

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 Informatică
1.4 Domeniul de studii	Calculatoare și Tehnologia Informației
1.5 Ciclul de studii	Licență
1.6 Programul de studiu / Calificarea	Ingineria Informației (în limba engleză)

2. Date despre disciplină

2.1 Denumirea disciplinei (ro)	Introducere în prelucrarea limbajului natural						
(en)	Introduction to Natural Language Processing						
2.2 Titularul activităților de curs	Lect.dr. Lupea Mihaiela-Ana						
2.3 Titularul activităților de seminar	Lect.dr. Lupea Mihaiela-Ana						
2.4 Anul de studiu	4	2.5 Semestrul	8	2.6. Tipul de evaluare	C	2.7 Regimul disciplinei	Optional DS
2.8 Codul disciplinei	MLE8151						

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

3.1 Număr de ore pe săptămână	5	Din care: 3.2 curs	2	3.3 seminar/laborator	1 LP 2 P
3.4 Total ore din planul de învățământ	70	Din care: 3.5 curs	28	3.6 seminar/laborator	42
Distribuția fondului de timp:					ore
Studiul după manual, suport de curs, bibliografie și notițe					15
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					10
Pregătire seminarii/laboratoare, teme, referate, portofolii și eseuri					15
Tutoriat					5
Examinări					10
Alte activități: proiecte de grup					
3.7 Total ore studiu individual			55		
3.8 Total ore pe semestru			125		
3.9 Numărul de credite			5		

4. Precondiții (acolo unde este cazul)

4.1 de curriculum	Limbaje formale, Structuri de date și algoritmi
4.2 de competențe	Cunoasterea 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. Competențele specifice acumulate

Competențe profesionale	<p>C3.1 Identificarea unor clase de probleme și metode de rezolvare caracteristice sistemelor informatice</p> <p>C3.2 Utilizarea de cunoștințe interdisciplinare, a tiparelor de soluții și a uneltelor, efectuarea de experimente și interpretarea rezultatelor lor</p> <p>C3.4 Evaluarea comparativă, inclusiv experimentală, a alternativelor de rezolvare, pentru optimizarea performanțelor</p> <p>C3.5 Dezvoltarea și implementarea de soluții informatice pentru probleme concrete</p> <p>C6.2 Utilizarea adecvată a metodelor pentru analiza semnalelor și algoritmilor fundamentali de inteligență artificială</p> <p>C6.4 Evaluarea cantitativă și calitativă a performanțelor sistemelor inteligente</p> <p>C6.5 Încorporarea metodelor de prelucrare a semnalelor și a soluțiilor specifice inteligenței artificiale în aplicații dedicate</p>
Competențe transversale	<p>CT1 Comportarea onorabilă, responsabilă, etică, în spiritul legii pentru a asigura reputația profesiei</p> <p>CT2 Identificarea, descrierea și derularea proceselor din managementul proiectelor, cu preluarea diferitelor roluri în echipă și descrierea clară și concisă, verbal și în scris, în limba română și într-o limbă de circulație internațională, a rezultatelor din domeniul de activitate</p> <p>CT3 Demonstrarea spiritului de inițiativă și acțiune pentru actualizarea cunoștințelor profesionale, economice și de cultură organizațională</p>

7. Obiectivele disciplinei (reieșind din grila competențelor acumulate)

7.1 Obiectivul general al disciplinei	<ul style="list-style-type: none"> • Introducerea principiilor de bază și problemelor specifice domeniului Prelucrării Limbajului Natural (PLN). • Cunoașterea stadiului actual al cercetării în domeniul prelucrării limbajului natural, pentru a efectua cercetări originale în acest domeniu.
7.2 Obiectivele specifice	<ul style="list-style-type: none"> • Studiul teoretic al modelelor formale (logici, gramatici formale, grafe), modelelor statistice (HMM- Hidden Markov Model), tehnicilor și algoritmilor proprii inteligenței artificiale aplicabili în domeniul PLN. • Proiecte practice: implementarea unor sisteme automate care rezolva task-uri specifice domeniului PLN (limbile română și engleză): detectarea părților de vorbire, analiza sintactică, dezambiguare, extragere cuvinte cheie, sumarizare de documente, rezolvarea anaferei, analiza sentimentelor și emoțiilor în texte, analiza cantitativa a textelor.

