UCL Université catholique de Louvain

LINGI2263

2015-2016

Computational Linguistics

5.0 credits	30.0 h + 15.0 h	1q

Teacher(s) :	Dupont Pierre ; Fairon Cédrick ;
Language :	Anglais
Place of the course	Louvain-la-Neuve
Inline resources:	> http://moodleucl.uclouvain.be/course/view.php?id=7865
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 :	Given the learning outcomes of the "Master in Computer Science and Engineering" program, this course contributes to the development, acquisition and evaluation of the following learning outcomes: -- INFO1.1-3 -- INFO2.3-4 -- INFO5.3-5 -- INFO6.1, INFO6.4 Given the learning outcomes of the "Master [120] in Computer Science" program, this course contributes to the development, acquisition and evaluation of the following learning outcomes: -- SINF1.M4 -- SINF2.3-4 -- SINF5.3-5 -- SINF6.1, SINF6.4 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% for practical works + 75% final exam (closed book) No possibility to present again practical works in the second session
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 http://moodleucl.uclouvain.be/course/view.php?id=7865
Bibliography :	Required slides available at: http://moodleucl.uclouvain.be/course/view.php?id=7865 1 textbook recommended: -- Speech and Language Processing (2nd Edition), D. Jurafsky and J.H. Martin, Prentice Hall, 2009.
Other infos :	Background: -- LSINF1121: Algorithmics and data structures
Faculty or entity in charge:	INFO

Programmes / formations proposant cette unité d'enseignement (UE)				
Intitulé du programme	Sigle	Credits	Prerequis	Acquis d'apprentissage
Master [120] in Linguistics	LING2M	5	-	
Master [120] in Statistics: General	STAT2M	5	-	
Master [120] in Computer Science and Engineering	INFO2M	5	-	
Master [120] in Computer Science	SINF2M	5	-	