

FIȘA DISCIPLINEI
Anul universitar 2017 – 2018

Decan,
Prof. dr. ing. Corneliu LAZĂR

1. Date despre program

1.1 Instituția de învățământ superior	Universitatea Tehnică „Gheorghe Asachi” din Iași
1.2 Facultatea	Automatică și Calculatoare
1.3 Departamentul	Calculatoare
1.4 Domeniul de studii	Calculatoare și tehnologia informației
1.5 Ciclul de studii ¹	Licență
1.6 Programul de studii	Tehnologia informației

2. Date despre disciplină

2.1 Denumirea disciplinei	Regăsirea informațiilor pe WEB						
2.2 Titularul activităților de curs	ș.l. dr. ing. Alexandru ARCHIP						
2.3 Titularul activităților de aplicații	ș.l. dr. ing. Alexandru ARCHIP						
2.4 Anul de studii ²	4	2.5 Semestrul ³	8	2.6 Tipul de evaluare ⁴	examen	2.7 Tipul disciplinei ⁵	DS

3. Timpul total estimat al activităților zilnice (ore pe semestru)

3.1 Număr de ore pe săptămână	4	din care 3.2 curs	2	3.3a sem.	-	3.3b laborator	2	3.3c proiect	-
3.4 Total ore din planul de învățământ ⁶	56	din care 3.5 curs	28	3.6a sem.	-	3.6b laborator	28	3.6c proiect	-
Distribuția fondului de timp ⁷									Nr. ore
Studiul după manual, suport de curs, bibliografie și notițe									20
Documentare suplimentară în bibliotecă, pe platformele electronice de specialitate și pe teren									20
Pregătire seminarii/laboratoare/proiecte, teme, referate și portofolii									12
Tutoriat ⁸									7
Examinări ⁹									5
Alte activități:									-
3.7 Total ore studiu individual ¹⁰	64								
3.8 Total ore pe semestru ¹¹	120								
3.9 Numărul de credite	5								

4. Precondiții (acolo unde este cazul)

4.1 de curriculum ¹²	<ul style="list-style-type: none"> Structuri de date – CTI.DD.201 Proiectarea algoritmilor – CTI.DD.209 Tehnologii internet – TI.DIS.306 Rețele de calculatoare – TI.DIS.307 Programarea clientului de web – TI.DIS.416
4.2 de competențe	<ul style="list-style-type: none"> Cunoștințe solide de structuri de date și algoritmi Abilitatea de a analiza, proiecta și implementa algoritmi eficienți Cunoștințe despre protocoalele de comunicații uzuale (TCP/IP, HTTP, DNS) și funcționarea rețelelor de calculatoare Abilitatea de dezvoltare o aplicație capabilă să comunice cu sisteme remote pe baza unui protocol de comunicație impus (client HTTP, client DNS) Dezvoltarea aplicațiilor utilizând un limbaj de nivel înalt (C++, Java, C#) și cunoștințe solide de OOP Abilitatea de a utiliza clase specializate (colecții, containere, clase generice) în dezvoltarea de aplicații

5. Condiții (acolo unde este cazul)

5.1 de desfășurare a cursului ¹³	<ul style="list-style-type: none"> Sală de curs dotată cu video-proiector și acces internet
5.2 de desfășurare a seminarului / laboratorului / proiectului ¹⁴	<ul style="list-style-type: none"> Sală de laborator cu sisteme de calcul și acces internet JDK minim 1.7, sau C/C++ Mediu IDE preferat: Eclipse Sistem de operare preferat: distribuție Linux

6. Competențele specifice acumulate¹⁵

		Număr de credite alocate disciplinei ¹⁶ :	5	Repartizare credite pe competențe ¹⁷
Competențe profesionale	CP1	Operarea cu fundamente științifice, ingineresti și ale informaticii		1
	CP2	Proiectarea componentelor hardware, software și de comunicații		1
	CP3	Soluționarea problemelor folosind instrumentele științei și ingineriei calculatoarelor		1
	CP4	Proiectarea și integrarea sistemelor informatice utilizând tehnologii și medii de programare		0.5
	CP5	Întreținerea și exploatarea sistemelor hardware, software și de comunicații		0.5
	CP6	Utilizarea sistemelor inteligente		0.5
	CPS1	-		-
	CPS2	-		-
Competențe transversale	CT1	Comportarea onorabilă, responsabilă, etică, în spiritul legii pentru a asigura reputația profesiei		0.1
	CT2	Identificarea, descrierea și derularea proceselor din managementul proiectelor, cu preluarea diferitelor roluri în echipă și descrierea clară și concisă, verbal și în scris, în limba română și într-o limbă de circulație internațională, a rezultatelor din domeniul de activitate		0.2
	CT3	Demonstrarea spiritului de inițiativă și acțiune pentru actualizarea cunoștințelor profesionale, economice și de cultură organizațională		0.2
	CTS	-		-

