**FACULTATEA CALCULATOARE, INFORMATICĂ ŞI MICROELECTRONICĂ**

**DEPARTAMENTUL INFORMATICA ŞI INGINERIA SISTEMELOR**

|  |  |  |
| --- | --- | --- |
| **APROBATĂ**  **la şedinţa DIIS nr.\_\_din\_\_\_\_\_\_\_\_\_\_\_**  **Șefa DIIS**  **Viorica SUDACEVSCHI, conf. univ., dr.**  **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |  | **APROBATĂ**  **la şedinţa Consiliului FCIM**  **nr.\_\_\_din\_\_\_\_\_\_\_\_\_**  **Președintele Consiliului FCIM**  **Dumitru Ciorba, conf. univ., dr.**  **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |

**Program de studiu:** 0613.5 Informatică Aplicată

**Denumirea unității de curs: INGINERIA LIMBAJULUI**

**Beneficiari:** Studenţii anului II, învățământ cu frecvență

###### Ciclul de învățământ: Studii superioare de Licenţă, ciclul I

**Numărul de credite ECTS: 3** ( 45 ore în auditoriu şi 45 ore de activităţi individuale ale studenţilor, 1 credit = 15 ore în auditoriu şi 15 ore de activităţi individuale ale studenţilor).

**Titularul unității de curs: conf. univ., dr. Victoria BOBICEV**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

semnătura titularului de curs

1. **PRELIMINARII**

INGINERIA LIMBAJULUI UMAN (Human Language Engineering) întruneşte necesităţile profesioniştilor şi cercetătorilor din toate domeniile procesării computaţionale a limbajului natural, atât din perspectiva lingvisticii teoretice sau descriptive, lexicologiei, cat şi a informaticii sau a ingineriei. Scopul este de a crea o punte între cercetările lingvisticii computaţionale tradiţionale şi implementarea aplicaţiilor practice în potenţiala utilizare în viața de zi cu zi. Tematica este destul de variata dar se axează pe procesarea textului și interacțiunea cu utilizatorul aplicând in practică analiza textelor prin metodele dezvoltate de regăsirea informaţiei, extragerea informației, interfeţe in formă de dialog şi alte tehnologii.

Scopul disciplinei este de a oferi studenţilor un cadru pragmatic pentru a studia procesarea limbajului în contextul învăţării programării. Cursul întruneşte necesităţile profesioniştilor şi cercetătorilor din toate domeniile procesării computaţionale a limbajului natural. Scopul este de a crea o punte între cercetările lingvisticii computaţionale tradiţionale şi implementarea aplicaţiilor practice în potenţiala utilizare în viaţa de zi cu zi. După terminarea acestui curs studenţii vor fi pregătiţi să creeze aplicații ce procesează informația prezentată in forma de text nestructurat și să organizeze o interacțiune cu utilizatorul prin dialog.

Obiectivele cursului:

* Manipularea şi analiza datelor lingvistice cu ajutorul unor module şi manipularea datelor lingvistice.
* Utilizarea conceptelor de bază din lingvistică computașională și procesarea limbajului natural în descrierea şi analiza textului.
* Utilizarea structurilor de date şi a algoritmilor de procesare a textelor.

Stocarea datelor lingvistice în formate standard şi utilizarea lor pentru spoririea performanţei aplicațiilor procesării textelor.

Cursul este orientat spre pregătirea specialiştilor de o calificare înaltă în domeniul procesării limbajului natural.

1. **PRECONDIŢII DE ACCES LA UNITATEA DE CURS/MODUL:**

Pentru a atinge obiectivele cursului studenţii trebuie să posede abilităţi de analiza automată a textelor, traducerea automată, regăsirea informaţiei şi procesarea şi generarea vorbirii, sisteme integrate, interfeţe multimodale, Tehnologii ale limbajului natural; Lingvistica Computaţională, instrumente de procesare a limbajului natural şi alte aplicaţii.

