COURSE SYLLABUS

Academic year 2024-2025

1. Programme Information

1.1. Higher education institution	**Lucian Blaga University of Sibiu**
1.2. Faculty	Faculty of Engineering
1.3. Department	Department of Computer Science and Electrical and Electronics Engineering
1.4. Field of study	Computer Science and Information Technology
1.5. Level of study[1]	MASTER
1.6. Programme of study/qualification	ADVANCED COMPUTING SYSTEMS

2. Course Information

2.1. Name of course	Advanced Methods in Text Mining			Code	ACS.103.RO
2.2. Course coordinator	Assoc. Prof. Ionel Daniel MORARIU, PhD				
2.3. Seminar/laboratory coordinator	associate professor Eng. Daniel Morariu, PhD				
2.4. Year of study[2]	1	2.5. Semester[3]	1	2.6. Evaluation form[4]	E
2.7. Course type[5]	O	2.8. The formative category of the course[6]			R

3. Estimated Total Time

3.1. Course Extension within the Curriculum – Number of Hours per Week					
3.1.a. Lecture	3.1.b. Seminar	3.1.c. Laboratory	3.1.d. Project	3.1.e. Other	Total
2	0	1	1	0	**4**
3.2. Course Extension within the Curriculum – Total Number of Hours within the Curriculum					
3.2.a. Lecture	3.2.b. Seminar	3.2.c. Laboratory	3.2.d. Project	3.2.e. Other	Total[7]
28	0	14	14	0	**56**

Time Distribution for Individual Study[8]	**Hours**
Learning by using course materials, references and personal notes	12
Additional learning by using library facilities, electronic databases and on-site information	10
Preparing seminars / laboratories, homework, portfolios and essays	56
Tutorial activities[9]	12
Exams[10]	4
3.3. Total Individual Study Hours[11] *(NOSI$_{sem}$)*	**94**
3.4. Total Hours in the Curriculum *(NOAD$_{sem}$)*	**56**
3.5. Total Hours per Semester[12] *(NOAD$_{sem}$ + NOSI$_{sem}$)*	**150**
3.6. No. of Hours / ECTS	**25**
3.7. Number of credits[13]	**6**

4. Prerequisites (if needed)

4.1. Courses that must be successfully completed first (from the curriculum)[14]	Knowledge of statistical techniques for data analysis, data mining, artificial intelligence and machine learning
4.2. Competencies	knowledge in some programing languages

5. Conditions (where applicable)

5.1. For course/lectures[15]	lecture + discussion, video-projector, whiteboard
5.2. For practical activities (lab/sem/pr/app)[16]	Develop and support the planned labs

6. Specific competencies acquired[17]

Number of credits assigned to the discipline[18]		6	Credits distribution by competencies[19]
6.1. Professional competencies	PC3	analyses massive groups of data	1
	PC11	develop data processing applications	1
	PC16	performs analytical mathematical calculations	1
	PC21	interpret current data	1
	PC23	present analysis results	1
6.2. Transversal competencies	TC4	works in teams	1

7. Course objectives (resulted from developed competencies)

7.1. Main course objective	Understanding the main concepts, algorithms and techniques of text mining. Learning the information retrieval system. Understanding the importance and applicability of the text mining field and its use in the processing of text documents
7.2. Specific course objectives	Knowledge of at least one software package specialized in natural language processing. Implementation in a programming language of discipline-specific algorithms.

8. Content

8.1. Lectures[20]		**Teaching methods[21]**	**Hours**
Lecture 1	Data mining process. Data mining vs. machine learning	*Exposition, board, discussions with students*	2
Lecture 2	Text mining. General architecture for a text mining system	*Exposition, board, discussions with students*	2
Lecture 3	Architecture of information extraction system	*Exposition, board, discussions with students*	2
Lecture 4	Text mining preprocessing. Bag-of-word representation.	*Exposition, board, discussions with students*	2
Lecture 5	Text mining preprocessing. Syntactic representation of documents	*Exposition, board, discussions with students*	2

