

Curs 2011-2012

Tecnologies de la Informació (12459)

Titulació / Estudis: Enginyeria en Informàtica i Enginyeria Tècnica en Informàtica de Sistemes**Curs:** 2n**Període:** 2n trimestre**Nombre de crèdits ECTS:** 3.6 crèdits**Nombre total d'hores de dedicació a l'assignatura:** 90**Llengua de docència:** català i castellà

1. Presentació de l'assignatura

Tecnologies de la Informació és una assignatura del 5è trimestre d'informàtica. Es fa quan els alumnes encara estan assumint les estructures de dades dels primers trimestres i les bases de dades del 4rt trimestre. L'objectiu principal és introduir l'alumne en els sistemes de recuperació d'informació, on a diferència del que ha vist fins el moment no hi ha una solució única i exacte sinó varies d'aproximades. S'introdueix doncs a l'alumne el tractament de la incertesa i de les solucions no úniques lligades al tractament del llenguatge natural i les percepcions de l'usuari.

L'assignatura es presenta als alumnes amb un objectiu finalista: fer un bon cercador de text o multimèdia. Aquest és el fil conductor que justifica l'ordre i la manera com es presenten els temes.

La recuperació de la informació tracta grans volums de dades i per això es tracten temes de complexitat computacional (rendiment, espai) de manera molt pràctica. Degut a que gran part de l'assignatura es fa sobre el tractament de texts, també s'avaluen mètodes de tractament i cerca de text. I s'introdueixen les mètriques com a mètode per mesurar la similitud entre el que demana l'usuari i els possibles resultats.

És una assignatura que introdueix conceptes nous que s'apliquen tant en les pràctiques com en els exercicis.

2. Prerequisits per al seguiment de l'itinerari formatiu

És imprescindible tenir assumits els coneixements de programació de primer i les estructures de dades.

3. Competències a assolir en l'assignatura

Competències generals

Instrumentals

- 1.Capacitat d'anàlisi i síntesi
- 2.Resolució de problemes
- 3.Capacitat de treballar amb informació no estructurada
- 4.Organització del temps i planificació

Interpersonals

5. Treball en grup

Sistemàtiques

- 6.Capacitat per aplicar el coneixement teòric a la pràctica.

Competències Específiques

- 1.Comprensió del funcionament d'un cercador.
- 2.Creació i ús d'índexs invertits
- 3.Eines d'anàlisi de text i documents
- 4.Cerca de text
- 5.Ús de mètriques
- 6.Tractament de documents XML

4. Objectius d'aprenentatge

Mitjançant l'estudi i la reflexió sobre els cercadors web que els alumnes ja coneixen, es pretén que els alumnes s'introdueixin en el tractament de grans volums de dades no estructurades i a treballar amb incertesa.

En els primers cursos han vist algorismes on el resultat correcte era únic i com estructurar les dades per aconseguir simplificar-ne el tractament. En aquest curs les dades són en llenguatge natural, les preguntes dels usuaris son imprecises, i el resultat a obtenir no és únic. L'objectiu és doncs que entenguin les diferents maneres d'abordar aquest problema.

Per altra banda, el fet de que normalment els cercadors treballin amb grans volums de dades també obliga al disseny d'estructures de dades i algorismes que permetin suportar-ho. Per accelerar els algorismes i reduir els índexs s'estudia com simplificar els algorismes. Es pretén que els alumnes discernixin entre complexitat i qualitat del resultat i en valorin els avantatges i inconvenients en cada cas. Per això necessiten desenvolupar un cert esperit crític.

5. Avaluació

5.1 Criteris generals d'avaluació

Pràctiques: avaluació continuada, per parelles, obligatòria. Avaluació individual obligatòria a l'examen final.

Seminaris: avaluació individual al seminari, exercicis optatius autoavaluats i exercicis obligatoris avaluats.

Presentació: avaluació treball per parelles i presentació individual, obligatoris.

Teoria: avaluació individual obligatòria a l'examen final.

Ponderació:

Teoria: 40%

Practica: 40%

Seminari: 10%

Presentació: 10%

$$NF = 0.4 \times NT + 0.4 \times NP + 0.1 \times NS + 0.1 \times NPr$$

NF = Nota final

NT = Nota teoria

NP = Nota pràctiques

NS = Nota seminaris

NPr = Nota presentació

Es requereix un mínim de 4,5 en cada una de les parts per poder fer el càlcul de la nota final. Es donarà l'opció de pujar nota (5%) als estudiants que participin en un estudi d'investigació i presentin un petit informe sobre els resultats obtinguts.