7. Obiectivele disciplinei (reieșind din grila competențelor specifice acumulate)

7.1 Obiectivul general al disciplinei	<ul style="list-style-type: none"> • Obținerea de cunoștințe și abilități pentru a proiecta și implementa un sistem de regăsire de informații în cadrul unei colecții de date
7.2 Obiective specifice	<ul style="list-style-type: none"> • recapitularea și consolidarea noțiunilor fundamentale legate de World Wide Web; • recapitularea și consolidare cunoștințelor legate de tehnologiile de bază: protocolul HTTP, sistemul DNS, limbajul HTML; • prezentarea noțiunilor fundamentale legate de arhitectura motoarelor de căutare; • arhitecturi și modele generale pentru roboți web, prezentarea REP (Robot Exclusion Protocol); • prezentarea și detalierea politicilor de revizitare și de paralelizare pentru roboții web; • definirea conceptului de data mining și prezentarea sumară a unor algoritmi fundamentali; • definirea și fundamentarea conceptului de web mining, precum și prezentarea în detaliu a unor algoritmi fundamentali pentru web mining.

8. Conținuturi

8.1 Curs ¹⁸	Metode de predare ¹⁹	Observații
<p>1. Noțiuni recapitulative – 2 ore</p> <ul style="list-style-type: none"> • rețele de calculatoare, modele de referință (OSI, TCP/IP) • protocol de comunicație (definiții și concepte specifice) • stiva de protocole TCP/IP și protocole fundamentale: IP, TCP, UDP • protocole specifice: HTTP, DNS • limbajul HTML <p>2. Noțiuni introductive – 2 ore</p> <ul style="list-style-type: none"> • dezvoltarea World Wide Web • definirea celor 3 module componente fundamentale pentru sistemele de regăsire de informații: colectarea documentelor, pregătirea acestor pentru căutare și realizarea căutării efective • modele și arhitecturi fundamentale pentru motoarele de căutare pe WEB <p>3. Indexarea colecțiilor de documente – 8 ore</p> <ul style="list-style-type: none"> • tipuri de indexare ce vizează colecții de documente text <ul style="list-style-type: none"> o index direct/invers, index calitativ/cantitativ/pozițional • modele de reprezentare a documentelor și metrice specifice: distanța bazată pe funcția cosinus • algoritmi specifici: BSBI, SPIMI • metode și modele de paralelizare <ul style="list-style-type: none"> o modelul MapReduce și aplicații <p>4. Funcții și metode de căutare a colecțiilor indexate – 2 ore</p> <ul style="list-style-type: none"> • modele specifice: modelul boolean, modelul vectorial 	Expunere, prelegere, prezentare la tablă a problematicii studiate, utilizare videoprojector, discuții cu studenții	

