SYLLABUS

Introduction to Natural Language Processing

University year 2025-2026

1. Information regarding the programme

1.1. Higher education institution	**Babeș-Bolyai University, Cluj-Napoca**
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	**Computer Science in Romanian**
1.7. Form of education	**Full time**

2. Information regarding the discipline

2.1. Name of the discipline	**Introduction to Natural Language Processing**					Discipline code	**MLE8151**
2.2. Course coordinator					**Lecturer Ph.D. Lupea Mihaiela**		
2.3. Seminar coordinator					**Lecturer Ph.D. Lupea Mihaiela**		
2.4. Year of study	**3**	2.5. Semester	**6**	2.6. Type of evaluation	**C**	2.7. Discipline regime	**Optional**

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/project	**1 lab+ 2pr**
3.4. Total hours in the curriculum	60	of which: 3.5 course	24	3.6 seminar/laboratory/project	**36**
Time allotment for individual study (ID) and self-study activities (SA)					**hours**
Learning using manual, course support, bibliography, course notes (SA)					15
Additional documentation (in libraries, on electronic platforms, field documentation)					14
Preparation for seminars/labs, homework, papers, portfolios and essays					14
Tutorship					5
Evaluations					7
Practical project:					10
3.7. Total individual study hours	**65**				
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, Graphs Algorithms
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.)

6.1. Specific competencies acquired [1]

Professional/essential competencies	• Use of software tools in an interdisciplinary context. • Use of artificial intelligence concepts and techniques to solve real-world problems.
Transversal competencies	• Application of organized and efficient work rules, of responsible attitudes towards the didactic-scientific field, to bring creative value to own potential, with respect for professional ethics principles and norms. • Use of efficient methods and techniques to learn, inform, research and develop the abilities to bring value to knowledge, to adapt at the requirements of a dynamical society and to communicate efficiently in Romanian language and in an international language

6.2. Learning outcomes

Knowledge	• The graduate has the necessary knowledge for using computers, developing software programs and applications, information processing. • The graduate has the necessary knowledge for literature review.
Skills	• The graduate is able to identify complex problems and examine related issues to develop solving options and implement solutions. • The graduate has the ability to choose and use programming paradigms (procedural, object-oriented, functional) to develop software applications appropriate for the specific domain of the application being developed. • The graduate has the ability to choose and use existing modules and environments for application development. • The graduate is able to write a scientific/technical report.
Responsibility and autonomy:	• The graduate is able to define/identify/understand research problems in computer science. • The graduate has the ability to understand and communicate information effectively. • The graduate has the necessary skills to use research support tools.

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.

[1] One can choose either competences or learning outcomes, or both. If only one option is chosen, the row related to the other option will be deleted, and the kept one will be numbered 6.

7.2 Specific objective of the discipline	• Apply and use formal models (logics, grammars, parsing), statistic models (HMM), artificial intelligence algorithms and techniques to solve different tasks at the syntactic level (POS-tagging, parsing, chunking), and semantic level (keyword extraction, document summarization, anaphora resolution, sentiment analysis, word sense disambiguation) in Natural Language Processing domain for English and Romanian languages. • Use of Large Language Models (LLMs) to solve NLP tasks.

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. Part of speech tagging WordNet and RoWordNet - knowledge structure, semantic relations, lexical relations	Exposure: description, explanation, examples, debate, dialogue	
Course 3. **Text representation and natural language models. Text classification.**	Exposure: description, explanation, examples, dialogue	
Course 4. Syntactic parsing - grammar rules for English - sentence level construction; - Cocke-Kasami-Yonger (CKY) algorithm;	Exposure: description, explanation, dialogue, examples.	
Course 5. Hidden Markov Model - Markov chains, Hidden Markov Model(HMM); - three canonical problems associated with HMM - the forward algorithm; Viterbi algorithm.	Exposure: description, explanation, examples, debate, dialogue	
Course 6. Keyword extraction - TextRank and RAKE algorithms	Exposure: description, explanation, examples, dialogue	
Course 7. Document summarization - approaches based on clustering and graphs.	Exposure: description, explanation, examples, dialogue	
Course 8. Students' presentations of NLP tasks and tools.	Evaluation	
Course 9. Sentiment analysis - opinion mining in social media - emotion analysis in literature	Exposure: description, explanation, debate, examples, dialogue	
Course 10. Anaphora resolution - Lapin and Lease algorithm - Mitkov's algorithm	Exposure: description, explanation, debate, examples, dialogue	
Course 11. Word Sense Disambiguation - dictionary and graph-based approaches.	Exposure: description, explanation, dialog, examples	
Course 12. Students' presentations of the practical projects	Evaluation	

Bibliography

1. J.ALLEN : Natural language understanding, Benjamin/Cummings Publisher, 2nd ed., 1995.
2. E. CHARNIAK: Statistical language learning, MIT press, 1996.
3. L. DENG, Y. LIU: Deep learning in Natural Language Processing, Springer Verlag, Singapore, 2018
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. Aplicatii in prelucrarea limbajului natural, Editura Albastra, Microinformatica, 2003, ISBN 973-650-100-01.
10. S. VAJJALA, B. MAJUMDER, A. GUPTA, H. SURANA: Practical Natural Language Processing. A Comprehensive Guide to Building Real-World NLP Systems, O'REILLY. 2020.

8.2 Seminar / laboratory	Teaching methods	Remarks
1. Study of platforms and libraries from different programming languages that offer preprocessing functions for texts in Romanian and English languages.	Documentation on electronic platforms, explanation, dialogue, case studies	The seminar/lab is structured as 2 hours classes every second week
2. Study of the LLMs (large language models) for English and Romanian languages. Work with WordNet, Romanian WordNet and WordNetSimilarity. Work with dedicated tools for keyword extraction, summarization, anaphora resolution, sentiment analysis.	Documentation on electronic platforms, dialogue, case studies	
3. Identify practical tasks in Romanian NLP. Choose the NLP task to be solved, study different approaches, choose the approach that will be implemented. Search for the input data specific to the chosen task.	Documentation on electronic platforms, dialogue, case studies	
4. Students' presentations of a NLP task/ tool.	Evaluation	
5. Discussions about the implementation of an NLP tool. Develop resources for Romanian NLP tasks	Explanation, dialogue, case studies	
6. Students presentations of the practical projects.	Evaluation	

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

Activity type	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Percentage of final 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/laboratory	Be able to implement course algorithms	Practical project - implementation of a NLP tool.	35%
	Be able to apply theoretical concepts in practical tasks	Develop resources for Romanian NLP tasks	20%
10.6 Activity	Activity during courses and seminars.	Active attendance	10%
10.7 Minimum standard of performance			
• The final grade to be at least 5 (from a scale of 1 to 10).			

11. Labels ODD (Sustainable Development Goals)[2]

Not applicable.

Date: 15.04.2025	Signature of course coordinator Lect. PhD. Mihaiela LUPEA	Signature of seminar coordinator Lect. PhD. Mihaiela LUPEA
Date of approval: ...		Signature of the head of department Assoc.Prof. PhD. Adrian STERCA

[2] Keep only the labels that, according to the *Procedure for applying ODD labels in the academic process*, suit the discipline and delete the others, including the general one for *Sustainable Development* – if not applicable. If no label describes the discipline, delete them all and write „*Not applicable*.".