

## FIȘA DISCIPLINEI

### 1. Date despre program

1.1 Instituția de învățământ superior	Universitatea Tehnică din Cluj-Napoca
1.2 Facultatea	Facultatea de Electronică, Telecomunicații și Tehnologia Informației
1.3 Departamentul	Comunicații
1.4 Domeniul de studii	Inginerie electronică, telecomunicații și tehnologii informaționale
1.5 Ciclul de studii	Master
1.6 Programul de studii / Calificarea	Inteligență artificială și prelucrări de semnale în electronică și telecomunicații (în limba engleză) / Master
1.7 Forma de învățământ	IF – învățământ cu frecvență
1.8 Codul disciplinei	IAPSET-E 18.40

### 2. Date despre disciplină

2.1 Denumirea disciplinei	Procesarea limbajului natural						
2.2 Aria de conținut	Inginerie electronică, telecomunicații și tehnologii Informaționale						
2.3 Responsabil de curs	Prof.dr.ing. Mircea Giurgiu (Mircea.Giurgiu@com.utcluj.ro)						
2.4 Titularul activităților de seminar / laborator / proiect	Prof.dr.ing. Mircea Giurgiu (Mircea.Giurgiu@com.utcluj.ro)						
2.5 Anul de studiu	2	2.6 Semestrul	1	2.7 Tipul de evaluare	Exam	2.8 Regimul disciplinei	DA/DO

### 3. Timpul total estimat

3.1 Număr de ore pe săptămână	2	din care: 3.2 curs	1	3.3 laborator	1
3.4 Total ore din planul de învățământ	28	din care: 3.5 curs	14	3.6 laborator	14
Distribuția fondului de timp					ore
Studiul după manual, suport de curs, bibliografie și notițe					24
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					16
Pregătire seminarii / laboratoare, teme, referate, portofolii și eseuri					24
Tutoriat					4
Examinări					4
Alte activități: .....					
3.7 Total ore studiu individual	72				
3.8 Total ore pe semestru	100				
3.9 Numărul de credite	4				

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

4.1 de curriculum	Prelucrarea numerică a semnalelor, Teoria informației și a codării, Tehnici de codare a datelor, Modelarea canalelor de comunicații
4.2 de competențe	Competențe de programare în Matlab / Python

## 5. Condiții (acolo unde este cazul)

5.1. de desfășurare a cursului	Aulă cu videoproiector / MS Teams
5.2. de desfășurare a seminarului / laboratorului / proiectului	Laborator cu rețea de calculatoare, acces la Internet, Matlab, PyCharm pentru Python, baze de date cu text în diferite limbi, toolkit-uri specific pentru prelucrarea limbajului natural

## 6. Competențele specifice acumulate

Competențe profesionale	<p><b>C4. Conceperea, implementarea și operarea serviciilor de prelucrare automată a textelor, bazate pe înțelegerea și aplicarea notiunilor fundamentale din domeniul comunicațiilor și transmisiunii informației</b></p> <p>C4.1 Rezolvarea de probleme practice utilizând cunoștințe generale privind tehnicile multimedia</p> <p>C4.2 Explicarea și interpretarea principalelor cerințe și tehnici specifice de abordare pentru transmisiile de text, date, și multimedia</p> <p>C4.3 Rezolvarea de probleme practice utilizând cunoștințe generale privind tehnicile multimedia</p> <p>C4.4 Utilizarea principalilor parametri specifici în evaluări bazate pe conceptul de calitate a serviciilor în comunicații</p> <p>C4.5 Dezvoltarea unor servicii simple de comunicații bazate pe prelucrarea automată a textelor</p> <p><b>C7. Conceperea, implementarea și testarea de sisteme și de diverse tipuri de aplicații (prelucrări de semnale, clasificare, regresie, detecție, procesarea limbajului natural, recunoaștere de forme) care se bazează pe tehnici de învățare automată sau de învățare profundă</b></p> <p>C7.1 Proiectare, implementare, testare și exploatare de aplicații folosind metode clasice de învățare automată</p> <p>C7.2 Proiectare, implementare, testare și exploatare de aplicații folosind rețele neuronale clasice și convoluționale</p>
Competențe transversale	<p>CT3. Capacitatea de se adapta în utilizarea și implementarea unor tehnologii multimedia pentru prelucrarea automată a limbajului natural, în special arhitecturi cu rețele neuronale profunde specifice pentru inteligența artificială (IA). Flexibilitate în gândire și lucru în echipă într-un domeniu multidisciplinar.</p>