<ul style="list-style-type: none"> • forme de căutare: booleană, vectorială • criterii de stabilire a relevanței rezultatelor relativ la interogările utilizatorilor <p>5. Roboți WEB – 8 ore</p> <ul style="list-style-type: none"> • principiul general de funcționare al unui robot WEB și arhitectură • prezentarea caracteristicilor esențiale pentru roboți WEB • prezentarea caracteristicilor recomandate pentru roboți WEB • protocoale specifice: REP – Robots Exclusion Protocol, Sitemap • pseudocod general și modalități de explorare a WEB-ului • definirea politicilor de revizitare a documentelor WEB • metode și modele de paralelizare/distribuire <p>6. Tehnici de analiză specifice Data mining – 2 ore</p> <ul style="list-style-type: none"> • definiții, scurt istoric • concepte generale • metode de analiză • exemple de aplicații practice <p>7. Web mining – 4 ore</p> <ul style="list-style-type: none"> • tipuri de date analizate (de conținut, de structură, de utilizare) și exemple practice de aplicare • algoritmi specifici: <ul style="list-style-type: none"> o algoritmi PageRank și HITS o determinarea tiparelor frecvente și a regulilor de asociere (Apriori) o partiționarea k-medii (k-Means Clustering) 		
<p>Bibliografie curs:</p> <ol style="list-style-type: none"> 1. Mitică Craus et al. (2005), Regăsirea informațiilor pe WEB, Editura Politehniun, Iași 2. Florin Leon (2008), Explorarea datelor Web, Editura Tehnopress, Iași 3. Two Crows Corporation (1999), Introduction to Data Mining and Knowledge Discovery, Third Edition, Two Crows Corporation 4. JeanMarc Adamo (2000), Data Mining for Association Rules and Sequential Patterns, Ed. SpringerVerlag, New York 5. Jiawei Han and Micheline Kamber, Data Mining – Concepts and Techniques, Second Edition, The Morgan Kaufmann Series in Data Management Systems, Ed. Morgan Kaufmann, 2006 <p>Alte resurse pentru curs și pentru laborator:</p> <ol style="list-style-type: none"> 1. http://www.google.com/insidesearch/howsearchworks/thestory/ 2. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze (2008), Introduction to Information Retrieval, Cambridge University Press (http://nlp.stanford.edu/IR-book/) 3. http://www.robotstxt.org/ 4. Raymond J. Mooney (2008), Information Retrieval and Web Search – note de curs (http://www.cs.utexas.edu/~mooney/ir-course/) 		
8.2a Seminar	Metode de predare ²⁰	Observații
-	-	-
8.2b Laborator	Metode de predare ²¹	Observații
<ol style="list-style-type: none"> 1. Recapitulare – socket-uri – 2 ore 2. Aplicație pentru parsarea documentelor HTML – 4 ore 3. Aplicații pentru indexarea colecțiilor de documente text – 4 ore 4. Aplicații pentru realizarea funcțiilor de căutare – 2 ore 5. Aplicații client pentru protocoale specifice – DNS – 2 ore 6. Aplicații client pentru protocoale specifice – HTTP – 2 ore 7. Aplicații de tip „robot Web” („Web Crawler”) – 4 ore 8. Analiza relevanței – algoritmul PageRank – 4 ore 9. Clusterizarea documentelor text – 4 ore 	Discuții, dezbateri, analiză și rezolvare de probleme	
8.2c Proiect	Metode de predare ²²	Observații
-	-	-
<p>Bibliografie aplicații (seminar / laborator / proiect):</p> <ol style="list-style-type: none"> 1. Mitică Craus et al. (2005), Regăsirea informațiilor pe WEB, Editura Politehniun, Iași 2. Florin Leon (2008), Explorarea datelor Web, Editura Tehnopress, Iași 3. Two Crows Corporation (1999), Introduction to Data Mining and Knowledge Discovery, Third Edition, Two Crows Corporation 4. JeanMarc Adamo (2000), Data Mining for Association Rules and Sequential Patterns, Ed. SpringerVerlag, New York 5. Jiawei Han and Micheline Kamber, Data Mining – Concepts and Techniques, Second Edition, The Morgan Kaufmann Series in Data Management Systems, Ed. Morgan Kaufmann, 2006 <p>Alte resurse pentru curs și pentru laborator:</p> <ol style="list-style-type: none"> 1. http://www.google.com/insidesearch/howsearchworks/thestory/ 2. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze (2008), Introduction to Information Retrieval, Cambridge University Press (http://nlp.stanford.edu/IR-book/) 3. http://www.robotstxt.org/ 		

9. Coroborarea conținuturilor disciplinei cu așteptările reprezentanților comunității epistemice, asociațiilor profesionale și angajatori reprezentativi din domeniul aferent programului²³

- Discipline similare se regăsesc în programele analitice ale unor universități din țară (Universitatea Politehnica din București, Facultatea de Automatică și Calculatoare, master anul II) și din străinătate (Stanford University, University of Stuttgart sau University of Munich).
- Conținutul disciplinei adresează probleme legate de aplicații și clienți Web, procesarea și indexarea documentelor text, realizarea operațiilor de căutare în cadrul unor colecții indexate și tehnici de analiză a datelor; aplicațiile practice ale acestor noțiuni se regăsesc în cadrul sistemelor de gestiune a conținutului Web (WEB CMS), sunt folosite în proiectarea aplicațiilor și site-urilor WEB sau își regăsesc utilitatea în sistemele de extragere de informații (ex.: analiza tiparelor de achiziție a produselor unui magazin virtual).
- Competențele dobândite vizează, de asemenea, dezvoltarea capacității de a proiecta și implementa sisteme de procesare a datelor de mari dimensiuni; aceste competențe sunt solicitate de către majoritatea angajatorilor din piață.