1. **COMPETENŢELE CARE URMEAZĂ A FI DEZVOLTATE**

Unitatea de curs prevede formarea următoarelor competenţe profesionale şi transversale:

***CP4.*** Soluţionarea problemelor reale folosind instrumentele matematicii şi informaticii aplicate.

* Înţelegerea conceptului de modelare a limbajului natural şi proiectarea sistemelor

automatizate de analiză şi procesare a textului.

* + Abilităţi de programare iniţială şi explorare a datelor din text.
  + Manipularea unui corpus mare de date, explorarea modelelor lingvistice şi testarea problemelor empirice.
  + Construirea unor sisteme robuste pentru realizarea sarcinilor lingvistice cu aplicaţii

tehnologice.

* + Abilităţi de utilizare a tehnicilor de modelare a datelor, de extragere a datelor şi de

acumulare a cunoştinţelor pentru analiza limbajului natural.

* + Abilităţi de utilizare a algoritmilor şi structurilor de date în softurile robuste de procesare

a limbajului.

* + Analiza cerinţelor pieţei şi tendinţelor contemporane privind dezvoltarea produselor soft

din domeniul informaticii aplicate.

***CT1.*** Realizarea proiectului de an cu utilizarea corectă a surselor bibliografice şi metodelor specifice, în condiţii de autonomie restrânsă şi asistenţă calificată, precum şi susţinerea acestora cu demonstrarea capacităţii de evaluare calitativă şi cantitativă a unor soluţii tehnice din domeniu.

***CT3.*** Identificarea nevoii de formare profesională, cu analiza critică a propriei activităţi de formare şi a nivelului de dezvoltare profesională şi utilizarea eficientă a resurselor de comunicare şi formare profesională (Internet, e-mail, baze de date, cursuri on-line etc.), inclusiv folosind limbi străine.

1. **ADMINISTRAREA UNITĂŢII DE CURS**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Codul disciplinei | Anul predării | Semestrul | Numărul de ore | | | | | Evaluarea | | |
| Prelegeri | | Seminare | Lucrări de laborator | Lucrul individual | Credite | Curentă | Finală |
| F.03.O.011 | Învăţământ cu frecvenţă | | | | | | | | | |
| II | III | 30 | - | | 15 | 45 | 3 | 2 atestări | examen |
|  | | | | | | | | | |

