Events beyond Languages and Schema Constraints: Representation, Detection and Perception

Alexandria Workshop 2017

10/19/2017

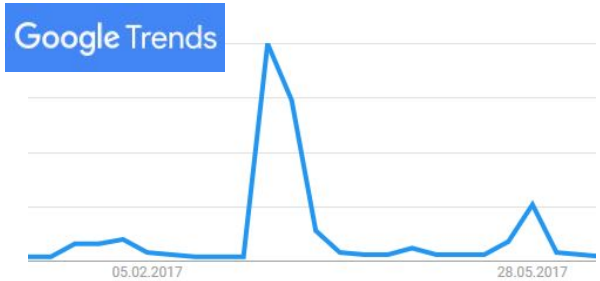
Simon Gottschalk

gottschalk@L3S.de,

Elena Demidova, Paolo Sottovia, Wolfgang Nejdl, Yannis Velegrakis

Trump says 'great' Merkel visit, amid 'fake news', but still hits Germany on NATO

Kanzlerin im Weißen Haus - Trump empfängt Merkel



Der Präsident der Vereinigten Staaten Donald Trump empfängt erstmals Bundeskanzlerin Angela Merkel zum Staatsbesuch.

<Angela_Merkel; visits; Donald_Trump>

Challenges

- Representation
  - What is an event?
  - How to describe it?

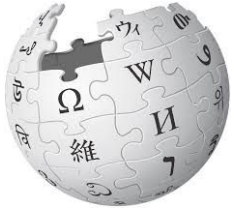
- Detection
  - How to detect important events?

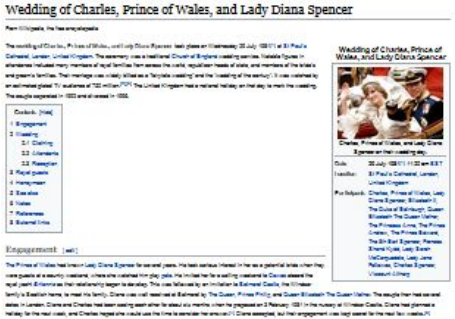
- Perception
  - How to provide an overview of different representations?

Events beyond Languages and Schema Constraints: Representation, Detection and Perception

The New York Times

LONDON, July 29 When Prince Charles and Lady Diana Spencer posed for their engagement photograph ....





WIKIDATA

Wedding of Charles and Diana
point in time: 29 July 1981

<*Charles*, spouse, *Diana*>
start time: 29 July 1981
end time: 28 August 1996

What is an Event?

- Wikipedia: “Event may refer to … “
- Merriam-Webster: “Something that happens“
- Oxford: “A thing that happens or takes place, especially one of importance.”
- TDT: “Something that happens at a particular time and place”

Event Examples

Text	Name	Relation	Location	Time
Donald Trump welcomes Angela Merkel...			White House	Mar 17, 2017
The wedding of … takes place at ...	*Wedding of Charles and Diana*	(<*Charles*, spouse, *Diana*>)	St Paul's Cathedral	29 July 1981
	financial crisis of 2007–08			start: 2007 end: 2008 (?)
The ...announces the first direct observation ...	First observation of gravitational waves		Washington, Hannover, ..?	14 September 2015 (announced: 11 February 2016)

Events beyond Languages and Schema Constraints: Representation, Detection and Perception

Event Identification: Approaches

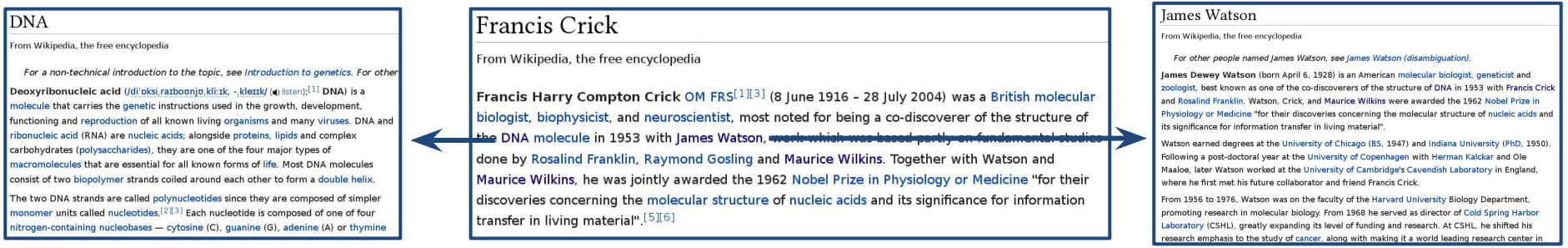
- Peaks in Activity
  - Wikipedia edits, Twitter posts, Google Trends, …
  - limited to the platforms' existence time
    ⇒ no "historical" events