10. Evaluare

Tip activitate	10.1 Criterii de evaluare	10.2 Metode de evaluare	10.3 Pondere din nota finală
10.4 Curs	<ul style="list-style-type: none"> • Cunoștințe teoretice și practice însușite: protocoale specifice, modele și arhitecturi; analiza datelor; algoritmi și arhitecturi; modele de căutare; 	Teste pe parcurs ²⁴ : -	-
		Teme de casă: 4 teme de casă	20 %
		Evaluare finală – examen: <ul style="list-style-type: none"> - modul indexare (pondere 33% din media de examen): probă tip grilă deschisă cu justificarea răspunsurilor selectate; problemă indexare; - modul căutare și web mining (pondere 33% din media de examen): probă de tip grilă deschisă cu justificarea răspunsurilor selectate; problemă căutare; - modul robot web (pondere 33% din media de examen): probă tip grilă deschisă cu justificarea răspunsurilor selectate; 	50 % (minim 5)
10.5a Seminar	-	-	-
10.5b Laborator	<ul style="list-style-type: none"> • Cunoașterea protocoalelor specifice și a modalităților de implementare; modele și arhitecturi; algoritmi specifici: analiză, implementare și testare; 	<ul style="list-style-type: none"> • lucrări experimentale, analiza rezultatelor obținute 	30 % (minim 5)
10.5c Proiect	-	-	-
10.5d Alte activități ²⁵	-	-	-
10.6 Standard minim de performanță ²⁶			
<ul style="list-style-type: none"> • abilitatea de a implementa un client simplist pentru un protocol existent (DNS, HTTP) • abilitatea de a implementa un robot web simplu • abilitatea de a implementa o aplicație care dezvoltă un index invers calitativ • abilitatea de a implementa metode de cautare booleană de bază (funcțiile logice OR, AND, NOT) 			

Data completării,

15.09.2014

Semnătura titularului de curs,

Ș.I. dr. ing. Alexandru ARCHIP

Semnătura titularului de aplicații,

Ș.I. dr. ing. Alexandru ARCHIP

Data avizării în departament,

25.09.2014

Director departament,

Prof. dr. ing. Petru CAȘCAVAL

¹ Licență / Master

² 1-4 pentru licență, 1-2 pentru master

³ 1-8 pentru licență, 1-3 pentru master

⁴ Examen, colocviu sau VP A/R – din planul de învățământ

⁵ DF - disciplină fundamentală, DID - disciplină în domeniu, DS – disciplină de specialitate sau DC - disciplină complementară - din planul de învățământ

⁶ Este egal cu 14 săptămâni x numărul de ore de la punctul 3.1 (similar pentru 3.5, 3.6abc)

⁷ Liniile de mai jos se referă la studiul individual; totalul se completează la punctul 3.7.

⁸ Între 7 și 14 ore

⁹ Între 2 și 6 ore

¹⁰ Suma valorilor de pe liniile anterioare, care se referă la studiul individual.

¹¹ Suma dintre numărul de ore de activitate didactică directă (3.4) și numărul de ore de studiu individual (3.7); trebuie să fie egală cu numărul de credite alocate disciplinei (punctul 3.9) x 24 de ore pe credit.

¹² Se menționează disciplinele obligatorii a fi promovate anterior sau echivalente

¹³ Tablă, videoproiector, flipchart, materiale didactice specifice etc.

¹⁴ Tehnică de calcul, pachete software, standuri experimentale, etc.

¹⁵ Competențele din Grilele G1 și G1bis ale programului de studii, adaptate la specificul disciplinei, pentru care se repartizează credite (www.rncis.ro sau site-ul facultății)

¹⁶ Din planul de învățământ

¹⁷ Creditele alocate disciplinei se distribuie pe competențe profesionale și transversale în funcție de specificul disciplinei

¹⁸ Titluri de capitole și paragrafe

¹⁹ Expunere, prelegere, prezentare la tablă a problematicii studiate, utilizare videoproiector, discuții cu studenții (pentru fiecare capitol, dacă este cazul)

²⁰ Discuții, dezbateri, prezentare și/sau analiză de lucrări, rezolvare de exerciții și probleme

²¹ Demonstrație practică, exercițiu, experiment

²² Studiu de caz, demonstrație, exercițiu, analiza erorilor etc.

²³ Legătura cu alte discipline, utilitatea disciplinei pe piața muncii

²⁴ Se va preciza numărul de teste și săptămânile în care vor fi susținute.

²⁵ Cercuri științifice, concursuri profesionale etc.

²⁶ Se particularizează la specificul disciplinei standardul minim de performanță din grila de competențe a programului de studii.