

## PROCESAREA LIMBAJULUI NATURAL

### 1. Date despre disciplină/modul

<b>Facultatea</b>	Calculatoare, Informatică și Microelectronică				
<b>Departamentul</b>	Informatică și ingineria sistemelor				
<b>Ciclul de studii</b>	Studii superioare de licență, ciclul I				
<b>Programul de studii</b>	0613.5 Informatica aplicată				
<b>Anul de studii</b>	<b>Semestrul</b>	<b>Tip de evaluare</b>	<b>Categoria formativă</b>	<b>Categoria de opționalitate</b>	<b>Credite ECTS</b>
III (învățământ cu frecvență);	5	E, PA	S – disciplina de domeniu profesional	A - unitate de curs opțională	6

### 2. Timpul total estimat

Total ore în planul de învățământ	Din care			
	Ore auditoriale		Lucrul individual	
	Curs	Laborator/seminar	Studiul materialului teoretic	Pregătire aplicații
180	30	30/30	30	60

### 3. Precondiții de acces la disciplină/modul

Conform planului de învățământ	Programarea calculatoarelor, Structuri de date și algoritmi, Programarea procedurală, Programarea interactivă, Baze de date, Lingvistica computațională.
Conform competențelor	Bazele programării procedurale, cunoștința cu algoritmizarea problemelor, inițierea în programarea orientată pe obiecte.

### 4. Condiții de desfășurare a procesului educațional pentru

Curs	Pentru prezentarea materialului teoretic în sala de curs este nevoie de proiector și calculator. La fel este de dorit și posibilitatea de a conecta calculatorul la internet pe parcursul lecțiilor.
Laborator/seminar	Studentii vor perfecta lucrările practice conform condițiilor și variantelor impuse de indicațiile metodice și vor prezenta profesorului în format electronic. Termenul de predare a lucrării de laborator – două săptămâni după finalizarea acesteia. Pentru predarea cu întârziere a lucrării aceasta se depunctează cu 1pct./două săptămâni de întârziere.

### 5. Competențe specifice acumulate

Competențele formate de această disciplină vor servi ca bază pentru formarea competențelor profesionale în cadrul unității de curs Procesarea Informației.

Disciplina prevede formarea următoarelor competențe profesionale și transversale:

#### CPL 1. Proiectarea aplicațiilor

- K1 Tehnici de modelare a cerințelor și tehnici de analiză a nevoilor.

- K2 Metodele de dezvoltare a software-ului și argumentarea acestora (de exemplu, prototipuri, metode agile, retroinginerie etc.).
- K3 Metricile care se referă la dezvoltarea aplicațiilor..

#### CPL 2. Proiectarea și dezvoltarea aplicațiilor

- K1 Programe/module software adecvate.
- K4 Tehnologiile de ultimă oră.

#### CPL 5. Implementarea soluțiilor

- K1 Tehnici de analiză a performanței.
- K2 Tehnicile legate de gestionarea problemelor (funcționare, performanță, compatibilitate).

### 6. Obiectivele unității de curs/modulului

Obiectivul general	Instruirea studenților cu scopul de a dezvolta abilitățile lor de a crea aplicații ce procesează informația prezentată în forma de text nestructurat, să organizeze o interacțiune cu utilizatorul prin dialog, să extragă cunoștințele necesare din text și să le organizeze în structuri accesibile pentru prelucrarea automată.
Obiectivele specifice	Utilizarea conceptelor de bază din lingvistică computațională și procesarea limbajului natural în descrierea și analiza textului. Stocarea datelor lingvistice în formate standard și utilizarea lor pentru sporirea performanței aplicațiilor procesării textelor. Manipularea și analiza datelor lingvistice cu ajutorul instrumentelor soft. Utilizarea structurilor de date și a algoritmilor de procesare a textelor.

### 7. Conținutul disciplinei/modulului

Tematica activităților didactice	Numărul de ore	
	învățământ cu frecvență	
<b>Tematica prelegerilor</b>		
T1. Procesarea limbajului natural, introducere.	2	
T2. Colectarea resurselor lingvistice, prelucrarea lor.	4	
T3. Metode de extragere și reprezentare a cunoștințelor.	4	
T4. Interogarea cunoștințelor codificate.	4	
T5. Aplicarea metodelor de învățare automată în PLN.	4	
T6. Digitizarea documentelor istorice.	2	
T7. Crearea resurselor lingvistice cu adnotarea sintactică.	4	
T8. Crearea resurselor lingvistice cu adnotarea semantică.	2	
T9. Modele lingvistice mari.	2	
T10. Probleme etice în crearea modelelor lingvistice.	2	
<b>Total prelegeri:</b>	<b>30</b>	
<b>Tematica lucrărilor practice</b>		
Pr1. Crearea bazei de cunoștințe în Prolog.	2	
Pr2. Baza de cunoștințe WordNet.	2	
Pr3. Baza de cunoștințe FrameNet.	2	
Pr4. Baza de cunoștințe ConceptNet.	2	
Pr5. Crearea ontologiei în Protege.	4	
Pr6. Codificarea cunoștințelor în RDF.	4	
Pr7. Interogarea cunoștințelor codificate în WikiData.	2	
Pr8. Crearea resurselor lingvistice cu adnotarea morfologică.	2	
Pr9. Crearea resurselor lingvistice cu adnotarea sintactică	4	
Pr10. Codificarea expresiilor idiomatice	2	
Pr11. Analiza sentimentelor și opiniilor în texte.	2	