- NLP-based event extraction
  - difficult to structure events and to rate their importance
  - applied on news (archives)
    ⇒ limited to a fixed set of articles
    ⇒ events reported from real-time perception view

Event Identification: Our Approach

- Find important events without any constraints on locations, fixed schemas or entity types

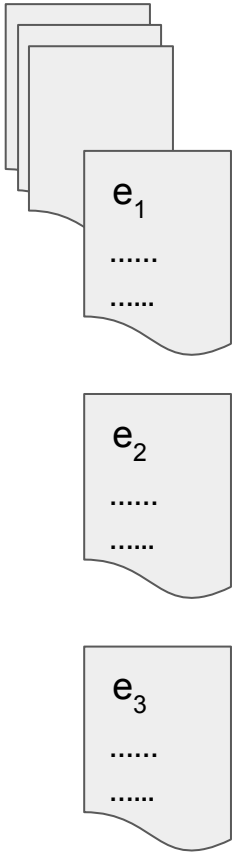
- Exploit the structure of Wikipedia
  - Each article focuses on one entity and links to others



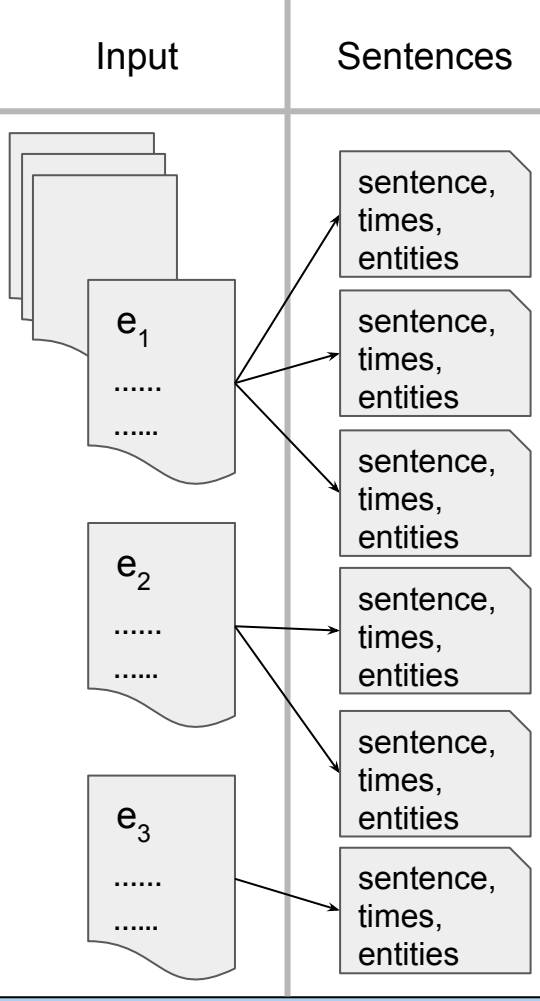
Our Approach: Co-references

Entity	Sentence	Time Expressions	Entities
DNA	Its molecular structure was identified by **James Watson** and **Francis Crick** in **1953**, whose model-building efforts were guided by X-ray diffraction data acquired by Rosalind Franklin.	1953	James Watson, Francis Crick
James Watson	James Dewey Watson (born April 6, 1928) is an American molecular biologist, geneticist and zoologist, best known as one of the co-discoverers of the structure of **DNA** in **1953** with **Francis Crick**.	April 6, 1928 1953	Molecular biology, Genetics, Zoologist, DNA, Francis Chrick
Francis Crick	Francis Harry Compton Crick (8 June 1916 – 28 July 2004) was a British molecular biologist, biophysicist, and neuroscientist, most noted for being a co-discoverer of the structure of the **DNA** molecule in **1953** with **James Watson**.	8 June 1916 – 28 July 2004 1953	British people, Molecular biology, Biophysics, Neuroscience, DNA, Molecule, James Watson

