

**FIȘA DISCIPLINEI****1. Date despre program**

1.1 Instituția de învățământ superior	Universitatea "Alexandru Ioan Cuza" din Iași
1.2 Facultatea	Facultatea de Informatică
1.3 Departamentul	DEPARTAMENTUL DE INFORMATICA
1.4 Domeniul de studii	Informatică
1.5 Ciclul de studii	Licență
1.6 Programul de studii / Calificarea	Informatică

2. Date despre disciplină

2.1 Denumirea disciplinei	Tehnici de procesarea limbajului natural						
2.2 Titularul activităților de curs	Lect. Dr. LIVIU-ANDREI SCUTELNICU						
2.3 Titularul activităților de seminar							
2.4 An de studiu	III	2.5 Semestrul	VI	2.6 Tip de evaluare*	E	2.7 Regimul disciplinei**	Op

*E – Examen / C – Colocviu / V – Verificare

**OB – Obligatoriu / OP – Opțional / F – Facultativ

3. Timpul total estimat (ore pe semestru și activități didactice)

3.1 Număr de ore pe săptămână	4	3.2 curs	2	3.3 seminar/laborator	2
3.4 Total ore din planul de învățământ	48	3.5 curs	24	3.6 seminar/laborator	24
Distribuția fondului de timp					ore
Studiu după manual, suport de curs, bibliografie și altele					20
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					8
Pregătire seminare/laboratoare, teme, referate, portofolii și eseuri					20
Tutoriat					0
Examinări					4
Alte activități					0
3.7 Total ore studiu individual*					52
3.8 Total ore pe semestru					100
3.9 Numărul de credite					4

4. Precondiții - De curriculum (dacă este cazul)

Structuri de date, Introducere în Inteligență Artificială, Învățare Automată.
Programare de nivel înalt, dezvoltarea sistemelor informatice.

5. Condiții (dacă este cazul)

5.1 De desfășurare a cursului	Sala de curs trebuie să dispună de videoproiector, conexiune la Internet și tablă pentru exemplificări
5.2 De desfășurare a seminarului/ laboratorului	Prezența obligatorie la laborator. Sala de laborator trebuie să dispună de conexiune la Internet și tablă pentru exemplificări. Studenții au nevoie de calculatoare pe care să fie instalat un limbaj ce le permite să rezolve temele de laborator (Java sau Python recomandate).

6. Obiective

Obiectiv general:

Obiectivul general al cursului este de a oferi studenților o bază solidă în domeniul prelucrării limbajului natural (PLN), echilibrând cunoștințele teoretice cu aplicabilitatea practică.

Obiective specifice:

OS1. Cunoașterea conceptelor fundamentale și a teoriilor din domeniul prelucrării limbajului natural;

OS2. Cunoașterea principiilor de dezvoltare a sistemelor dedicate limbajului, care includ preprocesarea textului, tokenizarea, etichetarea părților de vorbire, recunoașterea entităților denumite și aplicarea algoritmilor de învățare automată aplicate datelor textuale;

OS3. Cunoașterea metodelor de utilizare a instrumentelor și tehnologiilor de ultimă oră din domeniul lingvisticii computaționale;

OS4. Cunoașterea metodelor de aplicare a tehnicilor de învățare automată pentru rezolvarea problemelor specifice prelucrării limbajului natural, atât pentru sinteză, cât și pentru generare de text.

7. Competențe/Rezultate ale învățării

- Vor utiliza tehnici de bază de procesare a textului
- Vor cunoaște și vor putea aplica metode de clasificare a textelor
- Vor putea organiza semantic colecții de texte
- Vor cunoaște și vor utiliza anumite arhitecturi ale rețelelor neuronale pentru a rezolva sarcini variate pe date text
- Descrierea de concepte, teorii și modele folosite în domeniul de aplicare

8. Conținut

8.1 Curs	Metode de predare	Observații (ore și referințe bibliografice)
Introducere în IA și PLN	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Fundamente ale limbajului. Sintaxa și semantica	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Lexicologie și lexicografie. Dicționare și tezaure.	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Lexicologie și lexicografie. Corpusuri.	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Tehnici preliminare în PLN.	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Reprezentarea cuvintelor și semantica distribuțională.	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Tehnici de clasificare în PLN.	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Tehnologii de traducere automată.	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Interconectare de resurse.	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Prelucrarea vorbirii.	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Aplicații avansate în PLN.	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore
Provocări, etică și tendințe în PLN.	Expunere, prelegere participativă, exemplificare, aplicații.	2 ore

Bibliografie

- [1] Daniel Jurafsky and James Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. <https://www.cs.colorado.edu/~martin/slp.html>.
- [2] Eisenstein, Jacob. Introduction to natural language processing. MIT press, 2019. <https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf>
- [3] MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinriche. An introduction to Information Retrieval, 2008. <https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>
- [4] Bird, S., Klein, E., & Loper, E. (2009). "Natural Language Processing with Python". O'Reilly Media. <https://tjzhifei.github.io/resources/NLTK.pdf>
- [5] Scutelnicu A. (2021) - Representing and interconnecting lexical resources - Ph. D. Thesis. https://scdoc.info.uaic.ro/wp-content/uploads/2023/02/Rezumat_Teza_Doc_AndreiScutelnicu_EN.pdf.
- [6] A. V. Aho, M. S. Lam, R. Sethi, and J. D. Ullman, Compilers: Principles, Techniques, and Tools, Second Edition. Addison-Wesley, 2006 <https://dl.acm.org/doi/book/10.5555/1177220>
- [7] Moruz A., Scutelnicu A., Cristea D. (2018) - Interlinking and Extending Large Lexical Resources for Romanian. In V. Păiș, D. Gifu, D. Trandabăț, D. Cristea, D. Tufiș (eds.) Proceedings of the 13th International Conference on Linguistic Resources and Tools for Processing Romanian Language, Iași, pp. 125-132. <https://profs.info.uaic.ro/~dcristea/papers/paper%2014.pdf>.
- [8] Ivan Habernal - "Semantic Web Search using Natural Language", 2012, <https://core.ac.uk/download/pdf/295567817.pdf>

