Semantic Annotation for Multilingual Search

Shibamouli Lahiri
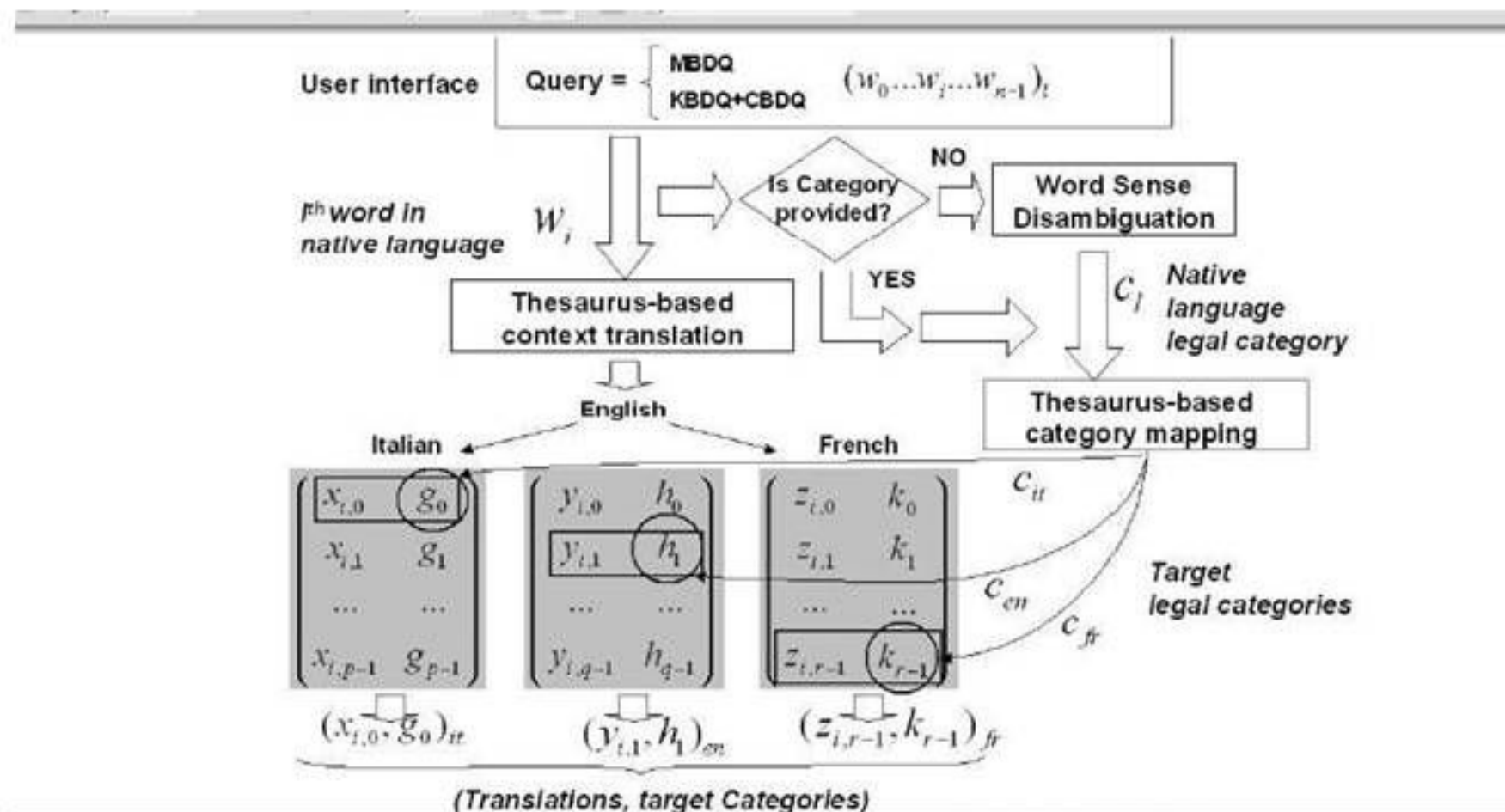
(shibamouli@cse.psu.edu)

Semantic Annotation

- Tags describe the content/meaning/sense of a particular piece of text, e.g. a sentence, a paragraph or a document

- Tags are given in natural language, so they need disambiguation ("jaguar" – an animal or a car)

Multilingual Access Modalities to Legal Resources Based on Semantic Disambiguation (Peruginelli and Francesconi, 2009)

- Categorization of metadata (tags)
- Metadata-based, keyword-based and category-based querying
- Query language-independent metadata (document title)
- English is pivot language

Multilingual Access Modalities to Legal Resources Based on Semantic Disambiguation (Peruginelli and Francesconi, 2009)

Semantic Annotation for Concept-Based Cross-Language Medical Information Retrieval (Volk et al., 2002)

- English – German parallel corpus of 9000 documents

- Concept annotation with EuroWordNet

- Concept annotation with Medical Schema

Semantic Annotation for Concept-Based Cross-Language Medical Information Retrieval (Volk et al., 2002)

	mAvP	Rel. Docs Retr.	AvP 0.1	P10
DE2EN-DE-token	0.0512	66	0.1530	0.1160
DE2EN-MT-PT2001	0.1184	376	0.3382	0.2520
DE2EN-EWN	0.0090	111	0.0311	0.0160
DE2EN-UMLS	0.1620	366	0.3724	0.2800
DE2EN-MeSH	0.1699	304	0.3888	0.2600
DE2EN-Semrel	0.0229	23	0.0657	0.0480
DE2EN-all-combined	0.1774	404	0.3872	0.2720
DE2EN-SimThes	0.2290	409	0.4492	0.3640
DE2EN-SimThes+all-comb.	0.2955	518	0.5761	0.4600

Table 3: Results of the cross-language runs: German queries and English documents

Other Pointers

- The TRUST project (English, French, Italian, Polish, Portuguese)

- Workshop on Multilingual Semantic Web (WWW 2010)

References

1. **Semantic Annotation, Indexing and Retrieval.** A. Kiryakov, B. Popov, D. Ognyanoff, D. Manov, A. Kirilov, M. Goranov. *International Semantic Web Conference (2003).*

2. **Multilingual Access Modalities to Legal Resources Based on Semantic Disambiguation.** G. Peruginelli and E. Francesconi. *3rd Workshop on Legal Ontologies and Artificial Intelligence Techniques (2009).*

3. **Semantic Annotation for Concept-Based Cross-Language Medical Information Retrieval.** M. Volk, B. Ripplinger, S. Vintar, P. Buitelaar, D. Raileanu, B. Sacaleanu. *International Journal of Medical Informatics (2002).*