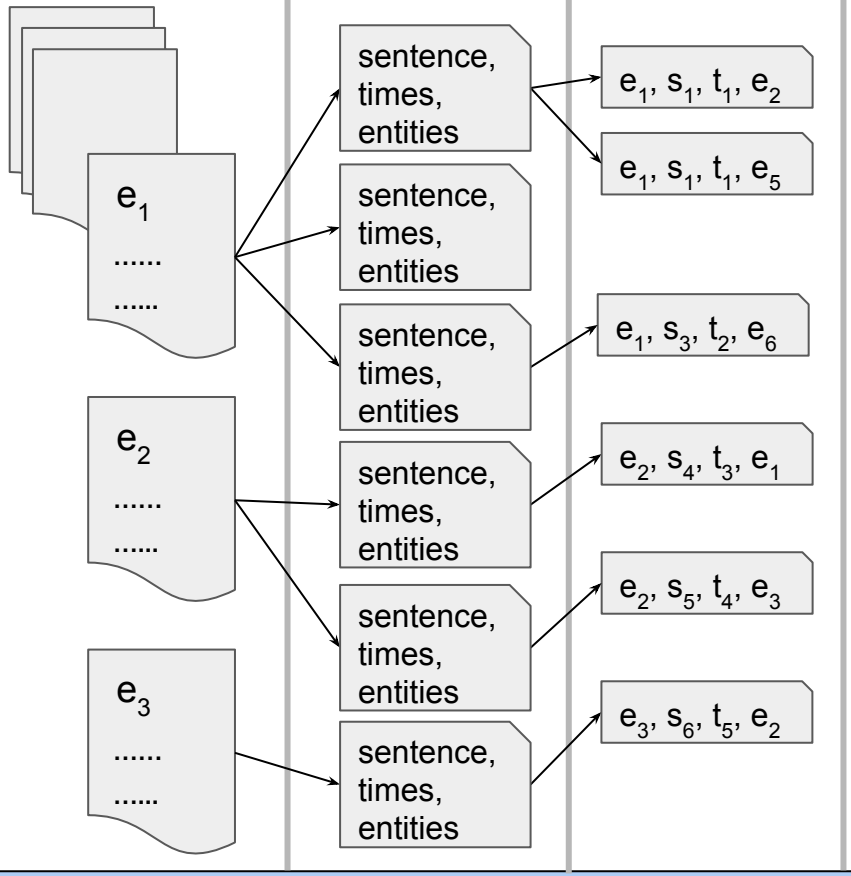
Input



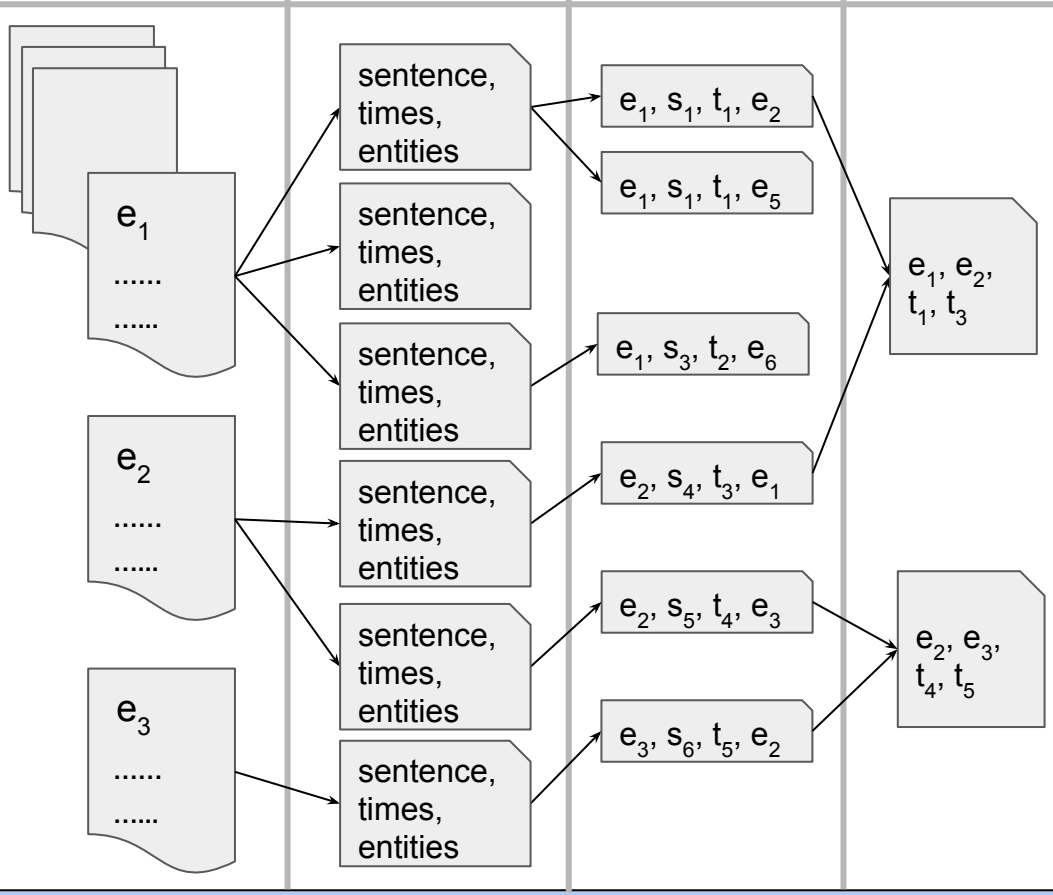
Input: A set E of entities, where each entity e_i is represented as a Wikipedia article



