

Advanced Methods for Corpus Processing

Lluís Padró <padro@lsi.upc.edu>

Lluís Màrquez <lluism@lsi.upc.edu>

German Rigau <german.rigau@ehu.es>



**Centre de Tecnologies i Aplicacions
del Llenguatge i la Parla**

Foreword

- **UPC: Artificial Intelligence**
 - EMNLP: Empirical Methods for NLP
 - 2001/2002-2002/2003
 - Lluís Padró, Lluís Màrquez, German Rigau
 - 2003/2004-2004/2005
 - Lluís Padró, Lluís Màrquez, Neus Català, German Rigau
 - 2005/2006
 - L. Padró, L. Màrquez, X. Farreres, J. Daudé, G. Rigau
- **EHU: NLP**
 - AMCM: Advanced Methods for Corpus Management
 - 2004/2005-2005/2006-2007/2008
 - Lluís Padró, Lluís Màrquez, German Rigau

Content

- Statistical methods for NLP (Lluís Padró)
 - 10h (Monday 28/1, Tuesday 29/1, Wednesday/1 30: 16h – 20h)
- Machine Learning for NLP (Lluís Màrquez)
 - 10h (Wednesday 30/1, Thursday 31/1, Friday 1/2: 16h – 20h)
- Ontologies & large-scale Knowledge Bases for NLP (German Rigau)
 - 9h (Tuesday 4/3, Thursday 6/3, Tuesday 11/3: 16h – 19h)
- Concluding remarks (German Rigau)
 - 3h (Thursday 13/3: 16h – 19h)

Content

- Statistical methods for NLP (Lluís Padró)
 - Mathematical Foundations
 - Linguistic Foundations
 - Corpora
 - Statistical Inference
 - Markov Models
 - Tagging
 - Maximum Entropy Models

Content

- Machine Learning for NLP (Lluís Màrquez)
 - Introduction: Machine Learning and Machine Learning for NLP
 - Machine Learning: Classical Methods from AI
 - Margin-based Machine Learning Algorithms
 - Machine Learning for NLP
 - Applications

Content

- Ontologies & Knowledge-based for NLP (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
- Concluding remarks (German Rigau)
 - Reasoning