



university of
groningen

faculty of arts

LIX001M05
Dr. A. Toral





1 /	Type of course unit, number of ECTS credit points and admission requirements	
	a.	Type: elective
	b.	ETCS credit points: 5
	c.	Admission requirements: BA in Information Science, Artificial Intelligence or a similar undergraduate programme.
2 /	Content of the course unit	
	The course provides a fundamental introduction to Natural Language Processing. Various important fundamental topics are treated, including Regular Expressions, Text Normalisation, Language Modeling, N-grams, Hidden Markov Models, Word Embeddings, Part-of-speech Tagging and Dependency Parsing.	
3 /	Position of the course unit in the degree programme	
	There are no dependencies between this and other courses of the degree programme.	
4 /	Learning outcomes of the course unit	
	The goal of the course is to ensure that students are familiar with a number of very fundamental techniques and algorithms in the area of natural language processing, in particular for regular expressions, statistical language models based on N-grams, part-of-speech tagging and dependency parsing.	
5 /	Mode of instruction and learning activities	
	The course is made of lectures and practical sessions: Weekly lectures introduce in detail the topics covered in the course. Weekly practical sessions complement the lectures. In these sessions students work on these topics by means of exercises. These exercises are to be compiled in a portfolio. A detailed weekly schedule is shown in section 10.	
6 /	Assesment	
	a.	Mode of assesment
		Students are assessed based on the portfolio and an exam.



b.	Assessment: duration, time and place; deadlines and procedures; perusal and resits
	The first portfolio is to be submitted via Nestor no later than 23:59 of Sunday 1st of March, local time. The second portfolio is to be submitted via Nestor no later than 23:59 of Sunday 22nd of March, local time. The written exam is on Friday April 3rd, 9-12
c.	Assesment criteria
	A random selection of the exercises included in the portfolio's will be graded. Missing exercises are graded as 0.
d.	Calculating preliminary and final marks
	Final mark = 25% portfolio 1 + 25% portfolio 2 + 50% exam. A student passes the course if the score the average for the portfolio's as well as the score for the exam is sufficient (at least 5.5).
e.	Conditions of taking exams
	The average grade for the portfolio's should be sufficient (at least 5.5).
f.	Example of tests
	Examples will be available on Nestor.

7 /	Cheating and plagiarism
	Cheating and plagiarism are subject to the provisions set down in the TER (Article 8.17 of Part A of the BA TER or Article 4.13 of Part A of the MA TER). The Board of Examiners is always informed in cases of suspected cheating or plagiarism.

8 /	Calculation of student workload						
	In hours: <table><tr><td>28</td><td>Lectures and labs (4 hours per week)</td></tr><tr><td>84</td><td>Self-study: reading and work on exercises (12 hours per week)</td></tr><tr><td>28</td><td>Exam preparation</td></tr></table>	28	Lectures and labs (4 hours per week)	84	Self-study: reading and work on exercises (12 hours per week)	28	Exam preparation
28	Lectures and labs (4 hours per week)						
84	Self-study: reading and work on exercises (12 hours per week)						
28	Exam preparation						
	Total = 140 (28+84+28).						



9 / Literature

Required:

Daniel Jurafsky and James H. Martin. Speech and Language Processing. An introduction to Natural Language Processing, Computational Linguistics and Speech Recognition (Third Edition). (Forthcoming) PDF of (draft) chapters available here:

<https://web.stanford.edu/~jurafsky/slp3/>

Additional reading materials (e.g. scientific articles) may be provided for some of the topics.

10 Weekly schedule

	Date	Topic	Reading (chapter)
1	3/2	Introduction, regular expressions, text normalisation	2
2	10/2	Language Modeling and N-grams	3
3	17/2	Vector Semantics and Word Embeddings	6
4	24/2	PoS tagging and HMMs	8
	1/3	Deadline Portfolio 1	
5	2/3	Sequence processing and Neural Networks	9
6	9/3	Dependency Grammar and Dependency Parsing	15
7	16/3	(To be decided) Wordnet, word senses, sentiment	19/21
		Discussion and Sample Exam	
	22/3	Deadline Portfolio 2	
	3/4	Exam	



11 / Copyright

Respect the copyright to the teaching material.

All teaching material is protected by copyright. Students may not make photocopies of teaching material, exams and lectures other than for their own study purposes. In addition, teaching material may not be further distributed in any format. Deliberate violation of copyright is a criminal offence. The University of Groningen will take appropriate measures upon detecting such violations.



The programme-level learning outcomes are described in the Teaching and Examination Regulations (TER) of the Master in Information Science.



Appendix 2. Marking tools

NA