

Web Search: Techniques, algorithms and Applications

German Rigau <german.rigau@ehu.es>



Content

- NLP Tools & Resources for Web Search (German Rigau)
 - 7h (**10-13/NOV**)
- Hands-on with NLP tools (Rodrigo Agerri)
 - 2h (**16/NOV**)
- Basic Techniques for Web Search (German Rigau)
 - 8h (**17-20/NOV**)
- Hands-on with search engine (Rodrigo Agerri)
 - 2h (**23/NOV**)
- Projects (students)
 - Discuss proposals (**20/NOV in class**)
=> **Deadline for title and one paragraph summary 23/NOV**
 - Discussion by email (23-26/NOV)
 - Oral presentations of design: 2h (**27/NOV**)
 - **Deadline for full report 18/DEC**

Content

- Basic Techniques for Web Search
 - Information Retrieval (IR)
 - Information Extraction (IE)
 - Question Answering (Q&A)
 - Clustering
 - Classification
 - Summarisation
 - Multilingüism
 - Cross-lingual Information Retrieval (CLIR)
 - Machine Translation (MT)

Content

- NLP Tools
 - Basic Tools
 - Tokenization
 - Sentence Splitting
 - Language Identifiers
 - Lemmatization, POS tagging
 - Named Entity Recognizers and Categorizers (NERC)
 - Parsing
 - Word Sense Disambiguation (WSD)
 - Semantic Role Labelling (SRL)

Content

- Resources
 - Words & Works
 - Ontologies:
 - Mikrokosmos
 - Large-scale Knowledge Bases:
 - WordNet & EuroWordNet
 - More large-scale resources
 - ConceptNet, Framenet, VerbNet, PropBank, ...
 - Building Wordnets
 - WordNet extensions:
 - SUMO ontology, eXtended WordNet, Meaning project
 - Reasoning

Evaluation

- Applications of Web Search:
 - Student/teacher topic
 - Short presentation
 - 10 minutes sharp, ~ 10 slides
 - Presentation: 27/11
 - Written report:
 - Format: <http://www.acl2013.org/site/call.html>
 - Deadline Report: 18/12
 - Two options
 - Report
 - < 2000 words
 - Short paper describing an experimental work
 - < 3000 words

Web Search: Techniques, algorithms and Applications

German Rigau <german.rigau@ehu.es>

