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	**Computers and Information Technology**
1.5 Study cycle	**Bachelor**
1.6 Study programme / Qualification	**Information Engineering**

2. Information regarding the discipline

2.1 Name of the discipline (en) (ro)			**Introduction to Natural Language Processing** **Introducere in prelucrarea limbajului natural**			
2.2 Course coordinator			**Lecturer Ph.D. Lupea Mihaiela-Ana**			
2.3 Seminar coordinator			**Lecturer Ph.D. Lupea Mihaiela-Ana**			
2.4. Year of study	**4**	2.5 Semester	**8**	2.6. Type of evaluation **C**	2.7 Type of discipline	**Optional DS**
2.8 Code of the discipline		MLE8151				

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

3.1 Hours per week	5	Of which: 3.2 course	2	3.3 seminar/laboratory	1 LP 2P
3.4 Total hours in the curriculum	70	Of which: 3.5 course	28	3.6 seminar/laboratory	42
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					15
Tutorship					5
Evaluations					10

3.7 Total individual study hours	55
3.8 Total hours per semester	125
3.9 Number of ECTS credits	5

4. Prerequisites (if necessary)

4.1. curriculum	Formal languages, 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	C3.1 Identifying classes of problems and solving methods that are specific to computing systems C3.2 Using interdisciplinary knowledge, solution patterns and tools, making experiments and interpreting their results C3.4 Comparatively and experimentaly evaluation of the alternative solutions for performance optimization C3.5 Developing and implementing information system solutions for concrete problems C6.2 Appropriate use of methods for signal analysis and fundamental artificial intelligence algorithms C6.4 Quantitative and qualitative evaluation of the performance of intelligent systems C6.5 Incorporating signal processing methods and artificial intelligence- specific solutions into dedicated applications
Transversal competencies	CT1 Honorable, responsible, ethical behavior, in the spirit of the law, to ensure the professional reputation. CT2 Identifying, describing and conducting processes in the projects management field, undertaking different team roles and clearly and concisely describing own profesional results, verbally or in writing, in Romanian and in an international language. CT3 Demonstrating initiative and pro-active behavior for updating professional, economical and organizational culture knowledge.

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

7.1 General objective of the discipline	- 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 implement a NLP tool.
7.2 Specific objective of the discipline	- Application and use of formal models (logics, grammars, graphs), statistic models (HMM), artificial intelligence algorithms and 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.		

Course 3. - part-of-speech (POS) tagging and lexical units in English and Romanian languages		
Course 4. Syntactic parsing - grammar rules for English - sentence level construction; - Cocke-Kasami-Yonger (CKY) algorithm;		
Course 5. Statistical parsing - Probabilistic Context-Free Grammars (PCFG); - Probabilistic CKY parsing of **PCFGs.**		
Course 6. Hidden Markov Model - Hidden Markov Model(HMM) – a statistical model - applications of HMM in real life and in POS tagging - the canonical problems associated with HMM and the algorithms that solve them		
Course 7. Word Sense Disambiguation - dictionary and graph-based approaches.		
Course 8. Document summarization - approaches based on clustering and graphs.		
Course 9. Anaphora resolution - Lapin and Lease algorithm - Mitkov's algorithm		
Course 10. Sentiment analysis - opinion mining in social media - emotion analysis in literature		
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. Y. GOLDBERG, Neural Network Methods for Natural Language Processing, Springer, 2017.
2. S. RAAIJMAKERS, Deep Learning for Natural Language Processing, 2022
3. U. KAMATH, J.LIU, J. WHITAKER, Deep Learning for NLP and speech recognisition., Springer, 2019, e-book, https://doi.org/10.1007/978-3-030-14596-5
4. H.HELBIG: Knowledge Representation and the Semantics of Natural Language, Springer, 2006.
5. D.JURAFSKY, J.MARTIN: Speech and language processing, Prentice Hall, 3rd edition, 2022.
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: Inteligenta artificiala. Aplicatii in prelucrarea limbajului natural, Editura Albastra, Microinformatica, 2003.

8.2 Seminar / laboratory	Teaching methods	Remarks
1. Working with WordNet, Romanian WordNet and WordNetSimilarity.	Explanation, dialogue, case studies	The seminar/lab is structured as 2 hours classes every second week
2. Working with dedicated parsers and taggers	Explanation,	

(Stanford, CST tools, Racai tools)	dialogue, case studies	
3. Students' presentations of a NLP domain and a corresponding tool.	Dialogue, debate	
4. Students' presentations of a NLP domain and a corresponding tool.	Dialogue, debate	
5. Working with dedicated tools for summarization, anaphora, co-reference resolution, sentiment analysis.	Explanation, dialogue, case studies	
6. Working with dedicated tools for information extraction and textual entailment.	Dialogue, debate	
7. Students' presentations of the practical projects.	Dialogue, debate	

Bibliography

1. Boros Tiberiu, Dumitrescu Stefan, Burtica Ruxandra. "NLP-Cube: End-to-End Raw Text Processing With Neural Networks," 2018.
2. https://opensource.adobe.com/NLP-Cube/index.html
3. http://wordnetweb.princeton.edu/
4. https://multiwordnet.fbk.eu/english/home.php
5. Resurse lingvistice in limba romana: www.racai.ro, nlptools.info.uaic.ro/Resources
6. https://demo.allennlp.org/
7. Rada Mihalcea: www.cs.unt.edu/~rada/downloads.html

8.3 Project	Teaching methods	Remarks
S1. Bibliographic study - the representation of texts for their use in tasks specific to the NLP field	Documentation on electronic platforms, dialogue, case studies	Students will work individually or in teams of 2 people to implement tools that solve practical tasks in the field of NLP in Romanian or English.
S2-S3. Study of platforms and libraries from different programming languages that offer preprocessing functions for texts in Romanian and English.	Documentation on electronic platforms, dialogue, case studies	
S4. Study of existing dedicated tools that solve NLP tasks in Romanian and English language.	Documentation on electronic platforms, dialogue, case studies	
S5. Identify practical tasks in the field of NLP, in Romanian, that can be solved by implementing course algorithms.	Documentation on electronic platforms, dialogue, case studies	
S6-S7. Choose the NLP task, study different approaches, choose the approach that will be implemented.	Documentation on electronic platforms, dialogue, case studies	
S8-S9. Search for the input data or create data sets specific to the chosen task.	Documentation on electronic platforms, dialogue, case studies	

S10-S12. Design and implementation of the system.	Documentation on electronic platforms, dialogue, case studies	
S13. Test the software system.	Dialogue, case studies	
S14. Final software system presentation	Evaluation	

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 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	20%
	- be able to implement course algorithms	Practical project - implementation of a NLP tool based on the studied methods.	35%
10.6	- activity during courses and labs	Attendance to activities, timely realization of the requested activities	10%
10.7 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

15.05.2022   Lecturer Ph.D. Lupea Mihaiela   Lecturer Ph.D. Lupea Mihaiela

Date of approval   Signature of the head of department

Prof. Ph.D. Dioşan Laura

24.05.2022