

Web Search: Techniques, algorithms and Aplications

Eneko Agirre <e.agirre@ehu.edu>
German Rigau <german.rigau@ehu.es>



Content

- Second slot: 17:45 – 19-45
 - EXCEPT: 15/11 **17:00-18:30**
- Basic Techniques for Web Search (Eneko Agirre)
 - 8h (12-15/11)
- NLP Tools & Resources for Web Search (German Rigau)
 - 8h (16,19,20,21/11)
- Hands-on with search engine (Arantxa Otegi)
 - 2h (22/11)
- Applications (Students)
 - Proposals: 2h (23/11)
 - Presentations: 2h (30/11)

Content

- Basic Techniques for Web Search (Eneko Agirre)
 - 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 (German Rigau)
 - 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 (German Rigau)
 - 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
- Two options:

Deadline to decide: 23/11

- Short presentation + report
 - ~ 20 slides + 2000 words, format <http://www.acl2011.org/call.shtml>
 - Presentation: 30/11
 - Report: 31/12
- Short paper describing an experimental work
 - ~ 3000 words, format: <http://www.acl2011.org/call.shtml>
 - Deadline : 31/12

Web Search: Techniques, algorithms and Aplications

Eneko Agirre <e.agirre@ehu.edu>
German Rigau <german.rigau@ehu.es>

