



COURSE PROGRAMME

1. Information about the programme

1.1 University	University "Alexandru Ioan Cuza" of Iasi
1.2 Faculty	Faculty of Computer Science
1.3 Department	Department of Computer Science
1.4 Domain	Computer Science
1.5 Cycle	Licence
1.6 Programme / Qualification	Computer Science (in English)

2. Information about the course

2.1 Course Name	Natural language processing techniques						
2.2 Course taught by	Lecturer PhD. LIVIU-ANDREI SCUTELNICU						
2.3 Seminary / laboratory taught by							
2.4 Year	III	2.5 Semester	VI	2.6 Type of evaluation*	E	2.7 Course type**	Op

*E – Exam / C – Colloquium / V – Verification

**OB – Obligatory / OP – Optionally / F – Facultative

3. Total hours (estimated per semester and activities)

3.1 Number of hours per week	4	3.2 course	2	3.3 seminary/laboratory	2
3.4 Total number of hours	48	3.5 course	24	3.6 seminary/laboratory	24
Distribution					hours
Individual study using textbooks, course notes, bibliography items, etc.					20
Supplementary study (library, on-line platforms, etc.)					8
Individual study for seminary/laboratory, homeworks, projects, etc.					20
Tutoring					0
Examination					4
Other activities					0
3.7 Total hours of individual activity*					52
3.8 Total hours per semester					100
3.9 Credit points					4

4. Pre-requisites - Curriculum (if necessary)

Data Structures, Artificial Intelligence, Machine Learning.
High-level programming, development of computer systems.

5. Conditions (if necessary)

5.1 Course	The classroom must have a video projector, Internet connection and blackboard for examples.
5.2 Seminary / Laboratory	Mandatory attendance at the laboratory. The laboratory room must have an Internet connection and a blackboard for examples. Students need computers with a language installed that allows them to solve the lab assignments (Java or Python recommended).

6. Objectives

General objective:

The overall objective of the course is to provide students with a solid foundation in the field of natural language processing (PLN), balancing theoretical knowledge with practical applicability.

Specific objectives:

O1. Knowledge of fundamental concepts and theories in the field of natural language processing;

O2. Knowledge of the principles of language systems development, including text preprocessing, tokenization, part-of-speech tagging, named entity recognition, and the application of machine learning algorithms to textual data;

O3. Knowledge of the methods of using cutting-edge tools and technologies in the field of computational linguistics;

O4. Knowledge of the methods of applying machine learning techniques to solve specific problems of natural language processing, both for synthesis and for text generation.

7. Specific competencies/Learning outcomes

- Vor utiliza tehnici de bază de procesare a textului
- Vor cunoaște și vor putea aplica metode de clasificare a textelor
- Vor putea organiza semantic colecții de texte
- Vor cunoaște și vor utiliza anumite arhitecturi ale rețelelor neuronale pentru a rezolva sarcini variate pe date text
- Descrierea de concepte, teorii și modele folosite în domeniul de aplicare

8. Contents

8.1 Course	Teaching methods	Remarks (number of hours, references)
AI and NLP – Introduction	Exposition, participatory lecture, exemplification, applications.	2 hours
Foundations of language. Syntax and semantics	Exposition, participatory lecture, exemplification, applications.	2 hours
Lexicology and lexicography. Dictionaries and thesauri	Exposition, participatory lecture, exemplification, applications.	2 hours
Lexicology and lexicography. Corporuses.	Exposition, participatory lecture, exemplification, applications.	2 hours
Preliminary techniques in NLP.	Exposition, participatory lecture, exemplification, applications.	2 hours
Word representation and distributional semantics.	Exposition, participatory lecture, exemplification, applications.	2 hours
Classification techniques in NLP.	Exposition, participatory lecture, exemplification, applications.	2 hours
Machine translation technologies.	Exposition, participatory lecture, exemplification, applications.	2 hours
Resource interconnection.	Exposition, participatory lecture, exemplification, applications.	2 hours
Speech processing.	Exposition, participatory lecture, exemplification, applications.	2 hours
Advanced applications in NLP.	Exposition, participatory lecture, exemplification, applications.	2 hours
Challenges, ethics and trends in NLP.	Exposition, participatory lecture, exemplification, applications.	2 hours

Bibliography
[1] Daniel Jurafsky and James Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. https://www.cs.colorado.edu/~martin/slp.html .
[2] Eisenstein, Jacob. Introduction to natural language processing. MIT press, 2019. https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf
[3] MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinriche. An introduction to Information Retrieval, 2008. https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf
[4] Bird, S., Klein, E., & Loper, E. (2009). "Natural Language Processing with Python". O'Reilly Media. https://tjzhifei.github.io/resources/NLTK.pdf
[5] Scutelnicu A. (2021) - Representing and interconnecting lexical resources - Ph. D. Thesis. https://scdoc.info.uaic.ro/wp-content/uploads/2023/02/Rezumat_Teza_Doc_AndreiScutelnicu_EN.pdf .
[6] A. V. Aho, M. S. Lam, R. Sethi, and J. D. Ullman, Compilers: Principles, Techniques, and Tools, Second Edition. Addison-Wesley, 2006 https://dl.acm.org/doi/book/10.5555/1177220
[7] Moruz A., Scutelnicu A., Cristea D. (2018) - Interlinking and Extending Large Lexical Resources for Romanian. In V. Păiș, D. Gîfu, D. Trandabăț, D. Cristea, D. Tufiş (eds.) Proceedings of the 13th International Conference on Linguistic Resources and Tools for Processing Romanian Language, Iași, pp. 125-132. https://profs.info.uaic.ro/~dcristea/papers/paper%2014.pdf .
[8] Ivan Habernal - "Semantic Web Search using Natural Language", 2012, https://core.ac.uk/download/pdf/295567817.pdf

