SYLLABUS

1.1 Higher education institution	**Babeş-Bolyai University**
1.2 Faculty	**Faculty of Mathematics and Computer Science**
1.3 Department	**Department of Mathematics and Computer Science of the Hungarian Line**
1.4 Field of study	**Computer Science**
1.5 Study cycle	**Master**
1.6 Study programme / Qualification	**Data Analysis and Modelling**

2. Information regarding the discipline

2.1 Name of the discipline	**Information retrieval** / Információ-visszakeresés / Regăsirea informaţiei						
2.2 Course coordinator					Lect. prof. dr. Szenkovits Annamária		
2.3 Seminar coordinator					Lect. prof. dr. Szenkovits Annamária		
2.4. Year of study	1	2.5 Semester	1	2.6. Type of evaluation	E	2.7 Type of discipline	Compulsory
2.8. Code of the discipline	MME8032						

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
3.4 Total hours in the curriculum	42	Of which: 3.5 course	28	3.6 seminar/laboratory	14
Time allotment:					hours
Learning using manual, course support, bibliography, course notes					54
Additional documentation (in libraries, on electronic platforms, field documentation)					30
Preparation for seminars/labs, homework, papers, portfolios and essays					64
Tutorship					6
Evaluations					4
Other activities:					-
3.7 Total individual study hours	158				
3.8 Total hours per semester	200				
3.9 Number of ECTS credits	8				

4. Prerequisites (if necessary)

4.1. curriculum	None
4.2. competencies	Algorithms, programming skills, basic math (algebra, probability theory, statistics)

5. Conditions (if necessary)

5.1. for the course	Video projector and blackboard/whiteboard
5.2. for the seminar /lab activities	Laboratory with computers; high level programming language environment(s) (e.g. Python, Julia); Matlab

6. Specific competencies acquired

Professional competencies	● Understanding the concepts, methods and models used in Information Retrieval (IR). ● Understanding the principles, design and implementation of data storage techniques, conversion between formats. ● Study and analysis of algorithms, that retrieve/extract information from textual databases.
Transversal competencies	● Responsible execution of lab assignments, research and practical reports. ● Application of efficient and rigorous working rules. ● Manifest responsible attitudes toward the scientific and didactic fields. ● Respecting the professional and ethical principles.

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

7.1 General objective of the discipline	● To present the field of IR, studying and analyzing the algorithms used in IR.
7.2 Specific objective of the discipline	● The basics of IR: ○ Basic concepts: document and term lists, document-term, term-document matrices, posting lists, indices ○ Building indices ○ Binary IR ○ Probabilistic models in IR ○ The Vector Space Model (VSM) ○ Supervised and unsupervised learning in IR ● Design and application of search engines

8. Content

8.1 Course	Teaching methods	Remarks
1. Introductory concepts, definitions, introduction to information retrieval systems. **keywords**: definition of information retrieval, boolean retrieval, inverted index (term-document incidence matrix, postings list)	interactive exposure, explanation, conversation, didactical demonstration	
2.1 The term vocabulary and postings lists. **keywords**: data structures for the term vocabulary (hashes, trees), wildcard queries, permuterm and k-gram indexes, edit distance **2.2** Dictionaries and tolerant retrieval **keywords**: data structures for the term vocabulary (hashes, trees), wildcard queries, permuterm and k-gram indexes, edit distance	interactive exposure, explanation, conversation, didactical demonstration	
3. Scoring, term weighting. The vector space model) **keywords**: ranked retrieval, Jaccard coefficient, bag of words model, tf-idf weights, vector space model	interactive exposure, explanation, conversation, didactical demonstration	
4-5. Elasticsearch and Kibana introduction **keywords:** Elasticsearch query DSL and analyzers, Kibana visualizations and maps	interactive exposure, explanation, conversation, didactical demonstration	
6. Evaluation of IR systems. **keywords:** precision and recall, F score, benchmark datasets	interactive exposure, explanation, conversation, didactical demonstration	
7. Language models in IR. Text classification and Naive Bayes	interactive exposure, explanation, conversation, didactical demonstration	
8. Vector Space Classification. Feature selection methods **keywords:** Classification methods in IR: Naive Bayes, Rocchio's algorithm, regularized least squares (RLS), support vector machines (SVM)	interactive exposure, explanation, conversation, didactical demonstration	

9-10. Introduction to Attention and Transformer models. Interpretability of Attention models	interactive exposure, explanation, conversation, didactical demonstration	
11. Introduction to the CLIP embedding space	interactive exposure, explanation, conversation, didactical demonstration	
12. Unsupervised learning in IR **keywords:** flat and hierarchical clustering	interactive exposure, explanation, conversation, didactical demonstration	
13-14. Project presentations, discussions.	interactive exposure, explanation, conversation	

Bibliography

[1] Manning C.D., Raghavan P., Schütze H. *Introduction to Information Retrieval*. Cambridge University Press, 2009.

[2] BAEZA-YATES R., RIBEIRO-NETO B. *Modern Information Retrieval*. Addison-Wesley, 1999. [3] VAN RIJSBERGEN C. J. *Information Retrieval* (2nd ed.). Butterworths, 1979.

[4] DOMINICH S. *The Modern Algebra of Information Retrieval*. Springer, 2008.

[5] BODON F. *Adatbányászati algoritmusok*. GNU Free Documentation License, 2010 (http://www.cs.bme.hu/~bodon/magyar/adatbanyaszat/tanulmany/adatbanyaszat.pdf).

8.2 Seminar / Laboratory	Teaching methods	Remarks
1. Introduction to Perl and/or Python programming.	documentation, explanation, conversation	
2. Famous classification algorithms in IR: Naive Bayes, Rocchio, SVM, MLPM etc.	documentation, explanation, conversation	
3-4. Classification in IR with transformer models (BERT).	documentation, explanation, conversation	
5-6. Elasticsearch indexing/search engine.	documentation, explanation, conversation	

7. Summary, project presentations.		Student presentations on selected related topics.

Bibliography

[1]–[5] +
[6] MANNING C. D., SCHÜTZE H. *Foundations of statistical language processing*. MIT Press, Cambridge, 1999.
[7] SEBASTIANI F. Machine Learning in Automated Text Categorization. ACM Computing Surveys, 2002, vol. 34, pp. 1–47.
[8] http://nlp.stanford.edu/IR-book/
[9] http://www.stanford.edu/class/cs276/

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

The content of the discipline is consistent with the course "Information Retrieval and Web Search" at Stanford University (http://web.stanford.edu/class/cs276/), and is based on the book "Introduction to Information Retrieval" by Manning, Raghavan and Schütze (http://nlp.stanford.edu/IR-book/, see also the bibliography above).

10. Evaluation

Type of activity	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Share in the grade
10.4 Course	Tests at the beginning of the courses	Written tests	15%
	Written exam at the end of the semester	Written exam	30%
10.5 Seminars/laboratory	Seminar/laboratory assignments during the semester	Evaluation of the programming assignments	30%
	Presentation of the software projects	Evaluation of the project	25%
10.6 Minimum performance standards			
At the presentation of the software projects and seminar/laboratory assignments during the semester, minimum half of the points needs to be collected.			

Date

19.09.2025

Date of approval

19.09.2025

Signature of course coordinator

Dr. Szenkovits Annamária

Signature of seminar coordinator

Dr. Szenkovits Annamária

Signature of the head of department

Dr. András Szilárd