

# Web Search: Techniques, algorithms and Applications

Eneko Agirre <[e.agirre@ehu.edu](mailto:e.agirre@ehu.edu)>  
German Rigau <[german.rigau@ehu.es](mailto:german.rigau@ehu.es)>



# Content

- Basic Techniques for Web Search (Eneko Agirre)
  - 8h (**10-13/NOV**)
- Hands-on with search engine (Arantxa Otegi & Eneko Agirre)
  - 2h (**18/NOV**)
- NLP Tools & Resources for Web Search (German Rigau)
  - 8h (**14,17,19,21/NOV**)
- Hands-on with NLP tools (Rodrigo Agerri & German Rigau)
  - 2h (**20/NOV**)
- Projects (students)
  - Discuss proposals (**24/NOV in class**)  
=> **Deadline for title and one paragraph summary 24/NOV**
  - Discussion by email or office hours (25-27/NOV)
  - Oral presentations of design: 2h (**28/NOV**)
  - **Deadline for full report 19/DEC**

# 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 minutes sharp, ~ 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 Aplications

Eneko Agirre <[e.agirre@ehu.edu](mailto:e.agirre@ehu.edu)>  
German Rigau <[german.rigau@ehu.es](mailto:german.rigau@ehu.es)>

