

FIȘA DISCIPLINEI

1. Date despre program

1.1 Instituția de învățământ superior	Universitatea Babeș-Bolyai Cluj-Napoca
1.2 Facultatea	Facultatea de Matematică și Informatică
1.3 Departamentul	Departamentul de Matematică și Informatică al Liniei Maghiare
1.4 Domeniul de studii	Calculatoare și tehnologia informației
1.5 Ciclu de studii	Master
1.6 Programul de studiu / Calificarea	Analiza datelor și modelare (în limba maghiară)

2. Date despre disciplină

2.1 Denumirea disciplinei	Information retrieval / Információ-visszakeresés / Regăsirea informației						
2.2 Titularul activităților de curs	Lect. dr. Szenkovits Annamária						
2.3 Titularul activităților de semianr	Lect. dr. Szenkovits Annamária						
2.4 Anul de studiu	1	2.5 Semestrul	1	2.6 Tipul de evaluare	E	2.7 Regimul disciplinei	obligatoriu
2.8. Codul disciplinei	MME8032						

3. Timpul total estimat (ore pe semestru al activităților didactice)

3.1 Număr de ore pe săptămână	3	din care: 3.2 curs	2	3.3 seminar/laborator	1
3.4 Total ore din planul de învățământ	42	din care: 3.5 curs	28	3.6 seminar/laborator	14
Distribuția fondului de timp:					ore
Studiul după manual, suport de curs, bibliografie și notițe					54
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					30
Pregătire seminarii/laboratoare, teme, referate, proiecte					64
Tutoriat					6
Examinări					4
Alte activități:					
3.7 Total ore de studiu individual	158				
3.8 Total ore pe semestru	200				
3.9 Numărul de credite	8				

4. Precondiții (acolo unde este cazul)

4.1 de curriculum	Nu sunt.
4.2 de competențe	Algoritmi, abilități de programare, matematică de bază (algebră, teoria probabilităților, statistică)

5. Condiții (acolo unde este cazul)

5.1 de desfășurare a cursului	Sală de curs dotată cu tablă și videoproiector.
5.2 de desfășurare a seminarului / laboratorului	Sală de calculatoare sau laptop propriu, Matlab și/sau Python, Julia.

6. Competențele specifice acumulate

Competențe profesionale	• Înțelegerea conceptelor, metodelor și modelelor utilizate în regăsirea informației (IR). • Înțelegerea principiilor, proiectării și implementării tehnicilor de stocare a datelor, conversia între formate. • Studiul și analiza algoritmilor care recuperează/extrag informații din baze de date textuale.
Competențe transversale	• Executarea responsabilă a sarcinilor de laborator, a cercetărilor și a rapoartelor practice. • Aplicarea unor reguli de lucru eficiente și riguroase. • Manifestarea unei atitudini responsabile față de domeniile științific și didactic. • Respectarea principiilor profesionale și etice.

7. Obiectivele disciplinei (reieșind din grila competențelor acumulate)

7.1 Obiectivul general al disciplinei	• Prezentarea domeniului IR, studierea și analizarea algoritmilor utilizați în IR.
7.2 Obiective specifice	• Noțiuni de bază privind IR: • Noțiuni fundamentale: liste de documente și termeni, matrice document-termen, matrice termen-document, liste de postare, indici • Crearea de indici • R binar • Modele probabilistice în IR • Modelul spațiului vectorial (VSM) • Învățare supravegheată și nesupravegheată în IR • Proiectarea și aplicarea motoarelor de căutare

8. Conținuturi

8.1 Curs	Metode de predare	Observații
1. Noțiuni introductive, definiții, introducere în sistemele de recuperare a informațiilor. Cuvinte cheie: definiția recuperării informațiilor, recuperarea booleană, index inversat (matricea de incidență termen-document, lista postărilor)	expunere interactivă, explicație, conversație, demonstrație didactică	
2.1 Termenul vocabular și liste de postări. Cuvinte cheie: structuri de date pentru vocabularul termenilor (hash-uri, arbori), interogări cu caractere generice, permuterm și indexuri k-gram, distanță de editare 2.2 Dicționare și recuperare tolerantă Cuvinte cheie: structuri de date pentru vocabularul termenilor (hash-uri, arbori), interogări cu caractere generice, permuterm și indexuri k-gram, distanță de editare	expunere interactivă, explicație, conversație, demonstrație didactică	

