

TimeTrails: A System for Exploring Spatio-Temporal Information in Documents

Jannik Strötgen and Michael Gertz
Database Systems Research Group, University of Heidelberg

Motivation

Information Extraction

- a lot of information only published in unstructured format → textual documents

Spatial and Temporal Information

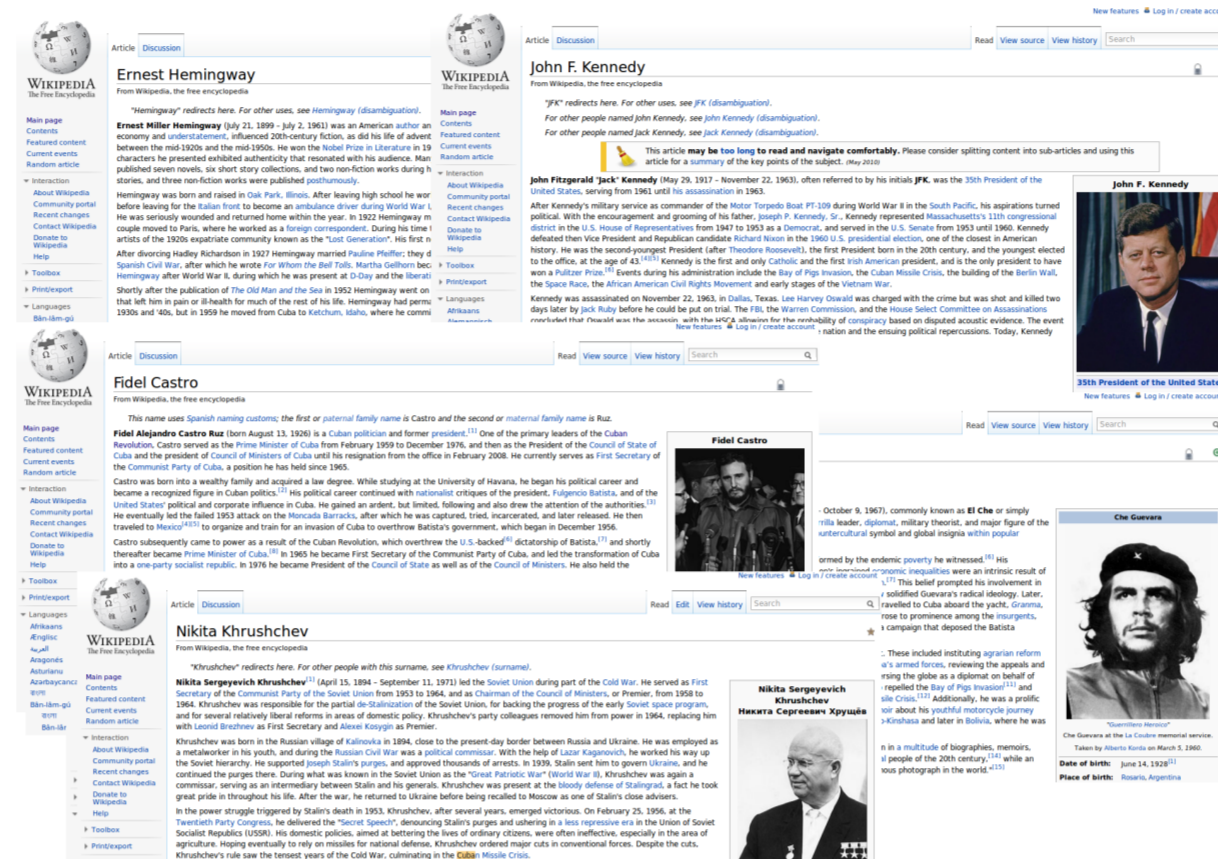
- widely spread in text documents
- can be extracted and normalized
- useful for search and exploration tasks

Events

- happen at specific place and time
- space/time as two dimensions of events
- co-occurrences of spatial and temporal expressions form events

Document Collection Exploration

- do documents talk about same events?



Some Wikipedia articles with a lot of spatial and temporal information. Are they talking about same events?

