



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	Informatică
1.4 Domeniul de studii	Informatică
1.5 Ciclul de studii	Masterat
1.6 Programul de studii / Calificarea	Informatică

2. Date despre disciplină

2.1 Denumirea disciplinei	Introducere în lingvistica computațională						
2.2 Titularul activităților de curs	Prof. Dr. Dan CRISTEA						
2.3 Titularul activităților de seminar	Prof. Dr. Dan CRISTEA						
2.4 An de studiu	I	2.5 Semestru	2	2.6 Tip de evaluare	E	2.7 Regimul disciplinei*	OB

* OB – Obligatoriu / OP – Opțional

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

3.1 Număr de ore pe săptămână	4	din care: 3.2 curs	2	3.3 seminar/laborator	2
3.4 Total ore din planul de învățământ	56	din care: 3.5 curs	28	3.6 seminar/laborator	28
Distribuția fondului de timp					ore
Studiu după manual, suport de curs, bibliografie și altele					28
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren					28
Pregătire seminarii/laboratoare, teme, referate, portofolii și eseuri					48
Tutoriat					-
Examinări					4
Alte activități					-
3.7 Total ore studiu individual					104
3.8 Total ore pe semestru					164
3.9 Număr de credite					6

4. Precondiții (dacă este cazul)

4.1 De curriculum	Cursul Structuri de date și algoritmi sau un curs de Algoritmă și programare Cursul de Lexicologie, morfologie și sintaxă computațională
4.2 De competențe	Cunoștințe de algoritmă și programare ajută dar nu sunt obligatorii

5. Condiții (dacă este cazul)

5.1 De desfășurare a cursului	- Studenții trebuie să fie pregătiți ca orele sau numai o parte din ele să fie ținute în regim online (pe perioada restricțiilor impuse de pandemie). - La prelegerile ținute în clasă, seminarii/laboratoare, studenții se vor prezenta cu telefoanele mobile închise.
-------------------------------	--



5.2 De desfășurare a seminarului/ laboratorului	<ul style="list-style-type: none">- Termenul predării temelor de seminar este stabilit de titularul de laborator, de comun acord cu studenții.- Nu se acceptă cererile de amânare decât pe motive obiectiv întemeiate.- Pentru predarea cu întârziere a temelor de seminar, acestea vor fi depunctate cu 0,5 pct./zi de întârziere.
--	---

6. Competențe specifice acumulate

Competențe profesionale	<p>C1. Cunoașterea teoriilor semnificative și a metodelor de lucru utilizate în lingvistica computațională.</p> <p>C2. Dezvoltarea de deprinderi de utilizare a unor aplicații de prelucrarea limbajului natural.</p> <p>C3. Abilitatea de a adapta și implementa modele existente la specificul limbii române.</p> <p>C4. Capacitatea de a utiliza limbajul științific al domeniului lingvisticii computaționale și al prelucrării limbajului natural în interacțiuni cu specialiști informaticieni și lingviști.</p>
Competențe transversale	<p>CT1. Identificarea rolurilor și responsabilităților într-o echipă pluridisciplinară și folosirea de tehnici de comunicare, relaționare și muncă eficientă în cadrul unei echipe.</p> <p>CT2. Aplicarea normelor și valorilor de etică profesională</p> <p>CT3. Participarea la proiecte cu caracter științific în condiții de autonomie și independență profesională și demonstrarea capacității de a identifica oportunități pentru propria formare profesională în viitor.</p> <p>CT4. Utilizarea eficientă a surselor informaționale și a resurselor de comunicare și formare profesională asistată, atât în limba română, cât și într-o limbă de circulație internațională.</p>

7. Obiectivele disciplinei (din grila competențelor specifice acumulate)

7.1 Obiectivul general	Inițierea în domeniul teoretic al lingvisticii computaționale (LC) și cel aplicativ al prelucrării limbajului natural (PLN).
7.2 Obiectivele specifice	La finalizarea cu succes a acestei discipline, studenții vor fi capabili să: <ul style="list-style-type: none">▪ explice caracteristicile modelelor folosite în LC/PLN ;▪ deprindă utilizarea metodelor simbolice, statistice și neurale aplicate în LC/PLN și adecvarea lor în diverse aplicații;▪ descrie conceptele și metodele utilizate în dezvoltarea aplicațiilor de LC/PLN;▪ utilizeze corect termenii specifici modelelor de natură simbolică, statistică, neurală;▪ adapteze și să implementeze modele simbolice, statistice și neurale pe baza unor specificații date pentru diverse probleme reale din domeniul LC/PLN;▪ analizeze, modeleze și evalueze soluția unei aplicații informatice simple din domeniul LC/PLN.

8. Conținut

	Curs	Metode de predare	Observații (ore și referințe bibliografice)
1	Niveluri de prelucrare aplicate limbajului natural; construcția de resurse, prelucrări inițiale ale documentelor	Expunere online, prelegere participativă	2 ore Dan Cristea
2	Processing Polish Parliamentary Data	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: dr. Maciej Ogrodniczuk, Linguistic Engineering Group,



