

# FIȘA DISCIPLINEI

## Tehnici de NLP

Anul universitar 2025-2026

### 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 Informatică
1.4. Domeniul de studii	Informatică
1.5. Ciclu de studii	Licență
1.6. Programul de studii / Calificarea	Inteligența artificială
1.7. Forma de învățământ	Cu frecvență

### 2. Date despre disciplină

2.1. Denumirea disciplinei	Introducere în prelucrarea limbajului natural			Codul disciplinei	MLE5208		
2.2. Titularul activităților de curs	Lect. dr. Lupea Mihaiela						
2.3. Titularul activităților de seminar	Lect. Dr. Lupea Mihaiela						
2.4. Anul de studiu	3	2.5. Semestrul	5	2.6. Tipul de evaluare	C	2.7. Regimul disciplinei	Compulsory

### 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/proiect	1 lab
3.4. Total ore din planul de învățământ	42	din care: 3.5. curs	28	3.6 seminar/laborator/proiect	14
<b>Distribuția fondului de timp pentru studiul individual (SI) și activități de autoinstruire (AI)</b>					<b>ore</b>
Studiul după manual, suport de curs, bibliografie și notițe (AI)					12
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					10
Pregătire seminare/ laboratoare/ proiecte, teme, referate, portofolii și eseuri					15
Tutoriat (consiliere profesională)					3
Examinări					5
Proiect practic					13
<b>3.7. Total ore studiu individual (SI) și activități de autoinstruire (AI)</b>				<b>58</b>	
<b>3.8. Total ore pe semestru</b>				<b>100</b>	
<b>3.9. Numărul de credite</b>				<b>4</b>	

### 4. Precondiții (acolo unde este cazul)

4.1. de curriculum	Limbaje formale, Structuri de date și algoritmi
4.2. de competențe	Cunoașterea unui limbaj de programare de nivel înalt.

### 5. Condiții (acolo unde este cazul)

5.1. de desfășurare a cursului	-
5.2. de desfășurare a seminarului/ laboratorului	Laborator cu calculatoare, medii de programare (.NET , Java)

## 6.1. Competențele specifice acumulate<sup>1</sup>

<b>Competențe profesionale/esențiale</b>	<ul style="list-style-type: none"><li>• Creează modele de date</li><li>• Utilizează biblioteci de software</li><li>• Creează softuri</li></ul>
<b>Competențe transversale</b>	<ul style="list-style-type: none"><li>• Dă dovadă de inițiativă</li><li>• Isi asumă responsabilitatea</li><li>• Gândește analitic</li></ul>

## 6.2. Rezultatele învățării

<b>Cunoștințe</b>	<ul style="list-style-type: none"><li>• Absolventul cunoaște și aplică tehnici bazate pe inteligența artificială pentru prelucrarea limbajului natural și exploatarea datelor de natură lingvistică.</li><li>• Absolventul cunoaște, înțelege și aplică conceptele și algoritmi fundamentali utilizați în inteligența artificială și este capabil să le evalueze pe baza unor metrici.</li><li>• Absolventul cunoaște și înțelege conceptele și tehnicile de reprezentare a cunoștințelor și le poate aplica în vederea rezolvării de probleme.</li></ul>
<b>Aptitudini</b>	<ul style="list-style-type: none"><li>• Absolventul este capabil să descrie formal problemele abordate din diferite domenii, modelându-le ca probleme care se pot aborda cu tehnici din sfera inteligenței artificiale.</li><li>• Absolventul este capabil să aplice algoritmi fundamentali de inteligență artificială pentru a rezolva probleme din lumea reală.</li><li>• Absolventul este capabil să evalueze, atât în mod cantitativ și cât și în mod calitativ, performanța sistemelor inteligente.</li></ul>
<b>Responsabilități și autonomie</b>	<ul style="list-style-type: none"><li>• Absolventul are cunoștințe necesare pentru revizuirea literaturii de specialitate și folosirea bazelor de date și bibliotecilor digitale internaționale de cercetare academică.</li><li>• Absolventul are cunoștințele necesare pentru selectarea și utilizarea procedurilor potrivite de instruire pentru a facilita procesul de asimilare a cunoștințelor.</li><li>• Absolventul este capabil să redacteze un raport științific.</li></ul>

