



Ontology-based Query Construction for GeoCLEF

Rui Wang & Günter Neumann

Saarland University & LT lab, DFKI

Saarbrücken, Germany





☆ Motivation

☆ Query Processing

- Topic Parsing
- Keyword Expansion
- Query Construction

☆ Results & Analysis

☆ Conclusion & Future Work





☆ The IR System

- Query Processing
- Document Retrieval
- Document Ranking

☆ Traditional IR vs. Geographic IR

- Geographic variation





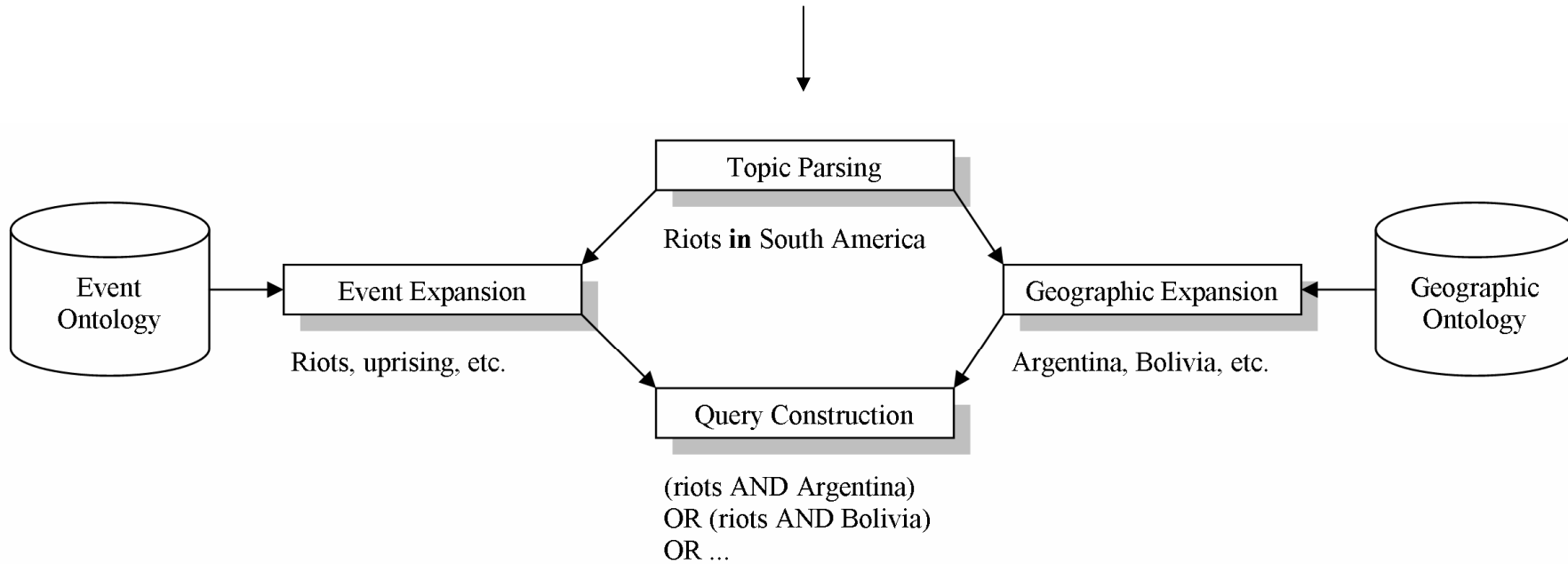
☆ Topic Parsing

☆ Ontology-based Keyword Expansion

- Geographic Expansion
- Event Expansion

☆ Query Construction





☆ Riots in South America

- Event: Riots, uprising, etc.
- Geographic: Argentina, Bolivia, etc.
- Query: (riots AND Argentina) OR (riots AND Bolivia) OR ...



☆ The *Event* part and the *Geographic* part

- *Riots in South American prisons*
- *Nobel Prize winners from Northern European countries*
-
- *Most visited sights in the capital of France and its vicinity*



<http://www.world-gazetteer.com/>

<http://www.geonames.org/>

Planet (i.e. Earth) --*part-of*-- Continent --*part-of*-- Country -- $\left\{ \begin{array}{l} \text{--}i\text{part-of}\text{-- City/Town/... (artificial)} \\ \text{--}i\text{part-of}\text{-- River/Island/... (natural)} \end{array} \right.$

☆ Other links:

- Subcontinent: *the Indian subcontinent, the Persian Gulf*, etc.
- Subcountry: *Lower Saxony, the Western USA*, etc.
- Organization: *the Organization for Economic Co-operation and Development (OECD)*, etc.
- Others: *Spanish islands*, etc.

☆ The *equal* relation:

- E.g. *the United Kingdom, the UK, Great Britain*, etc.





☆ Natural disasters

- Earthquakes: *San Francisco Earthquake (1906)*, *Good Friday Earthquake Earthquake (1964)*, etc.

☆ Human Events and others

- Nobel Prize winners: *Marie Curie (Russian Poland, Physics, 1903)*, *Albert Einstein (Germany, Physics, 1921)*, *Mother Teresa (Albania, Peace, 1979)*, etc.



- ☆ Used for populating the ontologies:
 - Online: Automatically from narratives tag using NER
 - Offline: Manually from Wikipedia

- ☆ Usage of core ontology:
 - If the geographic part contains the granularity of the basic terms, e.g. country, city, etc, the ontology will provide all the geographic terms at that level;
 - Otherwise, the ontology will provide all the geographic terms below the level of that term.



- ☆ Level 4 (1000)
 - the event ontology AND the geographic ontology
- ☆ Level 3 (100)
 - the event terms AND the geographic ontology
- ☆ Level 2 (10)
 - the event terms AND the geographic terms
- ☆ Level 1 (1)
 - the event terms OR the geographic terms
- ☆ Ontology
 - 1) from narratives (auto);
 - 2) from wikipedia (manual)
- ☆ Terms: from titles, but not narratives



☆ Run1 (M)

- Use queries from Level 1~4 and both ontologies are constructed with Wikipedia information

☆ Run2 (A)

- Similar to Run1, but both ontologies are constructed with narratives

☆ Run3 (M)

- Use queries from Level 1~3 and the ontology is constructed with Wikipedia information

☆ Run4 (A)

- Similar to Run3, but the ontology is constructed with narratives

☆ Run5 (A)

- Use queries from Level 1~2



Submissions	R-Prec	MAP
Run1 (M)	33.38% (1/68)	29.18% (3/68)
Run2 (A)	33.19% (2/68)	29.24% (2/68)
Run3 (M)	31.70% (3/68)	30.37% (1/68)
Run4 (A)	31.41% (4/68)	27.73% (6/68)
Run5 (A)	20.95% (58/68)	16.07% (68/68)

- ☆ Run2 is the best automatic
- ☆ Run1 and Run3 have the best R-Prec and MAP respectively
- ☆ Run5 (without query expansion) is rather poor



- ☆ The Query Expansion is important for Geographic IR
- ☆ Manually designed ontology does help
- ☆ The framework can be extended to other languages



☆ Lexical semantics

- For topic/query analysis, e.g., semantics of prepositions

☆ Document ranking

- Context window to control the distance between the event and the geographic term

☆ Multilinguality

- German



Done!
Thank you!

