University of Hagen at GeoCLEF 2005: Using Semantic Networks for Interpreting Geographical Queries

Johannes Leveling, Sven Hartrumpf, Dirk Veiel
Intelligent Information and Communication Systems (IICS)
University of Hagen (FernUniversität in Hagen)
58084 Hagen, Germany

GeoCLEF @ CLEF 2005, Vienna, Austria

Synonymy and Polysemy - A Problem in GIR?

- Compare charcteristics of concepts in a lexical net (GermaNet) and toponyms in the GeoNet Names Server (GNS) data

Characteristic	Resource	
	GermaNet	GNS (DE, normalized)
synsets total	41,777	95,993
synonyms in synsets	60,646	103,508
unique literals	52,251	80,808
synonyms per synset	1.45	1.08
word senses per literal	1.16	1.28

Problems Less Known for GIR

- **Toponyms in different languages**
  $\rightarrow$ access to data often requires translation
- **Name variants**
  "Köln", *"Cologne"*, *"Cöllen"*
- **Composite names**
  "Haren (Ems)", *"Frankfurt/Oder"*, or *"Freiburg im Breisgau"*
- **Semantic relations between toponyms and related concepts**
  "Scottish", *"Scotsman"*, or *"Scottish districts"* $\rightarrow$ *"Scotland"*
- **Temporal changes in toponyms**
  "the EU" $\rightarrow$ the European Union (before or after the expansion?)
- **Metonymic usage**
  "British soil" $\rightarrow$ *"Great Britain"*

Expanding Geographical Background Knowledge (I)

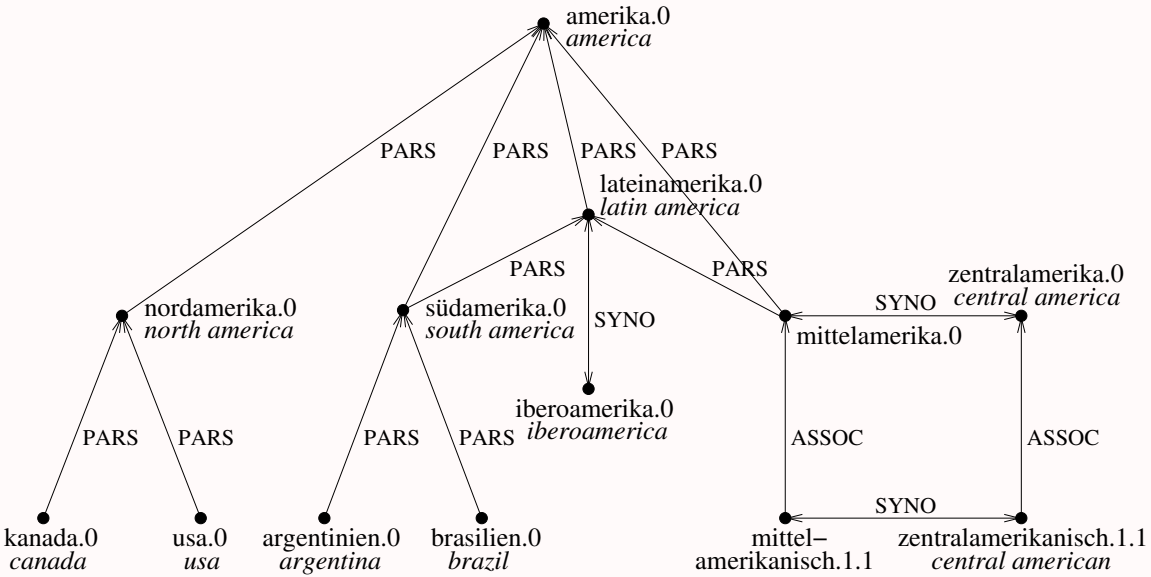
- **General idea:**
  Generate concept hypotheses and meronymy relations
- Concatenate some prefix with regular semantics in geography with the original concept
- Use typical prefixes:
  'Southeast'/*"Südost-"*, *'Central'*/*"Zentral-"*,*'Middle'*/*"Mittel-"*, *"Left"*/*"Links-"*, *"Upper"*/*"Ober-"*
- Reject concept hypotheses with a frequency less than a given threshold
- Add inferred meronymy relation to knowledge base

Expanding Geographical Background Knowledge (II)

- GNS data contains encoding of path to geographical entity
  → RC:1;CC1:AU;ADM1:09;GC:ADM1;FC:A
- Decode information and transform into a set of semantic relations
  → *“Wien”*/*‘Vienna’* is located
  - in *“Amerika oder Westeuropa”*/*‘America or Western Europe’*,
  - in *“Europa”*/*‘Europe’*,
  - in *“Österreich”*/*‘Austria’*, and
  - in the *“Bundesland Wien”*

  → *“Vienna”* is a name variant of *“Wien”*: EQU(“Vienna”, “Wien”)

Excerpt from the Semantic Network Representation (MultiNet) of our Geographical Knowledge Base



Some Important Semantic Relations

MultiNet Relation	Description
EQU(x,y)	name variants; endonyms, exonyms
LOC(x,y)	specifying locations; x takes place at y
PARS(x,y)	meronymy, holonymy; x is part of y
SUB(x,y)	subordination; x is a y
SYNO(x,y)	synonyms and near-synonyms
TEMP(x,y)	temporal specification
*NEAR(x,y)	x is close to y
*OUTSIDE(x,y)	x is not inside y
*SOUTH_OF(x,y)	x is south of y

Conclusion and Future Work

- We provided a basic architecture for further experiments in GIR
- The combination of traditional IR and question answering approach for GIR looks promising
- Future work:
  - Improve named entity recognition
  - Connect semantic networks and databases
  - Expand geographical background knowledge,
  - Investigate the role of semantic relations in geographical queries

References

[Hammer et al.(2005)Hammer, Dickmeiss, Levanto, and Taylor] Hammer, Sebastian; Adam Dickmeiss; Heikki Levanto; and Mike Taylor (2005). *Zebra – User's Guide and Reference*. Copenhagen.

[Hartrumpf(2005)] Hartrumpf, Sven (2005). Question answering using sentence parsing and semantic network matching. In *Multilingual Information Access for Text, Speech and Images: Results of the Fifth CLEF Evaluation Campaign* (edited by Peters, C.; P. D. Clough; G. J. F. Jones; J. Gonzalo; M. Kluck; and B. Magnini), volume 3491 of *Lecture Notes in Computer Science (LNCS)*, pp. 512–521. Berlin: Springer.

[Helbig(2005)] Helbig, Hermann (2005). *Knowledge Representation and the Semantics of Natural Language*. Berlin: Springer.

[Leveling and Hartrumpf(2005)] Leveling, Johannes and Sven Hartrumpf (2005). University of Hagen at CLEF 2004: Indexing and translating concepts for the GIRT task. In *Multilingual Information Access for Text, Speech and Images: Results of the Fifth CLEF Evaluation Campaign* (edited by Peters, C.; P. D. Clough; G. J. F. Jones; J. Gonzalo; M. Kluck; and B. Magnini), volume 3491 of *Lecture Notes in Computer Science (LNCS)*, pp. 271–282. Berlin: Springer.