Web Search:
Techniques, algorithms and Aplications

Eneko Agirre <e.agirre@ehu.edu>

German Rigau <german.rigau@ehu.es>

Content

- Basic Techniques for Web Search (Eneko Agirre)
  - 8h (13-16/02: 17:30h – 19:30h)

- NLP Tools & Resources for Web Search (German Rigau)
  - 8h (20-23/02: 17:30h – 19:30h)

- Applications (students)
  - Proposals: 2h (24/02: 17:30h - 19:30h)
  - Presentations: 4h (1/03: 17:30h – 19:30h)

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: 24/02
  - Short presentation + report
    - ~ 20 slides + 2000 words
    - Presentation: 1/03
    - Report: 4/04
  - Short paper describing an experimental work
    - ~ 3000 words
    - Format: http://www.acl2011.org/call.shtml
    - Deadline : 4/04

Web Search:
Techniques, algorithms and Aplications

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