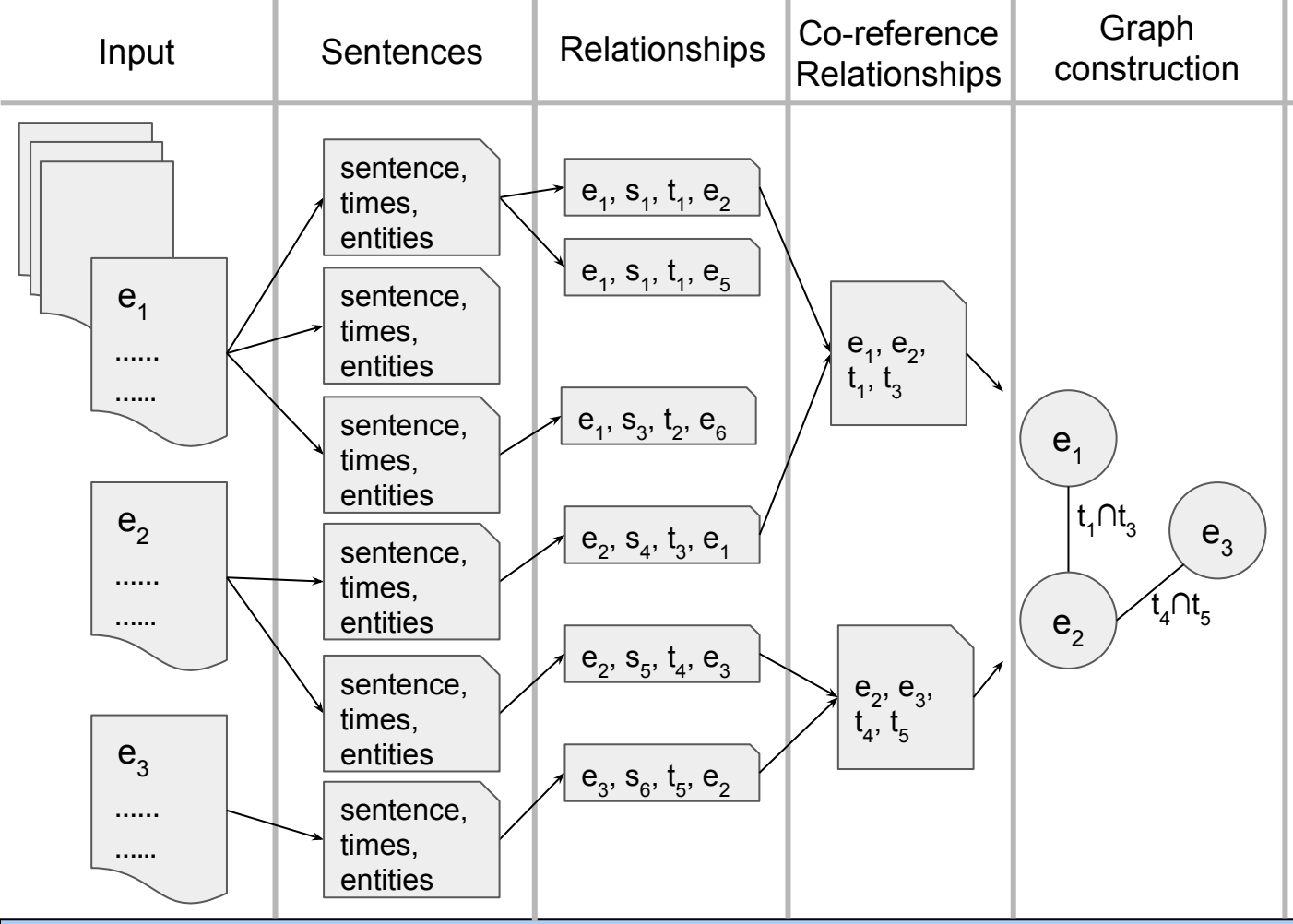
Preprocessing: Split each entity article into sentences and extract time mentions and other entity mentions



For each entity mention entity $e_j \in$ links(e_i) , create a relationship.



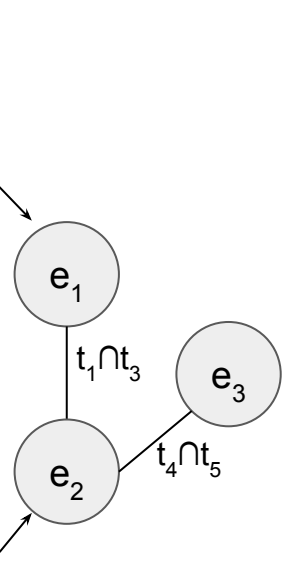
For each entity e_i, time span t_i
  For each e_j, $t_j \in \mathit{links}(e_i)$
    if $t_i \cap t_j \neq \emptyset$:
      emit tuple $\langle e_i, e_j, t_i \cap t_j \rangle$

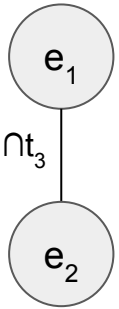


Create an entity graph where two entities are connected if they are part of the same event.