TimeTrails - Contributions

Extraction

- Spatio-temporal information as events

Storage

- spatio-temporal document profiles
- normalization: expression-independent
- PostGIS: spatio-temporal querying

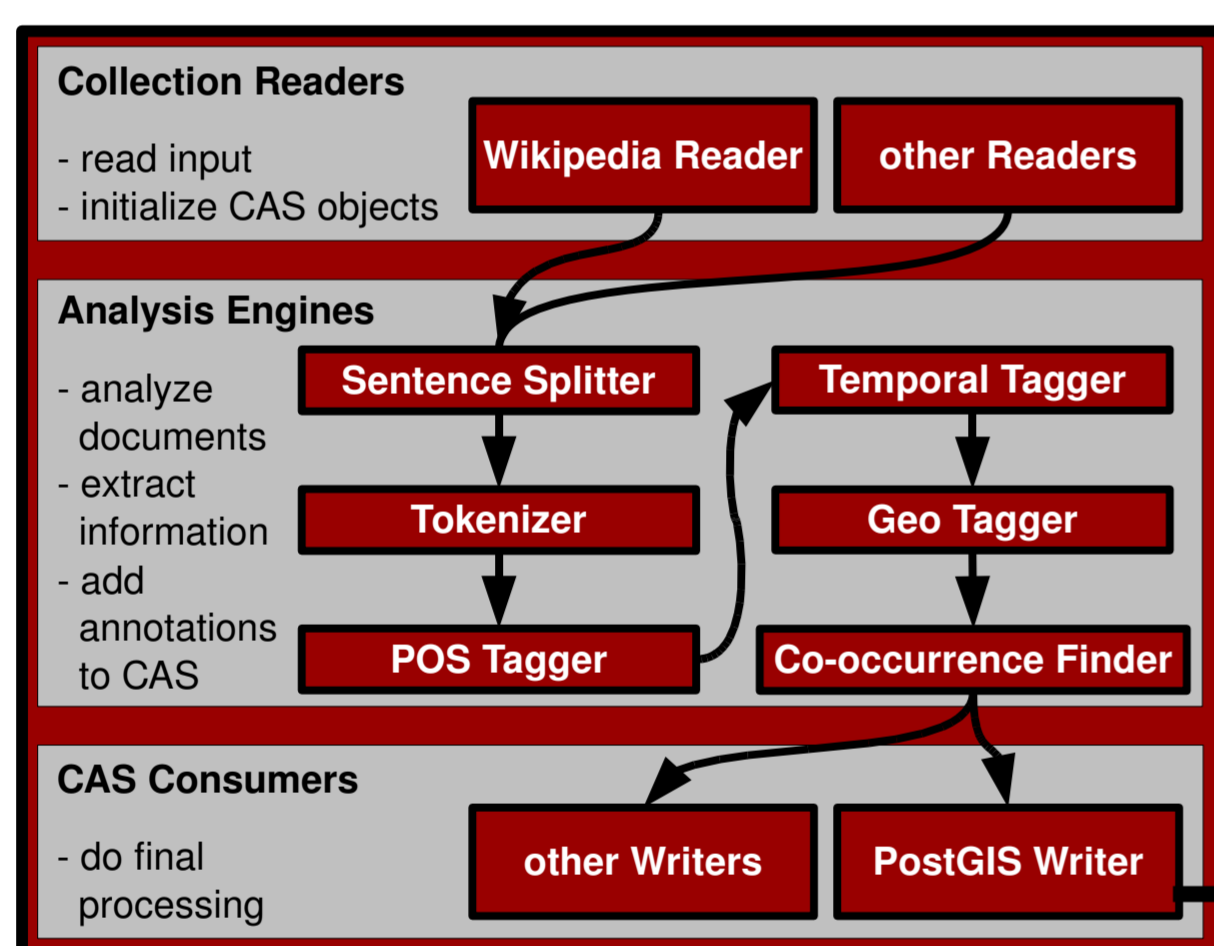
Querying

- temporal query constraints
- spatial query constraints

Exploration

- document trajectories
- multiple document visualization

Extraction and Storage



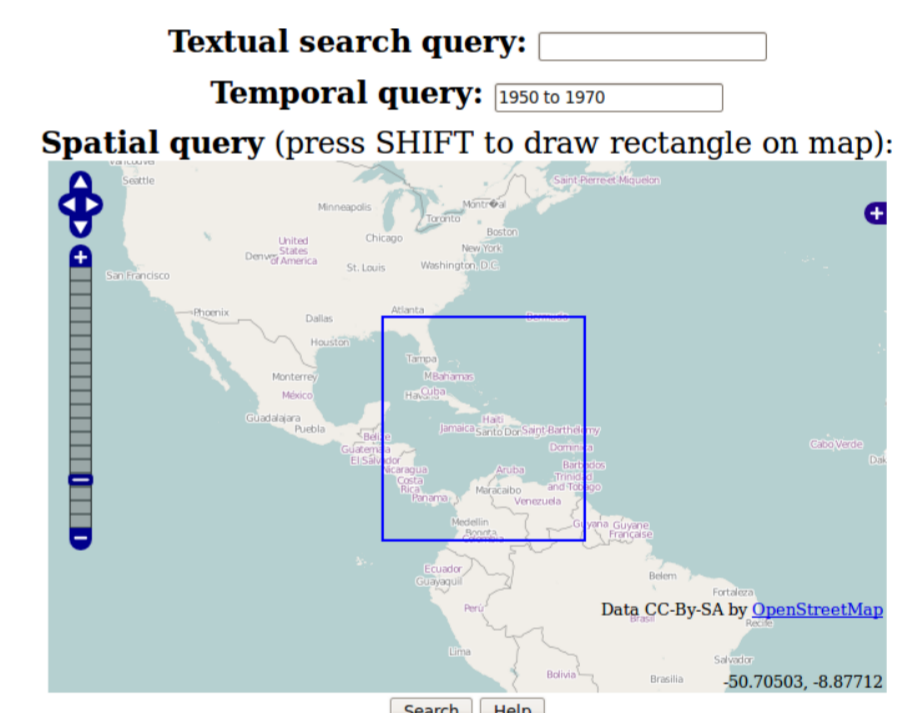
UIMA based Text Mining Pipeline

- all components use same data structure (CAS) → tools not built to be used together are easy to connect [1]

PostGIS Database

- all events are stored as tuples into spatio-temporal document profiles [2]
 $\langle \text{value}_t, \text{offset}_t, \text{value}_s, \text{offset}_s \rangle$

Querying



User interface to query the document collection with textual, temporal, and spatial constraints.

