic and semantic annotations
- ...exporting data in XML or RDF
- To execute the service, one should send an HTTP POST request, with the raw text in the body:
 - `curl -d "Enrycher was developed at JSI, a research institute in Ljubljana. Ljubljana is the capital of Slovenia."`
`http://enrycher.ijs.si/run`



Cross-linguality

<http://xling.ijs.si/>

Cross-linguality

How to operate in many languages?

- Cross-linguality is a set of functions on how to transfer information across the languages
 - ...having this, we can track information independent of the language borders
 - Machine Translation is expensive and slow, so the goal is to avoid machine translation to gain speed and scale
- The key building block is the function for comparing and categorization of documents in different languages
 - <http://XLing.ijs.si> is an open web service to bridge information across 100 languages

Languages covered by XLing (top 100 Wikipedia languages)

Afrikaans	Chinese	Hungarian	Malagasy	Simple English
Albanian	Chuvash	Icelandic	Malay	Slovak
Arabic	Croatian	Ido	Malayalam	Slovenian
Aragonese	Czech	Indonesian	Marathi	Spanish
Armenian	Danish	Irish	Nepali	Sundanese
Asturian	Dutch	Italian	Newar / Nepal Bhasa	Swahili
Azerbaijani	English	Japanese	Norwegian (Bokmål)	Swedish
Bashkir	Esperanto	Javanese	Norwegian (Nynorsk)	Tagalog
Basque	Estonian	Kazakh	Occitan	Tamil
Belarusian	Finnish	Kirghiz	Persian	Tatar
Belarusian (Taraškievica)	French	Korean	Piedmontese	Telugu
Bengali	Galician	Kurdish	Polish	Thai
Bishnupriya Manipuri	Georgian	Latin	Portuguese	Turkish
Bosnian	German	Latvian	Quechua	Ukrainian
Breton	Greek	Lithuanian	Romanian	Urdu
Bulgarian	Gujarati	Lombard	Russian	Uzbek
Burmese	Haitian	Low Saxon	Serbian	Vietnamese
Cantonese	Hebrew	Luxembourgish	Serbo-Croatian	West Frisian
Catalan	Hindi	Macedonian	Sicilian	Western Panjabi
Cebuano				Yoruba

XLing (<http://XLing.ijs.si>)

service for comparing and categorization of documents across 100 languages

The screenshot shows the XLing web interface with several annotations:

- Automatically Extracted Keywords**: Points to the keywords for the English document: **china economi econom growth bank market chines rate that global**.
- Similarity Between Two Documents**: Points to the similarity score **0.669834**.
- Automatically Extracted Keywords**: Points to the keywords for the Chinese document: **经济 中国 银行 增长 全球 企业 在 改革 政府 资本**.
- Chinese Text**: Points to the Chinese text content: **世界银行:中国经济“硬着陆”风险仍存**.
- English Text**: Points to the English text content: **World Bank cuts China and Thailand's growth forecasts**.
- Selection Of 100 Languages**: Points to the language selection dropdown menu.

The interface includes a **Report** button, language selection dropdowns for **English** and **Chinese**, and a **Dmoz cat** button. The text content is displayed in two columns, with the English text on the left and the Chinese text on the right. The Chinese text includes a quote from Andrew Burns, a BBC Chinese website reporter, and a list of related content and news topics.

Example: Cross-lingual News Recommendation

- What local media (e.g. German) is writing about the topic we are reading in English?
- Usual fear of publishers: are users we are sending away coming back?
 - ...evaluation shows they all come back

The screenshot shows the Bloomberg website interface. At the top, there's a navigation bar with links to Bloomberg.com, Businessweek.com, Bloomberg TV, and Premium. Below this is a 'MARKET SNAPSHOT' section with a table of stock indices for the U.S., Europe, and Asia. A large green banner promotes 'THE WEALTH WATCH NEWSLETTER' with a 'SIGN UP NOW' button. The main navigation bar includes links for HOME, QUICK, NEWS, OPINION, MARKET DATA, PERSONAL FINANCE, TECH, POLITICS, SUSTAINABILITY, TV, VIDEO, and RADIO. A search bar is located on the right. Below the navigation bar, there are several news snippets with images and headlines. The main article is titled 'Taxes to Rise for Workers as Budget Deal Still Elusive' by Kathleen Hunter & Roxana Tiron, dated Dec 31, 2012. It includes social media sharing buttons and a '243 COMMENTS' count. The article text discusses the U.S. budget deal and its impact on workers. To the right of the main article, there's a sidebar with a 'GET THE POLITICAL CAPITAL NEWSLETTER' sign-up button and a section titled 'DEUTSCH NACHRICHTEN' (German News). This section lists several headlines in German, including 'US-Kongress kommt am Silvestertag zu Sitzung über Budget zusammen', 'US-Haushaltsstreit steuert auf dramatisches Finale zu', 'Ägyptisches Pfund stürzt ab', 'Lästige Konkurrenz: Deutsche Bank neidisch auf Datenkrake Google', 'Die Frauen, der Tod und ein hilfloser Staat', and 'Erdogan spricht syrischen Flüchtlingen Mut zu'. A blue circle highlights the 'DEUTSCH NACHRICHTEN' section.

U.S.	EUROPE	ASIA
↑ DJIA	12,938.10	-158.20 -1.21%
↑ S&P 500	1,402.43	-15.67 -1.10%
↑ NASDAQ	2,960.31	-25.59 -0.86%

Taxes to Rise for Workers as Budget Deal Still Elusive

By Kathleen Hunter & Roxana Tiron - Dec 31, 2012 2:36 PM GMT+0100

With taxes set to increase for almost every U.S. worker at midnight, Congress hasn't reached a budget deal that Democrats and Republicans say is necessary to prevent a blow to the U.S. economy.

Private talks between Senate Majority Leader [Harry Reid](#) and Minority Leader [Mitch McConnell](#) that began Dec. 28 stalled yesterday because of disputes over income tax rates, the estate tax and other issues.

DEUTSCH NACHRICHTEN

- US-Kongress kommt am Silvestertag zu Sitzung über Budget zusammen [Cash Magazine](#)
- US-Haushaltsstreit steuert auf dramatisches Finale zu (2. AF) [Cash Magazine](#)
- Ägyptisches Pfund stürzt ab [Cash Magazine](#)
- Lästige Konkurrenz: Deutsche Bank neidisch auf Datenkrake Google [Die Welt](#)
- Die Frauen, der Tod und ein hilfloser Staat [Frankfurter Allgemeine Zeitung](#)
- Erdogan spricht syrischen Flüchtlingen Mut zu [Cash Magazine](#)

Example: Social Media Recommendation

(<http://aidemo.ijs.si/xlike/hsdemo/>)

- Usual problem publishers have what and when to publish to which social media channel?
 - ...the demo shows what Bloomberg articles should be the most relevant based on what is most 'popular' in a particular geographical area

The screenshot displays the 'XLike | Regional Demand Prediction' web application. The browser address bar shows the URL aidemo.ijs.si/xlike/hsdemo/. The page title is 'XLike | Regional Demand Prediction' with links for 1 hour, 3 hour, 6 hour, and 12 hour predictions. The content is organized into a grid of six panels, each representing a different geographical region: Germany, France, Spain, Italy, USA East coast, and USA West coast. Each panel lists recommended articles with their rank, title, and engagement metrics (Page views, Local views, Facebook, and Twitter).

Region	Rank	Article Title	Page views	Local views	Facebook	Twitter
Germany	[63]	German States Join Ranks Pressing Merkel to Spur Spending	2068	6%	0	25
	[46]	BayernLB to File Constitutional Complaint Over Hypo Alpe	38	15%	0	0
	[44]	Biggest German Public Pension Likes Buyout, Hedge Funds	128	10%	1	3
	[35]	Second Health Worker Tests Positive for Ebola in Texas	4518	1%	12	56
	[33]	EBay Merchants Lamenting Neglect Bid PayPal Good Riddance	1296	1%	1	2
France	[54]	North, South Korea Generals Deadlocked in First Talks Since 2007	59	3%	0	1
	[31]	EDF Said to Replace Proglio With Thales CEO Jean-Bernard Levy	389	25%	0	0
	[29]	German States Join Ranks Pressing Merkel to Spur Spending	2068	3%	0	25
	[28]	Second Health Worker Tests Positive for Ebola in Texas	4518	1%	12	56
	[28]	EBay Merchants Lamenting Neglect Bid PayPal Good Riddance	1296	1%	1	2
Spain	[64]	Second Health Worker Tests Positive for Ebola in Texas	4518	1%	12	56
	[57]	Low-Budget Shows Leap From Web to TV to Attract Big Money	48	2%	0	0
	[50]	Isn't There More to Worry About Than Headphones?: Opening Line	88	0%	0	2
	[40]	EBay Merchants Lamenting Neglect Bid PayPal Good Riddance	1296	0%	1	2
	[37]	Aldermore Bank Postpones London IPO Amid Market Selloff	29	3%	0	3
Italy	[63]	Cosmo Plans Management Move in Search of Irish Corporate Charm	37	3%	0	0
	[40]	EBay Merchants Lamenting Neglect Bid PayPal Good Riddance	1296	1%	1	2
	[38]	German States Join Ranks Pressing Merkel to Spur Spending	2068	3%	0	25
	[38]	Mercedes G-Class Woos Kardashians by Staying Stuck in Mud	1829	1%	3	4
	[34]	German Yields Drop to Records With Peers on Bets Slump Worsening	189	6%	0	0
USA East coast	[67]	Isn't There More to Worry About Than Headphones?: Opening Line	88	24%	0	2
	[56]	Second Health Worker Tests Positive for Ebola in Texas	4518	23%	12	56
	[37]	Gun Lobby Turns Back on Senate Democratic Allies	16	41%	0	0
	[37]	EBay Merchants Lamenting Neglect Bid PayPal Good Riddance	1296	19%	1	2
	[21]	Mercedes G-Class Woos Kardashians by Staying Stuck in Mud	1829	14%	3	4
USA West coast	[64]	Isn't There More to Worry About Than Headphones?: Opening Line	88	9%	0	2
	[37]	Second Health Worker Tests Positive for Ebola in Texas	4518	6%	12	56
	[35]	BofA Posts Surprise Profit After Mortgage Settlement	882	20%	0	5
	[33]	EBay Merchants Lamenting Neglect Bid PayPal Good Riddance	1296	6%	1	2
	[28]	Low-Budget Shows Leap From Web to TV to Attract Big Money	48	8%	0	0

News Reporting Bias

<http://aidemo.ijs.si/diversinews/>

News Reporting Bias example

UK SOLDIERS CLEARED IN IRAQI DEATH – SEVEN BRITISH SOLDIERS WERE ACQUITTED ON THURSDAY OF CHARGES OF BEATING AN INNOCENT IRAQI TEENAGER TO DEATH WITH RIFLE BUTTS. A JUDGE AT A SPECIALLY CONVENED MILITARY COURT IN EASTERN ENGLAND ORDERED THE ADJUDICATING PANEL TO RETURN 'NOT GUILTY' VERDICTS AGAINST THE SEVEN BECAUSE HE DID NOT BELIEVE THERE WAS SUFFICIENT EVIDENCE AGAINST THEM, THE MINISTRY OF DEFENCE SAID. . . .

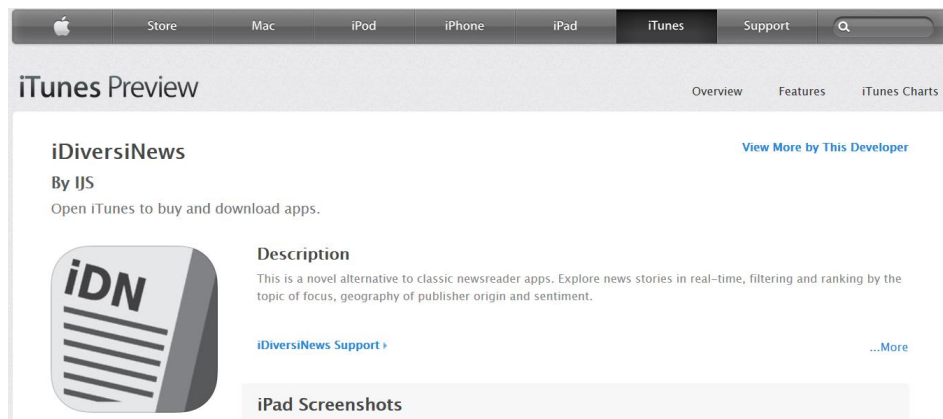
BRITISH MURDERERS IN IRAQ ACQUITTED – THE JUDGE AT A COURT-MARTIAL ON THURSDAY DISMISSED MURDER CHARGES AGAINST SEVEN SOLDIERS, FROM THE 3RD BATTALION, THE PARACHUTE REGIMENT, WHO'RE ACCUSED OF MURDERING IRAQI TEENAGER; CLAIMING THERE'S INSUFFICIENT EVIDENCE TO SECURE A CONVICTION, THE ASSOCIATED PRESS REPORTED THURSDAY. . . .

Detecting News Reporting Bias

- The task:
 - Given a news story, are we able to say from which news source it came?
- We compared **CNN** and **Aljazeera** reports about the same events from the war in Iraq
 - ...300 aligned articles describing the same story from both sources
- The same topics are expressed in both sources with the following keywords:
 - CNN with:
 - **Insurgents**, Troops, Baghdad, Iran, **Militant**, Police, **Suicide**, **Terrorist**, United, National, Hussein, **Alleged**, Israeli, Syria, Terrorism...
 - Aljazeera with:
 - Attacks, Claims, **Rebels**, Withdrawing, Report, **Fighters**, President, **Resistance**, Occupation, Injured, Army, Demanded, Hit, Muslim, ...

DiversiNews iPad App (1/2)

- DiversiNews iPad App is using newsfeed.ijs.si and enrycher.ijs.si services
- ...in its initial screen is shows list of current hot topics and current trending events



Hot Topics



Trending Events

DiversiNews iPad App (2/2)

- DiversiNews “***diversity search***” screen allows dynamic reranking of articles describing an event along three dimensions:
 - **Geography** – where is a content being published from
 - **Subtopics** – what are subtopics of an event
 - **Sentiment** – what are good and what are bad news
- For each query it provides
 - Automatically generated **summary**
 - List of corresponding **articles**

Geography

Subtopics

Sentiment

Summary

Articles

iPad 13:30 Not Charging

< Back Kerry attempts the improbable with Middle East peace deal - Financial Times

Related People and Organizations:
John Kerry, Prime Minister Benjamin Netanyahu, Bennett, West Bank, Minister Yuval Steinitz, Minister Naftali, Palestinian Authority, Jewish Home

Story Started: 1. feb. 2014 12:00
Last Update: 3. feb. 2014 12:45
News Items: 100
Sentiment: Neutral

Categories:
Middle East, Society, Regional, Warfare and Conflict, Issues

Focus on News Coming From

Focus on News About

REMARKS
FAIL ISRAEL
GROWING INTERNATIONAL
MOMENTARY EXPECTS
ARAB ABBAS PLAN
MINISTER SLAMS
CONTINUED ECONOMIC
PEACE ISRAEL
MAIN TV
WON
SUNDAY SIGNAL

Selection: BOYCOTT, MOMENTARY, EXPECTS, MR KERRY, ECONOMIC, GROWING

Focus on Specific News Sentiment

Summary per Your Settings

Google Summary: The apparently choreographed chorus of criticism drew a sharp response from the US state department, which denied that Mr Kerry - who is currently trying to draw up a framework agreement between Israel and the Palestinians - was behind an international embargo. While the US secretary of state's remarks were made against a backdrop of new EU regulations barring deals with Israeli businesses based in West Bank settlements, they provoked reactions that he was threatening Israel in on-going peace talks with the Palestinians. Naftali Bennett, the industry minister and leader of the far-Right Jewish Home party, said: "We expect of our friends in the world to stand by our side against the attempts to impose an anti-Semitic boycott on Israel, and not to be their mouthpiece."

Top News Items

View All News Items...