## 7. Obiectivele disciplinei (reieșind din grila competențelor acumulate)

<b>7.1 Obiectivul general al disciplinei</b>	<ul style="list-style-type: none"><li>• Introducerea principiilor de bază și problemelor specifice domeniului prelucrării limbajului natural (PLN).</li><li>• Cunoașterea stadiului actual al cercetării în domeniul prelucrării limbajului natural, pentru a implementa sisteme care rezolvă task-uri specifice PLN.</li></ul>
<b>7.2 Obiectivele specifice</b>	<ul style="list-style-type: none"><li>• Aplicarea și utilizarea de modele formale (logici, gramatici formale, grafe), modele statistice (HMM- Hidden Markov Model), tehnici și algoritmi proprii inteligenței artificiale pentru rezolvarea unor probleme specifice domeniului PLN (limbile română și engleză): detectarea părților de vorbire, analiza sintactică, dezambiguizare, extragere cuvinte cheie, sumarizare de documente, rezolvarea anaferei, analiza sentimentelor și emoțiilor în texte.</li><li>• Utilizarea de modele mari de limbaje (LLMs) pentru a rezolva task-uri PLN.</li></ul>

<sup>1</sup> Se poate opta pentru competențe sau pentru rezultatele învățării, respectiv pentru ambele. În cazul în care se alege o singură variantă, se va șterge tabelul aferent celeilalte opțiuni, iar opțiunea păstrată va fi numerotată cu 6.

## 8. Conținuturi

8.1 Curs	Metode de predare	Observații
1. Prelucrarea limbajului natural (PLN): principii de bază, probleme specifice, abordări și aplicații.	Prelegere teoretică, explicații, exemple, studiu de caz, dezbateri, dialog	
2. Detectarea părților de vorbire și a entităților lexicale în limbile engleză și română Baze de date lexicale pentru limbile engleză și română: WordNet și RoWordNet		
3. Reprezentare textelor și modele de limbaje naturale. Clasificarea textelor.		
4. Analiza sintactică a textelor - reguli gramaticale pentru construirea frazelor în limba engleză; gramatici independente de context - algoritmul Cocke-Kasami-Yonger		
5. HMM (Hidden Markov Model) și aplicarea lui la detectarea părților de vorbire în texte. Algoritmii "forward" și Viterbi.		
6. Extragerea cuvintelor cheie din texte -algoritmii TextRank și RAKE		
7. Sumarizarea documentelor folosind abordări bazate pe clustering și grafe		
8. Prezentări rapoarte de cercetare de către studenți	Evaluare	
9. Detectarea opiniilor, sentimentelor, emoțiilor din text	Prelegere teoretică, explicații, exemple, studiu de caz, dezbateri, dialog	
10. Rezolvarea anaferei în texte		
11. Dezambiguizarea cuvintelor în texte –metode bazate de dicționare și grafe		
12. Extragerea informației din texte		
13. Examen scris	Evaluare	
14. Prezentare proiecte de către studenți	Evaluare	
<p><b>Bibliografie</b></p> <ol style="list-style-type: none"> <li>1. J.ALLEN : Natural language understanding, Benjamin/Cummings Publisher, 2nd ed., 1995.</li> <li>2. E. CHARNIAK: Statistical language learning, MIT press, 1996.</li> <li>3. L. DENG, Y. LIU: Deep learning in Natural Language Processing, Springer Verlag, Singapore, 2018</li> <li>4. D.FEHRER et al: Description logics for natural language processing. In Proc. of the 1994 Description Logic Workshop (DL'94), 1994.</li> <li>5. H. HELBIG: Knowledge Representation and the Semantics of Natural Language, Springer, 2006.</li> <li>6. D.JURAFSKY, J.MARTIN: Speech and language processing, Prentice Hall, 2000.</li> <li>7. C.MANNING, H.SCHUTZE: Foundation of statistical natural language processing, MIT, 1999.</li> <li>8. R. MITKOV(ed): The Oxford Handbook of Computational Linguistics, Oxford University Press, 2003.</li> <li>9. D. TATAR: Inteligența artificială. Aplicații în prelucrarea limbajului natural, Editura Albastra, Microinformatică, 2003, ISBN 973-650-100-01.</li> <li>10. S. VAJJALA, B. MAJUMDER, A. GUPTA, H. SURANA: Practical Natural Language Processing. A Comprehensive Guide to Building Real-World NLP Systems, O'REILLY. 2020.</li> </ol>		
8.2 Seminar / laborator	Metode de predare	Observații
1. Studiul de platforme și biblioteci în diferite limbaje de programare care oferă funcții de preprocesarea textelor în limbile română/engleză.	Documentare pe platforme electronice, studii de caz, explicații	Laboratorul este structurat sub forma a 2 ore din 2 în 2 săptămâni.
2. Studiul modelelor mari de limbaje (LLM) pentru limba română și engleză. Studiul bazelor de date lexicale WordNet (limba engleză), RoWordNet (limba română) . Utilizarea de tool-uri existente pentru diferite task-uri: extragere cuvinte cheie, sumarizare, rezolvarea anaferei, analiza sentimentelor.		

