SYLLABUS

Applications of Computational Linguistics

University year 2025-2026

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	Computer Science
1.5. Study cycle	Master
1.6. Study programme/Qualification	Applied Computational Intelligence
1.7. Form of education	Full time

2. Information regarding the discipline

2.1. Name of the discipline	Applications of Computational Linguistics					Discipline code	MME8062
2.2. Course coordinator				Lecturer PhD. Dana Lupsa			
2.3. Seminar coordinator				Lecturer PhD. Dana Lupsa			
2.4. Year of study	2	2.5. Semester	3	2.6. Type of evaluation	E	2.7. Discipline regime	Compulsory

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

3.1. Hours per week	**4**	of which: 3.2 course	**2**	3.3 seminar/laboratory/project	**1 +1**
3.4. Total hours in the curriculum	56	of which: 3.5 course	28	3.6 seminar/laboratory/project	**28**
Time allotment for individual study (ID) and self-study activities (SA)					**hours**
Learning using manual, course support, bibliography, course notes (SA)					38
Additional documentation (in libraries, on electronic platforms, field documentation)					40
Preparation for seminars/labs, homework, papers, portfolios and essays					50
Tutorship					4
Evaluations					8
Other activities:					4
3.7. Total individual study hours	**144**				
3.8. Total hours per semester	**200**				
3.9. Number of ECTS credits	**8**				

4. Prerequisites (if necessary)

4.1. curriculum	Algorithmics, data structures, statistics
4.2. competencies	Ability to write computer programs in a high level programming language. Average programming skills

5. Conditions (if necessary)

5.1. for the course	
5.2. for the seminar /lab activities	Room with computers are needed; high level programming language environment.

6.1. Specific competencies acquired [1]

[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.

Professional/essential competencies	• understanding and working with basic concepts in computational intelligence; • assimilation of mathematical concepts and formal mode/s to understand, verify and validate software systems.
Transversal competencies	• capability of information analysis and synthesis; • professional communication skills; concise and precise description, both oral and written, of professional results, negotiation abilities;

6.2. Learning outcomes

Knowledge	• The graduate possesses the fundamental knowledge for modelling, being able to analyse real life problems and to translate them in concrete requirements and to design a corresponding software model • The graduate proves knowledge related to specifying the requirements of research activities in the domain of computer science in general and computational intelligence in particular and he/she understands the role of research in promoting progress
Skills	• The graduate has the ability of interdisciplinary vision between computer science subdomains in order to combine them in a software system computational intelligence • The graduate is able to carry on activities for education and training on different topics related to software systems and computational intelligence • The graduate can use specific language and terminology for the field of computational intelligence being able to communicate and interact with members of a team
Responsibility and autonomy:	• The graduate proves abilities to work independently in order to obtain knowledge necessary for designing, managing and evaluating research activities in the field of computational intelligence • The graduate has the ability to combine information in different ways in order to form a positive attitude towards its his/her own development

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

7.1 General objective of the discipline	• Understand how IR and MT systems works • Identify techniques for information retrieval, language translation
7.2 Specific objective of the discipline	• Design and implement a small to medium size information storage and retrieval system, or digital library. • Understand techniques for information retrieval and language translation • Types of text data. The influence on CL results

8. Content

8.1 Course	Teaching methods	Remarks
1. Ontologies		
2. Information retrieval. A first view. Vector space model (VSM)		

3. Boolean model. Extensions. 4. Information Storage and Retrieval: Inverted Index 5. 5.1 Statistical properties of texts. 5.2 Vector similarity: issues 6. 6.1 Dictionaries and tolerant retrieval 6.2 Meta-data, fields and zones 7. Evaluation in IR 8. Semantic aspects 9. Relevance feedback. Query expansion. 10. Probabilistic retrieval 11. Language models 12. Web as graph: link analysis 13. Efficient scoring and ranking: issues. Inexact top k retrieval 14. Machine translation (MT).	• Interactive exposure • Explanation • Conversation • Didactical demonstration	