8.2 Seminar / Laborator	Metode de predare	Observații (ore și referințe bibliografice)
Exemplificarea unor sisteme de prelucrarea a limbajului natural.	Problematizare. Explicații la tablă și cu suport electronic.	2 ore
Dicționare și lucrul cu acestea. Evaluarea parțială tema 1.	Problematizare. Explicații la tablă și cu suport electronic.	2 ore
Corpusuri și lucrul cu acestea. Evaluarea parțială tema 1.	Problematizare. Explicații la tablă și cu suport electronic.	2 ore
Predarea temei 1.	Prezentare	2 ore
Tokenizarea, Lemmatizare și POS Tagging. Parsarea dependențelor. Evaluarea parțială tema 2.	Problematizare. Explicații la tablă și cu suport electronic.	2 ore
Word Embeddings. Semantica vectorială și spații semantice. Evaluarea parțială tema 2.	Problematizare. Explicații la tablă și cu suport electronic.	2 ore
Analiza sentimentelor. Clasificarea tematică și categorisirea textului. Evaluarea parțială tema 2.	Problematizare. Explicații la tablă și cu suport electronic.	2 ore
Predarea temei 2.	Prezentare	2 ore
Exerciții de interconectare de resurse lexicale.	Problematizare. Explicații la tablă și cu suport electronic.	2 ore
Tehnici de detecție a semnalului vocal. Transpunerea automată a semnalului vocal în format text.	Problematizare. Explicații la tablă și cu suport electronic.	2 ore
Chatbots și asistenți virtuali. Discuții preliminare proiecte.	Problematizare. Explicații la tablă și cu suport electronic.	2 ore
Prezentare proiect final.	Prezentare	2 ore

Bibliografie

- [1] Daniel Jurafsky and James Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. <https://www.cs.colorado.edu/~martin/slp.html>.
- [2] Eisenstein, Jacob. Introduction to natural language processing. MIT press, 2019. <https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf>
- [3] MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinriche. An introduction to Information Retrieval, 2008. <https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>
- [4] Bird, S., Klein, E., & Loper, E. (2009). "Natural Language Processing with Python". O'Reilly Media. <https://tjzhifei.github.io/resources/NLTK.pdf>
- [5] Scutelnicu A. (2021) - Representing and interconnecting lexical resources - Ph. D. Thesis. https://scdoc.info.uaic.ro/wp-content/uploads/2023/02/Rezumat_Teza_Doc_AndreiScutelnicu_EN.pdf.
- [6] A. V. Aho, M. S. Lam, R. Sethi, and J. D. Ullman, Compilers: Principles, Techniques, and Tools, Second Edition. Addison-Wesley, 2006 <https://dl.acm.org/doi/book/10.5555/1177220>
- [7] Moruz A., Scutelnicu A., Cristea D. (2018) - Interlinking and Extending Large Lexical Resources for Romanian. In V. Păiș, D. Gifu, D. Trandabăț, D. Cristea, D. Tufiș (eds.) Proceedings of the 13th International Conference on Linguistic Resources and Tools for Processing Romanian Language, Iași, pp. 125-132. <https://profs.info.uaic.ro/~dcristea/papers/paper%2014.pdf>.
- [8] Ivan Habernal - "Semantic Web Search using Natural Language", 2012, <https://core.ac.uk/download/pdf/295567817.pdf>

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

Conținutul disciplinei este coroborat cu tematica concursurilor pentru ocuparea posturilor în companiile IT care au departament de cercetare/dezvoltare dedicat sistemelor de prelucrare a limbajului natural.

10. Evaluare

10.1 Evaluare continuă		Pondere (min. 30%)	40		
Curs	Forma de evaluare		Verificare orală		
	Pondere		40		
	Nepromovarea Evaluării continue determină nepromovarea Evaluării finale		Nu		
	Metode de evaluare	Detalii		Pondere	cu reexaminare
		Verificare orală periodică		50	Nu
Verificare orală periodică		50	Nu		
Seminar / Laborator	Forma de evaluare		Verificare practică		
	Pondere		60		
	Nepromovarea Evaluării continue determină nepromovarea Evaluării finale		Nu		
	Metode de evaluare	Detalii		Pondere	cu reexaminare
		Proiect		50	Nu
Proiect		50	Nu		

10.2 Evaluare finală		Pondere (max. 70%)	60	
		Forma de evaluare	Verificare scrisă finală	

10.3 Mențiuni (situații speciale în evaluare)

--	--	--	--	--

10.4 Standard minim de performanță

1. Studenții vor cunoaște noțiunile fundamentale ale domeniului Lingvistică Computațională, începând cu descrierea domeniului și a obiectivelor sale și continuând cu însușirea conceptelor de bază teoretice și practice.
2. Pentru promovare este necesară acumularea a minim 5 puncte din L și minim 5 puncte din E. Notele se stabilesc folosind formula ROUNDUP ($L*0.4 + E*0.6$)

Data completării,

Titular de curs,
Lect. Dr. LIVIU-ANDREI SCUTELNICU

Titular de seminar,

Data avizării în departament,

Director de departament,
Conf. Dr. ANDREI ARUSOAIIE