**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 the same events?

**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

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

---

**References**


---

**Contact Information**

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

This work is presented at VLDB 2010, 36th International Conference on Very Large Databases, 13-17 September 2010, Singapore.