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

7.1 Obiectivul general al disciplinei	Dezvoltarea de cunoștințe, deprinderi și competențe în utilizarea practică a metodelor și algoritmilor de prelucrare a limbajului natural folosind principii ale inteligenței artificiale (IA).
7.2 Obiectivele specifice	<ul style="list-style-type: none"> <li>✓ să cunoască problemele fundamentale în prelucrarea limbajului natural, precum și soluțiile de bază pentru aceste probleme</li> <li>✓ să utilizeze o serie de instrumente software, biblioteci și seturi de date utilizate frecvent în prelucrarea limbajului natural</li> <li>✓ să cunoască o serie de arhitecturi de rețele neuronale profunde (DNN – Deep Neural Networks) aplicate în prelucrarea limbajului natural</li> <li>✓ să dobândească o serie de competențe privind proiectarea și implementarea unor soluții specifice pentru clasificarea textelor, traducere automată, identificarea conținutului semnificativ din text, dezvoltarea unor sisteme interactive de tip ChatBot sau asistent virtual</li> </ul>

	✓ să dobândească deprinderi în utilizarea unor instrumente software specifice sistemelor de inteligență artificială aplicate în prelucrarea limbajului natural
--	--

## 8. Conținuturi

8.1 Curs	Metode de predare	Observații
1. Probleme fundamentale in prelucrarea limbajului natural. Concepte, fluxuri de prelucrare si aplicatii	Expunere (inclusiv online MS Teams), prezentare cu videoprojector, dezbateri, Prezentări ale studentilor.	Nu este cazul.
2. Metode de pre-procesare si codificare a textului (BoW, n-gram, DTM. TF-IDF, Levenstein, Word2Vec, reprezentare latentă), cu aplicatii specifice.		
3. Analizoare de limbaj si tehnici de extragere a informatiei semantice (arbori sintactici, lematizare, POS, NER, dizambiguare, similaritate texte, analiza sentimentelor)		
4. Metode uzuale de clasificarea automata a textelor (Naive Bayes, Regresii, LDA, SVM, CRF, NMF) si aplicatii (extragerea topicurilor, sumarizare, identificarea stirilor false, sisteme de recomandare)		
5. Modelarea statistica, modele de limbaj si aplicatii (HMM, corectia si predictia textului, traducerea automata dintr-o limba in alta)		
6. Metode de prelucrare a limbajului natural folosind retele neuronale profunde (clasificare, sisteme de tip intrebare-raspuns, retele RNN, (B)LSTM, BERT si Transformer)		
7. Modele largi de limbaj si integrarea cu sisteme de recunoastere automata a vorbirii (ChatGPT, NeMo, RIVA, asistenti virtuali si generarea automata a textului)		
<b>Bibliografie</b>		
<ol style="list-style-type: none"> <li>1. C. D. Manning and H. Schütze, "Foundations of Natural Language Processing" , 6th Edition, The MIT Press Cambridge, 2003.</li> <li>2. A. Kao, S. R. Poteet, "Natural language processing and text mining", London : Springer, 2007</li> <li>3. D. Jurafsky, J. Martin, "Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition, Ed 3, Pearson Education, 2023</li> <li>4. H. Lane, C. Howard, H. Hapke, "Natural language processing in action", MANNING Publications, 2019.</li> <li>5. ***, "Natural Language Processing Crash Course for Beginners: Theory and Applications of NLP using TensorFlow 2.0 and Keras", AI. Publishing, 2020</li> <li>6. D. Rao, B. McMahan, "Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning", O'Reilly Publ., 2022</li> <li>7. B. Steven, E. Loper, E. Klein, "Natural Language Processing with Python". O'Reilly Media Inc., 2009.</li> <li>8. A. Clark, C. Fox, S. Lappin, "The Handbook of Computational Linguistics and Natural Language Processing", Wiley-Blackwell, 2012</li> </ol>		

9. R. Arumugam, R. Shanmugamani, "Hands-on natural language processing with Python: A practical guide to applying deep learning architectures to your NLP application", PACKT publisher, 2018.
10. J. Hirschberg and C. D. Manning, @Advances in natural language processing. Science, 349 (6245):261–266, 2015. URL <https://www.sciencemag.org/content/349/6245/261.full>"
11. N.Indurkha, F.J.Damerou (eds), "Handbook of Natural Language Processing", Chapman & Hall, 2010.
12. J.Eisenstein, "Introduction to Natural Language Processing", MIT Press, 2019