Exploration

Document Trajectories

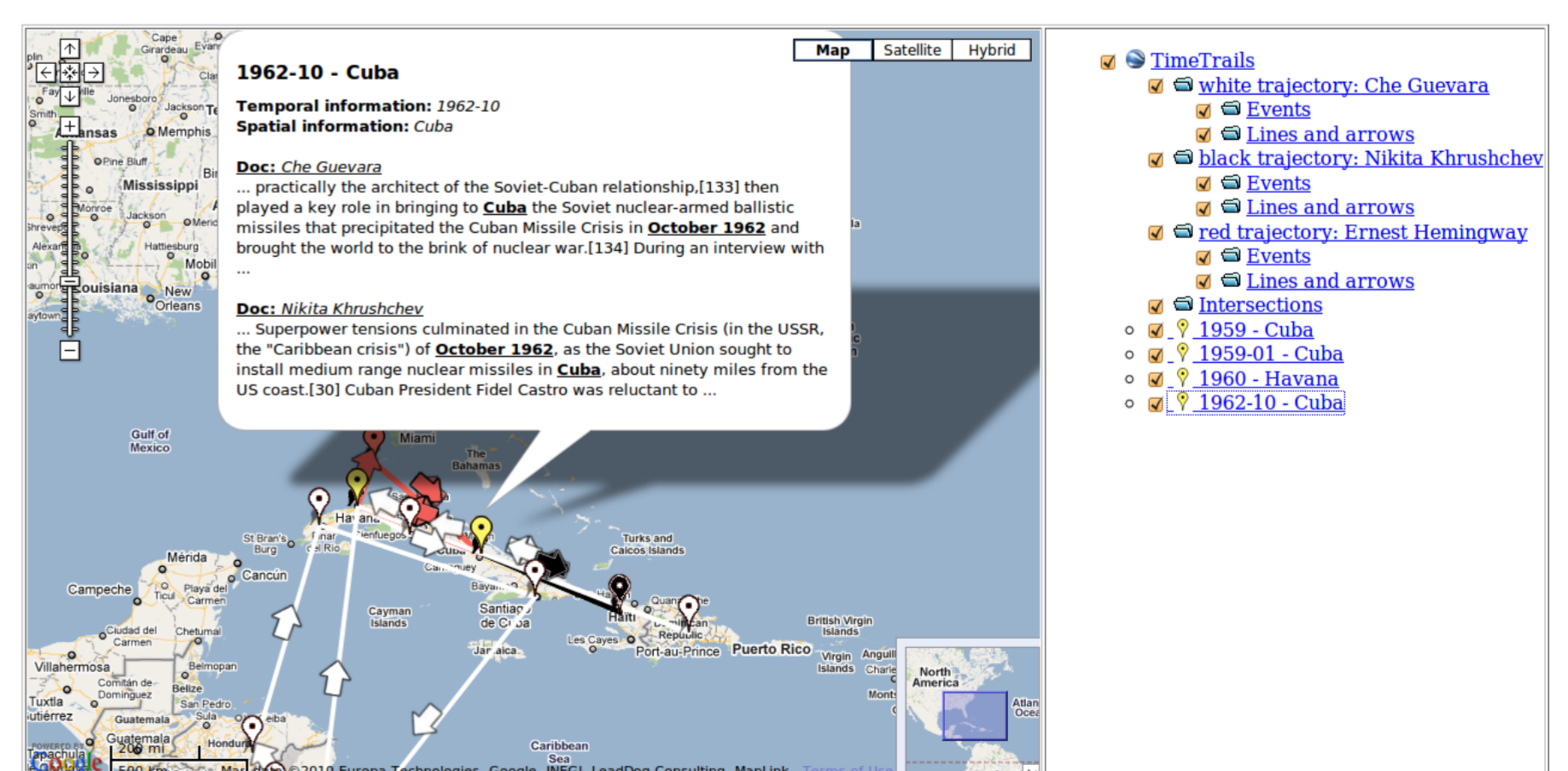
- textual documents visualized as document trajectories
- in the multiple document visualization (MDV) view intersections represent events described in more than one document

Query = text("*) AND time("1950 to 1970") AND window("88.058548121132" - "-59.933548122254");("2.951079525798" - "32.29393707526")
... [show internal details]

Results 1-10 of 119 for the given query.
MDV selected:

MDV	Document Title	Temporal distribution of events	Score	Events (total)	Events satisfying conditions (spatial / temporal / both)	Wikipedia Link
✓	Che Guevara (Category: People)	1928 - 2009	n/a	216	124 / 197 / 112	wiki link
□	Fidel Castro (Category: People)	1769 - 2010	n/a	148	91 / 57 / 32	wiki link
✓	Ernest Hemingway (Category: Literature and theatre)	1807 - 1996	n/a	160	46 / 35 / 14	wiki link
✓	Nikita Khrushchev	1804 - 2008	n/a	186	10 / 105 / 10	wiki link

User interface to explore the hit list with information on events.



Document trajectories of three documents with spatio-temporal snippets for an intersection (event 1962-10 / Cuba).

References

- [1] UIMA, <http://uima.apache.org/>
- [2] Jannik Strötgen and Michael Gertz: Extraction and Exploration of Spatio-Temporal Information in Documents. In *Proceedings of the 6th Workshop On Geographic Information Retrieval*, 2010.

Contact information

Jannik Strötgen
stroetgen@uni-hd.de
<http://dbs.ifi.uni-heidelberg.de/stixx>