

# Web Search: Techniques, algorithms and Applications

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



# Content

- First slot: 15:00h – 17:00
- Basic Techniques for Web Search (Eneko Agirre)
  - 8h (**11-14/11**)
- Hands-on with search engine (Arantxa Otegi & Eneko Agirre)
  - 2h (**15/11**)
- NLP Tools & Resources for Web Search (German Rigau)
  - 8h (**18-20/11, 22/11**)
- Hands-on with NLP (Rodrigo Agerri & German Rigau)
  - 2h (**25/11?**)
- Applications (students)
  - Proposals: **21/11 no class, 2h**  
=> **Deadline for the title and one paragraph summary 22/11**
  - Discussion by email or office hours (25-28/11)
  - Presentations: 2h (**29/11**)
  - **Deadline for reports 20/12**

# 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
  - Short presentation
    - ~ 10 slides
    - Presentation: **29/11**
  - Written report:
    - Format: <http://www.acl2013.org/site/call.html>
    - Deadline Report: **20/12**
    - Two options
      - Report
        - < 2000 words
      - Short paper describing an experimental work
        - < 3000 words

# Web Search: Techniques, algorithms and Applications

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