Kerry attempts the improbable with M...
Financial Times FT

Israel says Kerry boycott warning is 'a...
ish Independent

Row over accepting Israel as 'Jewish'
BBC Radio Lincolnshire

John Kerry labelled 'anti Semite' for w...
Daily Telegraph

Reset Settings

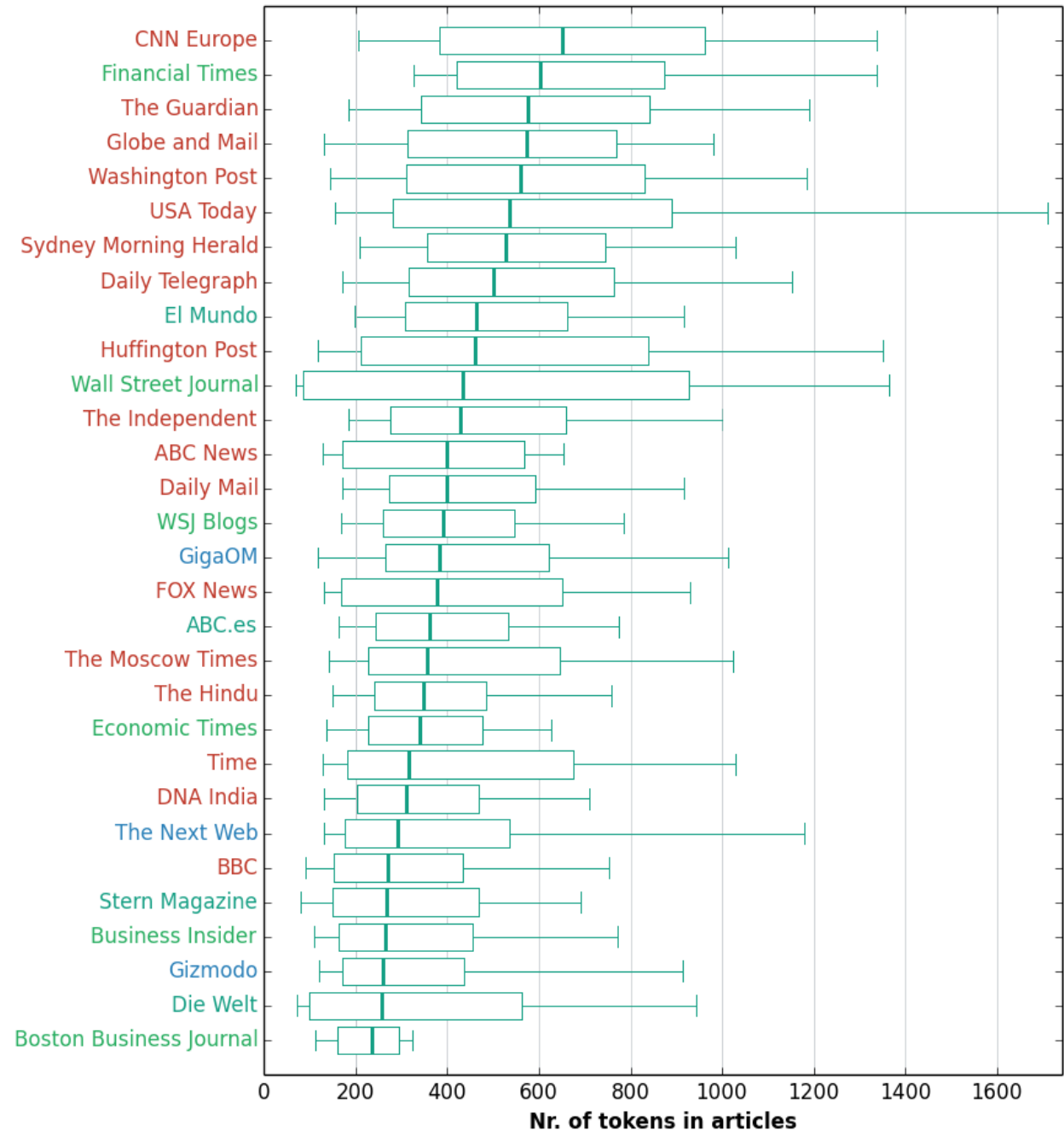
New Bias Experiment

- 30 news publishers selected
- All articles from Dec 15th 2014 – Aug 15 2014
- **General:** BBC, USA Today, Daily Mail, The Guardian, Globe and Mail, Huffington Post, Washington Post, CNN Europe, The Independent, Time, The Hindu, The Moscow Times,
- **Business related:** Business Insider, Boston Business Journal, Economic Times, Financial Times, WSJ Blogs
- **Tech related:** Gizmodo, GigaOM, The Next Web
- **Foreign:** ABC.es, El Mundo, Die Welt, Stern Magazine

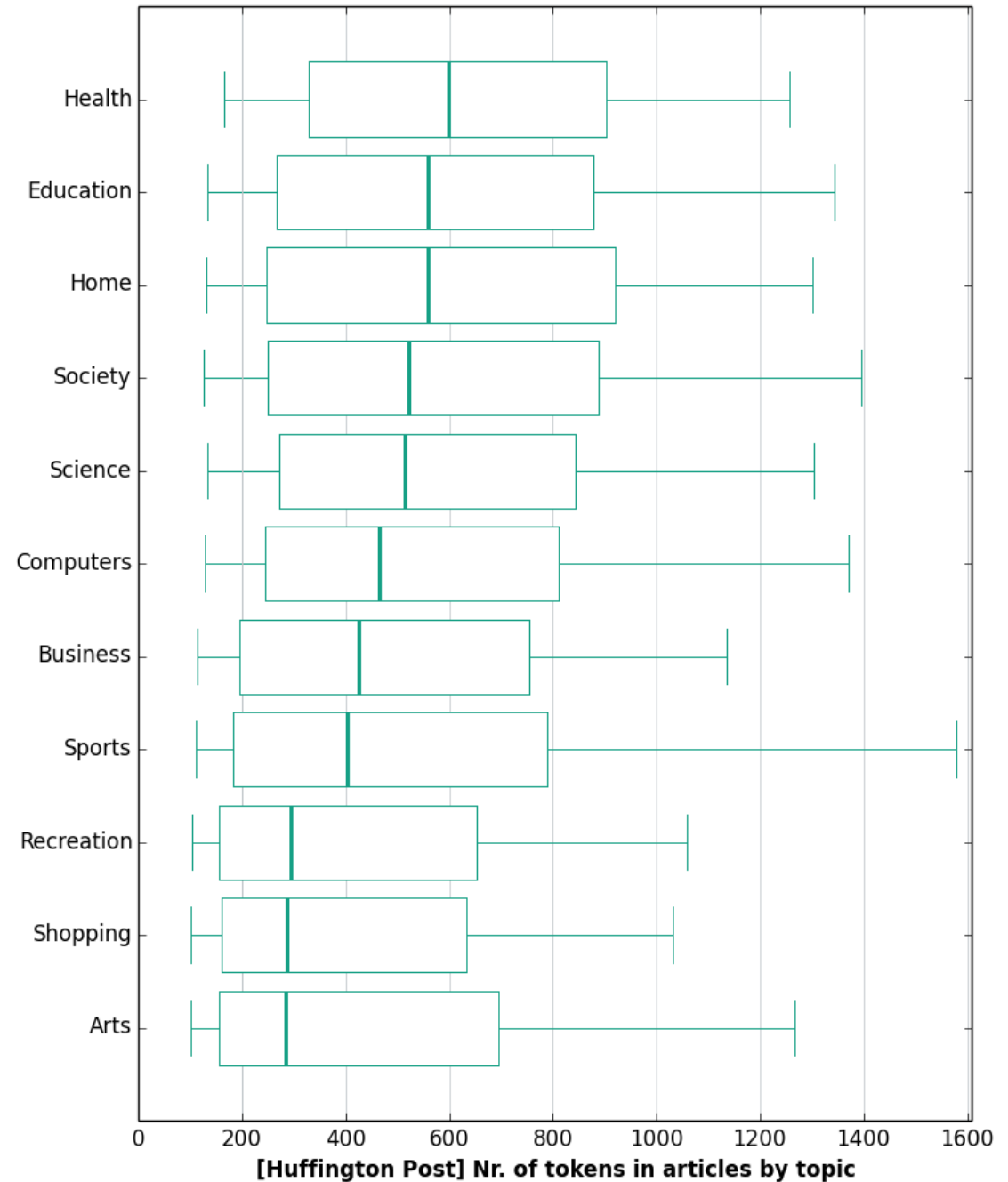
Dimensions of news bias to be observed

- Analysis of article length differences
- Analysis of grammatical differences
- Readability differences
- Newswire citations
- Geographical bias
- Topic (category) bias
- Speed of reporting
- Predicting article news source
- Similarity between news publishers

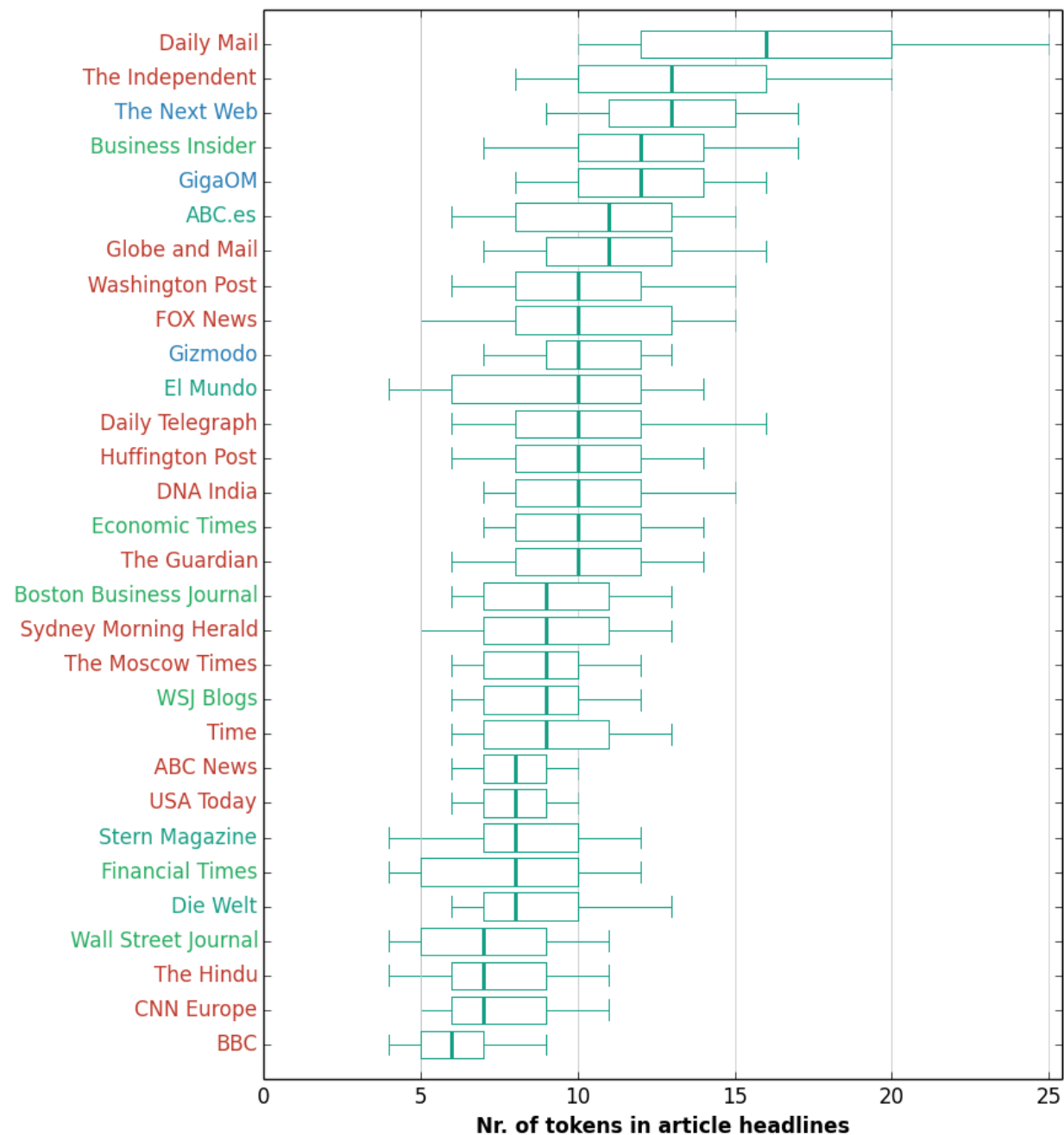
Article length differences



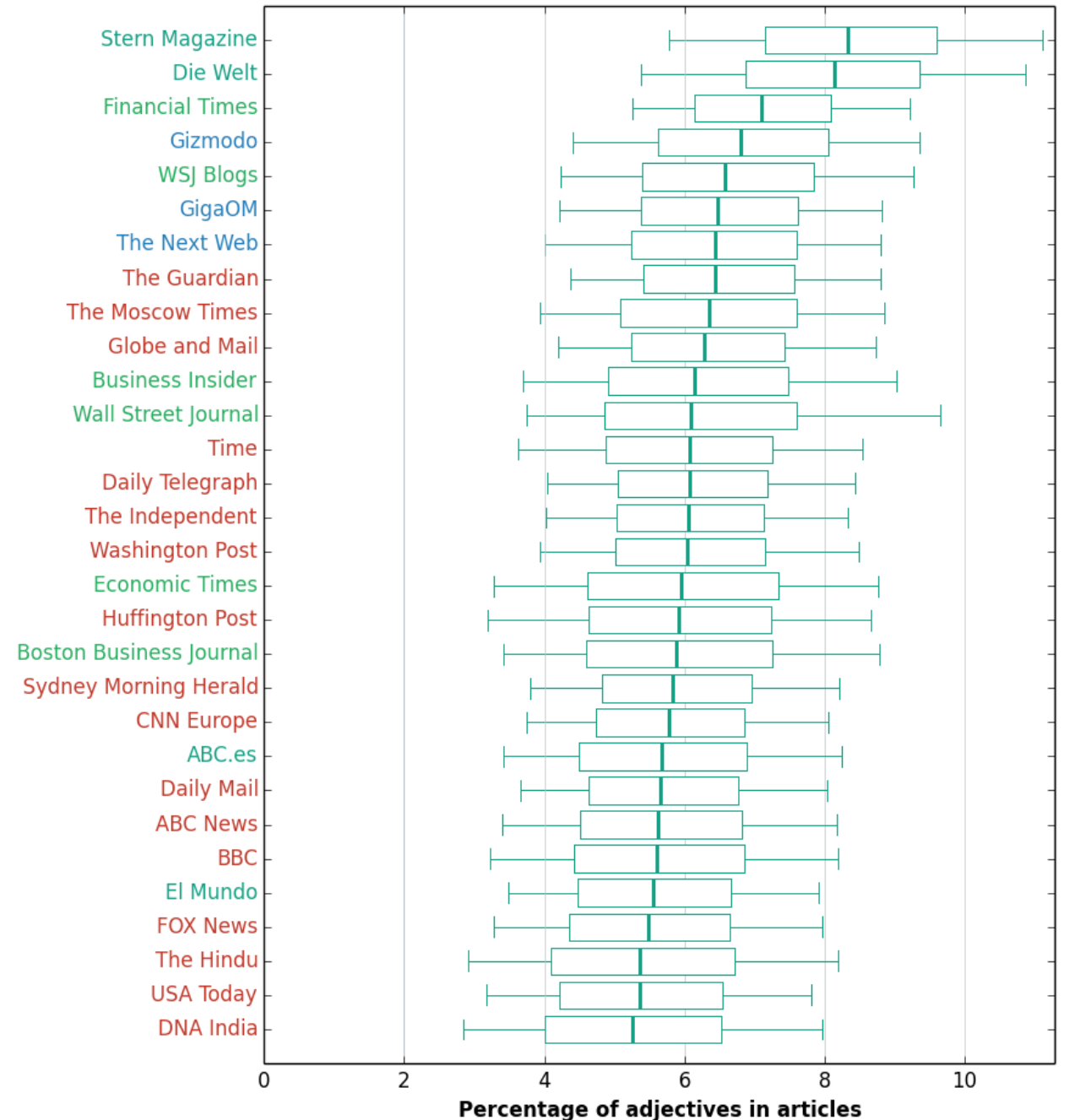
Number of tokens in articles per topic



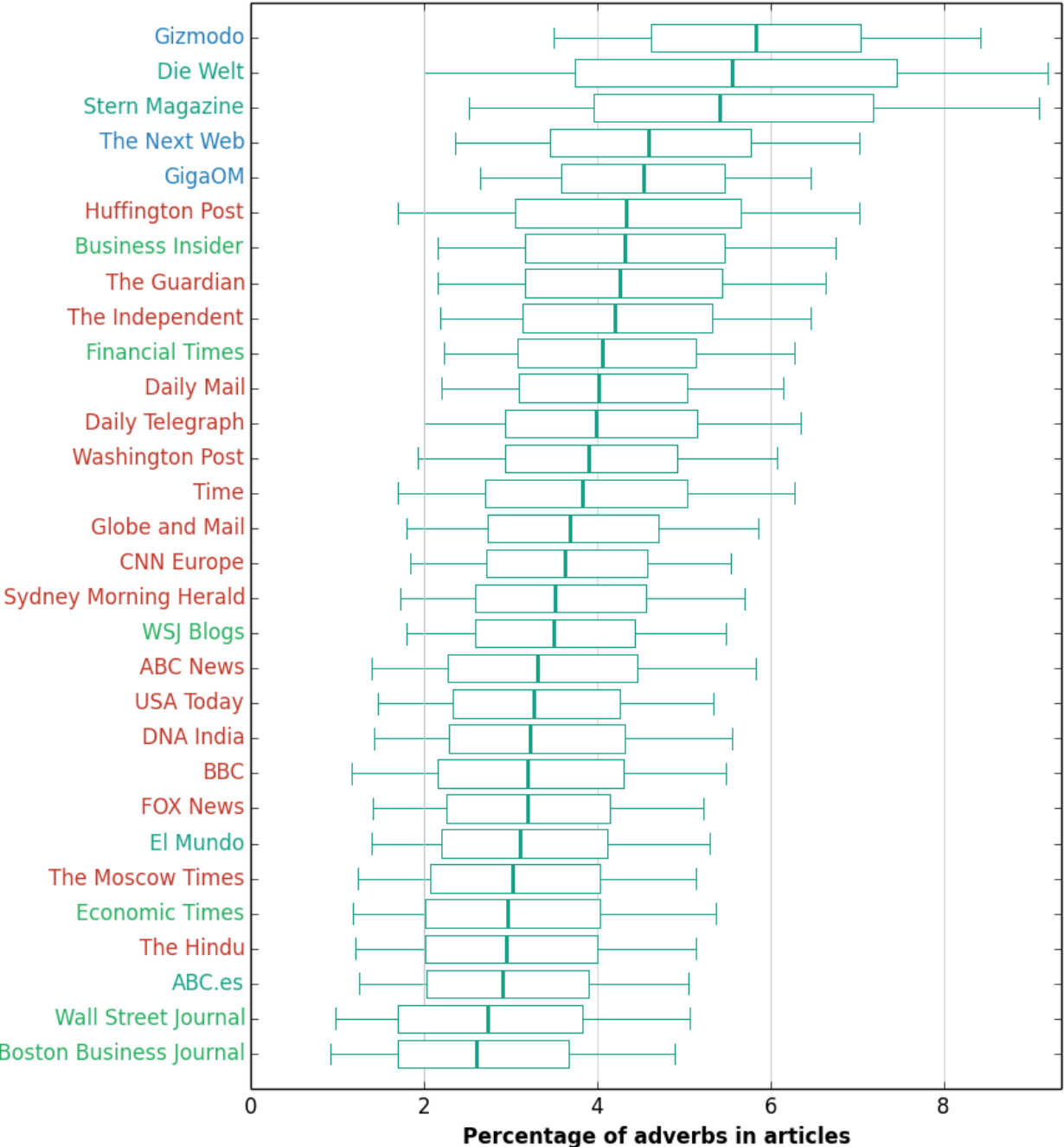
Number of tokens in article headlines



Grammatical differences (% of adjectives in articles)



Grammatical
differences
(% of adverbs in articles)