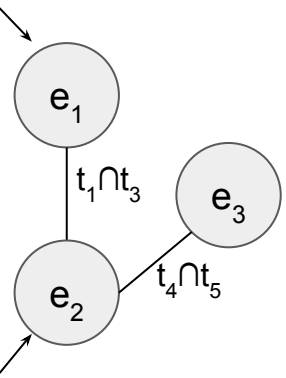
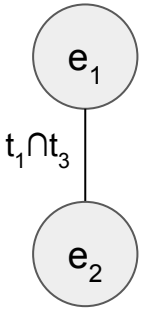
Graph construction

Co-reference subgraphs

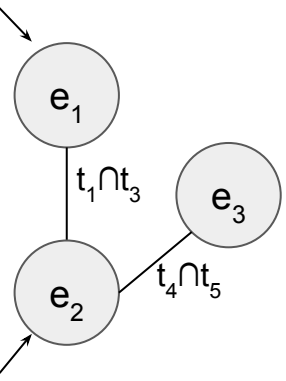
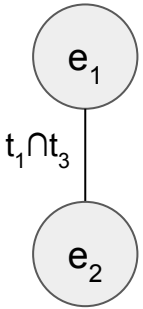
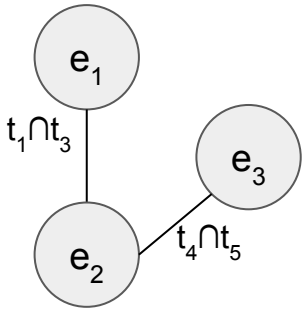




Restrictive Approach: Extract cliques only

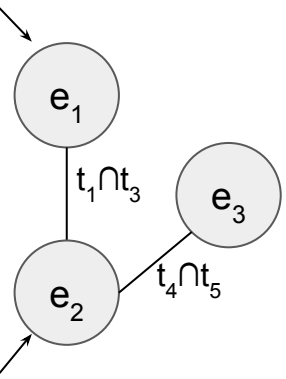
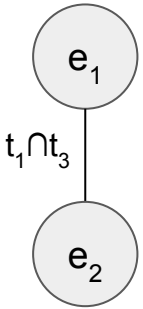
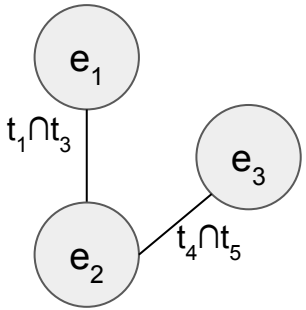
Graph construction	Co-reference subgraphs	Events
		$\{e_1,e_2\}$, $t_1 \cap t_3$

Restrictive Approach: Extract cliques only

Graph construction	Co-reference subgraphs	Events
		$\{e_1,e_2\}$, $t_1 \cap t_3$
		

Restrictive Approach:
Extract cliques only

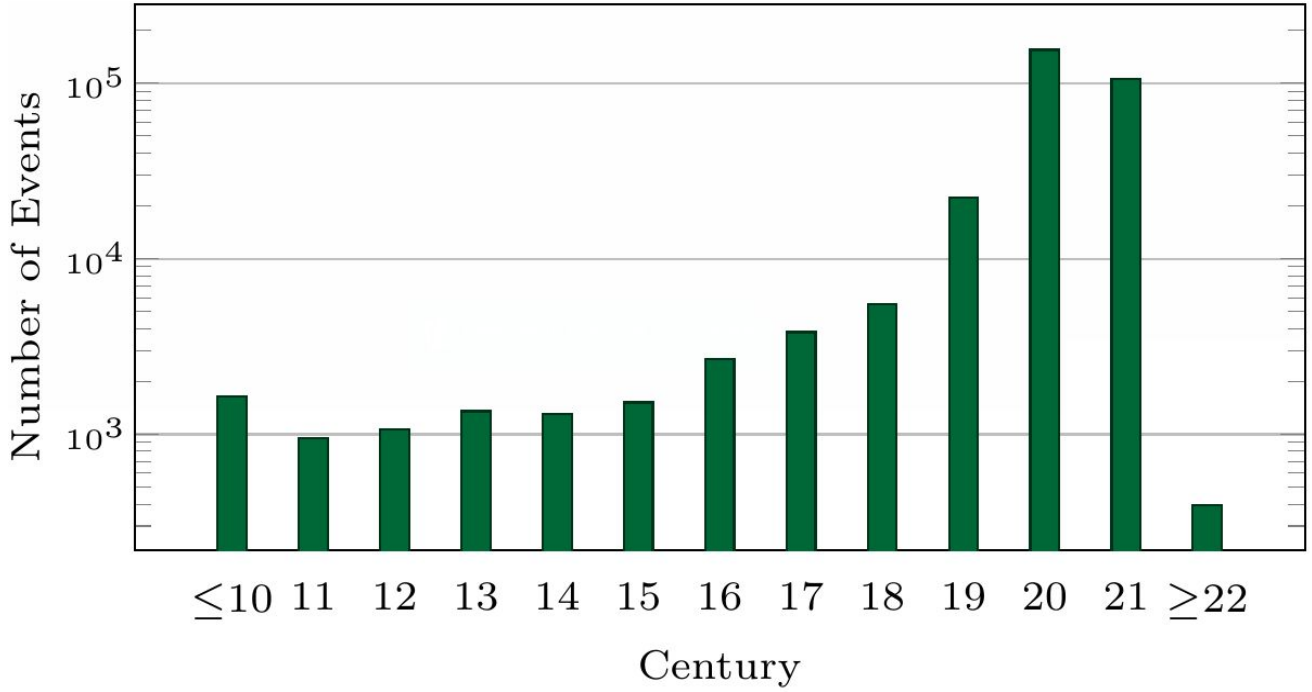
Relaxed Approach:
Connected components with temporal agreement (adapted BFS)