8.2 Seminary / Laboratory	Teaching methods	Remarks (number of hours, references)
Exemplification of natural language processing systems.	Problematization. Explanations on the blackboard and with electronic support.	2 hours
Dictionaries and working with them. Partial evaluation 1st homework.	Problematization. Explanations on the blackboard and with electronic support.	2 hours
Corpora and working with them. Partial evaluation 1st homework.	Problematization. Explanations on the blackboard and with electronic support.	2 hours
1st Homework – presentation.	Presentation	2 hours
Tokenization, Lemmatization and POS Tagging. Dependency Parsing. Partial evaluation 2nd homework.	Problematization. Explanations on the blackboard and with electronic support.	2 hours
Word Embeddings. Vector semantics and semantic spaces. Partial evaluation 2nd homework.	Problematization. Explanations on the blackboard and with electronic support.	2 hours
Sentiment analysis. Thematic classification and text categorization. Partial evaluation 2nd homework.	Problematization. Explanations on the blackboard and with electronic support.	2 hours
2nd Homework – presentation.	Presentation	2 hours
Exercises of interconnection of lexical resources.	Problematization. Explanations on the blackboard and with electronic support.	2 hours
Speech signal detection techniques. Automatic translation of the voice signal into text format.	Problematization. Explanations on the blackboard and with electronic support.	2 hours
Chatbots and virtual assistants. Draft preliminary discussions. Presentation of projects	Problematization. Explanations on the blackboard and with electronic support.	2 hours
Final project presentation.	Presentation	2 hours

Bibliography
[1] Daniel Jurafsky and James Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. https://www.cs.colorado.edu/~martin/slp.html .
[2] Eisenstein, Jacob. Introduction to natural language processing. MIT press, 2019. https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf
[3] MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinriche. An introduction to Information Retrieval, 2008. https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf
[4] Bird, S., Klein, E., & Loper, E. (2009). "Natural Language Processing with Python". O'Reilly Media. https://tjzhifei.github.io/resources/NLTK.pdf
[5] Scutelnicu A. (2021) - Representing and interconnecting lexical resources - Ph. D. Thesis. https://scdoc.info.uaic.ro/wp-content/uploads/2023/02/Rezumat_Teza_Doc_AndreiScutelnicu_EN.pdf .
[6] A. V. Aho, M. S. Lam, R. Sethi, and J. D. Ullman, Compilers: Principles, Techniques, and Tools, Second Edition. Addison-Wesley, 2006 https://dl.acm.org/doi/book/10.5555/1177220
[7] Moruz A., Scutelnicu A., Cristea D. (2018) - Interlinking and Extending Large Lexical Resources for Romanian. In V. Păiș, D. Gîfu, D. Trandabăț, D. Cristea, D. Tufiş (eds.) Proceedings of the 13th International Conference on Linguistic Resources and Tools for Processing Romanian Language, Iași, pp. 125-132. https://profs.info.uaic.ro/~dcristea/papers/paper%2014.pdf .
[8] Ivan Habernal - "Semantic Web Search using Natural Language", 2012, https://core.ac.uk/download/pdf/295567817.pdf

9. Coordination of the contents with the expectations of the community representatives, professional associations and relevant employers in the corresponding domain

The content of the discipline is corroborated with the theme of competitions for filling positions in IT companies that have a research/development department dedicated to natural language processing systems

10. Assessment and examination

10.1 Continuous assessment		Percentage (min. 30%)	40		
Course	Assessment type		Oral assessment		
	Percentage		40		
	Failure to pass the continuous assessment results in failure to pass the final assessment		No		
	Assessment methods	Details		Percentage	with reexamination
		Continuous oral assessment		50	No
Continuous oral assessment		50	No		
Seminary / Laboratory	Assessment type		Practical assessment		
	Percentage		60		
	Failure to pass the continuous assessment results in failure to pass the final assessment		No		
	Assessment methods	Details		Percentage	with reexamination
		Project		50	No
Project		50	No		

10.2 Final assessment		Percentage (max. 70%)	60	
		Assessment type	Final written assessment	

10.3 Special notes (special situations is assessment)

--	--

10.4 Minimum performance standard

- Students will possess the fundamental notions in the field of Computational Linguistics, starting with the description of the field and its objectives and continuing with the acquisition of basic theoretical and practical concepts.
- For graduation, a minimum of 5 points from L and a minimum of 5 points from the E are required. Grades are determined using the ROUNDUP formula ($L*0.4 + E*0.6$).

Date,

Course coordinator,
Lecturer PhD. LIVIU-ANDREI SCUTELNICU

Seminary coordinator,

Aproval date in the department,

Head of the department,
Assoc. Prof. PhD. ANDREI ARUSOAIE