Readability differences

- **Dale-Chall readability formula.**

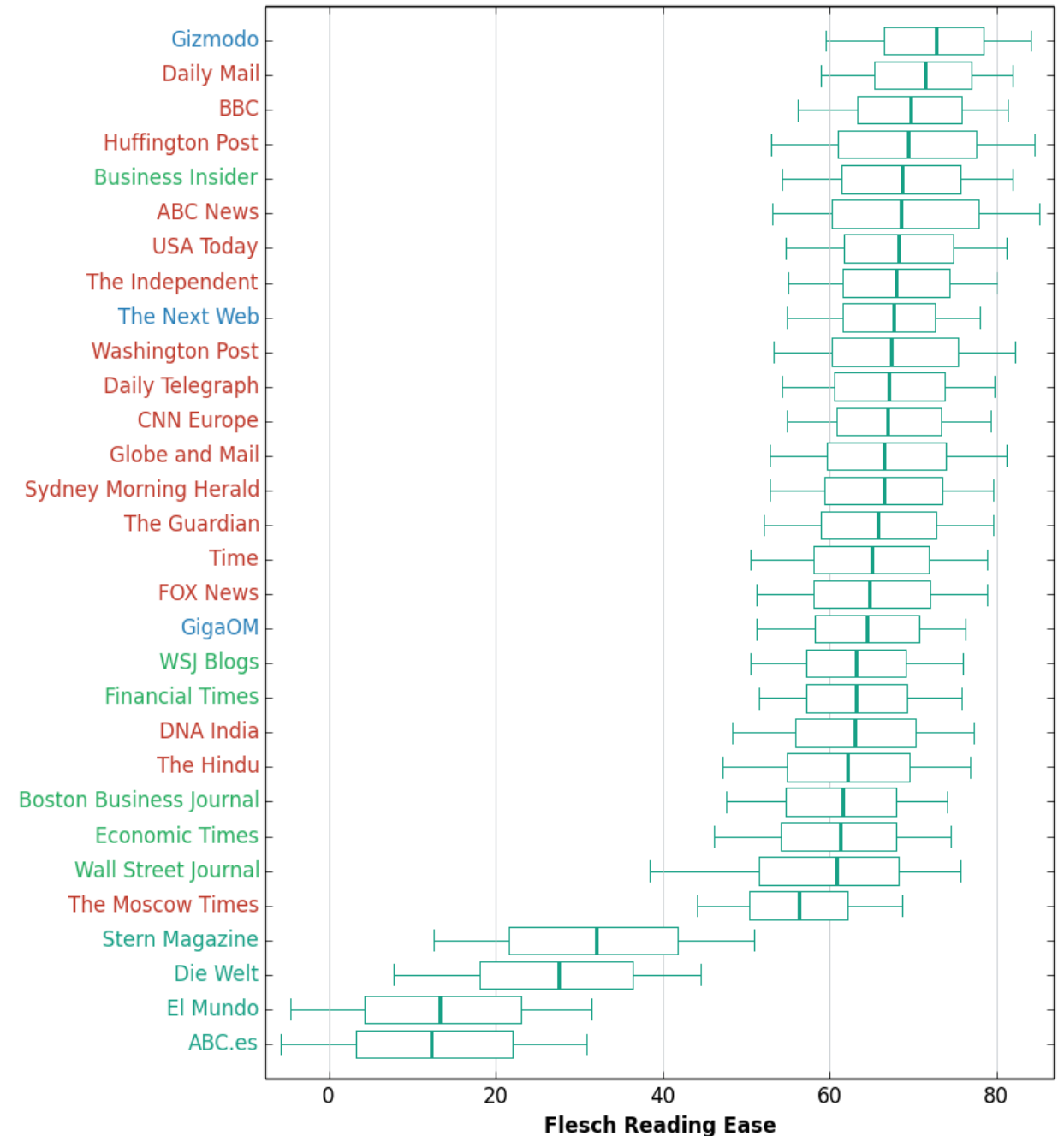
$$0,1579 \left(\frac{\text{difficult words}}{\text{words}} \times 100 \right) + 0,0496 \left(\frac{\text{words}}{\text{sentences}} \right)$$

- **Flesch-Kincaid Reading Ease**

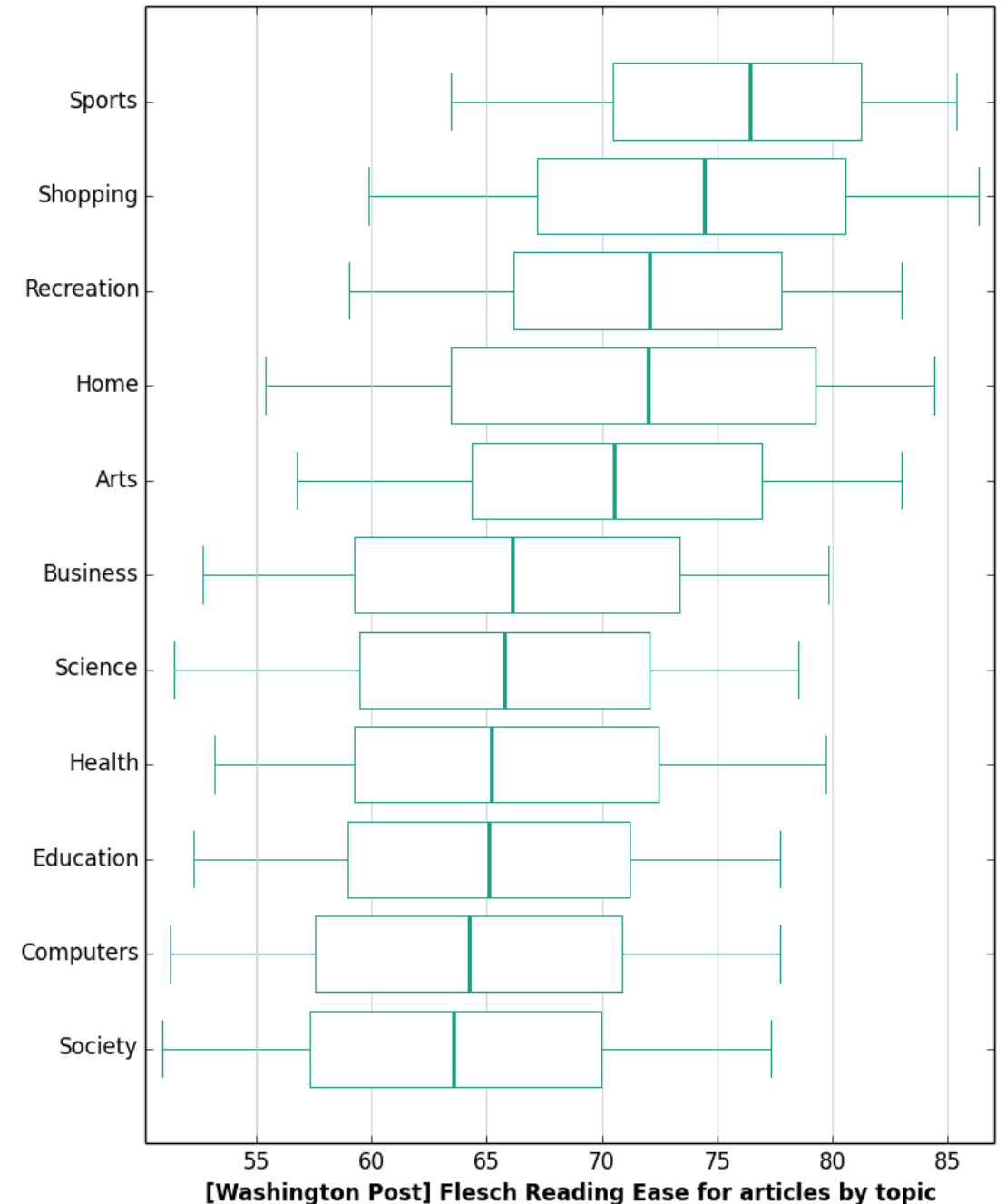
- This measure is a standard way for many US agencies for evaluating technical documents.

$$206,835 - 1,015 \left(\frac{\text{total words}}{\text{total sentences}} \right) - 84,6 \left(\frac{\text{total syllables}}{\text{total words}} \right)$$

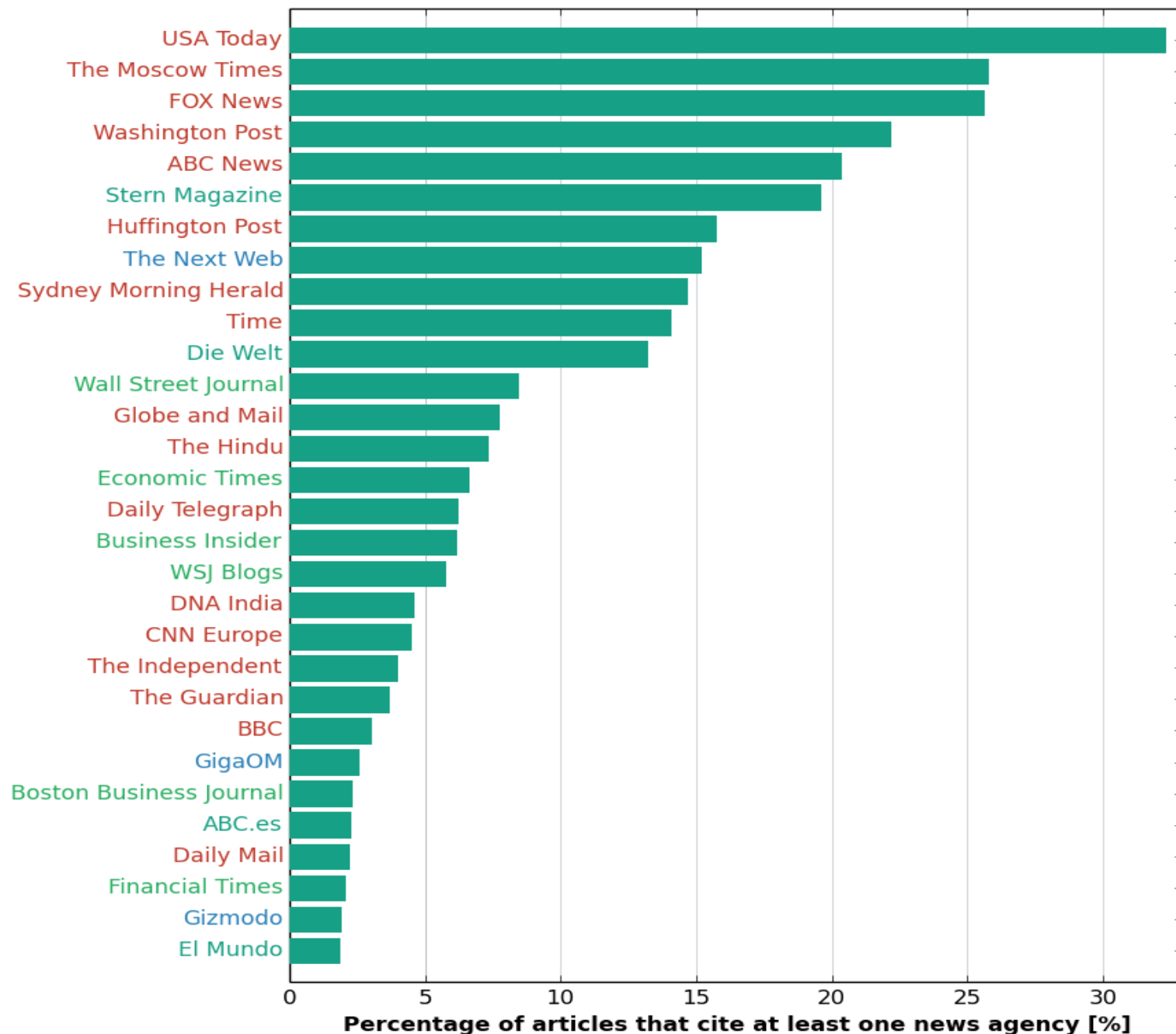
Readability in articles (Flesch-Kincaid test)



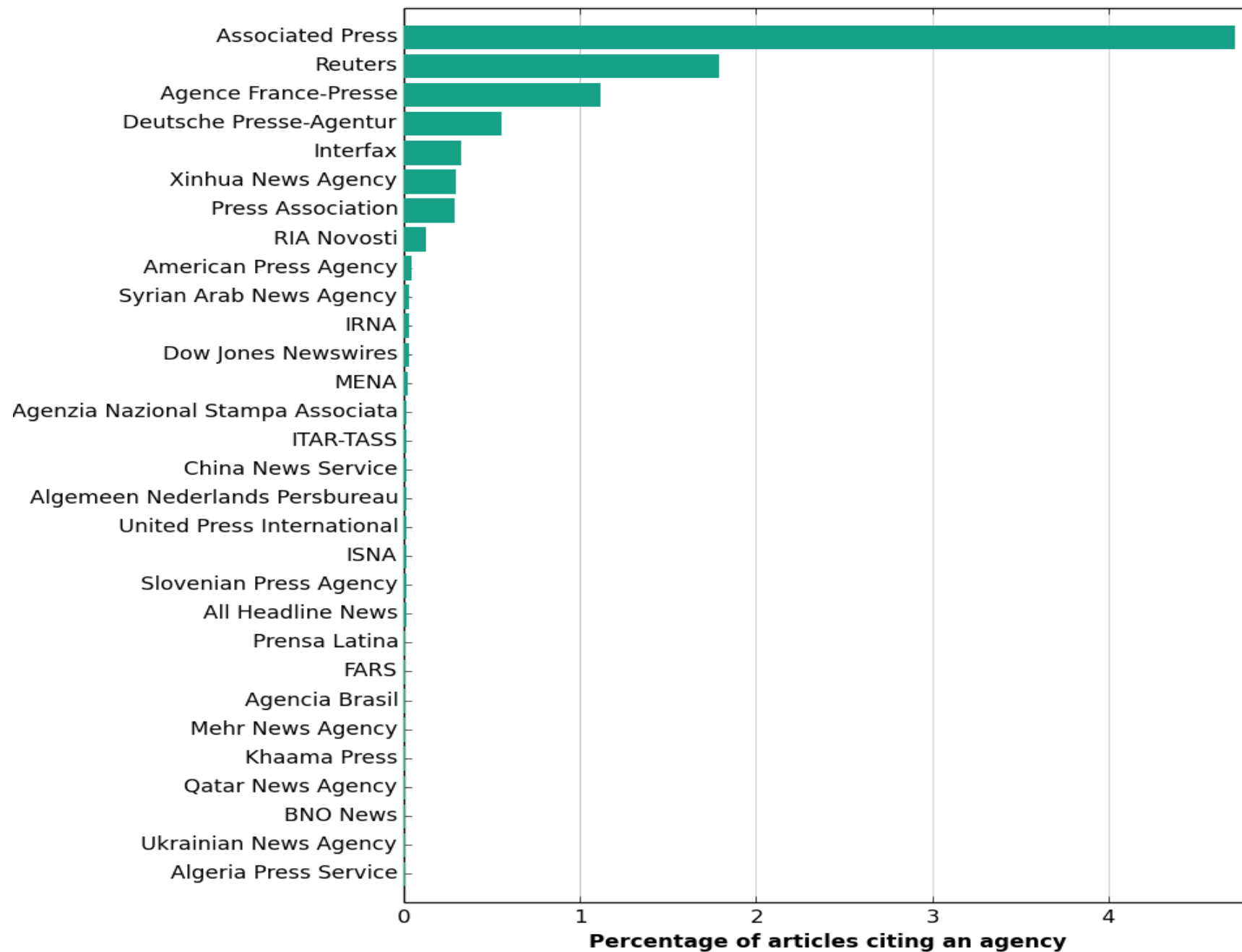
Readability in articles (Flesch-Kincaid test on Washington Post topics)



Publishers citing news agencies



Agencies Citations

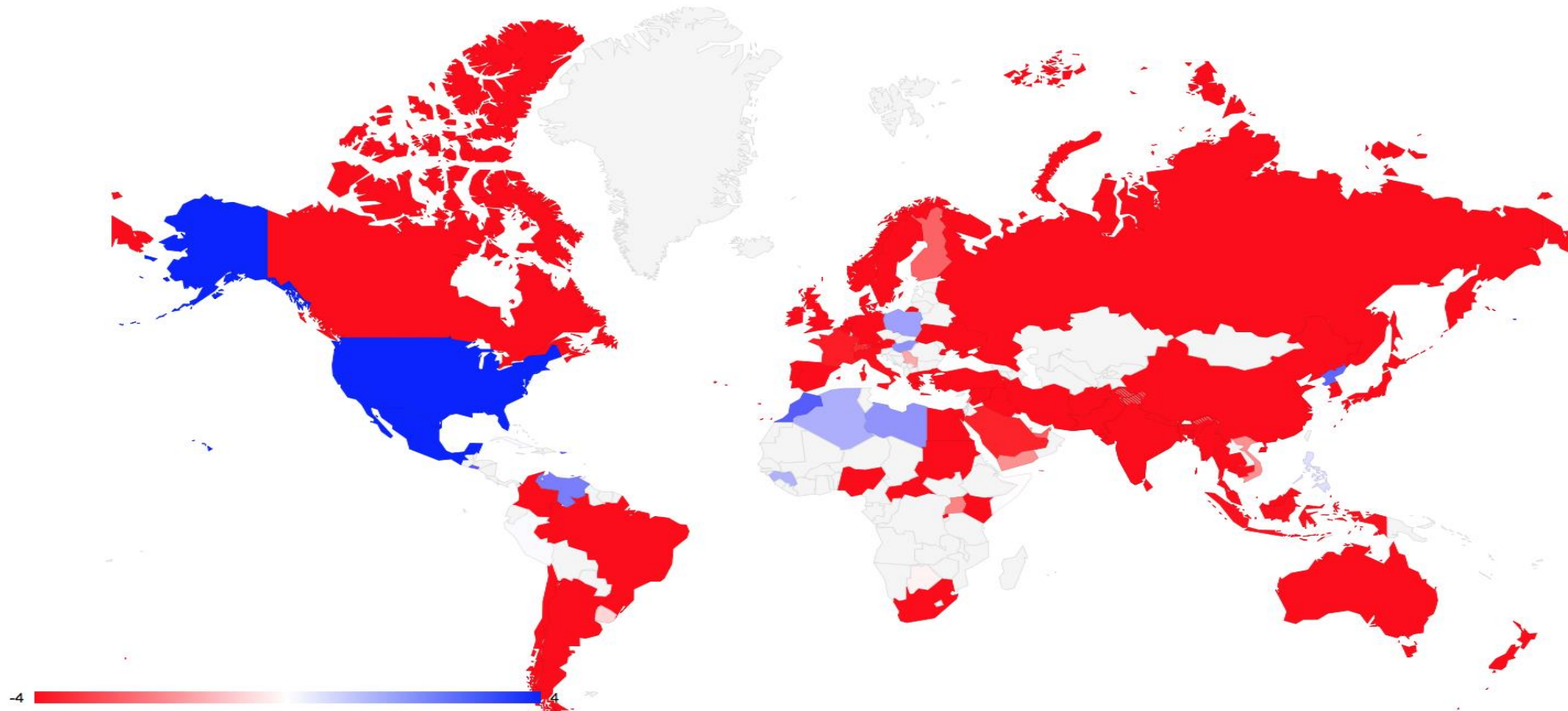


Geographical bias

- What events are being reported by news publishers?

