UCL Université catholique de Louvain

LINGI2263

2012-2013

Computational Linguistics

5.0 credits	30.0 h + 15.0 h	2q

Teacher(s) :	Fairon Cédrick ; Dupont Pierre (coordinator) ;
Language :	Anglais
Place of the course	Louvain-la-Neuve
Prerequisites :	-- algorithmics and preferably basic knowledge in machine learning (as provided by SINF1121 and ING2262)
Main themes :	-- Basics in phonology, morphology, syntax and semantics -- Linguistic resources -- Part-of-speech tagging -- Statistical language modeling (N-grams and Hidden Markov Models) -- Robust parsing techniques, probabilistic context-free grammars -- Linguistics engineering applications such as spell or syntax checking software, POS tagging, document indexing and retrieval, text categorization
Aims :	Students completing successfully this course should be able to -- describe the fundamental concepts of natural language modeling -- master the methodology of using linguistic resources (corpora, dictionaries, semantic networks, etc) and make an argued choice between various linguistic resources -- apply in a relevant way statistical language modeling techniques -- develop linguistic engineering applications Students will have developed skills and operational methodology. In particular, they have developed their ability to -- integrate a multidisciplinary approach to the edge between computer science and linguistics, using wisely the terminology and tools of one or the other discipline, -- manage the time available to complete mini-projects, -- manipulate and exploit large amounts of data. *The contribution of this Teaching Unit to the development and command of the skills and learning outcomes of the programme(s) can be accessed at the end of this sheet, in the section entitled "Programmes/courses offering this Teaching Unit".*
Evaluation methods :	-- 25% miniprojects -- 75% final exam
Teaching methods :	-- 12 lectures -- 3 miniprojects -- feedback sessions about the miniprojects
Content :	-- Linguistic essentials: morphology, part-of-speech, phrase structure, semantics and pragmatics -- Mathematical foundations: formal languages, and elements of information theory -- Corpus analysis: formating, tokenization, morphology, data tagging -- N-grams: maximum likelihood estimation and smoothing -- Hidden Markov Models: definitions, Baum-Welch and Viterbi algorithms -- Part-of-Speech Tagging -- Probabilistic Context-Free Grammars: parameter estimation and parsing algorithms, tree banks -- Machine Translation: classical and statistical methods (IBM models, Phrase-based models), evaluation -- Applications: SMS predictors, POS taggers, information extraction
Bibliography :	Main reference: -- Speech and Language Processing (2nd Edition), D. Jurafsky and J.H. Martin, Prentice Hall, 2009. This reference is highly recommended but not mandatory to follow the course. Mandatory material: The mandatory material for this course is defined as the set of documents and slides made available on the icampus website, together with the oral communications and talks given during the weekly lectures. None of this material can be consulted during the final examination (closed book exam). Additional references: -- Foundations of Statistical Natural Language Processing, C. Manning and H. Schutze, MIT Press, 1999. -- The Oxford Handbook of Computational Linguistics, Ruslan Mitkov (Editor), Oxford University Press, 2003. -- Ingénierie des langues. Sous la dir. de J.M. Pierrel, Hermes Science Publications, 2000.

Cycle and year of study :	> Master [120] in Computer Science and Engineering > Master [120] in Computer Science > Master [120] in Linguistics > Master [120] in Statistics: General
Faculty or entity in charge:	INFO