			Institute of Computer Science, Polish Academy of Sciences
3	Editare critică computerizată	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: dr. Violeta Seretan, Senior Researcher in Digital Humanities at the University of Lausanne, Switzerland
4	Semantic relations between nominals	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: dr. Vivi Năstase, Institut für Computerlinguistik, Universität Heidelberg, Germany
5	Machine translation evaluation from 'the other side of the pond'	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: dr. Constantin Orăsan, Professor of Language and Translation Technologies at the Centre of Translation Studies, University of Surrey, Great Britain
6	How search works	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: Oana Nikolov, Bing, USA
7	The evolution of deep learning models for question answering	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: conf. dr. ing. Trăian Rebedea, Faculty of Computer Engineering, University Politehnica of Bucharest
8	Text Mining at Character Level	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: conf. dr. Marius Popescu, Facultatea de Matematică și Informatică, Universitatea din București
9	Romanian text processing pipelines: the Deep Learning chapter	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: CS II dr. ing. Radu Ion, Institutul de Inteligență Artificială, Academia Română, București
10	Applications of NLP in mental healthcare	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: Valentin Tablan, chief AI officer at The IESO Research Lab, England
11	Conversie vorbire-text	Expunere online, prelegere participativă, exemplificare,	2 ore Invitat: conf. dr. ing.



		aplicații	Horia Cucu, Facultatea de Electronică, Telecomunicații și Tehnologia Informației, Universitatea Politehnică din București
12	An overview of speech synthesis: from concatenative to deep learning approaches	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: ș.l. dr. ing. Adriana Stan, Grupul de Prelucrarea Vorbirii, Universitatea Tehnică din Cluj-Napoca
13	Coreference resolution	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: prof. dr. Amalia Todirașcu, Groupe Linguistique, Langue, Parole, Faculte de Lettre, Universite de Strassbourg
14	The ultimate assets of the knowledge society: the Data	Expunere online, prelegere participativă, exemplificare, aplicații	2 ore Invitat: acad. Dan Tufiș, Institutul de Inteligență Artificială, Academia Română, București

Bibliografie

- R. Mitkov (2005). The Oxford Handbook of Computational Linguistics.
- Chris Manning and Hinrich Schütze (1999). Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA, <https://nlp.stanford.edu/fsnlp/>
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean (2014). Distributed Representations of Words and Phrases and their Compositionality: <https://arxiv.org/pdf/1310.4546.pdf>
- Chris Nicholson and Adam Gibson (2014). SkyMind: A Beginner's Guide to Word2Vec and Neural Word Embeddings: <https://skymind.ai/wiki/word2vec>
- TensorFlow: An open source machine learning framework for everyone: <https://www.tensorflow.org/>
- Natural Language Processing Fundamentals in Python, Natural Language Toolkit, de ex. <http://nltk.org/book>.
- Uday Kamath, John Liu, James Whitaker (2019). Deep Learning for NLP and Speech Recognition, Springer.
- Site-ul NLP-Group@UAIC-FII, la <http://nlptools.info.uaic.ro/>
- Altele vor fi anunțate în cursuri

Seminar

Seminarele/laboratoarele au ca obiectiv rezolvarea unui set de proiecte. Studenților li se indică pachete software și aplicații web cu care să construiască miniaplicații. Tehnologiile prezentate sunt exemplificate în aplicații de tipul: asistenți virtuali, regăsire de informații, analiza textelor din medicină, prelucrarea vorbirii etc.

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

Conținutul cursului are la bază experiența acumulată de un număr de cercetători de prestigiu din țară și străinătate, o parte din ei – foști absolvenți ai UAIC-FII. Lingvistica computațională și prelucrarea limbajului natural sunt considerate astăzi domenii cheie în tehnologiile informației. Corporații internaționale mari precum



Google, Bing, Microsoft, IBM, Facebook, Nuance etc. au investit substanțial în acest domeniu. Peste tot în lume, mii de start-up-uri se înființează pentru diverse aplicații și servicii de prelucrare a limbajului natural, esențiale pentru extragerea de informații din colecții imense de date structurate sau nestructurate. Tehnologia limbajului permite oamenilor să colaboreze, să învețe, să facă afaceri și să împărtășească cunoașterea dincolo de granițele limbajului, indiferent de competențele de utilizare a calculatorului.

10. Evaluare

10.1 Tip activitate	10.2 Criterii de evaluare	10.3 Metode de evaluare	10.4 Pondere în nota finală (%)
Curs	- Corectitudinea și completitudinea cunoștințelor; - Capacitatea de a înțelege și explica modele simbolice, statistice și neurale aplicate limbajului natural. - Utilizarea corectă a terminologiei domeniului - Creativitate	Teste practice (evaluare parțială – la cerere – și evaluare finală online), proiecte și tehnologii existente în lume	33%
Seminar/laborator	- Abilitatea de a construi aplicații de analiză a textelor folosind diferite pachete software.	Exerciții de seminar/laborator	33%
	Criterii ce vizează aspectele atitudinale: - Interesul pentru studiu individual - Interactivitatea, prezența activă, la curs și seminar	Proiect de cercetare	33%
		Participarea activă la cursuri și seminarii/laboratoare	bonus 10%

11. Standard minim de performanță

Pentru promovare trebuie îndeplinite simultan următoarele criterii:

- Utilizarea corectă a terminologiei domeniului (testată în exprimările din teză și în interacțiunile de la curs și laboratoare);
- Abilitatea de a explica un model simbolic, statistic ori neural aplicat PLN;
- Capacitatea de a elabora un model simbolic, statistic ori neural pentru o problemă reală de PLN;
- Minim 50% prezență la curs & seminar/laborator;
- Minim 30% puncte din temele de la seminar/laborator;
- Minim 40 puncte la teză.

Notele se stabilesc conform cu criteriile ECTS.

Data completării
martie 2021
actualizat iunie 2021

Titular de curs
Prof. Dr. Dan Cristea

Titular de seminar
Prof. Dr. Dan Cristea

Data avizării în departament

Director de departament
Prof.Dr. Dorel LUCANU