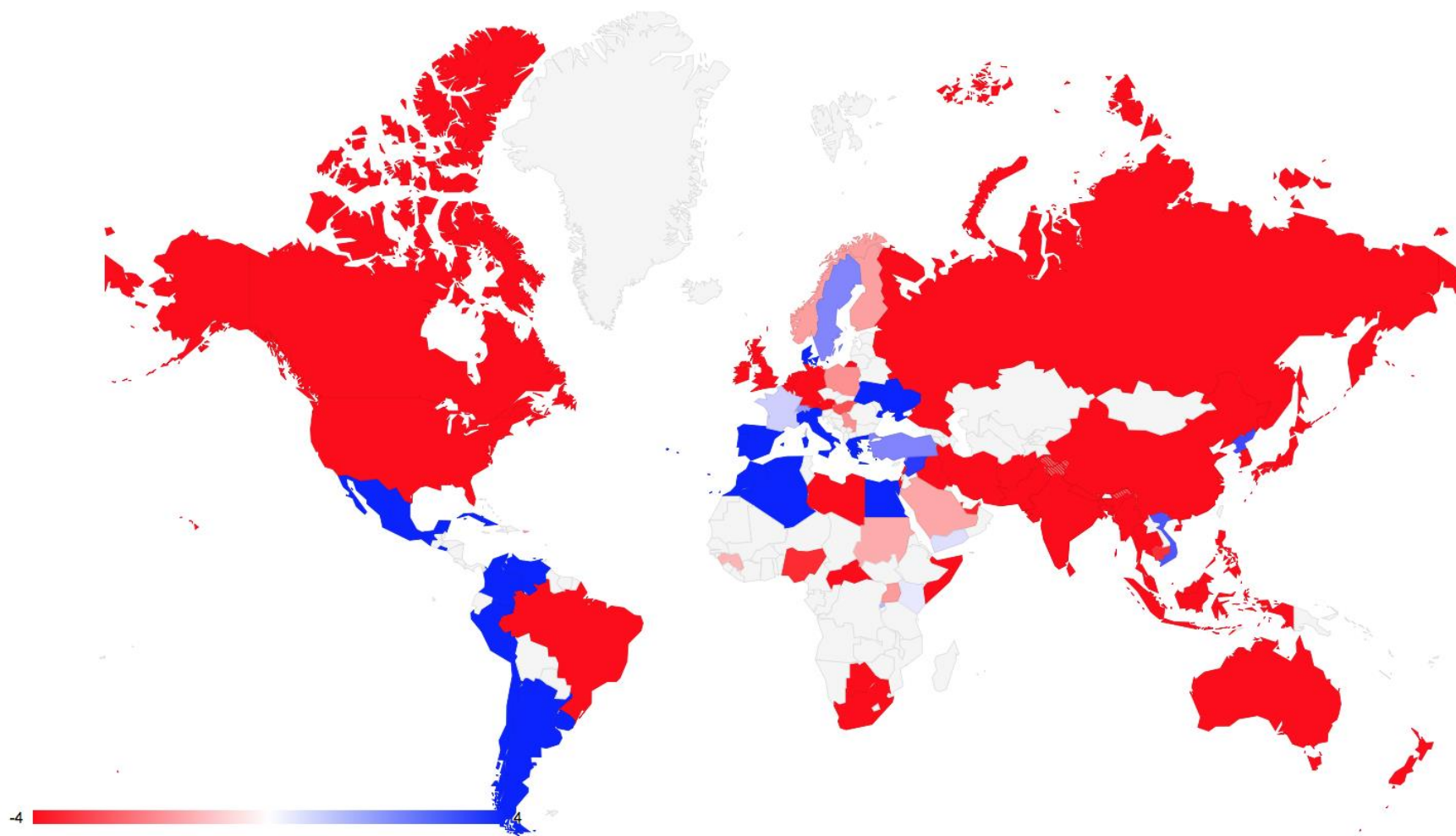
USA Today geographical coverage

USA Today



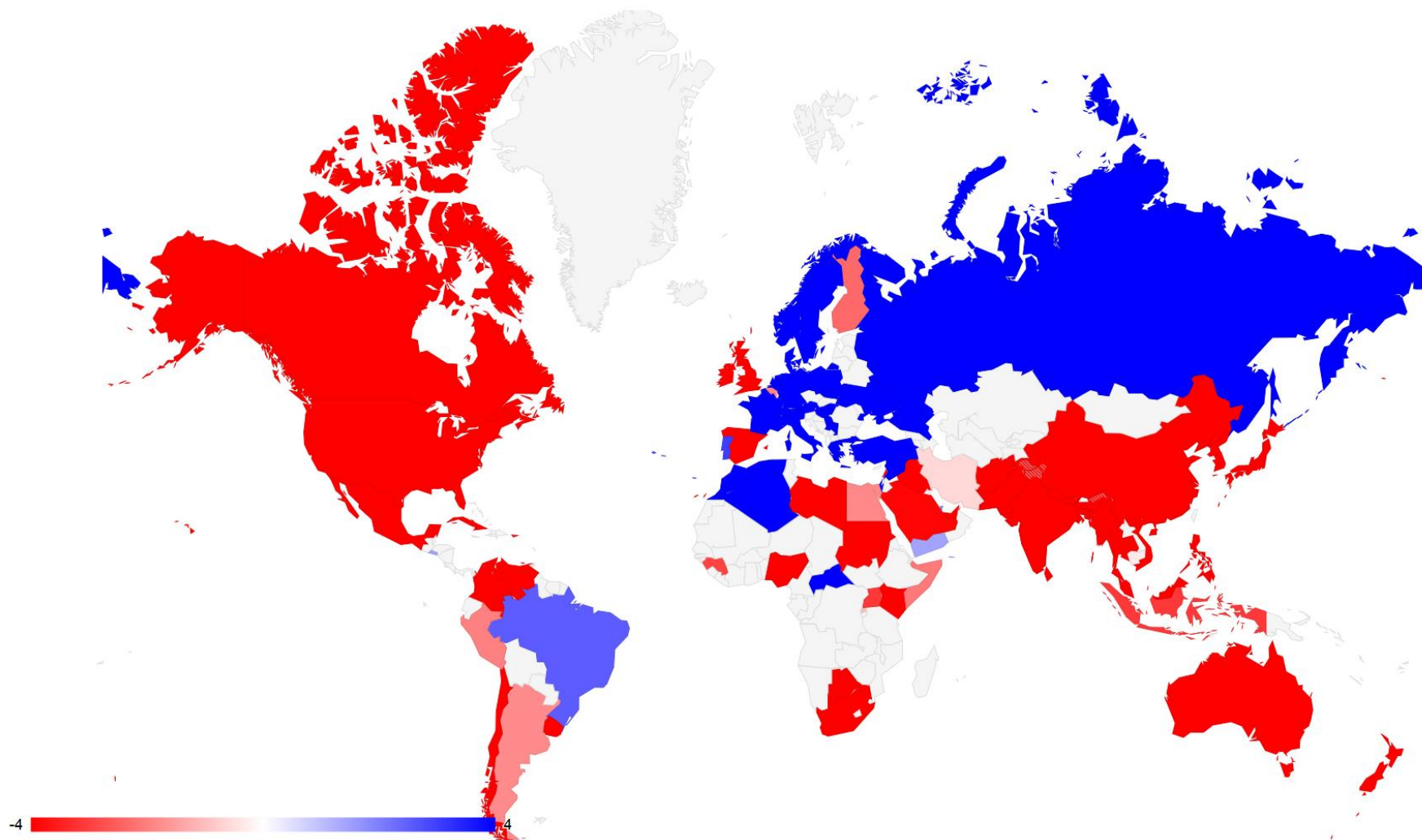
ABC.es geographical coverage

ABC.es



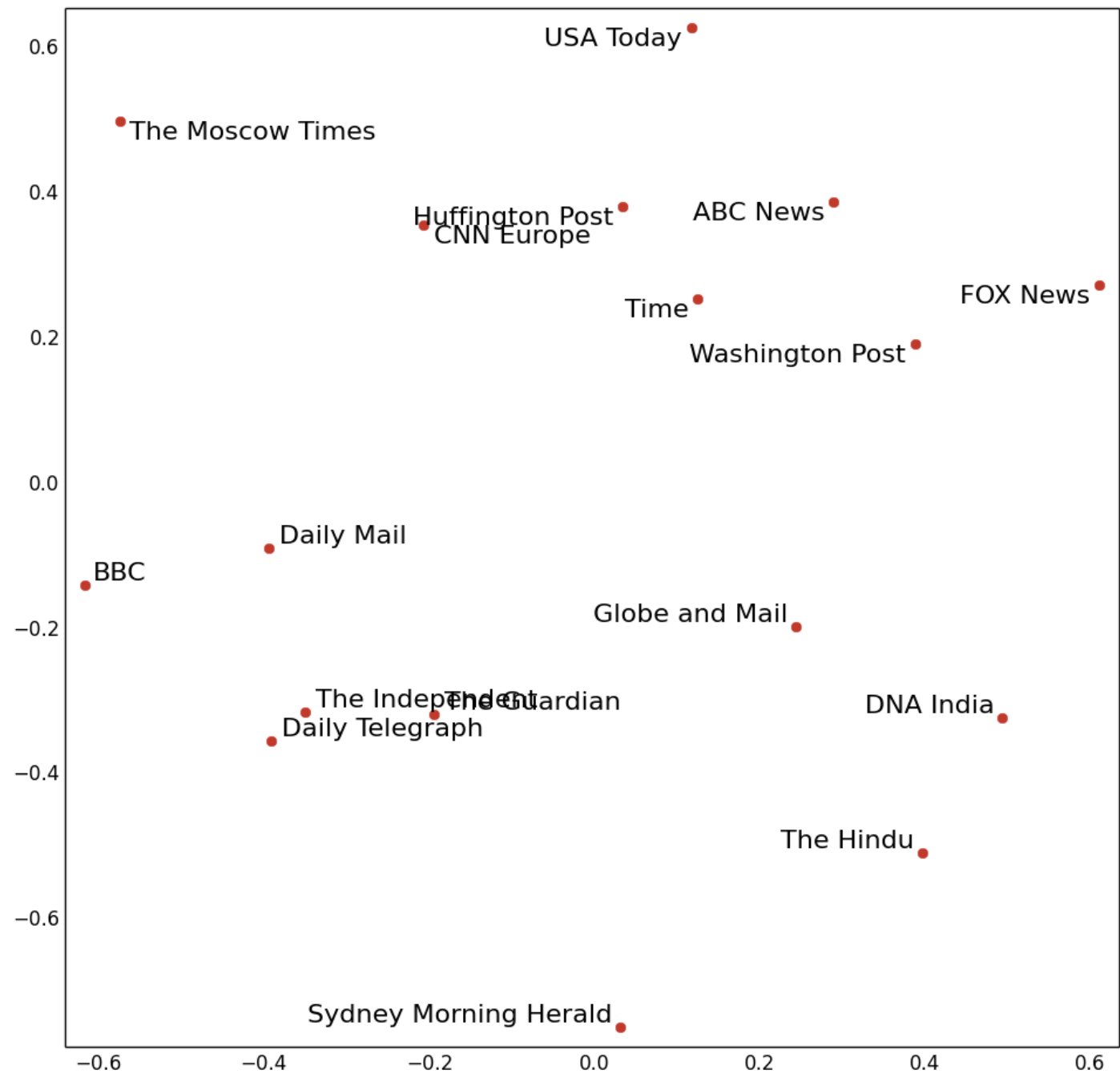
“Die Welt” geographical coverage

Die Welt



Separability between news sources

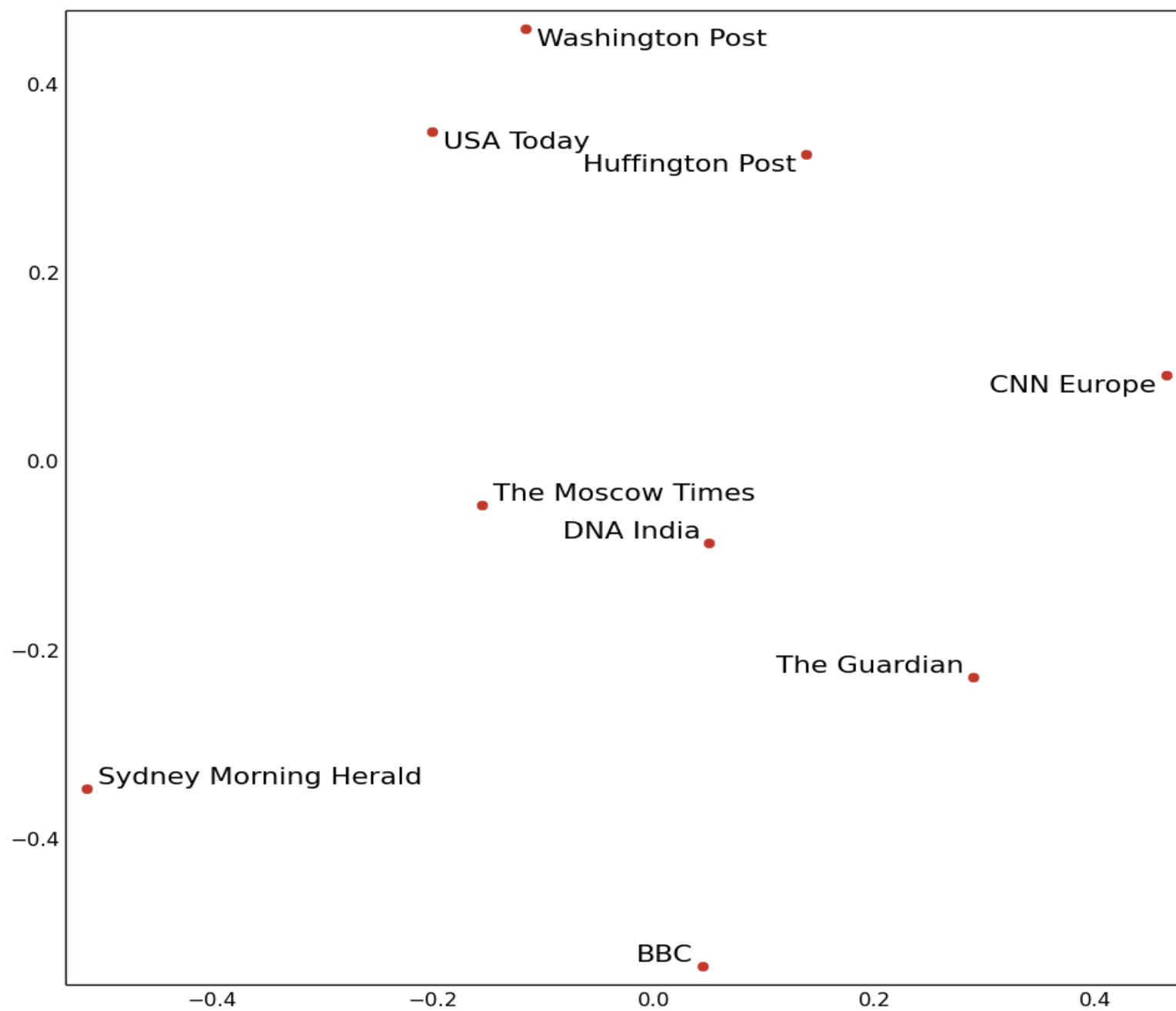
- How well we can determine which news publisher is the author of the article
- Tested on pairs of news publishers
- Used 10.000 articles from each news publisher
- Removed a custom-built set of tokens (like huffpost, mailonline, ...)
- Used keywords as features, used 1.000 top ranking features
- Linear SVM
- Used 70% articles for training, 30% for testing
- Use accuracy as a distance measure → use MDS for plotting



News source	Keywords
Time	selfie, uber, rapper, minister, researchers, Hillary, sex, party, relationships
Washington post	Maryland, redskins, Georgetown, Montgomery, reelection, capital, Arlington, officials, capitol
FOX News	Jerusalem, Christians, minister, dynasty, officials, prosecutors
Daily mail	Mr, uk, centre, behavior, football, Scotland, British, premier, Cameron
BBC	Correspondents, Belfast, council, Scotland, UK, Wales, Ireland
CNN	Editor, commentary, opinions, affiliate, according, London, official, Pakistan
Sydney Morning Herald	Australian, Australia, Melbourne, Fairfax, Canberra, rugby, Queensland, Brisbane,
Moscow times	Russian, Russia, Interfax, soviet, Barack, Petersburg, Siberia, Sergei, Kremlin, Duma, Yury, Oleg
DNA India	Mumbai, indian, singh, Bollywood, Kapoor, cricket, crore, Bangalore, lakh, Pune, Kumar, Tata, Gandhi

Separability on reports about Ukraine

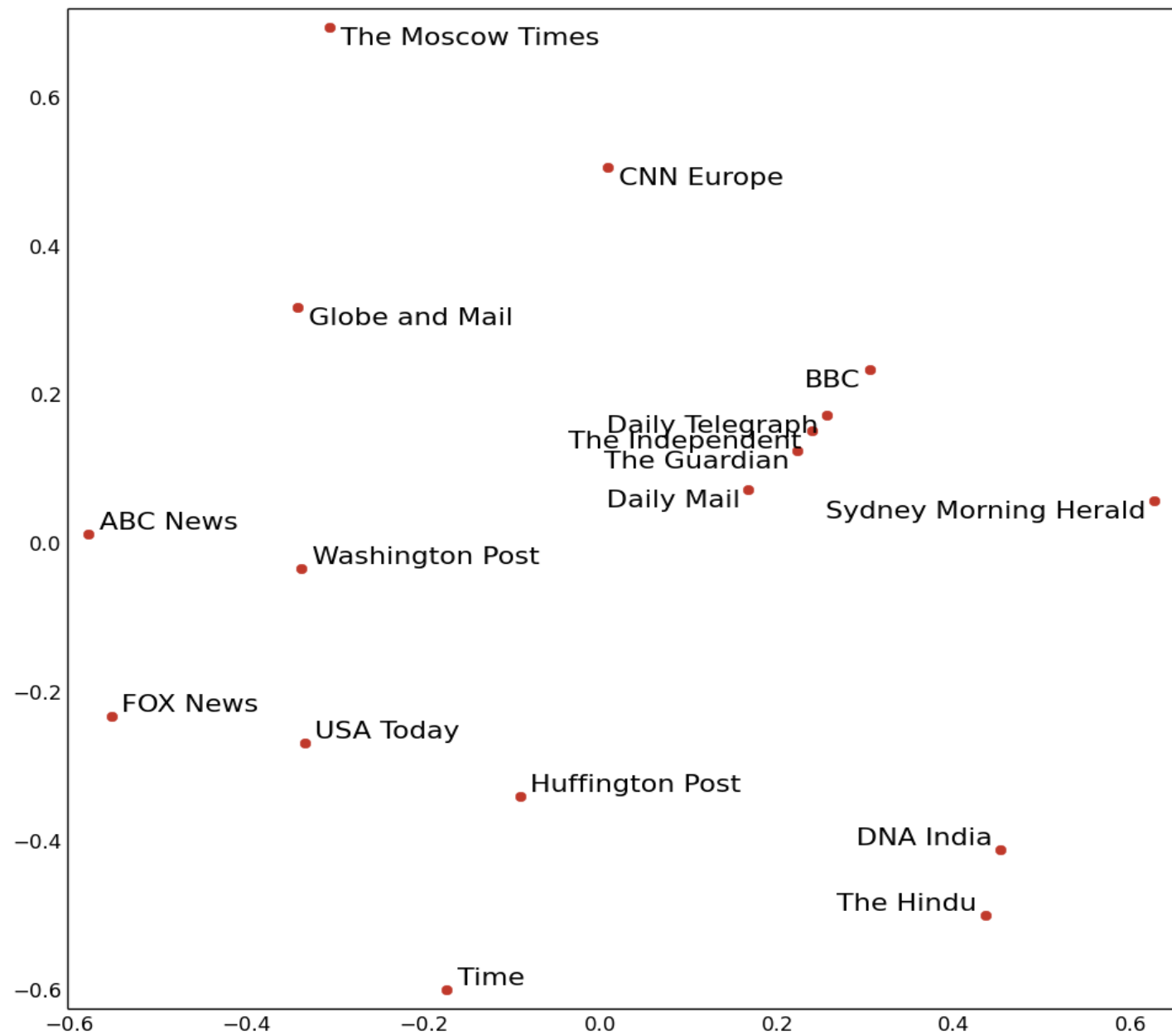
- Analyze articles reporting about Russia-Ukraine incident
- Selected articles
 - Mention Ukraine and Russia
 - between Jan 15th, 2014 and June 1st, 2014
 - Categorized into appropriate Dmoz category
- Selected publishers that reported about 500 articles



News source	Keywords
Non-Russian sources	troops, allies, invasion, aggression, excellency, insurgents, investigation, defense, ethnic, seized
The Moscow Times	overthrow, authorities, western, election, khodorkovsky, detained, annexation, pressure, seizure, fanning, imposed, domestic

Similarity between news publishers regarding the events they cover

- For each article we know the corresponding event in the Event Registry
- Compute a list of events covered by each news publisher
- Compute the pairwise similarity between the lists
- We used cosine similarity as the similarity measure



Event Representation

<http://eventregistry.org/>

What is an event?

