lingi2263

2020

UCLouvain

Computational Linguistics

Due to the COVID-19 crisis, the information below is subject to change, in particular that concerning the teaching mode (presential, distance or in a comodal or hybrid format).

5 credits	30.0 h + 15.0 h	Q1

Teacher(s)	Dupont Pierre ;Dupont Pierre (compensates Fairon Cédrick) ;Fairon Cédrick ;
Language :	English
Place of the course	Louvain-la-Neuve
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 1 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	**Due to the COVID-19 crisis, the information in this section is particularly likely to change.** The projects are worth 30 % of the final grade, 70 % for the final exam (closed-book). The projects **cannot** be implemented again in second session. The project grades are fixed at the end of the semester and included as such in the global score for the second session. The final exam is, by default, a written exam (on paper or, when appropriate, on a computer). These evaluation rules are subject to possible updates due to the sanitary situation. In particular, the relative weights between the projects and the final exam could be adapted. Such possible updates would be notified to the students by a general announcement posted on the Moodle site of this course.

Teaching methods	**Due to the COVID-19 crisis, the information in this section is particularly likely to change.**
	• Lectures • Practical projects implemented in Python. By default, lectures can be followed **face to face** in the auditorium announced in the official schedule. Depending on the number of registered students and the evolution of the sanitary situation, students will be able to follow the lectures as well **remotely** on *Teams*. Practical projects are submitted **on line** and evaluated on the *Inginious* platform.
Content	• Various levels of linguistic analysis • (Automated) corpus processing: formating, tokenization, data tagging • Probabilistic language models: N-grams, HMMs • Part-of-Speech Tagging • (Probabilistic) Context-Free Grammars: parameter estimation and parsing algorithms • Introduction to Machine Translation • Introduction to Deep Learning • Typical linguistic applications such as automated completion, POS taggers, parsing or machine translation.
Inline resources	http://moodleucl.uclouvain.be/course/view.php?id=7865
Bibliography	One recommended textbook - un ouvrage conseillé : • Speech and Language Processing (2nd Edition), D. Jurafsky and J.H. Martin, Prentice Hall, 2009.
Faculty or entity in charge	INFO

Force majeure

Teaching methods	Lectures are given **online** and can be followed **remotely**. Computing projects are submitted **online** on the Inginious platform.
Evaluation methods	The final exam is an open book exam to be made **individually online**. The material for this final exam is the same as in the normal situation (see "supports de cours"). The global grade for the course is based on the projects implemented during the semester (50 %) + on the individual final exam (50 %). The projects cannot be re-implemented for the second session. Hence, the project grade is fixed at the end of the semester.

Programmes containing this learning unit (UE)

Program title	Acronym	Credits	Prerequisite	Aims
Master [120] in Data Science : Statistic	DATS2M	5		🔍
Master [120] in Linguistics	LING2M	5		🔍
Master [120] in Computer Science and Engineering	INFO2M	5		🔍
Master [120] in Computer Science	SINF2M	5		🔍
Master [120] in Data Science Engineering	DATE2M	5		🔍
Master [120] in Data Science: Information Technology	DATI2M	5		🔍