1. **REZULTATELE ÎNVĂŢĂRII, CONŢINUTURI ŞI METODE DIDACTICE APLICATE**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Rezultatele învăţării.**  **Studentul trebuie:** | **Conţinuturi** | | **Metode de predare** | **Realizarea în timp (ore)\*** | | | |
| **Prelegeri** | **Lucrări de laborator** | **învăţământ cu frecvenţă** | | **învăţământ cu frecvenţă redusă** | |
| **prelegeri** | **l. lab** | **prelegeri** | **l. lab** |
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| ***să cunoască:***   * *Obiectivele şi scopul lingvisticii computaţionale şi procesării limbajului natural;* * *conceptele principale ale ingineriei limbajului natural;* * *metodele de bază aplicate in domeniile date.*   ***să fie capabil:***   * *Să definească conceptele de bază din domeniile date;* * *Să facă distincţie între subdomeniile ingineriei limbajului natural;* * *De a înțelege semnificația resurselor lingvistice.* | **Tema 1.**  Recapitulare. Lingvistica computaţională şi procesarea limbajului natural. | Lucrarea de laborator nr. 1.  Crearea sistemului automat de Întrebare-Răspuns. Etapa 1. Selectarea textului, formularea intrebărilor, analiza lor | **Pentru prelegere:** expunerea, conversaţia  **Pentru lucrare de laborator:** învăţarea prin cercetarea web-ului | **4** | **4** |  |  |
| ***să cunoască:***   * *Tipuri de interfețe in limbaj natural;* * *Metodologiile utilizate in crearea interfețor in limbaj natural;* * *Problemele ce sunt cercetate în procesul creării interfețelor in limbaj natural.*   ***să fie capabil:***   * *să descrie metodologii diverse utilizate in crearea interfețor in limbaj natural;* * *să analizeze problemele in crearea interfețor in limbaj natural;* * *să propună soluții in crearea interfețor in limbaj natural.* | Tema 2.  Interfețe în limbaj natural. Componentele interfețelor în limbaj natural. | **Lucrarea de laborator nr 1.**  continuare | **Pentru prelegeri:** expunerea, învăţare prin colaborare  **Pentru lucrare de laborator:** problematizarea; | **2** |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| ***să cunoască:***   * *tipurile stemelor de întrebare – răspuns;* * *Modulele sistemului de întrebare – răspuns;* * *abordările curente pentru cerarea sistemele de întrebare – răspuns.*   ***să fie capabil:***   * *să descrie modulele utilizate in crearea sistemelor de* *întrebare – răspuns;* * *să analizeze problemele in crearea sistemelor de întrebare – răspuns;* * *să propună soluții in crearea sistemelor de* *întrebare – răspuns.* | Tema 3  Sisteme de intrebare-răspuns (QA), modulele lor, tehnologii utilizate Clasificarea sistemelor de întrebare – răspuns. Arhitectura unui sistem QA. Procesarea întrebărilor pentru un Open domain QA. Clasificarea tipurilor de răspunsuri pentru sistemele de întrebare-răspuns deschise. | Lucrarea de laborator nr. 2  Crearea sistemului automat de Întrebare-Răspuns. Etapa 2: analiza sintactică și semantică a intrebărilor și extragerea cuvintelor cheie. | **Pentru prelegeri:** expunerea, învăţare prin colaborare  **Pentru lucrare de laborator:** problematizarea; tehnici de învăţare interactiv-creativă | **4** | **4** |  |  |
| ***să cunoască:***   * *cum este efectuată analiza morfologică a textelor;* * *codurile morfologice utilizate;* * *metodele de parsare sintactică.*   ***să fie capabil:***   * *să analizeze morfologic un text utilizând metodele existente;* * *să parseze un text utilizând aplicațiile deja existente;* * *să extragă cuvintele cheie din text analizat.* | **Tema 4**  Etapele analizei textului: morfologică, sintactică, semantică. Parsing parțial. | **Lucrarea de laborator nr.** **2.**  continuare | **Pentru prelegeri:** expunerea, învăţare prin colaborare  **Pentru lucrare de laborator:** problematizarea; tehnici de învăţare interactiv-creativă | **2** |  |  |  |