(abstract description)

- ...more practical question: what definition of is computationally feasible?
- In general, an event is something which “sticks out” from the rest of information in some kind of (high dimensional) data space
 - ...could be interpreted as an “anomaly”
 - ...densification of data points (e.g. many similar documents)
 - ...significant change of distribution (e.g. a trend on Twitter)
- In practice, the event could be:
 - A cluster of documents / change of a distribution in data
 - Detected in an unsupervised way
 - A fit to a pre-built model
 - Detected in a supervised way

How to represent an event?

- Baseline data for a news event is usually a cluster of documents
 - ...with some preprocessing we extract linguistic and semantic annotations
 - ...semantic annotations are linked to ontologies providing possibility for multiresolution annotations
- Three levels of event representation:
 - **Feature vector event representation:**
 - ...light weight representation that can be easily represented as a set of feature vectors augmented with external ontologies – suitable for scalable ML analysis
 - **Structured event representation:**
 - Infobox representation (slots filling) using open schema or event taxonomy
 - **Deep event representation**
 - Semantic representation linked to a world-model (e.g. CycKB common sense knowledge) – suitable for reasoning and diagnostics

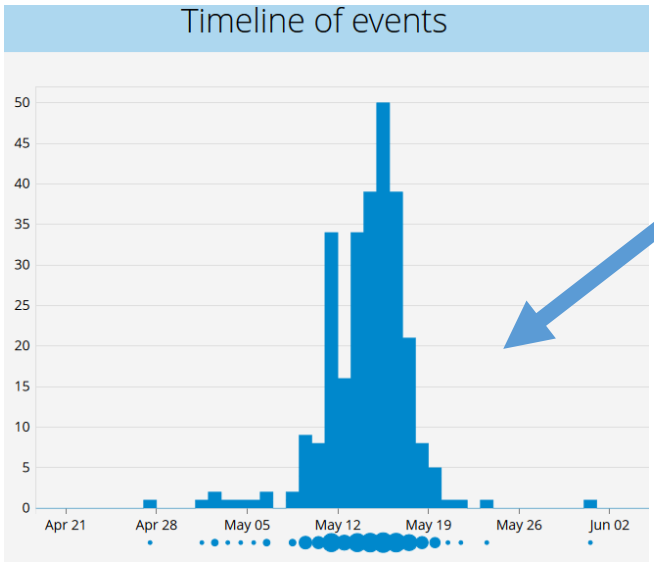
Feature vector event representation

- Feature vectors easily extractable from news documents:
 - **Topical dimension** – what is being talked about? (keywords)
 - **Social dimension** – which entities are mentioned? (named entities)
 - **Temporal aspect** – what is the time of an event? (temporal distribution)
 - **Geographical aspect** – where an event is taking place? (location)
 - **Publisher aspect** – who is reporting? (publisher identifiers)
 - **Sentiment/bias aspect** – emotional signals (numeric estimates)
- Scalable Machine Learning techniques can easily deal with such representation
 - ...in “Event Registry” system we use this representation to describe events

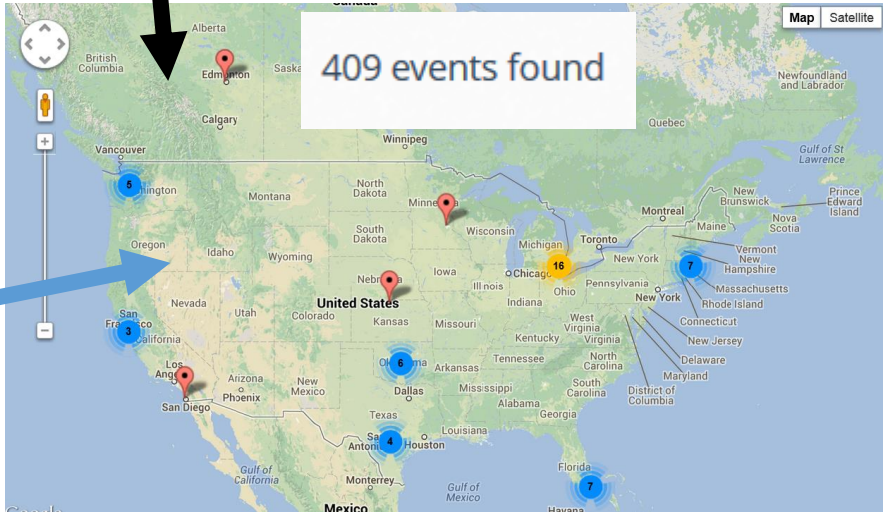
Example of “feature vector” event representation: Event Registry “Chicago” related events

Query:
“Chicago”

The screenshot shows the Event Registry search interface. The search bar contains the text "Chicago". Below the search bar, there are filters for "Event location" (city or country name), "Time of interest" (start date, end date, and a dropdown for "Any category"), and "Min. event coverage" (0 articles). The interface also has tabs for "Geographic locations" and "Timeline view". The "Geographic locations" tab is selected, showing a map of the United States with a red pin on Chicago. The "Timeline view" tab shows a bar chart of event frequency over time.



When?
(temporal distribution)



Where?
(geography)


The screenshot shows the Event Registry event viewer for the event "Emanuel says he'll seek 2nd term as Chicago mayor". The event details include the date "When: 8 May 2013", the number of articles "Nr. of articles: 9 (9 eng, 0 deu, 0 spa, 0 zho, 0 slv)", and a brief description of the event. Below the event details, there are two bar charts: "Event is about..." (Entities) and "Event is about..." (Keywords). The "Event is about..." (Entities) chart shows the frequency of various entities associated with the event, with "Rahm Emanuel" being the most frequent at 100. The "Event is about..." (Keywords) chart shows the frequency of various keywords associated with the event, with "School" being the most frequent at 35.

Who?
(named entities)

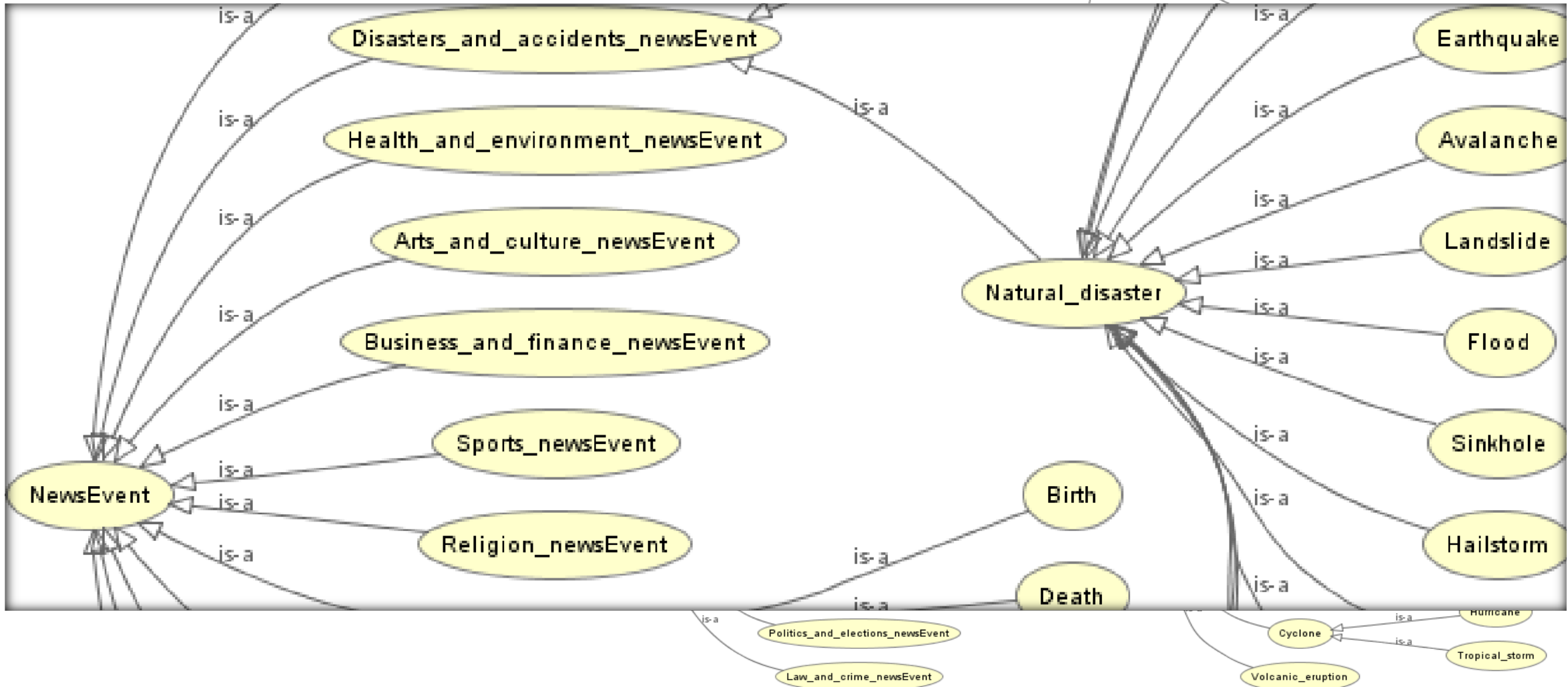
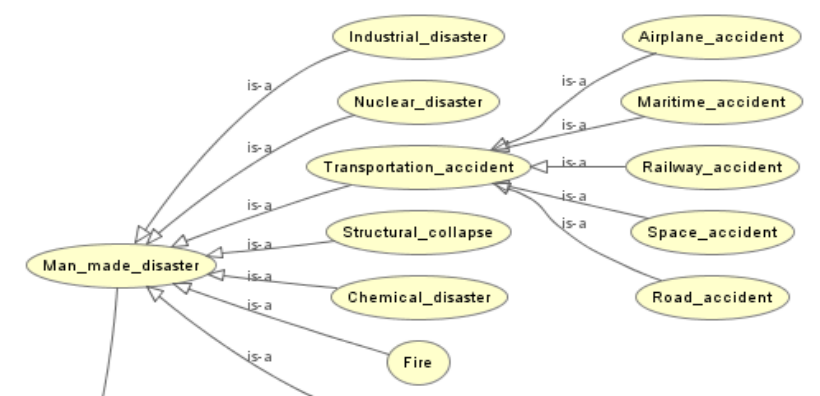
What?
(keyword/ topics)

Structured event representation

- Structured event representation describes an event by its “Event Type” and corresponding information slots to be filled
- Event Types should be taken from “Event Taxonomy”
- ...at this stage of development this level of representation still requires human intervention to achieve high accuracy (Precision/Recall) extraction
- Example on the right – Wikipedia event infobox:
 - **2011 Tōhoku earthquake and tsunami**

Date	11 March 2011 (3 years, 5 months and 2 weeks)
Origin time	14:46:24 JST (UTC+09:00)
Duration	6 minutes ^[1]
Magnitude	9.0 (M_w) ^{[2][3]}
Depth	30 km (19 mi)
Epicenter	 38.322°N 142.369°E
Type	Megathrust earthquake
Areas affected	Japan (primary) Pacific Rim (tsunami, secondary)
Total damage	Tsunami wave, flooding, landslides, fires, building and infrastructure damage, nuclear incidents including radiation releases
Max. intensity	IX - Violent
Peak acceleration	2.99 g
Tsunami	Up to 40.5 m (133 ft) in Miyako , Iwate , Tōhoku
Landslides	None
Foreshocks	7
Aftershocks	11,044 (as of 3 August 2014) ^[4]
Casualties	15,889 deaths, ^[5] 6,152 injured, ^[6] 2,609 people missing ^[7]

“Event Taxonomy” – preview to the current development



Prototype for event Infobox extraction: XLike annotation service

- The goal is to build a system for economically viable extraction of event infoboxes
 - ...using crowd-sourcing
 - ...aiming at high Precision & Recall for a small cost

The screenshot displays the 'Annotation interface' for the XLike event annotation crowdsourcing prototype. The browser address bar shows the URL 'aidemo.ijs.si/eventAnnotation?story_id=213'. The page title is 'XLike event annotation crowdsourcing prototype'. Below the title, it indicates 'Story id: 213 (25 articles)' and 'Logged in as: marko'. The main content area features a news article titled 'Bombing at Egypt police station kills 12'. The article text is annotated with blue boxes highlighting various entities and events. To the right of the article, there is a table titled 'Annotations' with columns for '#', 'entity name', and 'role'. The table lists 393 entities found, including 'Brotherhood', 'attack', 'group', 'government', 'Egypt', 'bombing', 'Mansoura', 'Morsi', 'violence', 'blast', 'Muslim Brotherhood', 'attacks', 'military', 'country', 'people', 'Cairo', 'city', 'Tuesday', 'security forces', 'building', 'groups', and 'security'. To the right of the table, there is a section titled 'Event type' with the event 'bombing' and a description 'An attack with an explosive.' Below this, there is a section titled 'Roles' with a list of suggested roles: VICTIM - Who was killed?, INJURED - Who was injured?, TARGET - Who was the bomb intended for?, ATTACKER - Which organization is suspected / admitted responsibility?, LOCATION - Where did the event happen?, EXPLOSIVES - What kind of explosives were used in attacks?, TIME - When did the attacks happen?, REPORTER - Who reported the bombings?, DAMAGE - What was damaged?, COMMENTATOR - Who commented on the event?, SECURITY - Who took care of security?, NUMBER KILLED - How many people were killed?, and NUMBER INJURED - How many people were injured?.

Annotation interface

aidemo.ijs.si/eventAnnotation?story_id=213

Most Visited Calls - Research Partici... Twitter Search World Cup 2014 Predi... Where to Learn Deep ... XLike -- Bloomberg Lo... Wikipedia based cross... XLIKE - Demonstration

XLike event annotation crowdsourcing prototype

Story id: 213 (25 articles) - Go to story: ...

Logged in as: marko (switch user)

Bombing at Egypt police station kills 12

Egypt PM calls bombing 'worst kind' of terrorism (807439)

MANSOURA, Egypt (AP) - Egypt's interim Prime Minister says the bombing of a police headquarters outside of Cairo that left 13 killed is the "worst kind of terrorism" against the state, vowing his government will face it "decisively" and legally.

Hazem el-Beblawi said Tuesday the bombing in Mansoura was part of a series of violations against Egyptians security, linking such attacks on security forces with the campaign of protests by supporters of ousted Islamist President Mohammed Morsi.

He called the attack a "maximum offense" to Egypt and will be dealt with decisively and according to the law, without elaborating.

El-Beblawi said his government has been working to implement a court order banning the Muslim Brotherhood, to which Morsi belongs. Pressure is mounting on el-Beblawi to declare the group a terrorist organization.

A powerful blast tore through a police headquarters in an Egyptian Nile Delta city early Tuesday, killing 13 people, wounding more than 100 and leaving victims buried under rubble in the deadliest bombing yet in a months-long wave of violence blamed on Islamic militants.

Investigators were trying to determine whether the blast, soon after midnight in the city of Mansoura, was from a car bomb or from explosives planted around the building. The explosion left a downtown street of the city strewn with piles of debris and charred cars.

Egypt has seen an escalating campaign of spectacular bombings and gun attacks, mainly against security forces, since the military ousted Islamist President Mohammed Morsi in July and launched a fierce crackdown on his Muslim Brotherhood. Most have been centered in the Sinai Peninsula, where multiple militant groups operate, but the insurgency has been spreading to the heavily populated Delta and the capital, Cairo.