3. Ponderarea termenilor. Modelul spațiului vectorial cuvinte cheie: recuperare ierarhică, coeficientul Jaccard, modelul bag of words, ponderi tf-idf, modelul spațiului vectorial	expunere interactivă, explicație, conversație, demonstrație didactică	
4-5. Introducere în Elasticsearch și Kibana cuvinte cheie: DSL și analizatori de interogări Elasticsearch, vizualizări și hărți Kibana	expunere interactivă, explicație, conversație, demonstrație didactică	
6. Evaluarea sistemelor IR. Cuvinte cheie: precizie recall, scor F, seturi de date de referință	expunere interactivă, explicație, conversație, demonstrație didactică	
7. Modele lingvistice în IR. Clasificarea textelor și Naive Bayes	expunere interactivă, explicație, conversație, demonstrație didactică	
8. Clasificarea în spațiul vectorial. Metode de selecție a caracteristicilor cuvinte cheie: Metode de clasificare în IR: Naive Bayes, algoritmul Rocchio, cele mai mici pătrate regularizate (RLS), mașini cu vectori de suport (SVM)	expunere interactivă, explicație, conversație, demonstrație didactică	
9-10. Introducere în modelele Attention și Transformer. Interpretabilitatea modelelor Attention	expunere interactivă, explicație, conversație, demonstrație didactică	
9. Filtre digitale – 5. - prelucrarea semnalelor stohastice	expunere interactivă, explicație, conversație, demonstrație didactică	
10. Analiză și estimare spectrală	expunere interactivă, explicație, conversație, demonstrație didactică	
11. Introducere în spațiul de încorporare CLIP	expunere interactivă,	

	explicație, conversație, demonstrație didactică	
12. Învățarea nesupravegheată în IR cuvintele cheie IR: gruparea flat și ierarhică	expunere interactivă, explicație, conversație, demonstrație didactică	
13-14. Prezentări de proiecte, discuții.	expunere interactivă, explicație, conversație, demonstrație didactică	

Bibliografie

- [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 / laborator	Metode de predare	Observații
1. Introducere în programarea Perl și/sau Python	documentație, explicație, conversație	
2. Algoritmi de clasificare renumiți în IR: Naive Bayes, Rocchio, SVM, MLPM etc.	documentație, explicație, conversație	
3-4. Clasificarea în IR cu modele de transformatoare (BERT).	documentație, explicație, conversație	
5-6. Motorul de indexare/căutare Elasticsearch.	documentație, explicație, conversație	
7. Rezumat, prezentări de proiecte.	documentație, explicație, conversație	

Bibliografie

- [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. Coroborarea conținuturilor disciplinei cu așteptările reprezentanților comunității epistemice, asociațiilor profesionale și angajatori reprezentativi din domeniul aferent programului.

Conținutul disciplinei este în concordanță cu cursul „Information Retrieval and Web Search” (Căutarea informațiilor și căutarea pe internet) de la Universitatea Stanford (<http://web.stanford.edu/class/cs276/>) și se bazează pe cartea „Introduction to Information Retrieval” (Introducere în căutarea informațiilor) de Manning, Raghavan și Schütze (<http://nlp.stanford.edu/IR-book/>, a se vedea și bibliografia de mai sus).

10. Evaluare

Tipul activității	10.1 Criterii de evaluare	10.2 Metode de evaluare	10.3 Ponderea în nota finală
10.4 Curs	Teste la începutul cursurilor	Teste scrise	15%
	Examen scris la sfârșitul semestrului	Examen scris	30%
10.5 Seminarii/laboratoare	Sarcini de seminar/laborator în timpul semestrului	Evaluarea sarcinilor de programare	30%
	Prezentarea proiectelor software	Evaluarea proiectului	25%
10.6 Standarde minime de performanță			
La prezentarea proiectelor software și a sarcinilor de seminar/laborator în timpul semestrului, trebuie obținută cel puțin jumătate din punctaj.			

Data completării
19.09.2025

Semnătura titularului de curs

Lect. dr. Szenkovits Annamária

Semnătura titularului de seminar

Lect. dr. Szenkovits Annamária

Data avizării în departament
19.09.2025

Semnătura directorului de departament,
Conf. univ. dr. András Szilárd