

Web Search: Techniques, algorithms and Applications

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



Content

- Introduction (German Rigau)
 - 2h (14/02: 17:30h - 19:30h)
- Tools & Resources (German Rigau)
 - 8h (15-18/02: 17:30h - 19:30h)
- Applications I (Eneko Agirre)
 - 6h (21-24/02: 17:30h - 19:30h)
- Applications II (students)
 - Proposals: 2h (25/02: 17:30h - 19:30h)
 - Presentations: 4h (3-4/03: 17:30h - 19:30h)

Content

- Introduction (German Rigau)
- 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

Content

- Applications I (Eneko Agirre) and II (students)
 - Information Retrieval (IR)
 - Information Extraction (IE)
 - Question Answering (Q\A)
 - Clustering
 - Classification
 - Summarisation
 - Multilingüism
 - Cross-lingual Information Retrieval (CLIR)
 - Machine Translation (MT)

Evaluation

- Student/teacher topic
- Two options:
 - Short presentation + homework
 - ~ 20 slides + 2000 words
 - Deadline : 3-4/03
 - Short paper describing an experimental work
 - ~ 3000 words
 - Format: <http://www.acl2011.org/call.shtml>
 - Deadline : 4/04

Web Search: Techniques, algorithms and Applications

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

