



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	Master
1.6 Programul de studii / Calificarea	Informatică/Master în Lingvistică Computațională

2. Date despre disciplină

2.1 Denumirea disciplinei	Prelucrarea statistică a limbajului natural						
2.2 Titularul activităților de curs	Conf. Dr. Diana TRANDABĂȚ						
2.3 Titularul activităților de seminar	Conf. Dr. Diana TRANDABĂȚ						
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					8

4. Precondiții (dacă este cazul)

4.1 De curriculum	Nu este cazul
4.2 De competențe	Nu este cazul

5. Condiții (dacă este cazul)

5.1 De desfășurare a cursului	- Studenții se vor prezenta la prelegeri, seminarii/laboratoare cu telefoanele mobile închise.
5.2 De desfășurare a seminarului/ laboratorului	- Termenul predării temelor de laborator 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 laborator, acestea vor fi depunctate cu 0,5 pct./zi de întârziere.



6. Competențe specifice acumulate

Competențe profesionale	C1. Cunoașterea caracteristicilor modelelor statistice și probabilistice. C2. Dezvoltarea de deprinderi de dezvoltare și utilizare corectă a unui model probabilistic. C3. Abilitatea de a adapta și implementa modele statistice pentru aplicații informatice specifice domeniului prelucrării limbajului natural. C4. Utilizarea adecvată a limbajului probabilistic pentru proiectarea și elaborarea unor programe complexe.
Competențe transversale	CT1. Identificarea rolurilor și responsabilităților într-o echipă pluridisciplinară și folosirea de tehnici de comunicare, relaționare și muncă eficientă în cadrul unei echipe. CT2. Aplicarea normelor și valorilor de etică profesională 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. 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.

7. Obiectivele disciplinei (din grila competențelor specifice acumulate)

7.1 Obiectivul general	Inițierea în statistică și probabilități și aplicarea conceptelor din curs la probleme reale din domeniul prelucrării limbajului natural.
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 statistice și probabilistice;▪ descrie conceptele și metodele utilizate în dezvoltarea aplicațiilor informatice folosind probabilități;▪ utilizeze corect sintaxa și termenii specifici modelelor probabilistice;▪ adapteze și să implementeze modele statistice pe baza unor specificații date pentru diverse probleme reale din domeniul prelucrării limbajului natural;▪ analizeze, modeleze și testeze soluția unei aplicații informatice simple.

8. Conținut

8.1	Curs	Metode de predare	Observații (ore și referințe bibliografice)
1.	Noțiuni generale despre desfășurarea cursului. Introducere în probabilități. Eveniment. Spațiul evenimentelor.	Expunere, prelegere participativă	2
2.	Evenimente independente. Probabilități condiționate. Teoria lui Bayes.	Expunere, prelegere participativă, exemplificare, aplicații	2
3.	Teoria informației. Variabilă aleatoare. Variație. Deviere standard. Expectanță.	Expunere, prelegere participativă, exemplificare, aplicații	2
4.	Modele de limbă. Evaluarea modelelor de limbă. Entropie. Perplexitate. Informație mutuală.	Expunere, prelegere participativă, exemplificare, aplicații	2
5.	Colocații. Medie și variație. <i>T</i> -test. Chi-square test.	Expunere, prelegere participativă, exemplificare, aplicații	2



6.	Gramatici regulate. Ierarhia lui Chomsky	Expunere, prelegere participativă, exemplificare, aplicații	2
7.	Arbori de parsare. Probabilitatea unui arbore și a unei propoziții.	Expunere, prelegere participativă, exemplificare, aplicații	2
8.	Algoritmul CKY. Varianta Probabilistică și cea cu reguli unare.	Discuții și dezbateri, aplicații, testare.	2
9.	Algoritmul Earley. Varianta probabilistică.	Expunere, prelegere participativă, modelare, aplicații	2
10 - 11.	Regăsirea de informații. Indexarea. TF-IDF. Model boolean. Spațiu vectorial. Index inversat. Map-reduce.	Expunere, prelegere participativă, modelare, aplicații	4
12 - 13.	Clasificare și clusterizare de texte. Naïve Bayes. k-NN. Evaluarea clasificării	Expunere, prelegere participativă, modelare, aplicații	4
14.	Dezambiguizare semantică. Algoritmul Lesk. Algoritmul Walker.	Expunere, prelegere participativă, modelare, aplicații	2