| ***să cunoască:***   * *abordări curente in domeniul de regăsire a informaţiei;* * *metode de definire a similarității între documente;* * *principiile de lucru a motoarelor de căutare.*   ***să fie capabil:***   * *să calculeze similaritatea între două documente utilizând metodele din domeniul de regăsire a informaţiei;* * *să evaluieze rezultatele de regăsire a textelor relevante interogării;* * *să creeze matricea termeni-documente pentru regăsire a informaţiei.* | Tema 5  Regăsirea informaţiei (IR). Componente ale sistemelor de regăsire a informaţiei. Abordări curente: indexarea documentelor; principii de funcționare a motoarelor de căutare. | **Lucrarea de laborator nr.** **3**  Crearea sistemului automat de Întrebare-Răspuns. Etapa 3: Utilizând lista cuvintelor cheie de creat partea aplicației care extrage din text fragmentele care pot să conțină răspunsul la întrebare. De sortat fragmentele după relevanță. | **Pentru prelegeri:** expunerea, învăţare prin colaborare  **Pentru lucrare de laborator:** problematizarea; tehnici de învăţare interactiv-creativă. | **2** | **4** |  |  |
|  |  |  |
| ***să cunoască:***   * *principiile de bază ale extragerii informației din texte;* * *tipuri de entități numite;* * *metodele de extragere a entităților numite.*   ***să fie capabil:***   * *să aplice expresii regulate pentru detectarea entităților numite;* * *să clasifice entitățile numite;* * *să elaboreze un soft ce detectează entitățile numite.* | tema 6.  Extragerea Informației (IE). Identificarea și clasificarea numelor (denumirilor). Construirea și utilizarea unui tagger (etichetator) de nume, denumiri. Recunoașterea și extragerea evenimentelor. | **Lucrarea de laborator nr.** **3.** continuare. | **Pentru prelegeri:** expunerea, învăţare prin colaborare  **Pentru lucrare de laborator:** problematizarea; tehnici de învăţare interactiv-creativă. | **2** |  |  |  |
| ***să cunoască:***   * *principiile de bază ale analizei semantice a textelor;* * *resursele semantice lexicale;* * *metodele și aplicațiile de analiza semantică a texelor.*   ***să fie capabil:***   * *să aplice resursele lexicale existente pentru analiza semantică a textelor;* * *să utilizeze aplicațiile existente pentru analiza semantică a textelor ;* * *să încadreze analiza semantică a textelor în aplicația sa practică.* | tema 7.  Analiza semantică a textului. | **Lucrarea de laborator nr.** **4**  Crearea sistemului automat de Întrebare-Răspuns. Etapa 4: Extragerea răspunsurilor din fragmentele găsite în lucrarea de laborator nr. 3 și formularea răspunsului complet conform gramaticii limbajului natural. | **Pentru prelegeri:** expunerea, învăţare prin colaborare  **Pentru lucrare de laborator:** problematizarea; tehnici de învăţare interactiv-creativă. Crearea aplicației. | **4** | **3** |  |  |
| ***să cunoască:***   * *definiția și tipurile de ontologii;* * *ontologii lexicale;* * *metodele de utilizare a ontologiilor lexicale.*   ***să fie capabil:***   * *să aplice resursele lexicale existente pentru analiza semantică a textelor;* * *să utilizeze aplicațiile existente pentru analiza semantică a textelor;* * *să încadreze ontologiile în aplicații practice.* | tema 8  Ontologia lexicală. WordNet. Verbnet. | **Lucrarea de laborator nr.** **4**  continuare | **Pentru prelegeri:** expunerea, învăţare prin colaborare  **Pentru lucrare de laborator:** problematizarea; tehnici de învăţare interactiv-creativă. | **6** |  |  |  |
| ***Total:*** |  |  |  | **30** |  |  |  |

