

Applications of Latent Entity Networks in Information Retrieval

Andreas Spitz, Michael Gertz

Heidelberg University, Institute of Computer Science
Database Systems Research Group

{spitz,gertz}@informatik.uni-heidelberg.de

Workshop Internationale Klima- und Energiediskurse
Darmstadt, May 26, 2017

Handelsblatt

ÜBERNAHME VON SECUSMART

Blackberry kauft „Merkel-Phone“

Autor: dpa

Datum: 29.07.2014 15:20 Uhr • Update: 29.07.2014, 17:15 Uhr

theguardian
Winner of the Pulitzer prize 2014

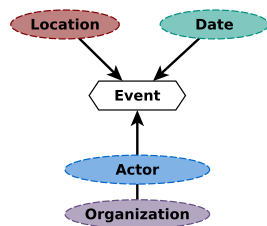
Germany opens inquiry into claims NSA tapped Angela Merkel's phone

Unexpected inquiry, announced by federal prosecutor, will determine if US actively listened in to calls

Motivation

Definition: Event

“Something that happens at a given place and time between a group of actors.”
[CSG⁺02]



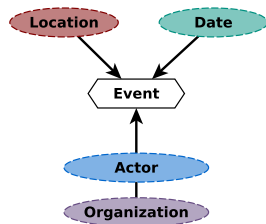
Motivation

Definition: Event

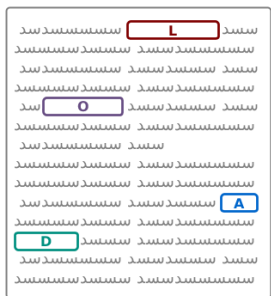
“Something that happens at a given place and time between a group of actors.”
[CSG⁺02]

For large document collections such as corpora of newspapers, how can we...

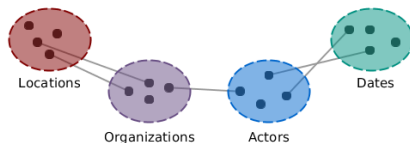
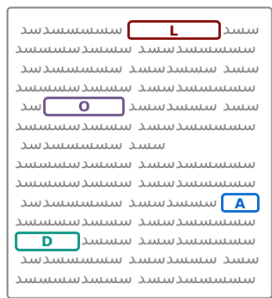
- obtain events from unstructured text?
- identify connections across documents?
- support entity-centric event search?



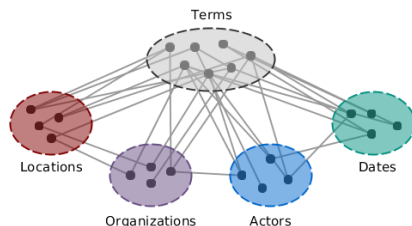
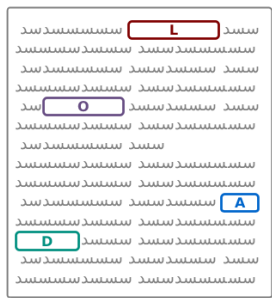
Information Network Extraction from Text



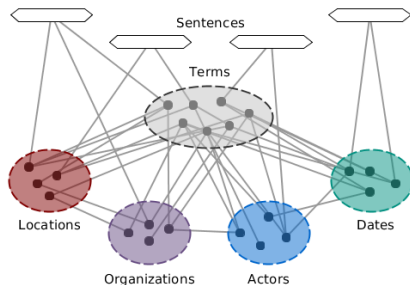
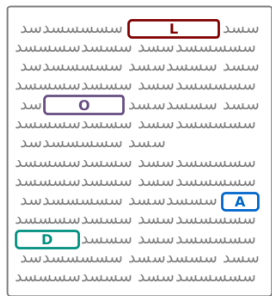
Information Network Extraction from Text



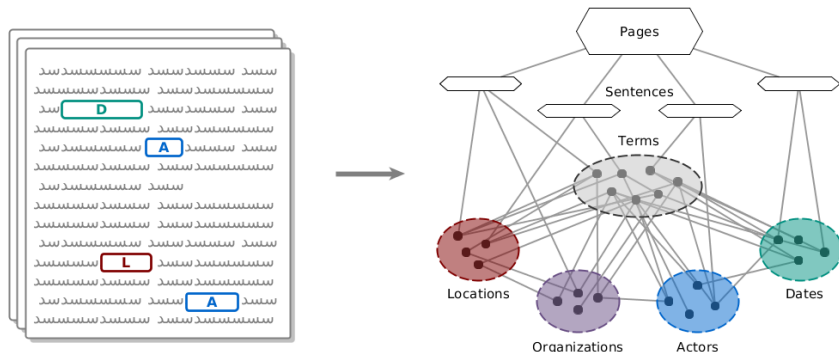
Information Network Extraction from Text



Information Network Extraction from Text

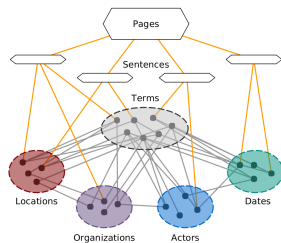


Information Network Extraction from Text



[SG16]

Edge Weight Generation

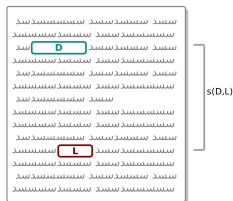
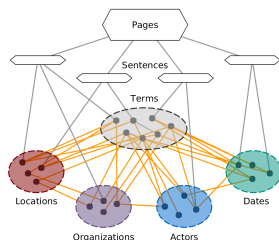


For edges (x, y) for which y is a page or sentence, count only (co-) occurrences:

$$\omega(x, y) = \begin{cases} 1 & \text{if } y \text{ contains } x \\ 0 & \text{otherwise} \end{cases}$$