Pr12. Testarea modelelor lingvistice mari.	2	
<b>Total lucrări de laborator/seminare:</b>	<b>30</b>	
<b>Tematica lucrărilor de laborator/seminarelor</b>		
LL1. Extragerea textelor și crearea corpusului.	6	
LL2. Digitizarea textelor vechi.	6	
LL3. Crearea resurselor lingvistice cu adnotarea sintactică.	6	
LL4. Analiza sentimentelor și opiniilor în texte	6	
LL5. Simplificarea textelor.	6	
<b>Total lucrări de laborator/seminare:</b>	<b>30</b>	

## 8. Referințe bibliografice

Principale	<ol style="list-style-type: none"> <li>1. Jacob Eisenstein, Introduction to Natural Language Processing (Adaptive Computation and Machine Learning series) Illustrated Edition. The MIT Press, 2019.</li> <li>2. Buchmann Robert Andrei, Ghiran Ana-Maria, Tehnologii semantice, Presa Universitara Clujeana, 2018.</li> <li>3. Natural Language Processing with Python. <a href="https://www.nltk.org/book/">https://www.nltk.org/book/</a></li> <li>4. C. D. Manning, P. Raghavan and H. Schtze, Introduction to Information Retrieval, U.K., Cambridge: Cambridge Univ. Press, 2008. <a href="https://nlp.stanford.edu/IR-book/html/htmledition/irbook.html">https://nlp.stanford.edu/IR-book/html/htmledition/irbook.html</a></li> <li>5. Bing Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Studies in Natural Language Processing, Cambridge University Press; 2nd edition, 2020.</li> </ol>
Suplimentare	<ol style="list-style-type: none"> <li>1. Deborah A. Dahl. Natural Language Understanding with Python: Combine natural language technology, deep learning, and large language models to create human-like language comprehension in computer systems. Packt Publishing, 2023.</li> <li>2. Dana Richards, Henry Hamburger, Logic And Language Models For Computer Science, 3rd Edition, World Scientific, 2017.</li> <li>3. AI. AI In Practice: An Introduction to Large Language Models : A short instructional guide written by AI (Instructional Guides by AI), 2023.</li> </ol>

## 9. Utilizarea IA generativă

<b>Permisiunea de utilizare</b>	<p>Utilizarea IA generative în cadrul temelor și proiectelor este permisă, cu condiția ca studenții să respecte următoarele reguli:</p> <ul style="list-style-type: none"> <li>• IA generativă poate fi utilizată pentru generarea de idei, structuri de text sau cod, dar toate materialele generate trebuie să fie revizuite și ajustate de către student pentru a se asigura că acestea corespund cerințelor academice.</li> <li>• Orice utilizare a IA generative trebuie să fie declarată în secțiunea de apendice a fiecărei lucrări, folosind fraza: "În timpul pregătirii acestei lucrări, autorul a utilizat [NUME INSTRUMENT / SERVICIU] în scopul [MOTIV]. După utilizarea acestui instrument/serviciu, autorul a revizuit și editat conținutul după cum a fost necesar și își asumă întreaga responsabilitate pentru conținutul lucrării."</li> </ul>
<b>Restricții de utilizare</b>	<p>Studenții nu trebuie să considere IA generativă ca o sursă de încredere pentru informații, deoarece nu oferă referințe clare sau surse documentate.</p> <ul style="list-style-type: none"> <li>• Nu este permisă citarea directă a conținutului generat de IA în lucrările academice ca și cum ar fi sursă primară.</li> <li>• Activitățile în care este interzis utilizarea IA generativă sunt specificare de profesor și sunt de regulă evaluări intermediare și finale sau care nu presupun activități de dezvoltare a competențelor profesionale.</li> </ul>

## 10. Evaluare

Periodică		Curentă	Studiu individual	Proiect/teză	Examen
EP 1	EP 2				
10%	10%	10%	---	40%	60%
<p>Standard minim de performanță</p> <p>Prezența la lecții; activitatea și calitatea pregătirii la / pentru prelegeri și lucrări de laborator; Obținerea notei minime de „5” la fiecare dintre atestări și lucrări de laborator;</p> <p>Obținerea notei minime de „5” la proiectul de an;</p> <p>Demonstrarea în lucrarea de examinare finală cunoștințelor teoretice și competențelor practice necesare pentru procesarea automată a textelor.</p>					