SYLLABUS

1. Information regarding the programme

1.1 Higher education institution	**Babeş Bolyai University**
1.2 Faculty	**Faculty of Mathematics and Computer Science**
1.3 Department	**Department of Computer Science**
1.4 Field of study	**Computer Science**
1.5 Study cycle	**Master**
1.6 Study programme / Qualification	**Computer Science/ Applied Computational Intelligence**

2. Information regarding the discipline

2.1 Name of the discipline	**Knowledge Based Systems and Language Technology**						
2.2 Course coordinator	**Lecturer Ph.D. Lupea Mihaiela**						
2.3 Seminar coordinator	**Lecturer Ph.D. Lupea Mihaiela**						
2.4. Year of study	**1**	2.5 Semester	**2**	2.6. Type of evaluation	**exam**	2.7 Type of discipline	**compulsory**

3. Total estimated time (hours/semester of didactic activities)

3.1 Hours per week	3	Of which: 3.2 course	2	3.3 seminar/laboratory	1
3.4 Total hours in the curriculum	42	Of which: 3.5 course	28	3.6 seminar/laboratory	14

Time allotment:	hours
Learning using manual, course support, bibliography, course notes	40
Additional documentation (in libraries, on electronic platforms, field documentation)	14
Preparation for seminars/labs, homework, papers, portfolios and essays	30
Tutorship	14
Evaluations	30
Other activities: individual project	30

3.7 Total individual study hours	158
3.8 Total hours per semester	200
3.9 Number of ECTS credits	8

4. Prerequisites (if necessary)

4.1. curriculum	Formal languages, Data structures, Machine learning
4.2. competencies	Programming skills in a high level programming language

5. Conditions (if necessary)

5.1. for the course	
5.2. for the seminar /lab activities	• Laboratory with computers; high level programming language environment (.NET or any Java environment a.s.o.)

6. Specific competencies acquired

Professional competencies	• Knowledge, understanding and use of basic concepts, tasks, applications of Language Technology - Natural Language Processing (NLP) domain. • Apply and use formal language concepts (grammars, parsing), statistic models (HMM) artificial intelligence algorithms (clustering, machine learning) and techniques (unsupervised, supervised) to solve different tasks at the syntactic level (POS-tagging, parsing, chunking), and semantic level (word sense disambiguation, information extraction, anaphora resolution) in NLP domain.
Transversal competencies	• The optimization of the search on Web, the interfaces in natural language and the recent aspects of text mining need a good understanding of the NLP domain.

7. Objectives of the discipline (outcome of the acquired competencies)

7.1 General objective of the discipline	• To provide the basic principles, technologies and applications of Language Technology (LT) and Knowledge based systems. • To understand the current state of the art in LT in order to realize original research in LT.
7.2 Specific objective of the discipline	• To apply and use formal language concepts, statistic models, artificial intelligence algorithms and techniques to solve different tasks at the syntactic level (POS-tagging, parsing, chunking), and semantic level (word sense disambiguation, information extraction, anaphora resolution) in NLP domain.

8. Content

8.1 Course	Teaching methods	Remarks
Course 1. Natural Language Processing (NLP): stages, domains, applications.	Exposure: description, explanation, examples, debate, dialogue	
Course 2. -WordNet: knowledge structure, semantic relations, lexical relations, applications, interfaces; corpora. - Part of speech tagging.	Exposure: description, explanation, examples, debate, dialogue	
Course 3. Syntactic parsing (1) - grammar rules for English - sentence level construction; - Cocke-Kasami-Yonger (CKY) algorithm;	Exposure: description, explanation, examples, debate, dialogue	