Annotations

393 entities found

#	entity name	role
91	Brotherhood	
79	attack	
78	group	
75	government	
64	Egypt	
61	bombing	
57	Mansoura	
54	Morsi	
53	violence	
48	blast	
41	Muslim Brotherhood	
40	attacks	
39	military	
39	country	
36	people	
35	Cairo	
34	city	
34	Tuesday	
33	security forces	
32	building	
29	groups	
29	security	

Event type

bombing
An attack with an explosive.

change

Roles

Suggested roles:

VICTIM - Who was killed?
INJURED - Who was injured?
TARGET - Who was the bomb intended for?
ATTACKER - Which organization is suspected / admitted responsibility?
LOCATION - Where did the event happen?
EXPLOSIVES - What kind of explosives were used in attacks?
TIME - When did the attacks happen?
REPORTER - Who reported the bombings?
DAMAGE - What was damaged?
COMMENTATOR - Who commented on the event?
SECURITY - Who took care of security?
NUMBER KILLED - How many people were killed?
NUMBER INJURED - How many people were injured?

Event sequences & Hierarchical events

- Once having events identified and represented we can connect events into “**event sequences**” (also called story-lines)
- “Event sequences” include events which are supposedly related and constitute a larger story
- Collection of interrelated events can be also organized in **hierarchies** (e.g. World Cup event consists from a series of smaller events)

An example event: Microsoft Windows 9



Windows 9 release date, price, features: why Microsoft can't wait to launch Windows 9 in 2015

WHEN: April 23, 2014

CATEGORIES: Shopping → Home and Garden → Windows , Home → Home Improvement → Windows and Doors , Computers → Software → Freeware

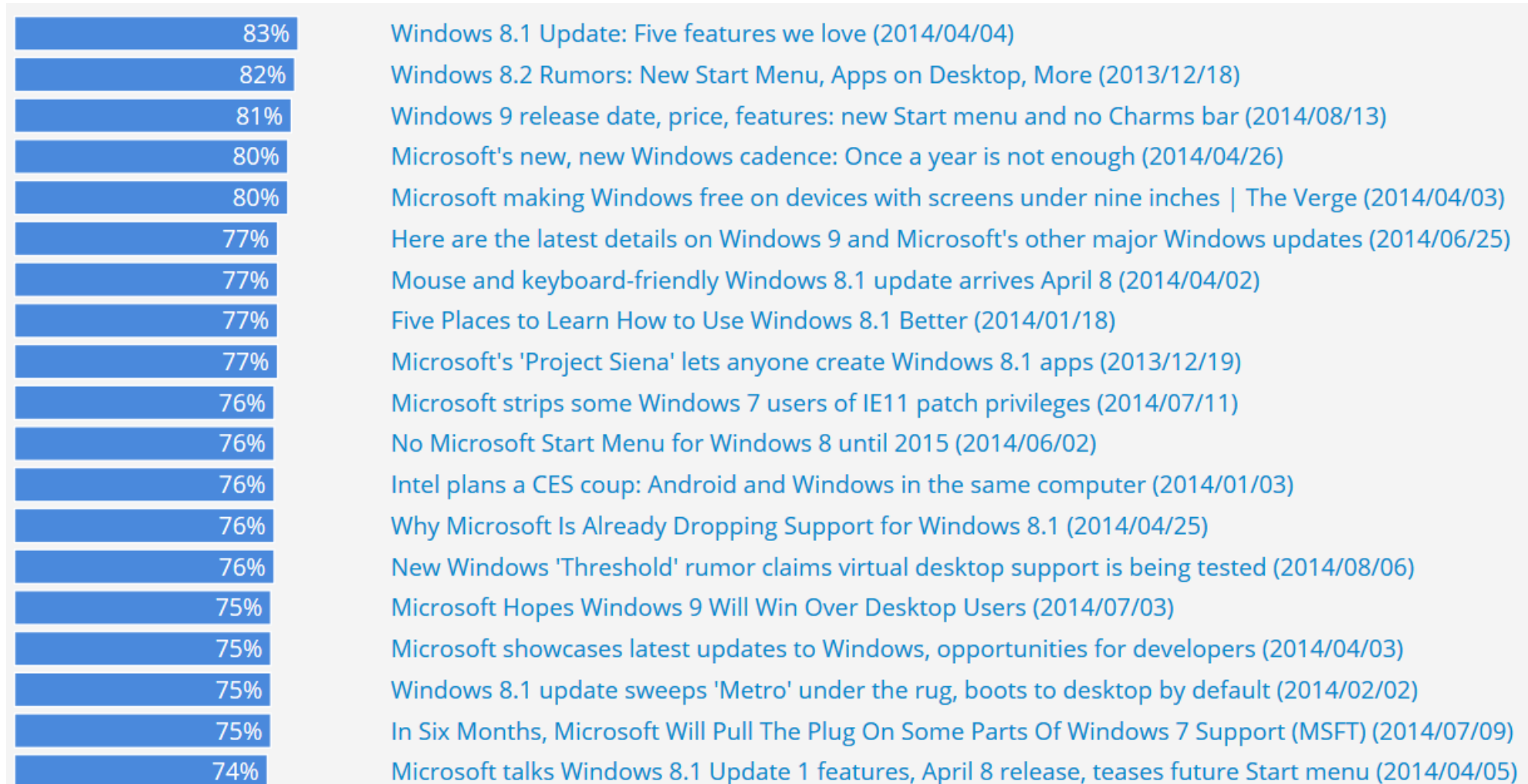
ENTITIES: Microsoft Twitter Google Apple Inc. Computerworld Gartner Facebook Universal Studios Nokia United States

KEYWORDS: Microsoft Windows History of Microsoft Windows Patch (computing) Software release life cycle Operating system
Desktop computer Electronics Company NeXT Software build

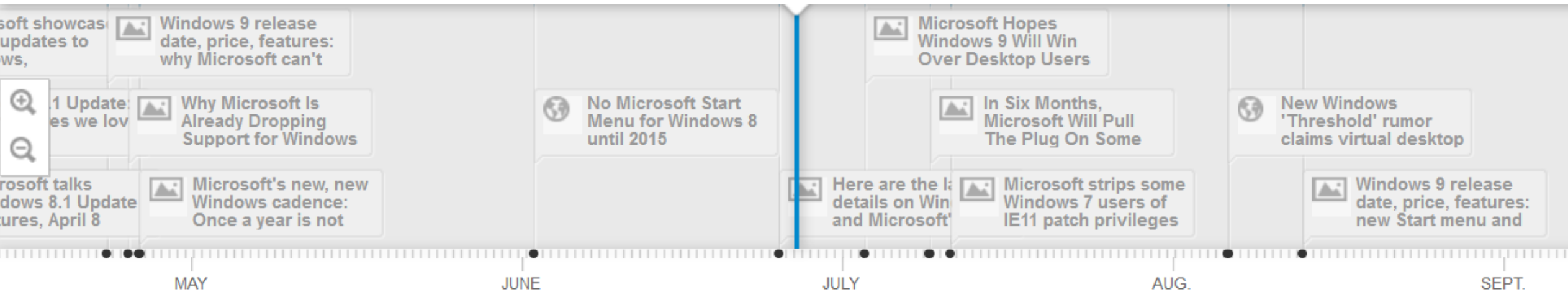
ARTICLES: 239 (239 eng, 0 ger, 0 spa, 0 chi, 0 slo)

We outline what we know about Windows 9 release date, Windows 9 price and Windows 9 features. It's fair to say that Windows 8 has not been a universal success. In many ways a necessary step not terribly well executed, Windows 8's attempt to stitch together desktop and mobile has been way too radical to encourage new users. And as the longer life of existing PCs and laptops combines with the growth of smartphones and tablets to continually slow PC sales, Window 8 has come to be seen as a dud....

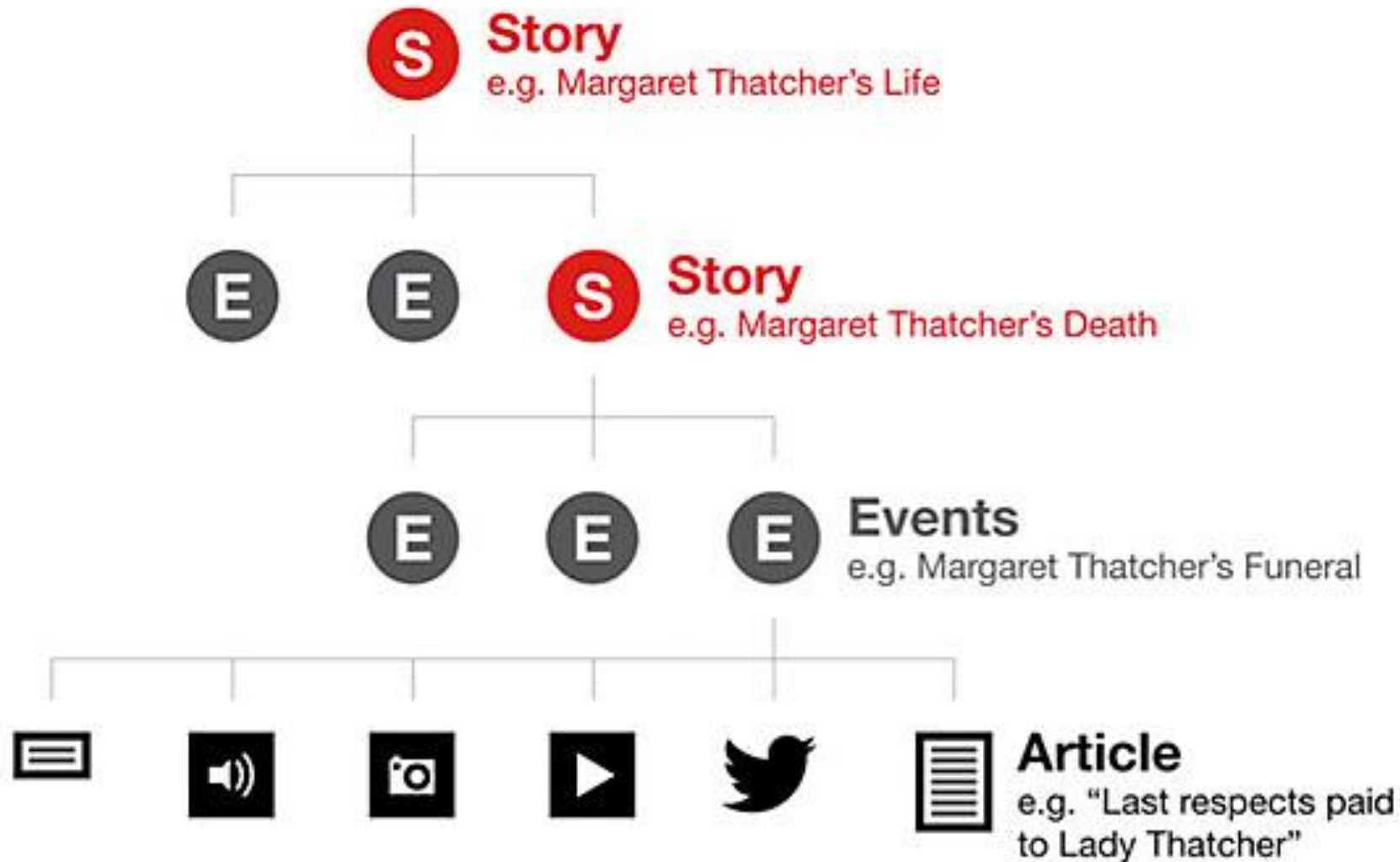
Similar events example: similar events to Microsoft Windows 9 event



