SYLLABUS

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

2. Information regarding the discipline

2.1 Name of the discipline	Applications of Computational Linguistics						
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 Type of discipline	**Compulsory**
2.8 Course code	**MME8062**						

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

3.1 Hours per week	3	Of which: 3.2 course	2	3.3 seminar/laboratory	1 sem + 1pr
3.4 Total hours in the curriculum	56	Of which: 3.5 course	28	3.6 seminar/laboratory	28
Time allotment:					hours
Learning using manual, course support, bibliography, course notes					30
Additional documentation (in libraries, on electronic platforms, field documentation)					44
Preparation for seminars/labs, homework, papers, portfolios and essays					30
Tutorship					20
Evaluations					20
Other activities:					

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	
4.2. competencies	• Average programming skills • Knowledge of data structures

5. Conditions (if necessary)

5.1. for the course	•
5.2. for the seminar /lab activities	•

6. Specific competencies acquired

Professional competencies	• Knowledge, understanding and use of IR concepts and their algorithms • Knowledge, understanding and use of of MT systems
Transversal competencies	• Perform Internet-based research. • Ability to use techniques specific to information retrieval and machine translation.

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	• Explain basic information storage and retrieval concepts. • Describe what issues are specific to efficient information retrieval. • Perform Internet-based research. • Design and implement a small to medium size information storage and retrieval system, or digital library. • Understand techniques for information retrieval, language translation • Study the usage of corpora. Identify examples of corpora for MT

8. Content

8.1 Course	Teaching methods	Remarks
1. Ontologies	Exposure, examples, discussion	
2. Boolean and set models	Exposure, examples, discussion	
3. Elements of VSM models	Exposure, examples, discussion	
4. Inverted Index	Exposure, examples, discussion	
5. Statistical properties of texts.	Exposure, description, explanation, examples, case studies	

6. 6.1 Dictionaries and tolerant retrieval 6.2 Meta-data, fields and zones	Exposure, description, explanation, examples, case studies	
7. Evaluation in IR	Exposure, description, explanation, examples, case studies	
8. Semantic aspects	Exposure, description, explanation, examples, case studies	
9. Relevance feedback. Query expansion.	Exposure, description, explanation, examples, case studies	
10. Probabilistic retrieval	Exposure, examples, discussion	
11. Language models	Exposure, examples, discussion	
12. Web as graph: link analysis		
13. Efficient scoring and ranking: issues. Inexact top k retrieval	Description, examples, case studies, discussion	
14. Machine translation (MT).	Exposure, description, explanation, examples, case studies, discussion	

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/

8.2 Seminar / laboratory	Teaching methods	Remarks
1,2. Ontologies Examples. Problems Survey of available information sources. Choose paper topics and schedule presentations	Interactive exposure Explanation Conversation Didactical demonstration	
3,4: Retrieval and ranking examples. Experiments	Dialogue, examples	
5: Applications of theoretical techniques. Experiments.	Dialogue, case studies, examples	
6. Research report presentations	Conversation	
7: Recent research in CL	Interactive exposure Conversation	

Bibliography

1. 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
2. https://protege.stanford.edu/
3. http://www.mt-archive.info/
4. http://www.statmt.org/

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 course respects ACM Curriculla Recommendations for Computer Science studies

10. Evaluation

Type of activity	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Share in the grade (%)
10.4 Course	- the basic principle of the domain; - apply the studied concepts	Written exam (in the regular session)	30%
10.5 Seminar/lab activities	- apply the studied concepts - make experiments	Oral presentation / interaction	20%
	Researh report	Evaluation of the research report	50 %
			Other activities evaluated as bonus points
10.7 Minimum performance standards			
➢ At least grade 5 (from a scale of 1 to 10) for the final computed grade			

Date

Signature of course coordinator

Lect. Phd. Dana Lupsa

Signature of seminar coordinator

Lect. Phd. Dana Lupsa

Date of approval

Signature of the head of department

Assoc. prof. phd. Sterca Adrian