1. **SUGESTII PENTRU ACTIVITATEA INDIVIDUALĂ A STUDENŢILOR**

Pe parcursul semestrului, studenţii realizează activităţi individuale, care includ:

- studiul literaturii obligatorii conform listei surselor bibliografice prezentate în curriculă;

- realizarea sarcinelor, propuse în cadrul lucrărilor de laborator;

Pe parcursul semestrului studenţilor li se propune tematica pentru elaborarea referatelor în scopul aprofundării cunoştinţelor teoretice.

| **Nr. crt.** | **Tema** | **Conținut activitate individuală** | **Durata, ore** | **Forma de control** | **Termeni de control (perioada)** |
| --- | --- | --- | --- | --- | --- |
| 1 | T1  LL1 | Studierea exemplelor plasate pe platforma online, indicate in link-uri utile.  Selectarea textului in internet  Pregătire intrebărilor  *Studiul temei: sisteme de intrebare-răspuns.* | 4  4 | Verificare la examen  Verificare îndeplinire sarcină | Examen  2 săptămâni de la data stabilirii sarcinii |
| 2 | T2 | Studierea interfețelor om – calculator in limbaj natural | 2 | Verificare la examen | Examen |
| 3 | T3  LL2 | Studierea modulelor sistemelor de intrebare-răspuns  Studierea codurilor morfologice, analizatorilor morfologice, parserilor, metodelor de parsing | 4  3 | Verificare la examen  Verificare îndeplinire sarcină | Examen  2 săptămâni de la data stabilirii sarcinii |
| 4 | T4 | Studierea codurilor morfologice, analizatorilor morfologice, parserilor, metodelor de parsing | 4 | Verificare la examen | Examen |
| 5 | T5  LL3 | *Studierea modalităților de lucru a motoarelor de căutare.*  Crearea codului in Python pentru indeplinirea sarcinii lucrarii de laborator | 2  4 | Verificare la examen  Verificare îndeplinire sarcină | Examen  2 săptămâni de la data stabilirii sarcinii |
| 6 | T6 | Studierea surselor de entități numite  *Studiul temei: expresii regulate*. | 4 | Verificare la examen | Examen |
| 7 | T7  LL | Studierea metodelor de reprezentare semantică a sensului textului. Studierea metodelor de analiza semantică.  Crearea codului in Python pentru indeplinirea sarcinii lucrarii de laborator | 4 | Verificare la examen  Verificare îndeplinire sarcină | Examen  2 săptămâni de la data stabilirii sarcinii |
| 8 | T8 | Studierea resurselor semantice lexicale: wordnet, verbnet, framenet. | 6 | Verificare la examen | Examen |
|  |  | **Total:** | 45 |  |  |

1. **EVALUAREA UNITĂȚII DE CURS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Forma de învățământ | Periodică | | Curentă | Lucrul individual | Examen final |
| Atestarea 1 | Atestarea 2 |
| Cu frecvență | 15% | 15% | 15% | 15% | 40% |
| Cu frecvență redusă | 25% | | | 25% | 50% |
| Standard minim de performanţă | | | | | |
| Prezenţa şi activitatea la prelegeri şi lucrări de laborator  Obţinerea notei minime de „5” la fiecare dintre evaluări şi lucrări de laborator | | | | | |

**VIII. LISTA DE SUBIECTE PENTRU EVALUĂRI PERIODICE ŞI CEA FINALĂ**

**Chestionar pentru atestarea I**

1. Ingineria limbajului: definiţie, scop.
2. Conceptul de resurse lingvistice.
3. Noţiune de lingvistica computațională.
4. Noțiune de procesarea limbajului natural.
5. Metodologii utilizate in cadrul lingvisticii computaționale.
6. Tipurile de bază de abordări in procesarea limbajului natural.
7. Cea mai mare problemă în ingineria limbajului.
8. Regăsirea Informației (IR).
9. Principalele componente ale unui sistem IR.
10. Principalele componente ale sistemului de intrebare-răspuns.
11. Sisteme de interacțiune intre utilizator și calculator.
12. Tipuri de sisteme de interacțiune intre utilizator și calculator.
13. Extragerea informației (IE).
14. Sistemele de întrebare-răspuns (QA).
15. Metodele de căutare într-un sistem IR.
16. Noțiunea de cuvinte cheie.
17. Metodele folosite în dezvoltarea sistemelor de tip QA.
18. Clasificarea întrebărilor.
19. Extragerea de informații.
20. Extragerea răspunsului.
21. Modulul de procesare a întrebării.

**Chestionar pentru atestarea a II-a**

1. Motorul de căutare din sistemul QA.
2. Etapele de procesarea documentelor într-un sistem QA.
3. Rezultatul extragerii și evaluării răspunsurilor candidate.
4. Extragerea termenilor din text și clasificarea lor.
5. Ontologii și tezaure.
6. Criteriile de extragere a termenilor.
7. Metode de evaluare a sistemelor de extragere a informației.
8. Metode de extragere a informației.
9. Expresii regulate.
10. Expresii regulate în Python.
11. Tipuri de entități numite.
12. Wordnet
13. Ontologia Wordnet
14. Utilizarea Wordnet in procesarea textelor
15. Wordnet pentru limba română
16. Verbnet, utilizarea lui.
17. Metode de analiza semantică a textelor.
18. Reprezentarea sensului textului.
19. Probleme in reprezentarea sensului textului.
20. Resurse lexicale existente, utilizarea lor.