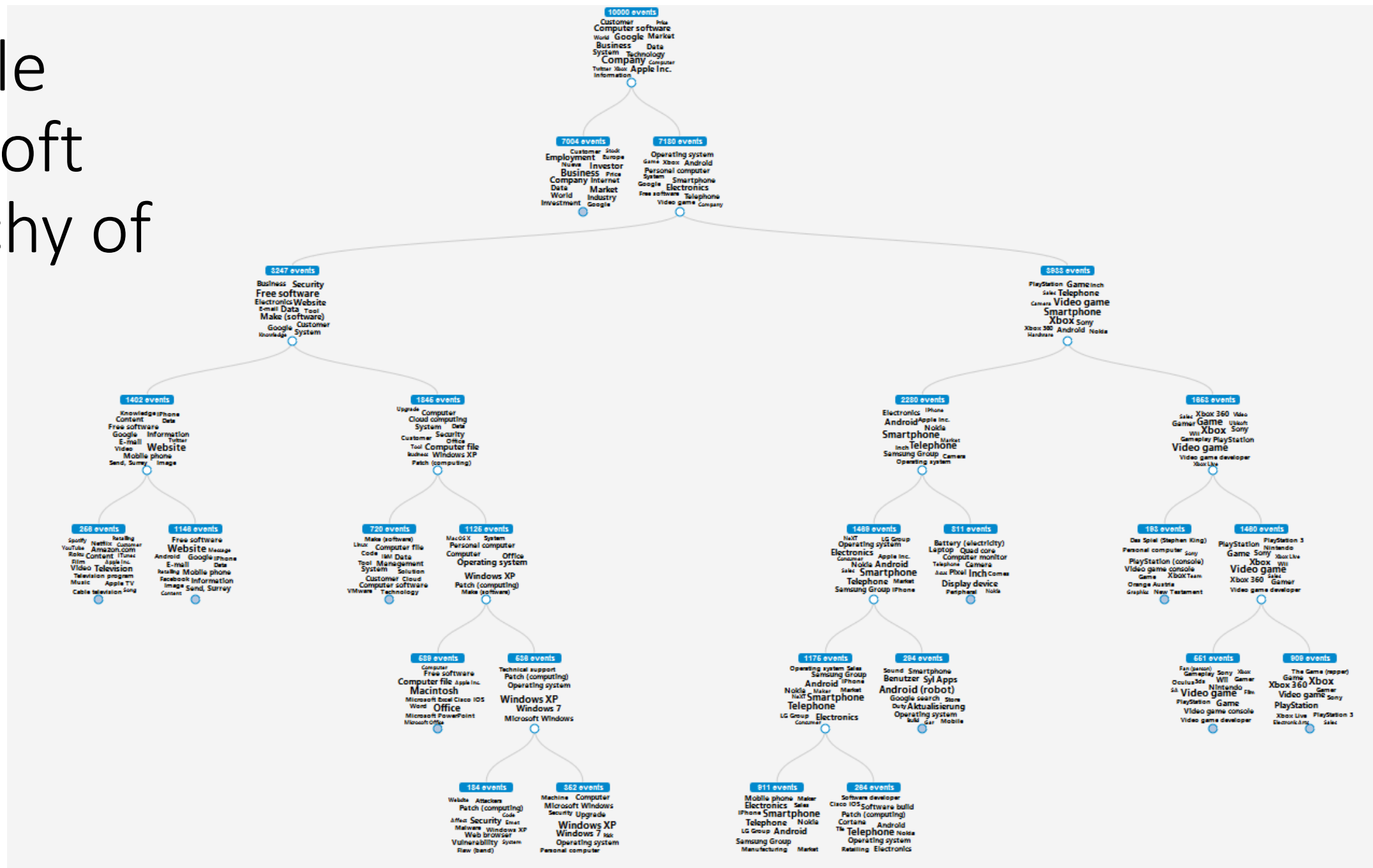
Event sequence identification



Hierarchy of events



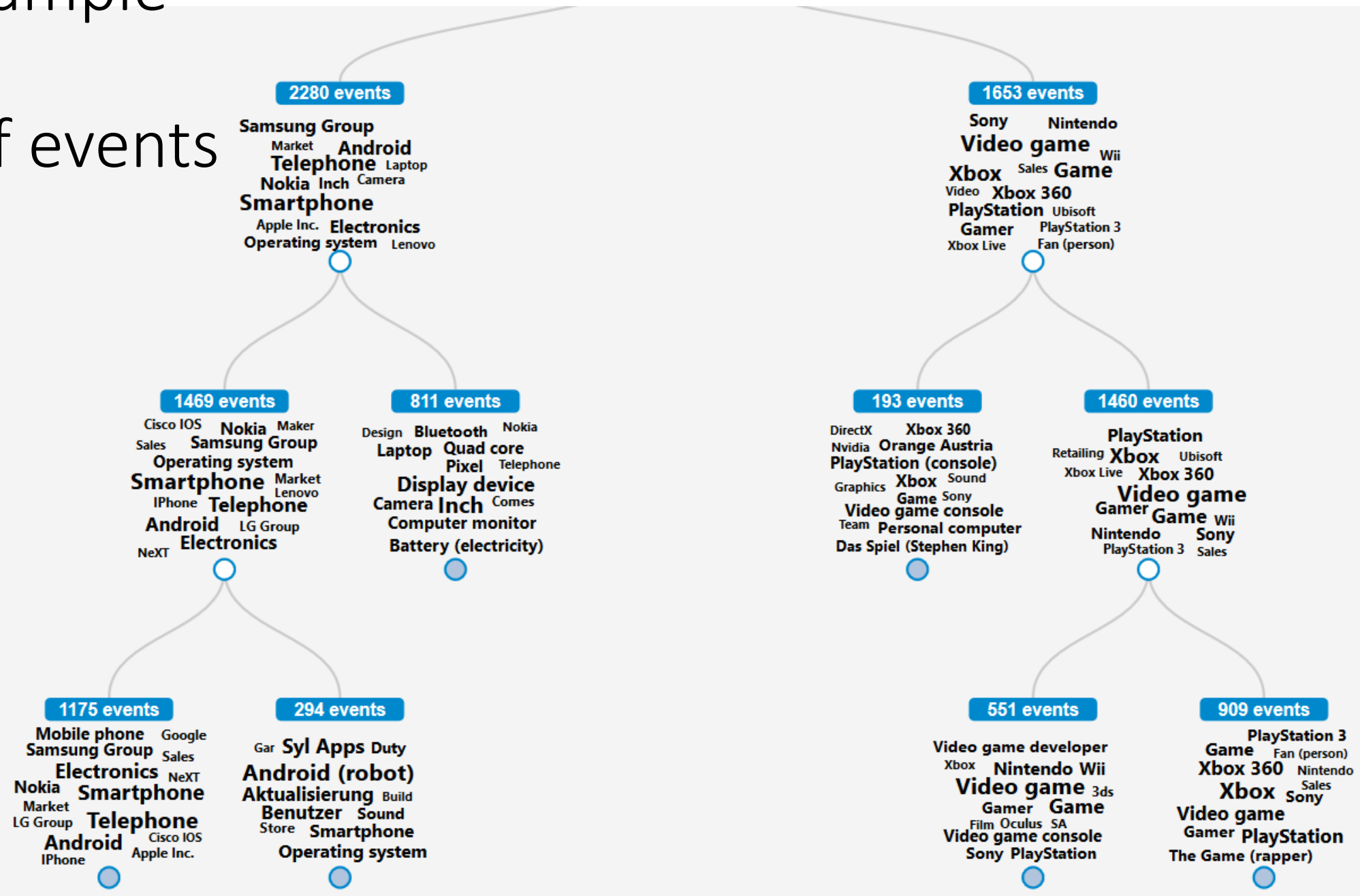
Example Microsoft hierarchy of events



Zoom-in Example

Microsoft

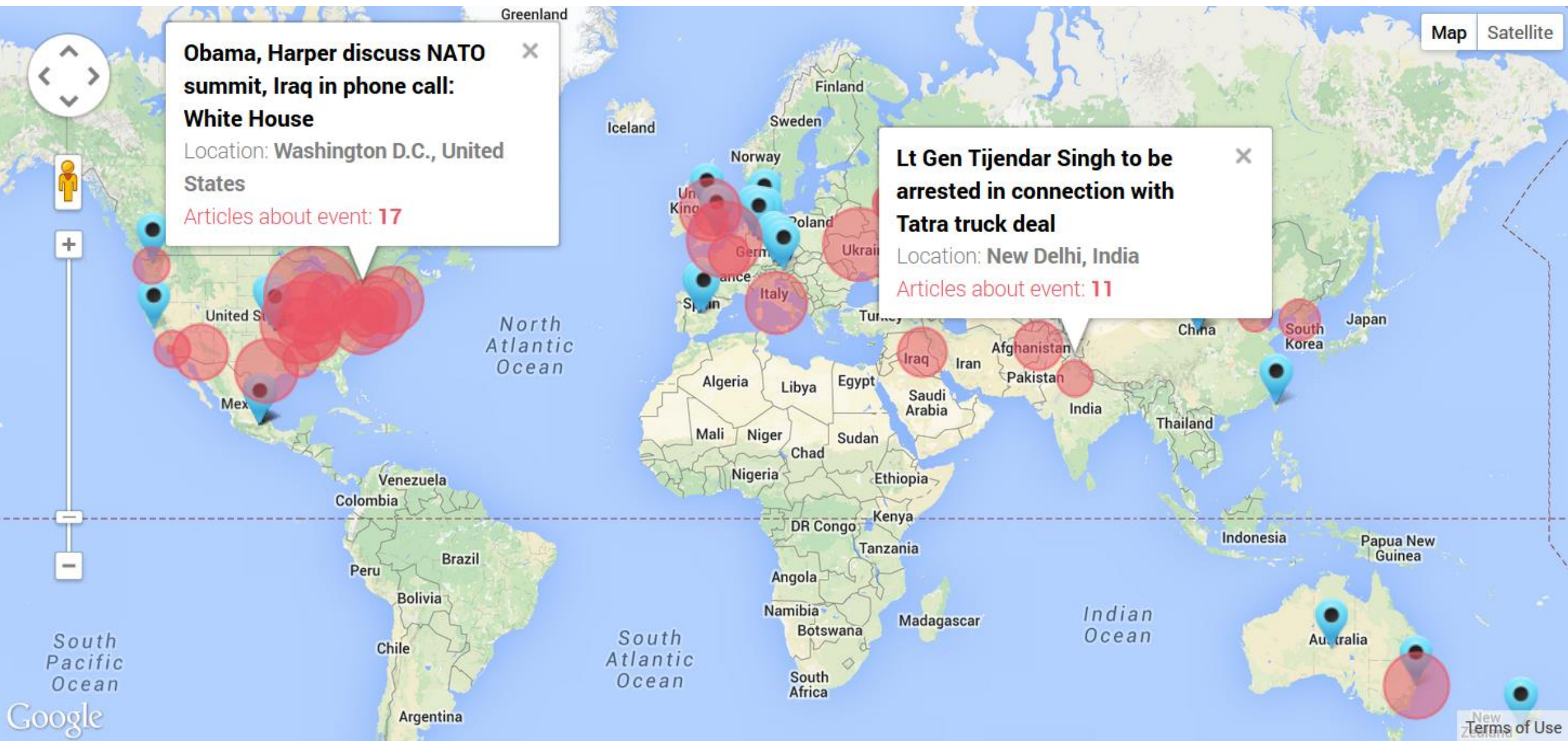
hierarchy of events



Event Visualization

<http://eventregistry.org/>

Live Event tracking with <http://EventRegistry.org/>

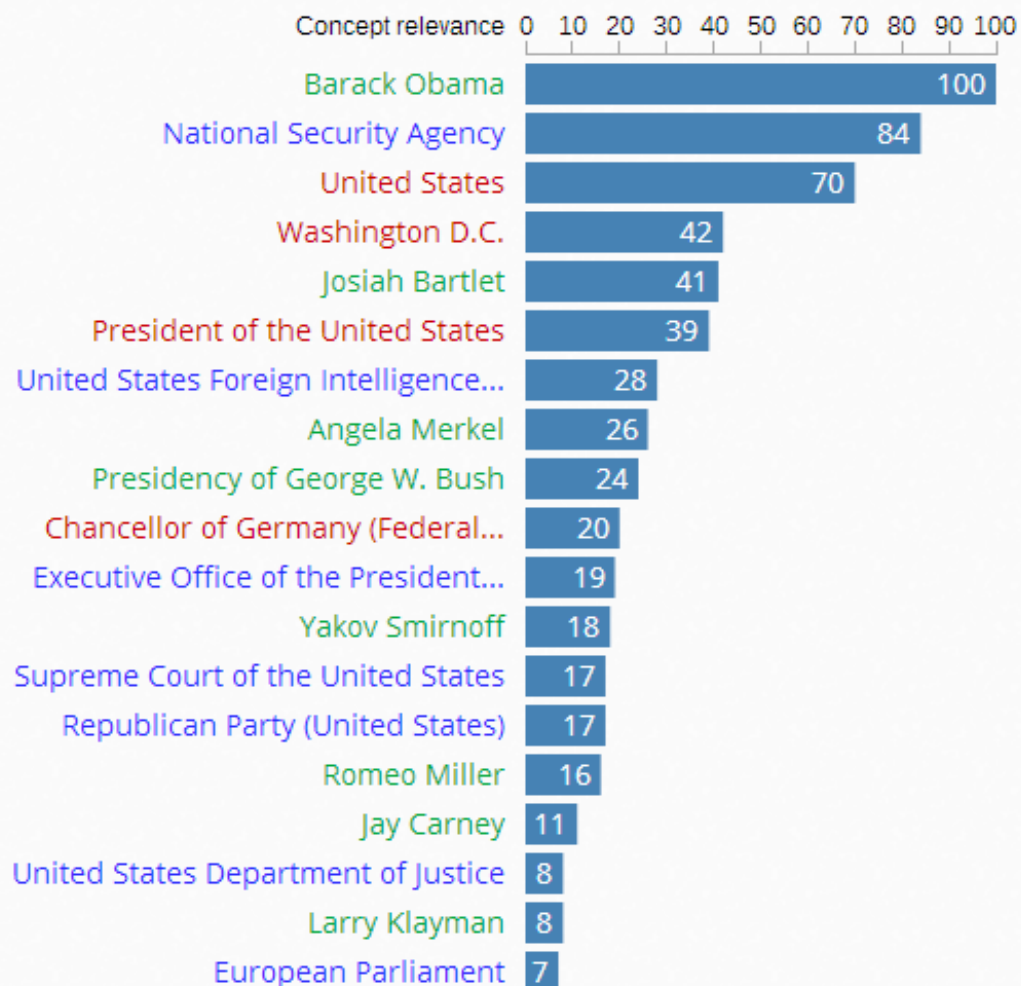


The White House...

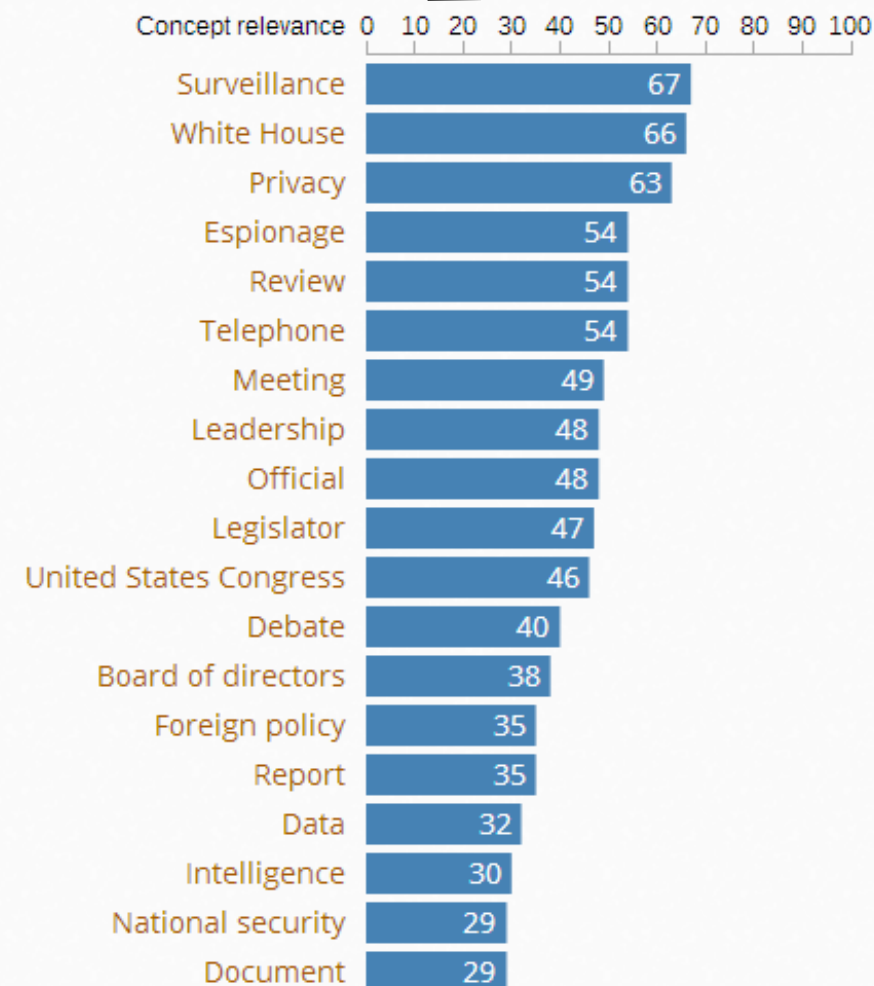
Nr. of articles: 105 (89 eng, 5 ger, 11 spa, 0 chi, 0 slo)

Event description through entities and Semantic keywords

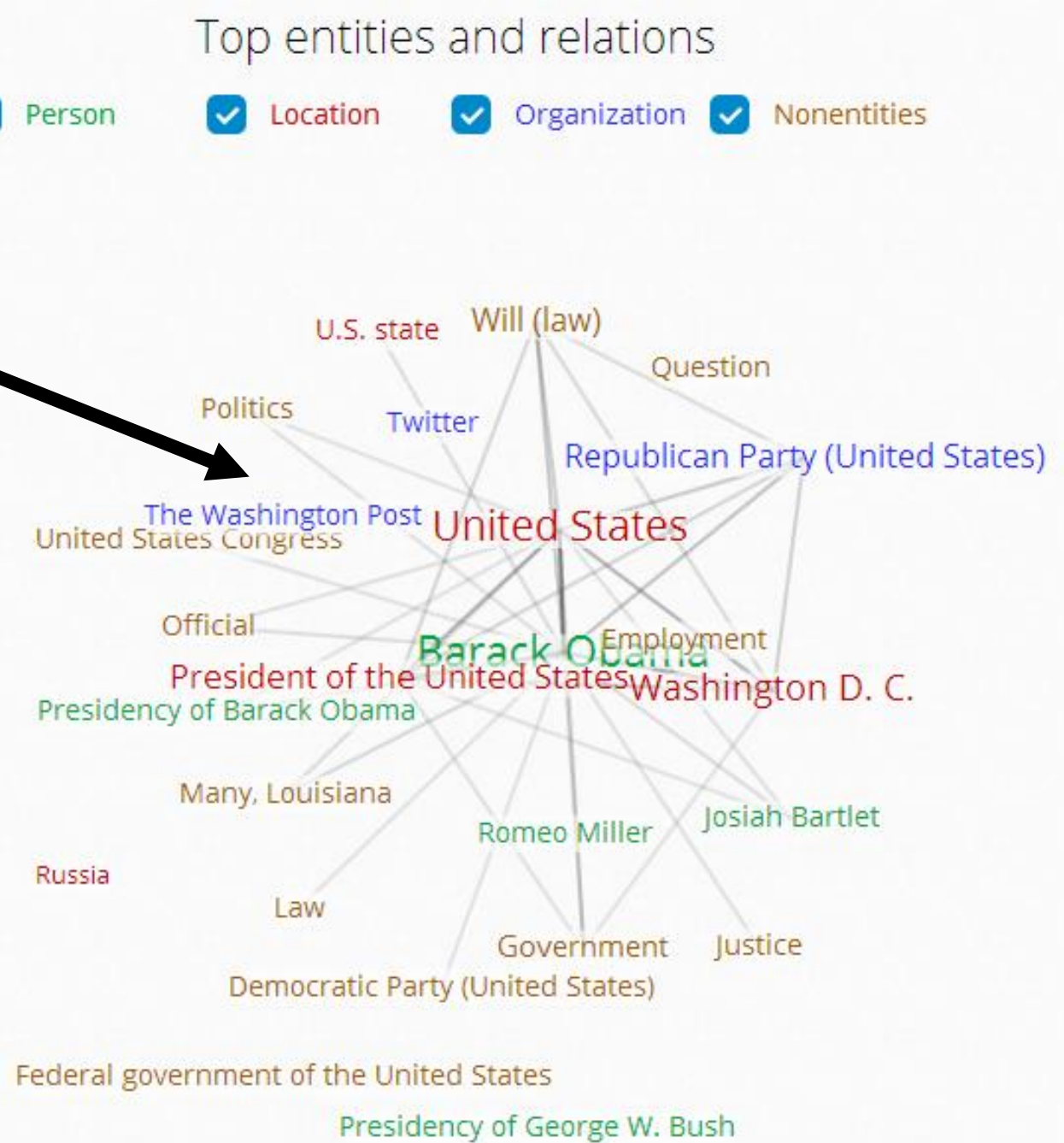
ENTITIES



KEYWORDS



Collection of events
described through
Entity relatedness



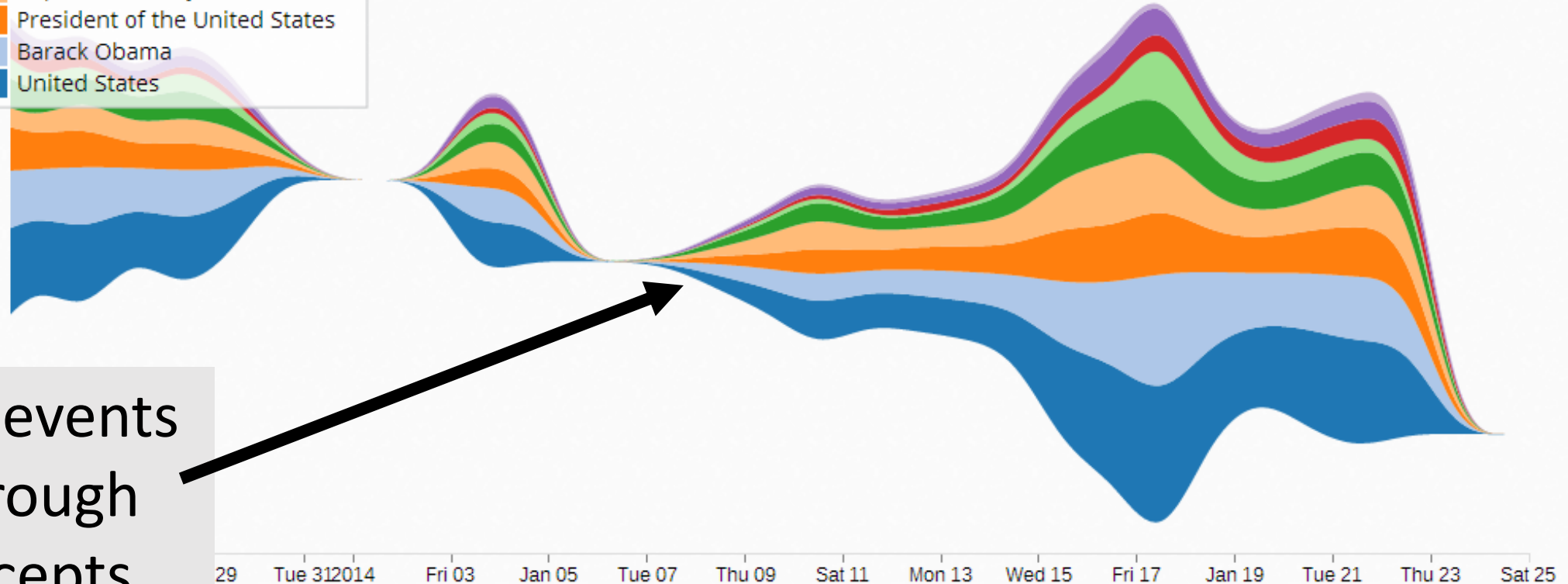
Trending of concepts in search results

☒ Person ☒ Location ☒ Organization ☐ Nonentities

☐ Normalize by article count

United States Barack Obama President of the United States Republican Party (United States) Washington D. C. National Security Agency
Russia U.S. state Syria