Course 4. Syntactic parsing (2) - Earley algorithm and the chart parsing approach. - Dependency trees	Exposure: description, explanation, examples, debate, dialogue	
Course 5. Statistical parsing -Probabilistic Context-Free Grammars (PCFG); - Probabilistic CKY (Cocke-Kasami-Yonger) parsing of PCFGs.	Exposure: description, explanation, examples, debate, dialogue	
Course 6. Hidden Markov Model (1) - Markov chains, Hidden Markov Model(HMM); - three canonical problems associated with HMM - the forward algorithm; the Viterbi algorithm	Exposure: description, explanation, examples, debate, dialogue	
Course 7. Hidden Markov Model (2) - the Baum-Welch algorithm for HMM; - applications to part–of-speech tagging.	Exposure: description, explanation, examples, debate, dialogue	
Course 8. Information extraction - tasks, applications, tools	Exposure: description, explanation, examples, debate, dialogue	
Course 9. Word Sense Disambiguation (1) - unsupervised (by clustering); - dictionary based approach (Lesk, Yarowsky, bilingual dictionaries).	Exposure: description, explanation, examples, debate, dialogue	
Course 10. Word Sense Disambiguation (2) - machine learning approach; - the bootstraping algorithm	Exposure: description, explanation, examples, debate, dialogue	
Course 11. Anaphora resolution (1) - hard constraints and preferences - Hobb's algorithm, Lapin and Lease algorithm	Exposure: description, explanation, examples, debate, dialogue	
Course 12. Anaphora and co-reference resolution (2) - Mitkov's algorithm - tools for co-reference resolution	Exposure: description, explanation, examples, debate, dialogue	
Course 13. First-order logic and description logics for natural language processing.	Exposure: description, explanation, examples, debate, dialogue	
Course 14. **Students' presentations of the practical project.**	Debate, dialog	

Bibliography

1. J.ALLEN : Natural language understanding, Benjamin/Cummings Publisher, 2nd ed., 1995.
2. E. CHARNIAK: Statistical language learning, MIT press, 1996.
3. B.CARPENTER: ALE: The attribute logic engine. User's guide. Carnegie Mellon University,1994.
4. D.FEHRER et al: Description logics for natural language processing. In Proc. of the 1994 Description Logic Workshop (DL'94), 1994.
5. H. HELBIG: Knowledge Representation and the Semantics of Natural Language, Springer, 2006.
6. D.JURAFSKY, J.MARTIN: Speech and language processing, Prentice Hall, 2000.
7. C.MANNING, H.SCHUTZE: Foundation of statistical natural language processing, MIT, 1999.

8. R. MITKOV(ed): The Oxford Handbook of Computational Linguistics, Oxford University Press, 2003.
9. D.TATAR: Inteligenta artificiala: demonstrare automata de teoreme, prelucrarea limbajului natural, Editura Albastra, Microinformatica, 2001.
10. D. TATAR: Inteligenta artificiala. Aplicatii in prelucrarea limbajului natural, Editura Albastra, Microinformatica, 2003, ISBN 973-650-100-01.

8.2 Seminar / laboratory	Teaching methods	Remarks
1. Working with WordNet and Romanian WordNet	Explanation, dialogue, case studies	The seminar/lab is structured as 2 hours classes every second week
2. Students' presentations of a NLP domain and a corresponding tool	Dialogue, debate	
3. Working with dedicated parsers and taggers (Stanford, CST tools, Racai tools)	Explanation, dialogue, case studies	
4. Students' presentations of the theoretical paper	Dialogue, debate	
5. Working with dedicated tools for information extraction	Explanation, dialogue, case studies	
6. Working with dedicated tools for anaphora and co-reference resolution	Explanation, dialogue, case studies	
7. Students' presentations of the practical project.	Dialogue, debate	

Bibliography

1. Rada Mihalcea: www.cs.unt.edu/~rada/downloads.html
2. Resurse lingvistice in limba romana: www.racai.ro

9. Corroborating the content of the discipline with the expectations of the epistemic community, professional associations and representative employers within the field of the program

- The course respects the IEEE and ACM Curricula Recommendations for Computer Science studies;
- The course exists in the studying program of all major universities in Romania and abroad;
- The optimization of the search on Web, the interfaces in natural language and the recent aspects of text mining need a good understanding of Natural Language Processing.

10. Evaluation

Type of activity	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Share in the grade (%)
10.4 Course	- know the theoretical concepts of the domain; - apply the course methods, algorithms in problem solving	Written exam	40%
10.5 Seminar/lab activities	- know to write an overview of a specific domain	Presentation of a NLP domain and a corresponding tool;	15%
	know to synthesize and compare different approaches/results of the same studied subject.	Theoretical paper based on recent research papers in NLP domain;	20%
	- be able to implement course algorithms	Practical project - implementation of a NLP tool based on the studied methods	25%

10.6 Minimum performance standards
➢ At least grade 5 (from a scale of 1 to 10) at all four evaluation stages.

Date	Signature of course coordinator	Signature of seminar coordinator
10.05.2013	Lecturer PhD. Lupea Mihaiela	Lecturer PhD. Lupea Mihaiela

Date of approval	Signature of the head of department
.....	Prof. PhD. Pârv Bazil