8. Conținuturi

8.1 Curs	Metode de predare	Observații
Curs 1. Prelucrarea limbajului natural: principii de bază, probleme specifice, abordări și aplicații.	Prelegere teoretică, explicații, exemple, studiu de caz, dezbateri, dialog	
Curs 2. Baze de date lexicale pentru limbile engleză și română		

- WordNet și RoWordNet - structura, relații semantice și relații lexicale - WordNetSimilarity tool – masuri de similaritate între cuvinte		
Curs 3. Detectarea părților de vorbire și a entităților lexicale în limbile engleză și română		
Curs 4. Analiza sintactică - reguli gramaticale pentru construirea frazelor în limba engleză; gramatici independente de context - algoritmul Cocke-Kasami-Yonger de analiză sintactică		
Curs 5. Analiza sintactică probabilistică - gramatici probabilistice independente de context în studiul limbajului natural - algoritmul probabilistic al lui Cocke-Kasami-Yonger		
Curs 6. Modelul Markov cu stări ascunse (HMM) - HMM (Hidden Markov Model) – model statistic - aplicații ale HMM în viața de zi cu zi și în detectarea părților de vorbire - problemele canonice asociate cu HMM și algoritmi care le rezolvă		
Curs 7. Dezambiguarea cuvintelor în texte - identificarea sensurilor cuvintelor dintr-un text: abordări bazate pe dicționare și grafe		
Curs 8. Sumarizarea (rezumarea) documentelor - abordări bazate pe clusterizări și grafe.		
Curs 9. Rezoluția anaferei - algoritmul lui Lapin și Lease - algoritmul lui Mitkov		
Curs 10. Analiză pe bază de sentimente a textelor - detectarea opiniilor și sentimentelor în texte media - analiza pe bază de emoții a textelor literare		
Curs 11. Extragerea informației din texte		
Curs 12. Implicare textuală		
Curs 13. Analiza cantitativă a textelor literare		
Curs 14. Prezentarea proiectelor de către studenți		

Bibliografie

1. Y. GOLDBERG, Neural Network Methods for Natural Language Processing, Springer, 2017.
2. S. RAAIJMAKERS, Deep Learning for Natural Language Processing, 2022
3. U. KAMATH, J.LIU, J. WHITAKER, Deep Learning for NLP and speech recognition., Springer, 2019, e-book, <https://doi.org/10.1007/978-3-030-14596-5>
4. H.HELBIG: Knowledge Representation and the Semantics of Natural Language, Springer, 2006.
5. D.JURAFSKY, J.MARTIN: Speech and language processing, Prentice Hall, 3rd edition, 2022.
6. M.LUPEA, M.RUKK, I.I.POPESCU, G.ALTSMANN, Some Properties of Rhyme, Studies in Quantitative Linguistics 26, RAM-Verlag, 2017.
7. C.MANNING, H.SCHUTZE: Foundation of statistical natural language processing, MIT, 1999.
8. R.MITKOV(ed): The Oxford Handbook of Computational Linguistics, Oxford University Press, 2003.

9. I.I.POPESCU, M.LUPEA, D.TATAR, G.ALTMANN, Quantitative Analysis of Poetic Texts, DE GRUYTER MOUTON, 2015.

10. D. TATAR: Inteligența artificială. Aplicații în prelucrarea limbajului natural, Editura Albastra, Microinformatica, 2003.