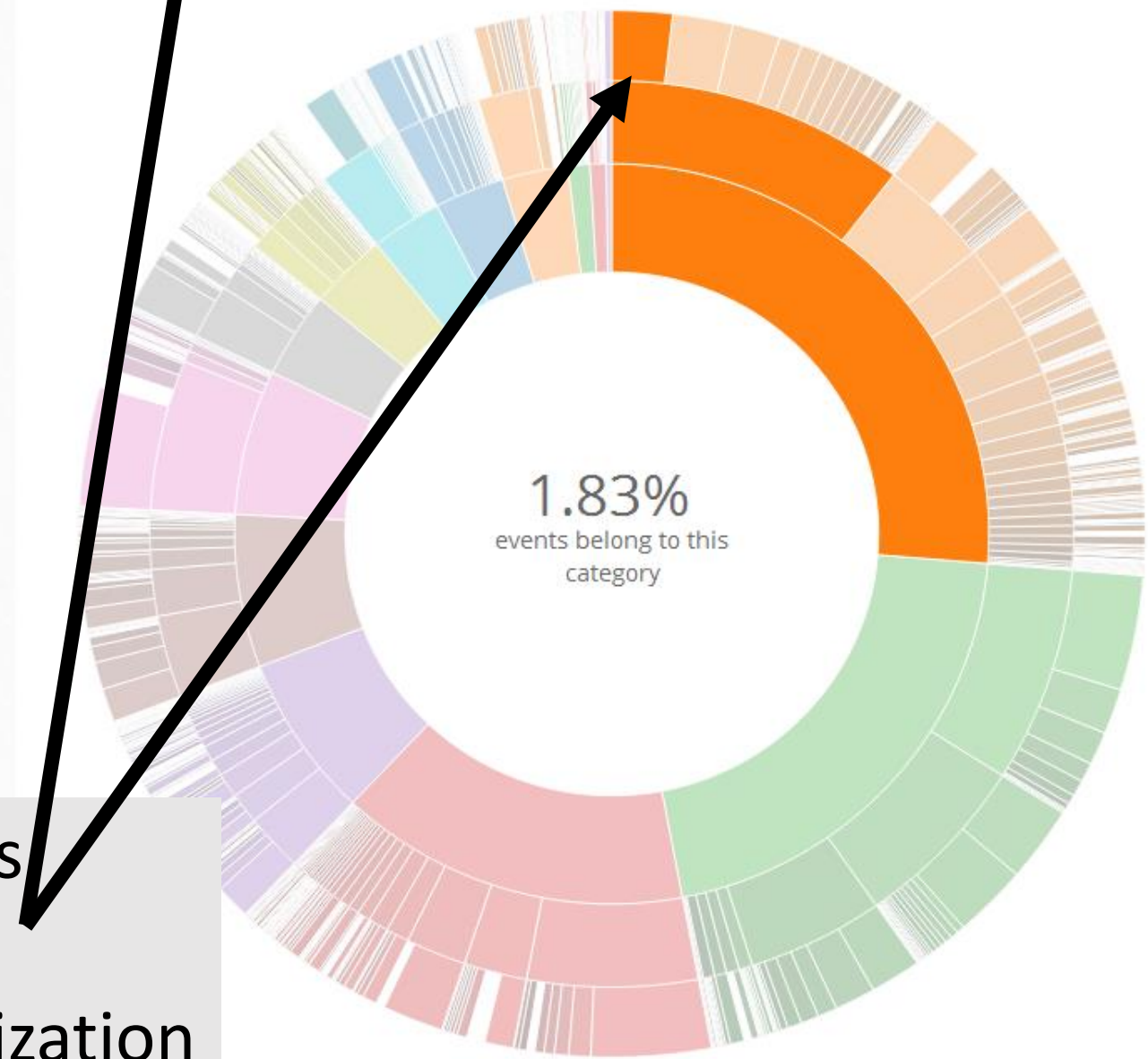
Syria
U.S. state
Russia
National Security Agency
Washington D. C.
Republican Party (United States)
President of the United States
Barack Obama
United States



Collection of events described through trending concepts

Categories of events

Society > Issues > Labor 1.83%



- Society
- Regional
- Business
- Health
- Science
- Home
- Recreation
- Computers
- Sports
- Shopping
- Reference
- Arts
- Games
- News

Collection of events described through three level categorization

Obama ponders limiting NSA access to phone records

15 9 Jan 2014

Washington D.C., United States

Washington -- President Barack Obama is expected to rein in spying on foreign leaders and is considering restricting National Security Agency access to the government's surveillance program.

Obama could unveil his highly anticipated decisions as early as next week. On Thursday, the president is expected to discuss his review with congressional lawmakers, while his top lawyer plans to meet with privacy groups. Representatives from tech companies are meeting with White House staff on Friday.

The White House...

Nr. of articles: 105 (89 eng, 5 ger, 11 spa, 0 chi, 0 slo)



Events identified across languages



Articles

89 eng11 spa5 ger0 chi0 slo

89 eng11 spa5 ger

Obama ponders limiting NSA access to phone records

Washington -- President Barack Obama is expected to rein in spying on foreign leaders and is considering restricting National Security Agency access to the government's surveillance program.

Obama could unveil his highly anticipated decisions as early as next week. On Thursday, the president is expected to discuss his review with congressional lawmakers, while his top lawyer plans to meet with privacy groups. Representatives from tech companies are meeting with White House staff on Friday.

The White House...

Jan. 9, 12:51

WWW.DETROITNEWS.COM

view with

Obama likely to adjust US spying on foreign leaders, nearing decision on intelligence changes

WASHINGTON - President Barack Obama is expected to restrict National Security Agency access to Americans' phone records and rein in spying on foreign leaders, according to people familiar with a White House review of the government's surveillance programs.

Obama could unveil his highly anticipated decisions as early as next week. On Thursday, the president is expected to discuss his review with congressional lawmakers, while his top lawyer plans to meet with privacy groups. Representatives ...

Jan. 9, 07:57

CKGL AM 570

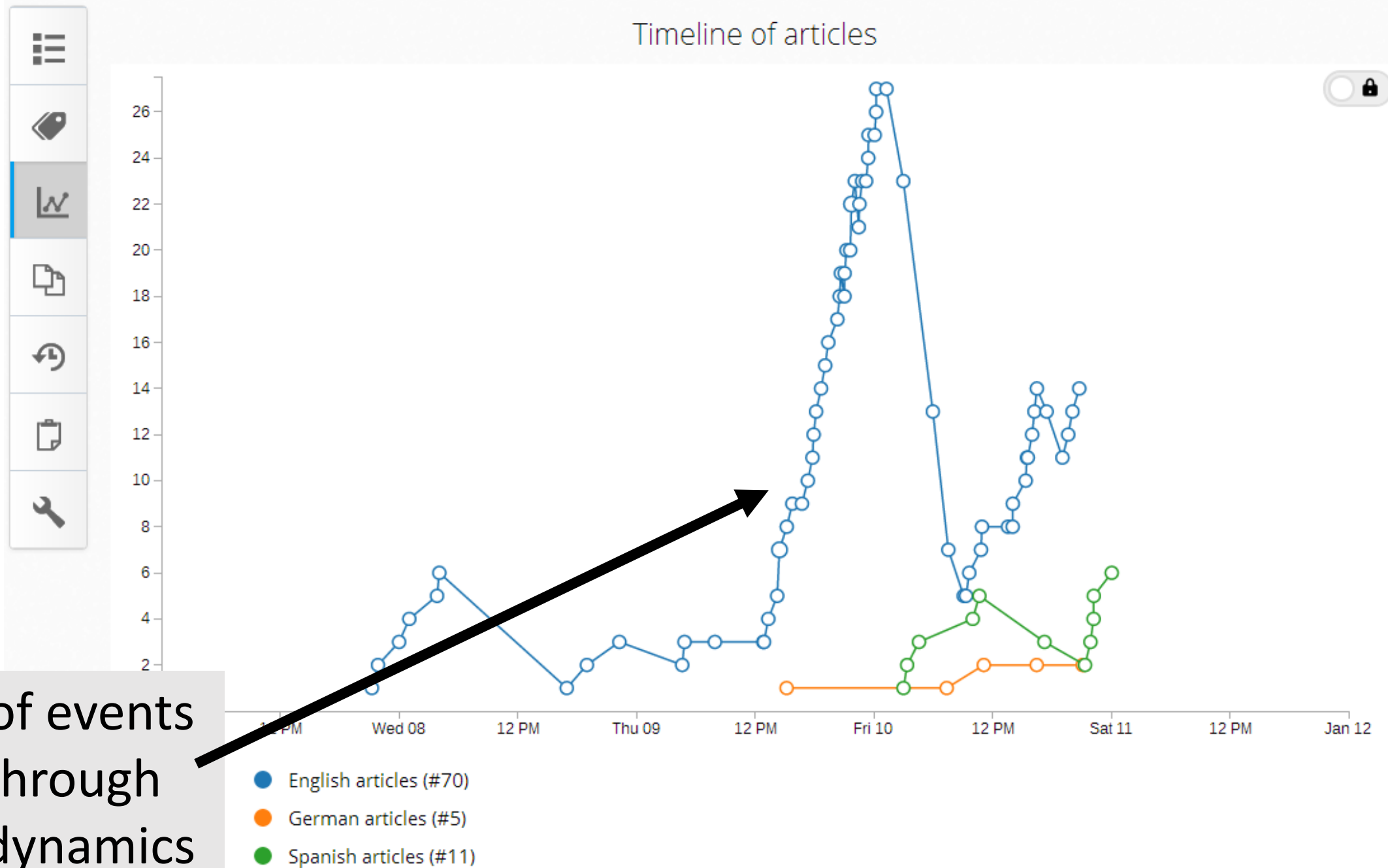
Obama might limit how much the NSA can spy on you

President Barack Obama speaks at a press conference on Dec. 20, 2013. (Pablo Martinez Monsivais/AP)

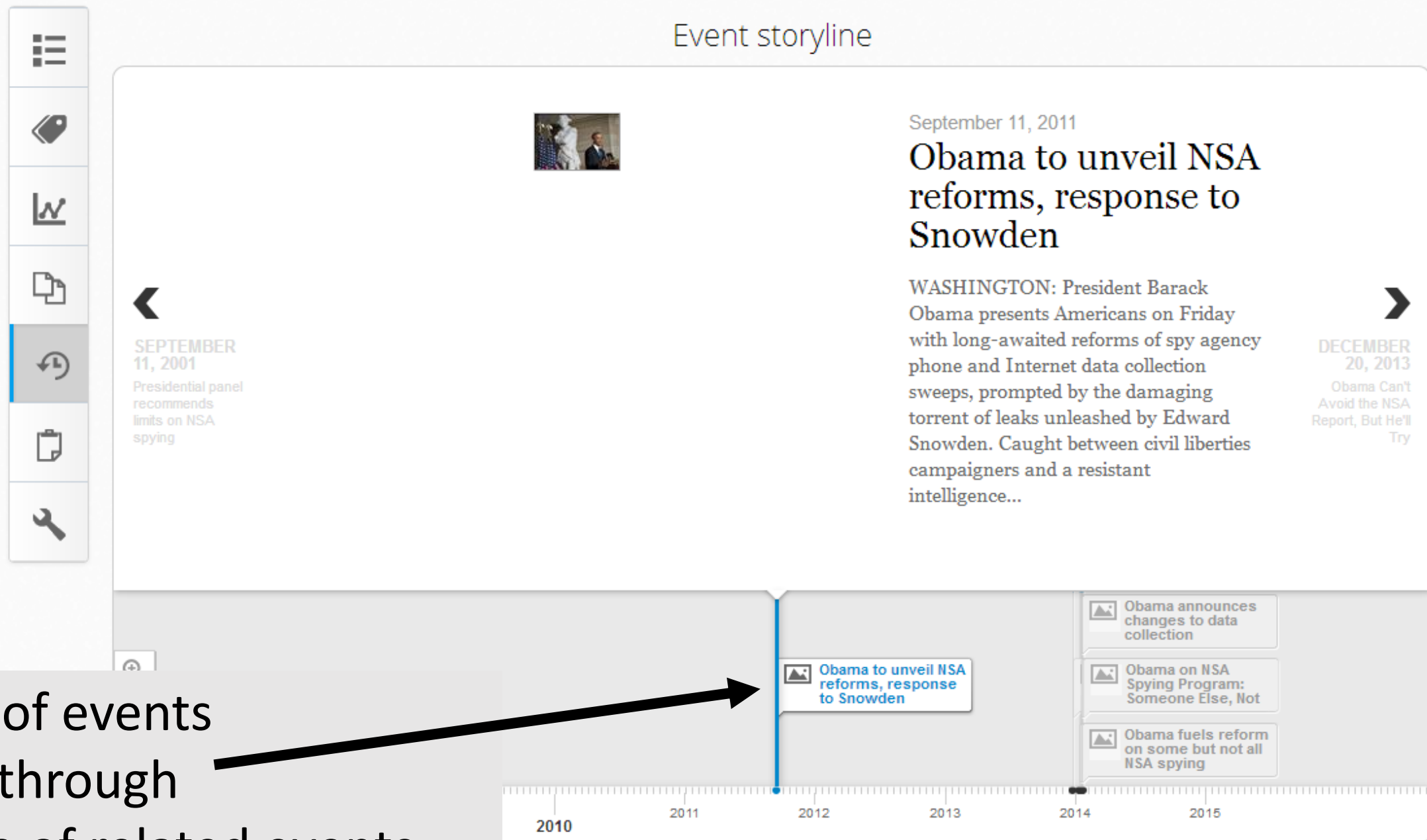
Jan. 10, 14:04

WWW.IMPACTOUSA.COM

Nr. of articles: 105 (89 eng, 5 ger, 11 spa, 0 chi, 0 slo)



Event storyline



Collection of events described through a story-line of related events

Event Registry API

Python code to access “Event Registry”

- <https://github.com/gregorleban/event-registry-python>

GitHub This repository Search Explore Features Enterprise Blog

 gregorleban / event-registry-python

Obtain data from Event Registry using Python

13 commits
 1 branch
 0 releases
 1 contributor

branch: master

[event-registry-python](#) / +

- fixes for some requests

gregorleban authored 18 days ago

latest commit 774afb72e3

.gitignore	Initial commit	a month ago
EventRegistry.py	- fixes for some requests	18 days ago
Examples.py	- fixes for some requests	18 days ago
LICENSE	Initial commit	a month ago
README.md	- beauty changes	26 days ago

README.md

GitHub

This repository search

Explore [Features](#) [Enterprise](#) [Blog](#)

[Sign up](#) [Sign in](#)

gregorleban

event-registry-python

★ Star

🔗 Fork

branch: master

event-registry-python / Examples.py

🔍

📄

gregorleban

10 days ago · fixes for some requests

1 contributor

67 lines (67 sloc) · 2.66 kb

Raw

Clone

History

```
1 from eventregistry import *
2
3 @E -> EventRegistry(host = "http://eventregistry.org", logging = True)
4
5
6 Q = QueryBuilder()
7 Q.addCondition(E['getEventTypes']('0000')) # get events related to 0000
8 # Q.addCondition(E['getCategory']('activity issues')) # did not related to issues in activity
9 # Q.addCondition(E['getCategory']('00')) # did not have been reported by SEC
10 Q.addCondition(E['getEventTypes']('0000')) # return only 0000 events
11 Q.addCondition(E['getEventTypes']('page = 0, sortBy = 10, sortType = "desc", sortOrder = True, sortOrder = 1)) # get
12 # E -> EventRegistry()
13
14 Q = QueryBuilder()
15 Q.addCondition(E['getEventTypes']('0000'))
16 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
17 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
18 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
19 # E -> EventRegistry()
20
21 # get info about event with url "123"
22 Q = QueryBuilder()
23 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
24 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
25 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
26 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
27 # E -> EventRegistry()
28
29 # get info about event with url "123"
30 Q = QueryBuilder()
31 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
32 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
33 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
34 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
35 # E -> EventRegistry()
36
37 # get info about event with url "123"
38 Q = QueryBuilder()
39 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
40 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
41 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
42 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
43 # E -> EventRegistry()
44
45 # get info about event with url "123"
46 Q = QueryBuilder()
47 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
48 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
49 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
50 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
51 # E -> EventRegistry()
52
53 # get info about event with url "123"
54 Q = QueryBuilder()
55 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
56 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
57 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
58 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
59 # E -> EventRegistry()
60
61 # get info about event with url "123"
62 Q = QueryBuilder()
63 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
64 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
65 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
66 Q.addCondition(E['getEventTypes']('0000', 0, 10), sortOrder = 10, sortType = "desc", sortOrder = 1))
67 # E -> EventRegistry()
```

Searching for events using Python

```
>>> from EventRegistry import *
>>> er = EventRegistry()
>>> q = QueryEvents()
# get events related to Barack Obama
>>> q.addConcept(er.getConceptUri("Obama"))
# and are related to issues in society
>>> q.addCategory(er.getCategoryUri("society issues"))
# and have been reported by the BBC
>>> q.addNewsSource(er.getNewsSourceUri("bbc"))
# return event details for first 30 events
>>> q.addRequestedResult(RequestEventsInfo(page = 0, count = 30))
# execute query and obtain results
>>> res = er.execQuery(q)
```

Result of the query

```
'events': { 'resultCount': 122,  
  'results': [  
    {'articleCounts': {'eng': 54.0, 'total': 54.0},  
      'categories': [{...}],  
      'concepts': [{...}, ...],  
      'eventDate': '2014-08-29',  
      'eventDateEnd': '',  
      'multiLingInfo': { 'eng': {  
        'title': ..., 'summary': ...}},  
      'uri': '1211229', 'wgt': 9.0}  
    }, ...  
  ]  
}
```

Getting info for a particular event

```
# get information about event with ID 123
```

```
>>> q = QueryEvent("123");
```

```
# return concept labels in 3 languages
```

```
>>> q.addRequestedResult(RequestEventInfo(["eng", "spa", "slv"]))
```

```
# get 10 most central articles
```

```
>>> q.addRequestedResult(RequestEventArticles(0, 10))
```

```
# get information how articles about the event were trending
```

```
>>> q.addRequestedResult(RequestEventArticleTrend())
```

```
# get top keywords
```

```
>>> q.addRequestedResult (RequestEventKeywordAggr())
```

```
>>> eventRes = er.execQuery(q);
```

Searching for articles

```
>>> q = QueryArticles();  
# articles should be from a particular time period  
>>> q.setDateLimit(datetime.date(2014, 4, 16), datetime.date(2014, 4, 28))  
# they should mention apple  
>>> q.addKeyword("apple")  
# they should also mention iphone  
>>> q.addKeyword("iphone")  
# get top 30 articles that match criteria  
>>> q.addRequestedResult(RequestArticlesInfo(page=0, count = 30));  
>>> res = er.execQuery(q)
```

Event Registry exports event data through API and RDF/Storyline ontology

- API to search and export event information
 - Export of all the system data in JSON
- Event data is exported in a structured form
 - BBC Storyline ontology
 - <http://www.bbc.co.uk/ontologies/storyline/2013-05-01.html>
- SPARQL endpoint:
 - <http://eventregistry.org/rdf/search>
 - <http://eventregistry.org/rdf/event/{eventID}>
 - <http://eventregistry.org/rdf/article/{articleID}>
 - <http://eventregistry.org/rdf/storyline/{storylineID}>
 - Example: <http://eventregistry.org/rdf/event/1234>

Future / Follow-up projects

Some of the follow-up projects/research

- Understanding global social dynamics
 - How global society functions?
- Integrating text-based media with TV channels
 - ...requires speech recognition, video processing, visual object recognition, face recognition, ...
- Event prediction / Event-Consequence prediction
 - ...requires understanding of causality in the social dynamics and much more
- Micro-reading / Machine-reading
 - ...full understanding of individual documents – the goal for 10+ years