Graph construction	Co-reference subgraphs	Events	
		$\{e_1,e_2\}$, $t_1 \cap t_3$	Restrictive Approach: Extract cliques only
		$\{e_1,e_2,e_3\}$, $t_1 \cap t_3 \cap t_4 \cap t_5$	Relaxed Approach: Connected components with temporal agreement (adapted BFS)

Example Events

- <{Eric Cradock, Lew Hayman, Léo Dandurand, Montreal Alouettes}, 1946>
- <{Bosaso, Ilaria Alpi, Italy, Miran Hrovatin, Mogadishu, Rai 3, Somalia}, 1994-03-20>
- <{Boom in the production of wheat, Epidemic, Stemrust, 1687 Peru earthquake, Peru, Chilean wheat cycle}, 1687>

Temporal Distribution



Clusters

#	Size	Key terms and entity types
1	25,149	Terms: released, album, band, studio, song Types: Person, Creator, Artist, Musician, Artifact
2	23,287	Terms: directed, film, starring, american, written Types: Movie, Person, Psychological Feat., (Social) Event
10	8,508	Terms: music, won, award, awarded, song Types: Person, Creator, Artist, Musician, Singer
11	7,971	Terms: party, elected, election, defeated, won Types: Person, Leader, Abstraction, (Social) Group
21	5,237	Terms: company, acquired, purchased, million, announced Types: Abstraction, (Social) Group, Organization, Institution

Events beyond Languages and Schema Constraints: Representation, Detection and Perception

Neutral Point of View (NPOV)

“All encyclopedic content on Wikipedia must be written from a neutral point of view (NPOV), which means representing fairly, proportionately, and, as far as possible, without editorial bias, all of the significant views that have been published by reliable sources on a topic.”

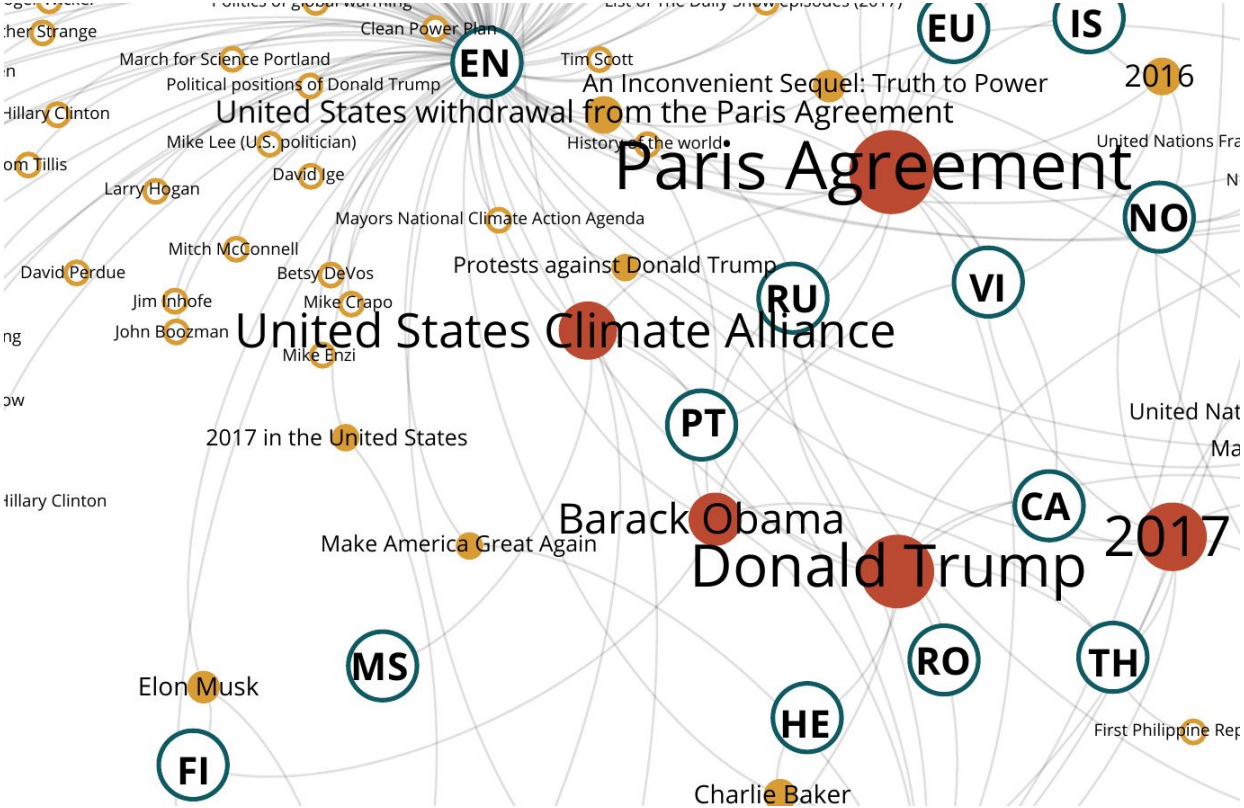
https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view

