

Light Weight Semantics & Language Independent Document Representation

Marko Grobelnik
Jozef Stefan Institute
Ljubljana, Slovenia
(also on behalf of CycCorp Europe)

Light Weight Semantics

- One of the big next goals in search is “**understanding the text**”
 - Full text understanding is still not feasible
 - ...either too expensive or too complex
 - ...with today's technology it is possible to exploit and deal with the structure of the text:
 - Key technologies are: *Computation Linguistics + Machine Learning + Semantic Web (KRR)*
 - In brief, the idea is to: *deep parse text, co-reference resolution, fact extraction, connecting & verifying facts to domain ontology*
 - ...compare approaches from Leskovec et al (AAAI2005), Witbrock et al (AAAI2005), Etzioni (WWW2004) and more
 - ...resulting search over deeply processed documents enables **qualitative advantage** over traditional keyword based search