Bibliografie

- Kao Anne, Poteet R. Stephen (Eds) (2007) - Natural Language Processing and Text Mining, Springer
- Manning, H. Schütze. Foundations of Statistical Natural Language Processing. MIT Press 1999.
- R. Mitkov. The Oxford Handbook of Computational Linguistics, 2005
- D. Jurafsky, J. Martin. Speech and Language Processing, 2009

Seminar/Laborator

Laboratoarele urmează îndeaproape noțiunile predate la curs. Printre obiectele de studiu se numără paradigmele majore din domeniul prelucrărilor statistice. Studenții vor căpăta experiență în construirea de modele probabilistice, lucrând cu date reale, și vor învăța cum să formuleze și să investigheze diferite direcții de cercetare. Cursurile și laboratoarele vor varia de la prelegeri asistate de videoproiector, la prezentări și discuții, studii de caz, programare, experimente cu diferite instrumente existente, iar în a doua jumătate a semestrului lucru la proiecte.

Bibliografie

- C. Manning, H. Schütze. Foundations of Statistical Natural Language Processing. MIT Press 1999.
- R. Mitkov. The Oxford Handbook of Computational Linguistics, 2005
- D. Jurafsky, J. Martin. Speech and Language Processing, 2009

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

Conținutul cursului este coroborat cu conținutul cursurilor similare de la diferite universități internaționale de prestigiu. În contextul big data, modelele statistice și probabilistice sunt esențiale pentru extragerea de informații din colecții imense de date structurate sau nestructurate.

10. Evaluare



Tip activitate	10.1 Criterii de evaluare	10.2 Metode de evaluare	10.3 Pondere în nota finală (%)
10.4 Curs	<ul style="list-style-type: none">- Corectitudinea și completitudinea cunoștințelor;- Capacitatea de a înțelege și explica modele statistice și de probabilități.- Înțelegerea modului de aplicare a probabilităților în diverse probleme practice de prelucrare a limbajului natural.- Calitatea formulării răspunsurilor	Teste practice (evaluare parțială și evaluare finală), proiecte	30%
10.5 Seminar/ Laborator	<ul style="list-style-type: none">- Abilitatea de a construi un model probabilistic care rezolvă probleme de procesarea limbajului.- Folosirea corectă a sintaxei și termenilor specifici.- Însușirea deprinderilor de lucru în echipă și contribuire susținută la diversele etape din dezvoltarea unui proiect.	Exerciții de laborator, teme	30%
	Criterii ce vizează aspectele atitudinale: <ul style="list-style-type: none">- Calitatea programelor/ modelelor elaborate- Interesul pentru studiu individual	Proiect	30%
		Participarea activă la laboratoare	10%
10.6 Standard minim de performanță			
Pentru promovare trebuie îndeplinite simultan următoarele criterii: <ul style="list-style-type: none">▪ Utilizarea corectă a sintaxei specifice modelărilor statistice și probabilistice;▪ Abilitatea de a explica un model probabilistic;▪ Capacitatea de a elabora un model probabilistic pentru o problemă reală de prelucrarea limbajului natural;▪ Minim 50% prezență la laborator;▪ Minim 15 puncte din temele de la laborator;▪ Minim 15 puncte suma punctajelor testelor de la curs▪ Minim 50 puncte în total: (prezențe + teste + teme + bonus + proiect) Notele se stabilesc conform cu criteriile ECTS			

Data completării
28 septembrie 2018

Titular de curs
Conf.Dr. Diana Trandabăț

Titular de seminar
conf.Dr. Diana Trandabăț

Data avizării în departament

Director de departament
Prof.Dr. Dorel LUCANU