Case Study: Analysis Types and Feature Groups

- Brexit (B)
- US withdrawal from the Paris Agreement (P)

	Texts	Multimedia	Edits	Links	Categories
Content	B & P	B			
Temporal	B	P	B		
Network	P	P		B	B
Controversy			B		

Mentions of the Paris Agreement Withdrawal



Case Study: Analysis Types and Feature Groups

- Brexit (B)
- US withdrawal from the Paris Agreement (P)

	Texts	Multimedia	Edits	Links	Categories
Content	B & P	B			
Temporal	B	P	B		
Network	P	P		B	B
Controversy			B		

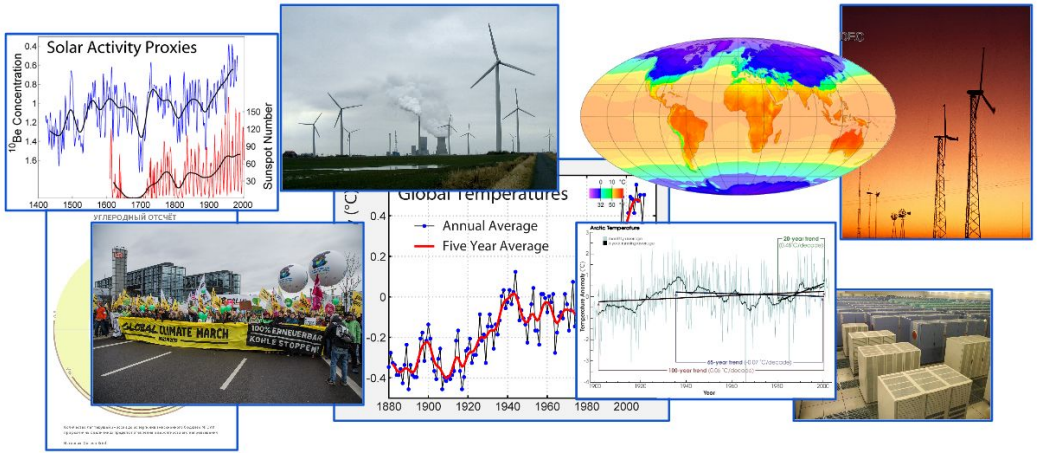


Image use before the US withdrawal

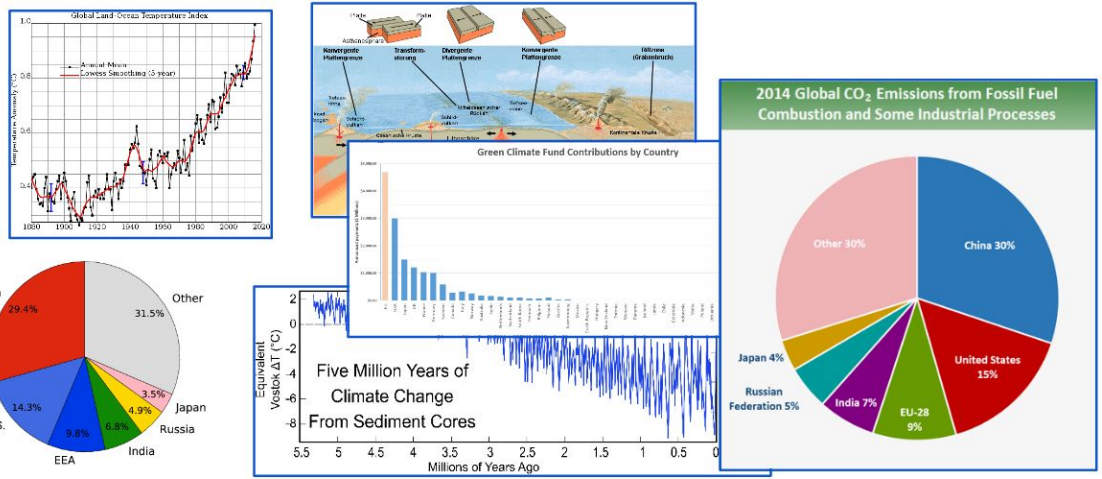


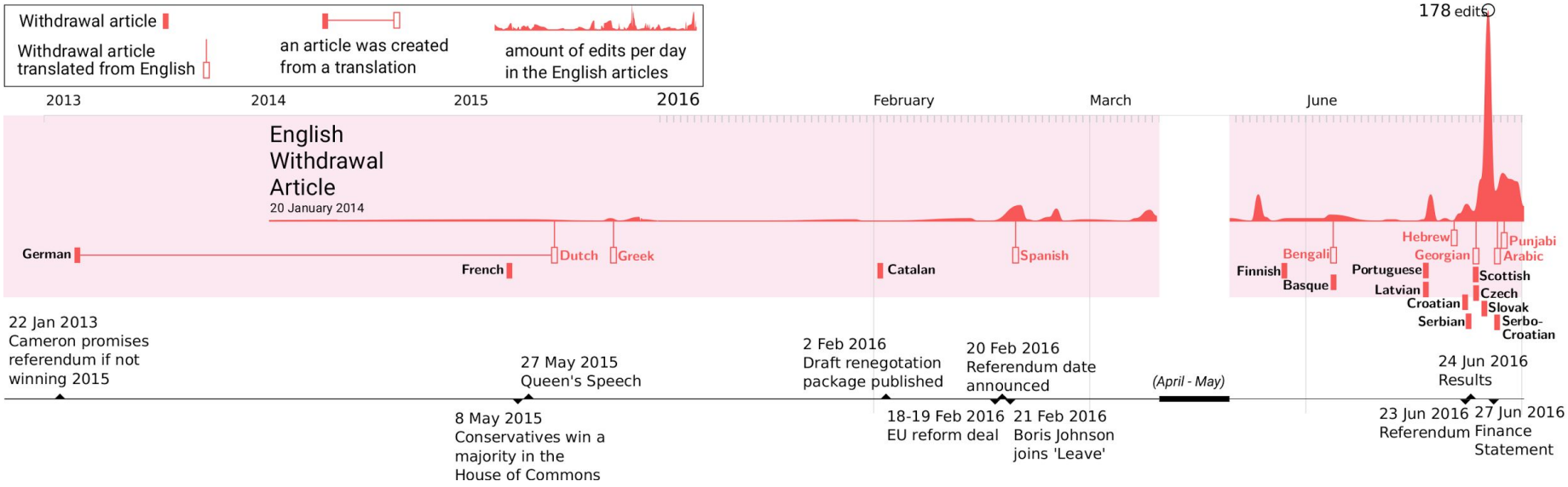
Image use after the US withdrawal

Case Study: Analysis Types and Feature Groups

- Brexit (B)
- US withdrawal from the Paris Agreement (P)

	Texts	Multimedia	Edits	Links	Categories
Content	B & P	B			
Temporal	B	P	B		
Network	P	P		B	B
Controversy			B		

Brexit: Article Creations across Languages



Conclusion

- Representation
  - Several types of events that require individual representations to acquire a complete overview of happenings
- Detection
  - Detection of important, historical events by exploiting temporal co-references between entities
- Perception
  - The coverage and description of events depends heavily on the temporal and language perspective

Events beyond Languages and Schema Constraints: Representation, Detection and Perception

Thank you —
Questions?

Simon Gottschalk
gottschalk@L3S.de

References

Gottschalk, Simon, and Elena Demidova. "MultiWiki: interlingual text passage alignment in Wikipedia." *ACM Transactions on the Web* (2016): 1-31.

Gottschalk, Simon, et al. "Ongoing Events in Wikipedia: A Cross-lingual Case Study." *Web Science '17*.

Rogers, Richard. *Digital methods*. MIT press, 2013.

Kuzey, Erdal, Jilles Vreeken, and Gerhard Weikum. "A fresh look on knowledge bases: Distilling named events from news." *CIKM '*14.

Allan, James, Ron Papka, and Victor Lavrenko. "On-line new event detection and tracking." *SIGIR* '98.

A. Abujabal and K. Berberich. Important Events in the Past, Present, and Future. In *WWW '15*.

A. Spitz and M. Gertz. Terms over LOAD: Leveraging Named Entities for Cross-Document Extraction and Summarization of Events. In *SIGIR '16*.