Bibliography

1. ALLEN, J.F. *Natural Language Semantics*, Wiley-Blackwell; 1 edition, 2001
2. D. Arnold, L. Balkan, S. Meijer, R. Humphreys, L. Sadler, *Machine Translation: An Introductory Guide, Manchester*, UK: NEC Blackwell, 1994.
3. R. BAEZA-YATES, B. RIBEIRO-NETO , *Modern Information Retrieval*, Addison-Wesley, 1999
4. E. CHARNIAK: *Statistical language learning*, MIT Press, 1996.
5. O. DAMERON, *Ontology-based methods for analyzing life science data*. Bioinformatics, Univ. Rennes , 2016
6. C.MANNING, H.SCHUTZE, *Foundation of statistical natural language processing*, MIT, 1999.
7. C. MANNING, P. RAGHAVAN, H. SCHUTZE, *Introduction to Information Retrieval*, Cambridge University Press, 2008.
8. R. MITKOV ed., *The Oxford Handbook of Computational Linguistics* (Oxford Handbooks in Linguistics), 2005
9. http://protege.stanford.edu
10. https://meshb.nlm.nih.gov/
11. G.M. Linders, M.M. Louwerse, *Zipf's law revisited*: Spoken dialog, linguistic units, parameters, and the principle of least effort. Psychon Bull Rev (2022)
12. Hugh E. Williams, Justin Zobel , Dirk Bahle, *Fast phrase querying with combined indexes*, ACM Transactions on Information Systems, Volume 22, Issue 4, 2004 pp 573–594
13. Singhal A., Buckley C., Mitra M., *Pivoted Document Length Normalization*, SIGIR 1996
14. Andrei Z. Broder, David Carmel, Michael Herscovici, Aya Soffer, Jason Zien, *Efficient Query Evaluation using a Two-Level Retrieval Process,* 2003
15. T. Suel , S. Ding, *Faster top-k document retrieval using block-max indexes*. SIGIR'2011.
16. Huahai, Yunyao , 2021 , T-Wand: Beat Lucene in Less Than 600 Lines of Code, https://yyhh.org/blog/2021/11/t-wand-beat-lucene-in-less-than-600-lines-of-code/
17. DAN JURAFSKY, JAMES H. MARTIN, Speech and Language Processing (3rd ed. draft) (https://web.stanford.edu/~jurafsky/slp3/)

8.2 Seminar / laboratory	Teaching methods	Remarks
1,2: Knowledge representation in ontologies. Examples	Discussion, examples, experiments	
3,4: Information in text; retrieval and ranking. Experiments. Examples	Discussion, examples, experiments	
5,6: Applications of theoretical techniques discussed in class . Experiments.	Discussion, examples, experiments	
7: Recent research in CL	Discussion, experiments	

Bibliography

1. R. Mitkov (Ed), Oxford Handbook of Computational Linguistics. Oxford University Press, 2003.
2. C.D. Manning, P. Raghavan, H. Schütze,. Introduction to Information Retrieval. Cambridge, England: Cambridge University Press, 2008.

http://nlp.stanford.edu/IR-book/html/htmledition/irbook.html
3. http://www.mt-archive.info/
4. http://www.statmt.org/
5. https://protege.stanford.edu/

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

- Two fields of CL with immediate applications in real life are IR and MT. IR systems are used on an everyday basis by a wide variety of users. The Internet has proven to be a huge stimulus for MT, with hundreds of millions of pages of text and an increasingly global -- and linguistically diverse – public.
- The content of the discipline is consistent with the similar disciplines from other Romanian universities and universities from abroad

10. Evaluation

Activity type	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Percentage of final grade
10.4 Course	- know the basic principle of the domain; - apply the studied concepts	Written exam (in the regular session)	30%
10.5 Seminar/laboratory	- apply the studied concepts - make experiments and solve problem	Oral presentation	20%
		Research report (presentations and experiments)	50 %
			Other activities evaluated as bonus points
10.6 Minimum standard of performance			

- Each student has to prove that (s)he acquired an acceptable level of knowledge and understanding of the field, that (s)he is capable of stating this knowledge in a coherent form, that (s)he has the ability to establish certain connections and to use the knowledge in solving different problems.
- Final grade has to be at least 5.

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

Not applicable.

Date:
...

Signature of course coordinator

Lect. PhD. Lupsa Dana

Signature of seminar coordinator

lect. PhD. Lupsa Dana

[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.*".

Date of approval:
...

Signature of the head of department

Assoc.prof.phd. Adrian STERCA