8.2 Seminar / laborator	Metode de predare	Observații
1. - Cunoașterea bazelor de date lexicale WordNet (limba engleză), RoWordNet (limba română) - Utilizarea instrumentului WordNetSimilarity.	Lucrarea de laborator Explicația, conversația	Laboratorul este structurat sub forma a 2 ore din 2 în 2 săptămâni.
2. Utilizarea unor instrumente dedicate pentru analiză sintactică (Stanford parser, CST tools, Racai tools)	Lucrarea de laborator Explicația, conversația	
3. Prezentările referatelor de către studenți	Dialogul, dezbateră	
4. Prezentările referatelor de către studenți	Dialogul, dezbateră	
5. Utilizarea unor instrumente dedicate pentru: sumarizarea textelor, rezoluția anaferei și analiza sentimentelor.	Lucrarea de laborator Explicația, conversația	
6. Utilizarea unor instrumente dedicate pentru: extragerea informației implicare textuală.	Dialogul, dezbateră	
7. Prezentarea proiectelor practice de către studenți.	Dialogul, dezbateră	

Bibliografie

1. Boros Tiberiu, Dumitrescu Stefan, Burtica Ruxandra. "NLP-Cube: End-to-End Raw Text Processing With Neural Networks," 2018.
2. <https://opensource.adobe.com/NLP-Cube/index.html>
3. <http://wordnetweb.princeton.edu/>
4. <https://multiwordnet.fbk.eu/english/home.php>
5. Resurse lingvistice în limba română: www.racai.ro, nlptools.info.uaic.ro/Resources
6. <https://demo.allennlp.org/>
7. Rada Mihalcea: www.cs.unt.edu/~rada/downloads.html

8.3 Proiect	Metode de predare	Observatii
S1. Studiu bibliografic - reprezentarea textelor pentru utilizarea lor în task-uri specifice domeniului NLP	Documentare pe platforme electronice, conversația	Studentii vor lucra individual sau în echipe de 2 persoane pentru implementarea unor tool-uri care rezolvă task-uri practice din domeniul PLN în limbile română sau engleză.
S2-S3. Studiul platformelor și bibliotecilor din diferite limbaje de programare care oferă funcții de preprocesare a textelor în limbile română și engleză.	Documentare pe platforme electronice, conversația	
S4. Studiul instrumentelor dedicate (tool-uri) existente care rezolvă task-urile PLN studiate la cursuri pentru limbile română și engleză.	Documentare pe platforme electronice, conversația	
S5. Identificarea de task-uri practice din domeniul PLN în limba română care pot fi rezolvate prin implementarea algoritmilor de la cursuri.	Documentare pe platforme electronice, conversația, studii de caz	

S6-S7. Alegerea task-ului PLN, studiul diferitelor abordari teoretice pentru rezolvarea sa, alegerea abordarii ce va fi implementata.	Documentare pe platforme electronice, conversatia, studii de caz	
S8-S9. Cautarea datelor de intrare necesare sau crearea unor seturi de date specifice task-ului ales.	Documentare pe platforme electronice, conversatia, studii de caz	
S10-S12. Proiectarea si implementarea sistemului.	Documentare pe platforme electronice, Conversatia, studii de caz	
S13. Testarea sistemului soft dezvoltat.	Conversatia, studii de caz	
S14. Prezentarea sistemului dezvoltat	Evaluare	

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	- redactarea unui referat teoretic	Raport teoretic - prezentarea unei probleme specifice domeniului și a unui instrument existent care o rezolvă	35%
10.5 Seminar/laborator	- aplicarea unor concepte teoretice in rezolvarea unor probleme practice din domeniu	Dezvoltarea unor resurse lingvistice specifice limbii romane	20%
	- implementarea unor algoritmi prezentati la curs	Proiect practic – implementarea unui sistem care rezolva un task din domeniului prelucrării limbajului natural.	35%
10.6 Activitate	- activitatea la curs și laborator/seminar	Prezența la activități, realizarea la timp a activităților solicitate	10%
10.7 Standard minim de performanță			
<ul style="list-style-type: none"> • Pentru promovare e necesar ca nota finală să fie minim 5. 			

Data completării

15.05.2022

Semnătura titularului de curs

Lect. dr. Lupea Mihaiela



Semnătura titularului de seminar

Lect.dr. Lupea Mihaiela



Data avizării în departament

Semnătura directorului de departament

Prof. Ph.D. Dioşan Laura

24.05.2022

A handwritten signature in black ink, appearing to read 'Dioşan', with a vertical line extending downwards from the end of the signature.