[SG16]

Edge Weight Generation



For edges (x, y) for which y is a page or sentence, count only (co-) occurrences:

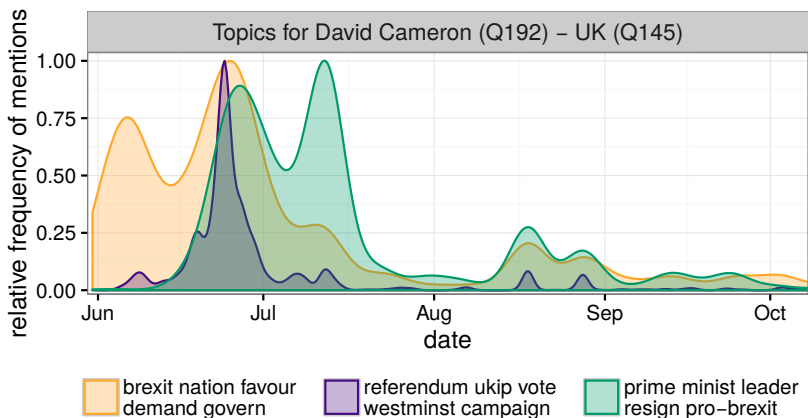
$$\omega(x, y) = \begin{cases} 1 & \text{if } y \text{ contains } x \\ 0 & \text{otherwise} \end{cases}$$

For edges (x, y) between entity types and terms, aggregate co-occurrence instances I : sum over similarities derived from sentence distances s .

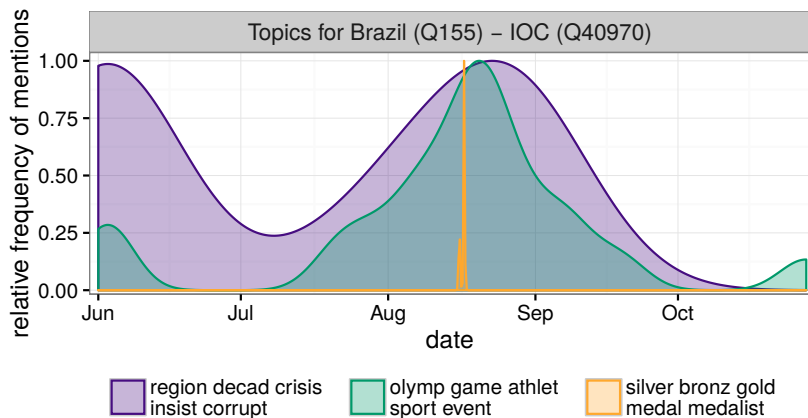
$$\omega(x, y) := \sum_{i \in I} \exp(-s(x, y, i))$$

[SG16]

Entity Topics: Brexit



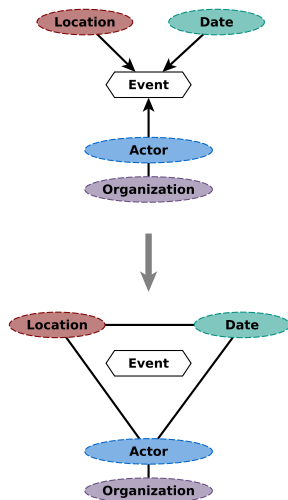
Entity Topics: Olympic Games



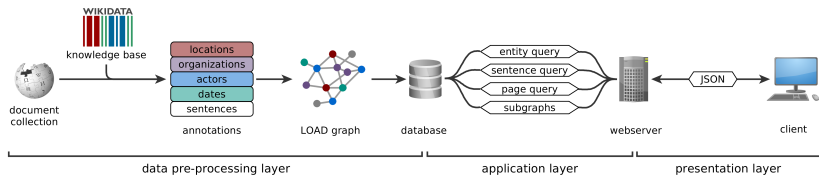
Event Extraction and Search

Intuition:

- Events correspond to patterns in the network (e.g., triangular structures)
- Participating entities can be used to complete events



Event and Entity Search and Exploration



EVELIN:

Exploration of Event and Entity Links in Information Networks

Available for:

Wikipedia: <http://evelin.ifi.uni-heidelberg.de/>

News Corpus: <http://evelin.ifi.uni-heidelberg.de:7777>

[SAG17]

Summary

Latent Entity Networks:

- fast entity and event exploration
- can support most entity-related Information Extraction tasks
- can be extended to any kind of entity
- scalable and fast
- language-agnostic with entity linking

Available for download:

- Wikipedia latent entity networks
- Code for generating latent entity networks
- Code for the query interface



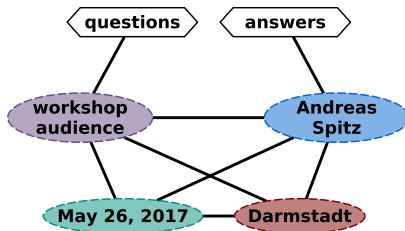
`http://dbs.ifi.uni-heidelberg.de/index.php?id=load`

Available for download:

- Wikipedia latent entity networks
- Code for generating latent entity networks
- Code for the query interface



<http://dbs.ifi.uni-heidelberg.de/index.php?id=load>



Bibliography I



Christopher Cieri, Stephanie Strassel, David Graff, Nii Martey, Kara Rennert, and Mark Liberman.

Corpora for topic detection and tracking.

In *Topic Detection and Tracking*. Springer, 2002.



Andreas Spitz, Satya Almasian, and Michael Gertz.

Evelin: Exploration of event and entity links in implicit networks.

In *WWW*, 2017.



Andreas Spitz and Michael Gertz.

Terms over LOAD: Leveraging named entities for cross-document extraction and summarization of events.

In *SIGIR*, 2016.