**Chestionar pentru examen**

1. Ingineria limbajului: definiţie, scop.
2. Conceptul de resurse lingvistice.
3. Noţiune de traducere automată.
4. Regăsirea Informației (IR).
5. Principalele componente ale unui sistem IR.
6. Interacțiunea cu sistemul de regăsire a informației.
7. Factorii ce influenţează eficienţa unui sistem IR.
8. Extragerea informației (IE).
9. Sistemele de întrebare-răspuns (QA).
10. Metodele de căutare într-un sistem IR.
11. Identificarea prin cuvinte cheie.
12. Metodele folosite în dezvoltarea sistemelor de tip QA.
13. Din câte module distincte este alcătuit un sistem QA?
14. Clasificarea întrebărilor.
15. Extragerea de informații.
16. Extragerea răspunsului.
17. Modulul de procesare a întrebării.
18. Analiza întrebării.
19. Motorul de căutare din sistemul QA.
20. Procesarea documentelor într-un sistem QA.
21. Rezultatul extragerii și evaluării răspunsurilor candidate.
22. Extragerea termenilor și clasificarea lor automată.
23. Ontologii și tezaure.
24. Extragerea Entităților Numite
25. Interfața cu utilizatorul .
26. Proiectarea interfețelor cu utilizatorul .
27. Tipuri de interfețe in limbaj natural.

**IX. REFERINŢE BIBLIOGRAFICE PRINCIPALE**

* + - 1. Jurafsky, D., J. H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall, 2000; [www.cs.colorado.edu/~martin/slp.html](http://www.cs.colorado.edu/~martin/slp.html) .
      2. Mitkov Ruslan. The Oxford Hadbook on Computational Lingustics. <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199276349.001.0001/oxfordhb-9780199276349>
      3. Igor A. Bolshakov and [Alexander Gelbukh](http://www.gelbukh.com/). Computational Linguistics. Models, Resources, Applications. In [www.Gelbukh.com/clbook](http://www.cicling.org/gelbukh/clbook/).
      4. Jurafsky, D., J. H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice-Hall, 2000; [www.cs.colorado.edu/~martin/slp.html](http://www.cs.colorado.edu/~martin/slp.html) .
      5. Mitkov Ruslan. The Oxford Hadbook on Computational Lingustics. <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199276349.001.0001/oxfordhb-9780199276349>
      6. Igor A. Bolshakov and [Alexander Gelbukh](http://www.gelbukh.com/). Computational Linguistics. Models, Resources, Applications. In [www.Gelbukh.com/clbook](http://www.cicling.org/gelbukh/clbook/).
      7. An introduction to NLP. <https://www.cs.bham.ac.uk/~pxc/nlp/index.html>
      8. Victoria Bobicev, Baze de cunoştinţe lingvistice : Note de curs, editura universitara, 2012.
      9. Florentina Hristea, Introducere in procesarea limbajului natural cu aplicatii in Prolog, Editura Universitatii din Bucuresti, 2000.
      10. Dan Cristea - "Curs de Lingvistică computaţională", Facultatea de Informatică, Universitatea "Alexandru Ioan Cuza" Iaşi, 2003.

**X. REFERINŢE BIBLIOGRAFICE SUPLIMENTARE**

1. <https://profs.info.uaic.ro/~dcristea/teaching.html>
2. Scutelnicu Liviu Andrei, PLN <https://sites.google.com/site/andreiscutelnicu/nlp>
3. Prelucrarea Limbajului Natural <https://muhaz.org/prelucrarea-limbajului-natural.html>