Lecture 6	Advanced methods of feature selections.	*Exposition, board, discussions with students*	2
Lecture 7	Text Categorization and Clustering	*Exposition, board, discussions with students*	2
Lecture 8	Evaluation of learning algorithms	*Exposition, board, discussions with students*	2
Lecture 9	Classification algorithm –Support vector Machine	*Exposition, board, discussions with students*	2
Lecture 10	Word sense disambiguation.	*Exposition, board, discussions with students*	2
Lecture 11	Part-of-speech tagging	*Exposition, board, discussions with students*	2
Lecture 12	Text summarization	*Exposition, board, discussions with students*	2
Lecture 13	Question answering	*Exposition, board, discussions with students*	2
Lecture 14	Word embedding.	*Exposition, board, discussions with students*	2
		Total lecture hours:	**28**

8.2. Practical activities (8.2.a. Seminar[22]/ 8.2.b. Laboratory[23]/ 8.2.c. Project[24])		**Teaching methods**	**Hours**
Act.1	Lab 1. Word and PoS extraction from Brown Corpus	*Exercise*	2
Act.2	Pr. 1. Word sense disambiguation system	*Practical demonstration*	2
Act.3	Lab 2. Reducing the number of PoS	*Exercise*	2
Act.4	Pr. 2. Part of speech tagging system	*Practical demonstration*	2
Act.5	Lab 3. Representing the relation between PoS	*Exercise*	2
Act.6	Pr. 3. Text summarization system	*Practical demonstration*	2
Act.7	Lab 4. Non context method for PoS	*Exercise*	2
Act.8	Pr. 4. Expression extraction system	*Practical demonstration*	2
Act.9	Lab 5. Naïve Bayes classifier for PoS	*Exercise*	2
Act.10	Pr. 5. Sentiment analysis system	*Practical demonstration*	2
Act.11	Lab 6. Evaluation methods for learning algorithms	*Exercise*	2
Act.12	Pr. 6. Question Answering system	*Practical demonstration*	2
Act.13	Lab 7. Viterbi algorithm for PoS	*Exercise*	2
Act.14	Pr. 7. Word embedding system	*Practical demonstration*	2
		Total seminar/laboratory hours:	28

9. Bibliography

9.1. Recommended Bibliography	D. Morariu, *Text Mining Methods based on Support Vector Machine*, Matrix ROM, 2008
	Ronen Feldman, James Sanger, *The Text Mining Handbook. Advanced approached in analyzing unstructured data*. Cambridge University Press, 2007
	Dan Jurafsky and James H. Martin, Speech and Language Processing (3rd ed. draft), published online, 2020, https://web.stanford.edu/~jurafsky/slp3/
	Jiawei Han, Micheline Kamber and Jian Pei, *Data Mining: Concepts and Techiques*, The Morgan Kaufmann Series, 2011
	David Grossman, Ophir Frieder, Information Retrieval Algorithms and Heuristics, Springer, 2004
9.2. Additional Bibliography	Christopher Bishop, *Pattern Recognition and Machine Learning*, Editura Springer, 2006

	Ruslan Mitkov, *The oxford Handbook of Computational Lingvistics*, Oxford university press, 2003

10. Conjunction of the discipline's content with the expectations of the epistemic community, professional associations and significant employers of the specific study program[25]

Approaching Big Data and artificial intelligence systems and regular discussions are held in a formal and informal setting with the representatives of the profile companies.

11. Evaluation

Activity Type	11.1 Evaluation Criteria	11.2 Evaluation Methods		11.3 Percentage in the Final Grade	Obs.[26]
11.4a Exam / Colloquy	• Theoretical and practical knowledge acquired (quantity, correctness, accuracy)	Tests during the semester[27]:	20%	60%	
		Homework:	0%		
		Other activities[28]:	0%		
		Final evaluation:	40%		
11.4b Seminar	• Frequency/relevance of participation or responses	Evidence of participation, portfolio of papers (reports, scientific summaries)		0%	
11.4c Laboratory	• Knowledge of the equipment, how to use specific tools; evaluation of tools, processing and interpretation of results	• Written questionnaire • Oral response • Laboratory notebook, experimental works, reports, etc. • Practical demonstration		10%	
11.4d Project	• The quality of the project, the correctness of the project documentation, the appropriate justification of the chosen solutions	• Self-evaluation, project presentation • Critical evaluation of a project		30%	
11.5 Minimum performance standard[29]					5

The Course Syllabus will encompass components adapted to persons with special educational needs (SEN – people with disabilities and people with high potential), depending on their type and degree, at the level of all curricular elements (skills, objectives, contents, teaching methods, alternative assessment), in order to ensure fair opportunities in the academic training of all students, paying close attention to individual learning needs.

