Irina Temnikova: processing patent texts

Important: because

There are more than 40 million patent texts worldwide

Searching is a non-trivial task: every inventor is its own lexicographer, aiming to describe the invention and claims as vague as possible and as broad as possible

Definitions are too descriptive: features of the subjects are documented, but direct terminological references are avoided:
e.g. computer = a system having storage for storing data files, an output device to display …. a component that modifies input data and controls flow of data …

Idea – to add metadata (indices) to patents, using external encyclopedic resources, e.g. Wikipedia

Research on patent search

Experiments with a subset of English Wikipedia (about 1500 articles) and a subset of patents prepared by TU Vienna

Together with Ivelina Nikolova, PhD student at IICT-BAS

Measuring similarity between patent text and Wiki-article (if “similar”/”close”, the article title is assigned as an index, and its multilingual translations are assigned as well)

“Gold standard”: manually selected closest Wiki-articles for about 100 patents

Results: the closest Wiki-article is identified by most IR approaches but there is (relatively much) noise: other articles are found as well

Papers published August-September 2013

1. I. Temnikova, K. B. Cohen: *Recognising Sublanguages in Scientific Journal Articles through Closure Properties*. In Proc. Bio-NLP 2013 Workshop, associated with the 51th Annual Conference of ACL, pp. 72-79, in the ACL Anthology

2. I. Temnikova, I. Nikolova, W. Baumgarther, G. Angelova and K. B. Cohen: *Closure Properties of Bulgarian Clinical Texts*. In Proc. RANLP 2013, pp. 664-671, in the ACL Anthology

3. I. Temnikova, N. Hailu, G. Angelova and K. B. Cohen: *Measuring Closure Properties of Patent Sublanguages*. In Proc. RANLP 2013, pp. 672-679, in the ACL Anthology

4. I. Nikolova, I. Temnikova, and G. Angelova: *Can Patent Search be enriched with Wikipedia Articles Keywords?* In Proc. RANLP 2013, pp. 570-576, in the ACL Anthology