8.2 Laborator	Metode de predare	Observații
1. Instrumente software dedicate prelucrării automate a limbajului natural și aplicații cu acestea	Experimente, problematizarea, lucru în echipă	Nu e cazul
2. Prelucrări de bază ale textului și tehnici de reprezentare și codare		
3. Sisteme de identificare automată a topicului și sisteme de recomandare bazate pe analiza textului		
4. Identificarea automată a sentimentelor din text		
5. Detectarea plagiarismului din text folosind similaritatea semantică		
6. Dezvoltarea unui ChatBot pentru un sistem de tip întrebare - răspuns		
7. Integrarea recunoașterii automate a vorbirii într-un sistem bazat pe ChatBot		

#### Bibliografie

1. D. Jurafsky, J. Martin, "Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition, Ed 3, Pearson Education, 2023
2. H. Lane, C. Howard, H. Hapke, "Natural language processing in action", MANNING Publications, 2019.
3. \*\*\*, "Natural Language Processing Crash Course for Beginners: Theory and Applications of NLP using TensorFlow 2.0 and Keras", AI. Publishing, 2020
4. D. Rao, B. McMahan, "Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning", O'Reilly Publ., 2022
5. NLTK Book: <http://www.nltk.org/book/>
6. All Stanford NLP tools: <http://nlp.stanford.edu/software/index.shtml>
7. Topic modelling: <https://radimrehurek.com/gensim/>
8. Toolkit Spacy: <https://spacy.io/usage/spacy-101>
9. Instrumente de la Universitatea din Stanford: <https://stanfordnlp.github.io/CoreNLP/>
10. Servicii Google: <https://cloud.google.com/natural-language>
11. Toolkit PyTorch: <https://pytorchnlp.readthedocs.io/en/latest/>
12. Resurse și toolkit ChatGPT: <https://openai.com/>

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

Competențele dobândite vor fi folosite în următoarele ocupații conform COR (Clasificarea Ocupațiilor din România): Inginer telecomunicații; Proiectant inginer de sisteme și calculatoare; Inginer proiectant comunicații; Inginer sisteme de securitate; Dezvoltator de aplicații multimedia; Consultant pentru sisteme comunicații.

## 10. Evaluare

Tip activitate	10.1 Criterii de evaluare	10.2 Metode de evaluare	10.3 Pondere din nota finală
10.4 Curs	Performanța studenților raportată la obiectivele educaționale	Probă scrisă	60% (T)
10.5 Laborator	Performanța studenților în rezolvarea problemelor practice, derularea experimentelor și interpretarea rezultatelor	Derulare experimente, evaluare intermediară și rapoarte de laborator	40% (L)

### 10.6 Standard minim de performanță

#### Nivel calitativ:

##### Cunoștințe minimale:

- ✓ Înțelegerea conceptelor privind metodele și algoritmi de baza în prelucrarea a limbajului natural (n-gram, TF-IDF, BoW, POS, lematizare, segmentarea textelor. procesari în spațiul latent, procesare statistică, antrenare și evaluare rețele neuronale profunde de diferite tipuri)
- ✓ Înțelegerea structurilor de rețele neuronale profunde specifice tehnicilor de inteligență artificială aplicate în prelucrarea limbajului natural (RNN, (B)LSTM, BERT, Ro-BERT, mecanismul de atenție și Transformers)

##### Competențe minimale:

- ✓ Să poată utiliza un ansamblu de instrumente software dedicate prelucrării limbajului natural (biblioteci Python, Spacy, Naïve Bayes, Regresii, CRF, LDA, SVM, DNN în diferite variante)
- ✓ Să poată dezvolta aplicații specifice pentru: clasificarea automată a textelor, detectarea automată a similarității în text și a plagiarismului, identificarea conținutului malicios (stiri false, sms și eMail phishing), dezvoltarea sistemelor de recomandare bazate pe analiza textului, dezvoltarea de sisteme automate de tip întrebare-răspuns, dezvoltarea de ChatBot și asistenți virtuali în diferite contexte, integrarea sistemelor de recunoaștere automată a vorbirii și de sinteză din text a vorbirii cu prelucrarea și înțelegerea automată a limbajului natural.

#### Nivel cantitativ:

- Efectuarea tuturor lucrărilor de laborator,
- Nota finală este  $0,6 \cdot T + 0,4 \cdot L$ , cu  $0,6 \cdot T + 0,4 \cdot L \geq 5,00$  ( $T \geq 5,00$  și  $L \geq 5,00$ )

Data completării:	Titulari	Titlu Prenume NUME	Semnătura
21.06.2024	Curs	Prof.dr.ing. Mircea GIURGIU	
	Aplicații	Prof.dr.ing. Mircea GIURGIU	

Data avizării în Consiliul Departamentului COM 10.06.2024	Director Departament Comunicații Prof.dr.ing. Virgil DOBROTĂ
Data aprobării în Consiliul Facultății ETTI 11.07.2024	Decan Prof.dr.ing. Ovidiu POP