

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	Bachelor
1.6 Study programme / Qualification	Artificial Intelligence

2. Information regarding the discipline

2.1 Name of the discipline (en) (ro)	NLP (Natural Language Processing) Techniques Tehnici de NLP						
2.2 Course coordinator	Lecturer Ph.D. Lupea Mihaiela-Ana						
2.3 Laboratory coordinator	Lecturer Ph.D. Lupea Mihaiela-Ana						
2.4. Year of study	3	2.5 Semester	1	2.6. Type of evaluation	E	2.7 Type of discipline	compulsory
2.8 Code of the discipline	MLE5208						

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 lab
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					15
Additional documentation (in libraries, on electronic platforms, field documentation)					10
Preparation for seminars/labs, homework, papers, portfolios and essays					18
Tutorship					5
Evaluations					10
3.7 Total individual study hours	58				
3.8 Total hours per semester	100				
3.9 Number of ECTS credits	4				

4. Prerequisites (if necessary)

4.1. curriculum	Data structures
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	<ul style="list-style-type: none"> • Assimilation of mathematical concepts and formal models to understand, verify and validate software systems ; • Advanced ability to approach, model and solve phenomena and problems from natural language and economy using fundamental knowledge from mathematics and computer science; • Ability to approach and solve complex problems using various techniques of computational intelligence; • Proficient use of methodologies and tools specific to programming languages and software systems.
Transversal competencies	<ul style="list-style-type: none"> • Eitic and fair behavior, committment to professional deontology • Team work capabilities; able to fulfill different roles • Professional communication skills; concise and precise description, both oral and written, of professional results , negotiation abilities; • Antepreneurial skills; working with economical knowledge; continuous learning • Good English communication skills

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

7.1 General objective of the discipline	<ul style="list-style-type: none"> • To introduce the basic principles, domains and tasks in Natural Language Processing (NLP) • To understand the current state of the art in order to realize an overview of a specific domain in NLP and to use NLP techniques to solve NLP tasks.
7.2 Specific objective of the discipline	<ul style="list-style-type: none"> • Application and use of formal models (logics, grammars, graphs), statistic models, artificial intelligence algorithms and NLP techniques to solve specific tasks in NLP domain, for English and Romanian languages. • Practical projects: implementation of automated systems that solve tasks specific to the NLP field (Romanian and English): part-of-speech tagging, parsing, chunking, word sense disambiguation, keyword extraction, document summarization, anaphora resolution, sentiment and emotion analysis, quantitative analysis of texts

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 and RoWordNet - knowledge structure, semantic relations, lexical relations. - WordNetSimilarity tool and similarity measures for words. - part-of-speech (POS) tagging and lexical units in English and Romanian languages		
Course 3. Text representation and text classification		

Course 4. Syntactic parsing - Probabilistic Context-Free Grammars (PCFG); - Probabilistic CKY parsing of PCFGs.		
Course 5. Keyword extraction - TextRank and RAKE algorithms		
Course 6. Students' presentations of a NLP task/ tool.		
Course 7. Document summarization - approaches based on clustering and graphs.		
Course 8. Sentiment analysis - opinion mining in social media - emotion analysis in literature		
Course 9. Anaphora resolution - Lapin& Lease algorithm and Mitkov's algorithm		
Course 10. Word Sense Disambiguation - dictionary and graph-based approaches.		
Course 11. Information extraction		
Course 12. Textual entailment		
Course 13. Quantitative analysis of literary texts		
Course 14. Students' presentations of the practical projects		

Bibliography

1. J.ALLEN : Natural language understanding, Benjamin/Cummings Publisher, 2nd ed., 1995.
2. E.CHARNIAK: Statistical language learning, MIT press, 1996.
3. D.FEHRER et al: Description logics for natural language processing. In Proc. of the 1994 Description Logic Workshop (DL'94), 1994.
4. H.HELBIG: Knowledge Representation and the Semantics of Natural Language, Springer, 2006.
5. D.JURAFSKY, J.MARTIN: Speech and language processing, Prentice Hall, 2000.
6. M.LUPEA, M.RUKK, I.I.POPESCU, G.ALTMANN, Some Properties of Rhyme, Studies in Quantitative Linguistics 26, RAM-Verlag, 2017.
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. I.I.POPESCU, M.LUPEA, D.TATAR, G.ALTMANN, Quantitative Analysis of Poetic Texts, DE GRUYTER MOUTON, 2015.
10. D. TATAR: Inteligența artificială. Aplicații în prelucrarea limbajului natural, Editura Albastra, Microinformatica, 2003.

8.2 Seminar / laboratory	Teaching methods	Remarks
1. Working with WordNet, Romanian WordNet and WordNetSimilarity. Working with dedicated parsers and taggers (Stanford, CST tools, Racai tools) Working with dedicated tools for summarization, anaphora, co-reference resolution, sentiment analysis.	Explanation, dialogue, case studies	The seminar/lab is structured as 2 hours classes every second week
2. Study of platforms and libraries from different programming languages that offer preprocessing functions for texts in Romanian and English.	Explanation, dialogue, case studies	
3. Students' presentations of a NLP task/ tool.	Dialogue, debate	
4. Identify practical tasks in Romanian NLP, that can be solved by implementing course algorithms.	Explanation, dialogue, case studies	

Choose the NLP task, study different approaches, choose the approach that will be implemented.		
5. Search for the input data or create data sets specific to the chosen task. Develop resources for Romanian NLP tasks	Explanation, dialogue, case studies	
6. Design and implementation of the NLP tool. Develop resources for Romanian NLP tasks	Explanation, dialogue, case studies	
7. Students' presentations of the practical projects.	Dialogue, debate	
Bibliography		
1. Rada Mihalcea: www.cs.unt.edu/~rada/downloads.html		
2. Resurse lingvistice in limba romana: www.racai.ro , nlptools.info.uaic.ro/Resources		

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

<ul style="list-style-type: none"> • 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	20%
	- know to write an overview of a specific NLP task	Theoretical report – presentation of a NLP task.	35%
10.5 Seminar/lab activities	- be able to apply theoretical concepts in practical tasks	Develop resources for Romanian NLP tasks	10%
	- be able to implement course algorithms	Practical project - implementation of a NLP tool based on the studied methods.	35%
10.6 Minimum performance standards			
➤ The final grade to be at least 5 (from a scale of 1 to 10).			

Date

Signature of course coordinator

Signature of seminar coordinator

25.04.2023

Lecturer Ph.D. Lupea Mihaiela

Lecturer Ph.D. Lupea Mihaiela

Date of approval

Signature of the head of department

.....

Prof. Ph.D. Dioşan Laura