3. Identificarea de task-uri PLN pentru limba romana. Alegerea task-ului care va fi rezolvat, studiul diferitelor abordari de rezolvare si alegerea metodei care va fi implementata. Cautare de date de intrare pentru task-ul ales.		
4. Prezentarea de catre studenti a referatului teoretic despre un task sau tool PLN.	Evaluare	
5. Discutii despre implementarea unui PLN tool.	Explicatii, dialog, studiu de caz	
6. Dezvoltarea de resurse lingvistice pentru limba romana.		
7. Prezentarea proiectelor de către studenți	Evaluare	
Bibliografie 1. Rada Mihalcea: <a href="http://www.cs.unt.edu/~rada/downloads.html">www.cs.unt.edu/~rada/downloads.html</a> 2. Resurse lingvistice in limba romana: <a href="http://www.racai.ro">www.racai.ro</a>		

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

- Cursul respectă Recomandările IEEE și ACM Curricula pentru studii de informatică;
- Cursul există în programul de studii al tuturor universităților majore din România și din străinătate;
- Optimizarea căutării pe Web, interfețele în limbaj natural și aspectele recente ale procesării textelor necesită o bună înțelegere a prelucrării limbajului natural.

### 10. Evaluare

Tip activitate	10.1 Criterii de evaluare	10.2 Metode de evaluare	10.3 Pondere din nota finală
10.4 Curs	Cunoasterea conceptelor teoretice ale domeniului. Aplicarea metodelor predate in rezolvarea problemelor specifice.	Examen scris	20%
	Redactarea unui referat teoretic	Raport teoretic - prezentarea unei probleme specifice domeniului și a unui instrument existent care o rezolvă	30%
10.5 Seminar/ laborator	Implementarea unor algoritmi prezentati la curs	Proiect practic – rezolvarea unei probleme specifice domeniului prelucrării limbajului natural.	35%
	Aplicarea unor concepte teoretice in rezolvarea unor probleme practice din domeniu PLN	Dezvoltarea unor resurse lingvistice specifice limbii romane	15%
10.6 Standard minim de performanță			
<ul style="list-style-type: none"> <li>• Nota finala sa fie cel puțin nota 5 (pe o scara de 1 la 10).</li> </ul>			

### 11. Etichete ODD (Obiective de Dezvoltare Durabilă / Sustainable Development Goals)<sup>2</sup>

*Nu se aplică.*

Data completării:  
15.04.2025

Semnătura titularului de curs  
Lect. dr. Mihaiela LUPEA

Semnătura titularului de seminar  
Lect. dr.. Mihaiela LUPEA

<sup>2</sup> Păstrați doar etichetele care, în conformitate cu *Procedura de aplicare a etichetelor ODD în procesul academic*, se potrivesc disciplinei și ștergeți-le pe celelalte, inclusiv eticheta generală pentru *Dezvoltare durabilă* - dacă nu se aplică. Dacă nicio etichetă nu descrie disciplina, ștergeți-le pe toate și scrieți "Nu se aplică".

Data avizării în departament:

...

Semnătura directorului de departament

Conf.dr. Adrian STERCA