Filling Date: 13.09.2024

Department Acceptance Date: 16.09.2024

	Academic Rank, Title, First Name, Last Name	Signature
Course Teacher	Assoc. Prof. Ionel Daniel MORARIU, PhD	
Study Program Coordinator	Prof. Adrian FLOREA, PhD	

Head of Department	Assoc. Prof. Radu George CREȚULESCU, PhD	
Dean	Prof. Maria VINȚAN, PhD	

[1] *Bachelor / Master*
[2] *1-4 for bachelor, 1-2 for master*
[3] *1-8 for bachelor, 1-3 for master*
[4] *Exam, colloquium or VP A/R - from the curriculum*
[5] *Course type: R = Compulsory course; E = Elective course; O = Optional course*
[6] *Formative category: S = Specialty; F = Fundamental; C = Complementary; I = Fully assisted; P = Partially assisted; N = Unassisted*
[7] *Equal to 14 weeks x number of hours from point 3.1 (similar to 3.2.a.b.c.)*
[8] *The following lines refer to individual study; the total is completed at point 3.37.*
[9] *Between 7 and 14 hours*
[10] *Between 2 and 6 hours*
[11] *The sum of the values from the previous lines, which refer to individual study.*
[12] *The sum (3.5.) between the number of hours of direct teaching activity (NOAD) and the number of hours of individual study (NOSI) must be equal to the number of credits assigned to the discipline (point 3.7) x no. hours per credit (3.6.)*
[13] *The credit number is computed according to the following formula, being rounded to whole neighbouring values (either by subtraction or addition*

$$No.credits = \frac{NOCpSpD \times C_C + NOApSpD \times C_A}{TOCpSdP \times C_C + TOApSdP \times C_A} \times 30\ credits$$

Where:

- NOCpSpD = Number of lecture hours / week / discipline for which the credits are calculated
- NOApSpD = Number of application hours (sem./lab./pro.) / week / discipline for which the credits are calculated
- TOCpSdP = Total number of course hours / week in the Curriculum
- TOApSdP = Total number of application hours (sem./lab./pro.) / week in the Curriculum
- C_C/C_A = Course coefficients / applications calculated according to the table

Coefficients	Course	Applications (S/L/P)
Bachelor	2	1
Master	2,5	1,5
Bachelor - foreign language	2,5	1,25

[14] *The courses that should have been previously completed or equivalent will be mentioned*
[15] *Board, video projector, flipchart, specific teaching materials, online platforms, etc.*
[16] *Computing technology, software packages, experimental stands, online platforms, etc.*
[17] *Competences from the Grids related to the description of the study program, adapted to the specifics of the discipline*
[18] *From the curriculum*
[19] *The credits allocated to the course are distributed across professional and transversal competences according to the specifics of the discipline*
[20] *Chapter and paragraph titles*
[21] *Exposition, lecture, board presentation of the studied topic, use of video projector, discussions with students (for each chapter, if applicable)*
[22] *Discussions, debates, presentations and/or analyses of papers, solving exercises and problems*
[23] *Practical demonstration, exercise, experiment*
[24] *Case study, demonstration, exercise, error analysis, etc.*
[25] *The relationship with other disciplines, the usefulness of the discipline on the labour market*
[26] *CPE – Conditions Exam Participation; nCPE – Does Not Condition Exam Participation; CEF - Conditions Final Evaluation; N/A – not applicable*
[27] *The number of tests and the weeks in which they will be taken will be specified*
[28] *Scientific circles, professional competitions, etc.*
[29] *The minimum performance standard in the competence grid of the study program is customized to the specifics of the discipline, if applicable*