

COURSE DESCRIPTION

1. Program details

University	"Lucian Blaga" University of Sibiu
Faculty	Engineering Faculty
Department	Department of Computer Science and Electrical and Electronics Engineering
Main field of study	Computer Engineering and Information Technology
Level of education	RESEARCH MASTER
Specialization	ADVANCED COMPUTING SYSTEMS

2. Course details

Course title	Advanced Methods in Text Mining			
Course code	Type of course	Year of study	Semester	Number of credits
mACS.103.S0	obligatory	1	1	6
Evaluation type	Type of course (FD=fundamental discipline.; DD=domain discipline; SD=specialized discipline; CD=complementary discipline)			
Exam	SD			
Course instructor	associate professor eng. Daniel Morariu, PhD			
Seminar/lab/project instructor	associate professor eng. Daniel Morariu, PhD			

3. Estimated time

Course duration in the curriculum – number of hours per week				
Lecture	Seminar	Lab	Project	Total
2	-	1	1	4
Course duration in the curriculum - Total of hours curriculum				
Lecture	Seminar	Lab	Project	Total (<i>NOAD_{sem}</i>)
28	-	14	14	56

Distribution of hours for individual study		No. hours
Individual study using course handbooks, bibliography and notes		25
Additional documentation in library and on specialized electronic platforms		6
Preparing seminars / labs, homework, essays and portfolios		20
Tutoring		2
Exam preparation		3
Total hours for individual study (<i>NOSI_{sem}</i>)		56
Total hours per semester (<i>NOAD_{sem}</i> + <i>NOSI_{sem}</i>)		112

4. Prerequisites (if applicable)

curriculum	Knowledge of statistical techniques for data analysis, data mining, artificial intelligence and machine learning
competencies	knowledge in some programming languages

5. Conditions (if applicable)

course materials	Active participation in classes, lecture + discussion, video-projector, whiteboard
sem/lab/project materials	Develop and support the planned labs.

6. Specific competences acquired

Professional competence	Knowledge and proper operation of discipline specific advanced concepts. Ability to integrate knowledge gained from other disciplines ability to work with and integrate specific sources of information. Understanding of innovative and advanced architectures for information processing and their research methodologies A unified view, holistic on the computer engineering
Transversal competences	The positive reaction to suggestions, requirements, didactic tasks, satisfaction to respond Involvement in scientific activities related to course Accepting a value assigned to an object, phenomenon, behavior etc. Ability to collaborate with experts in other fields

7. Objectives (based on the specific grid for the accumulated competences)

General objective	Understand the main concepts, algorithms and text mining techniques Preprocessing text mining methodology and some knowledge in natural language processing. Understand the importance and applicative scope of text mining and its use in the processing of text documents
Specific objectives	Knowledge at least one specialized software in text mining. Implement an entire text mining process.

8. Contents

Course	No. hours	
Course 1	Data mining process. Data mining vs. machine learning	2
Course 2	Text mining. General architecture for a text mining system	2
Course 3	Architecture of information extraction system	2
Course 4	Text mining preprocessing. Bag-of-word representation.	2
Course 5	Text mining preprocessing. Syntactic representation of documents	2
Course 6	Advanced methods of feature selections.	2
Course 7	Text Categorization and Clustering	2
Course 8	Evaluation of learning algorithms	2
Course 9	Classification algorithm –Support vector Machine	2
Course 10	Word embedding.	2
Course 11	Word sense disambiguation.	2
Course 12	Part-of-speech tagging	2
Course 13	Text summarization	2
Course 14	Question answering	2
Total course hours:		28

Laboratory		No. hours
Lab 1	Word extraction from Reuters database	1
Lab 2	Eliminating the stop-words and stemming	1
Lab 3	Vector Space Model representation	1
Lab 4	Information Gain for feature selection	1
Lab 5,6	Hierarchical Agglomerative Clustering algorithm for text documents	2
Lab 7,8	Evaluation methods for learning algorithms	2
Lab 9,10	Kernel method for classifying algorithm. SVM Algorithm.	2
Lab 11,12	WordNet for part-of-speech tagging	2
Lab 13,14	SVM for classifying documents	2
Total lab hours:		14

Project		No. hours
Project 1	Word sense disambiguation system	2
Project 2	Part of speech tagging system	2
Project 3	Text summarization system	2
Project 4	Expression extraction system	2
Project 5	Sentiment analysis system	2
Project 6	Question Answering system	2
Project 7	Word embedding system	2
Total lab hours:		14

Teaching methods

The presentation of the main ideas with discussions on the themes. Laboratory that emphasizes the concepts presented in the course. PowerPoint slides for course.	Language of instruction	English
---	-------------------------	---------

References

Recommended reading	D. Morariu, <i>Text Mining Methods based on Support Vector Machine</i> , Matrix ROM, 2008
	Ronen Feldman, James Sanger, <i>The Text Mining Handbook. Advanced approached in analyzing unstructured data</i> . Cambridge University Press, 2007
	Jiawei Han, Micheline Kamber and Jian Pei, <i>Data Mining: Concepts and Techniques</i> , The Morgan Kaufmann Series, 2011
	David Grossman, Ophir Frieder, <i>Information Retrieval Algorithms and Heuristics</i> , Springer, 2004
More references	Christopher Bishop, <i>Pattern Recognition and Machine Learning</i> , Editura Springer, 2006
	Ruslan Mitkov, <i>The oxford Handbook of Computational Linguistics</i> , Oxford university press, 2003

9. Linking course content with expectations of the epistemic community representatives, professional associations and employers' representatives in the field related to the program

--

10. Evaluation

Type	Evaluation criteria	Evaluation methods	Percentage in final grade	Obs.*
Course	Semester exam	written	40%	RFE
	course participation		5%	nREP
	course report	oral/written	15%	RFP
Lab	Project activity	Oral	30%	RFP
	Lab activity	oral presentation	10%	RFE
Minimum standard of performance				
50% result after summing weighted scores in column 4				

(*) REP – required for exam participation; nREP – not required for exam participation; RFE – required for final evaluation.

Date of completion:

Date of approval in the Department:.....

	Position, title, first name, surname	Signature
Course instructor	associate professor eng. Daniel Morariu, PhD	
Head of department	Prof. eng. Daniel VOLOVICI, PhD	