5.2. Concreció per competències

Competències a assolir en l'assignatura

Indicador d'assoliment

Procediment d'avaluació

Temporalització

Competències generals	Respondre amb encert les preguntes de l'examen	Examen	Final de trimestre.
1.Capacitat d'anàlisi i síntesi			
2.Resolució de problemes	Resoldre correctament els exercicis proposats en el temps requerit	Exercicis Pràctiques Treball + Presentació	Exercicis a resoldre a les sessions de seminari. Pràctiques a fer durant el curs

- 3.Capacitat de treballar amb informació no estructurada
- 4.Organització del temps i planificació
5. Treball en grup
- 6.Capacitat per aplicar el coneixement teòric a la pràctica.

Realització de les pràctiques

Preparació i realització d'una presentació a classe

Treball a realitzar durant el curs i presentació a classe en una de les últimes sessions de teoria.

6. Continguts

6.1. Blocs de contingut

- Pre-processament i emmagatzematge d'informació textual
- Models per a la recuperació d'informació textual
- Avaluació i retroalimentació en la recuperació d'informació textual
- Especialitzacions i usos de sistemes de recuperació d'informació

6.2. Organització i concreció dels continguts

Bloc de contingut 1. - Pre-processament i emmagatzematge d'informació textual

Conceptes	Procediments
1.Processament de text (stopwords, stemming, diccionaris)	1. Descripció teòrica dels conceptes
2.Índex invertit	2. Exercicis autoavaluats i seminari
3.Arbres i vectors de sufixes	3.Practica de programació de pre-processament.
	4.Pràctica de programació d'un índex invertit

Bloc de contingut 2. - Models per a la recuperació d'informació textual

Conceptes	Procediments
1. Model Booleà	1. Descripció teòrica dels conceptes
2. Model Vectorial	2. Exercicis autoavaluats i seminari
3.Model Probabilístic	3. Pràctica de programació del model Booleà i el Page Rank.
4.PageRank i HITS	4. Exercicis dels càlculs necessaris pel model Vectorial
5.Crawling	
6.Popularitat	

Bloc de contingut 3. - Avaluació i retroalimentació en la recuperació d'informació textual

Conceptes	Procediments
1.Precisió i recall	1. Descripció teòrica dels conceptes
2.Retroalimentació o feedback	2. Exercicis autoavaluats i seminari
3.Algorisme Rocchio	3. Exercicis dels càlculs de les mesures de precisió i recall per avaluar un sistema.
4.Expansió de la consulta	

Bloc de contingut 4. - Especialitzacions i usos de sistemes de recuperació de l'informació

Conceptes	Procediments
1.Classificació	1. Descripció teòrica mitjançant presentacions dels estudiants.
2.Browsing	2. Clarificació de conceptes al seminari
3.Recomenació	
4.Personalització	
5.Contextualització	
6.Multimedia IR	

7. Metodologia

A les classes de teoria es presentaran els conceptes teórico-pràctics fonamentals de l'assignatura. L'estudiant haurà de complementar aquesta activitat amb una lectura detinguda dels seus propis apunts i el material addicional que el professor hagi proporcionat. El treball addicional requerit fora de l'aula per part de l'estudiant serà aproximadament d'una hora per cada sessió de (2h) de teoria.

Les dues últimes sessions de teoria es dedicaran a presentacions fetes pels estudiants. Cada estudiant haurà de preparar individualment o per parelles un dels temes teòrics proposats pel professor i exposar-l'ho a classe. L'objectiu és que els estudiants descobreixin i utilitzin altres fonts d'informació, que desenvolupin la seva capacitat d'entendre i sintetitzar aquesta informació i que reforcin les seves habilitats de comunicació al transmetre el que han après.

A les classes de seminari, l'estudiant posarà en pràctica els conceptes i tècniques presentades a la sessió de teoria, mitjançant la resolució de problemes. El professor actuarà com a moderador i resoldrà els dubtes que hagin sorgit. L'objectiu és consolidar els fonaments per tal que posteriorment l'estudiant pugui resoldre problemes de major complexitat. El treball addicional requerit fora de l'aula per part de l'estudiant serà aproximadament de dues hores per cada sessió de (2h) de seminari.

A les classes de pràctiques, es programaran les tècniques presentades a les classes de teoria. L'objectiu de les classes de pràctiques és que els estudiants construeixen en grups de dues o tres persones un sistema de recuperació de la informació textual. Per a un millor assoliment dels objectius d'aprenentatge, els conceptes s'aniran aplicant gradualment. Es començarà amb una aplicació sencilla que s'anirà millorant i complicant gradualment fins a obtenir el cercador final. Les practiques inacabades caldrà entregar-les al final de la sessió, i després de discutir-les i corregir-les amb el professor haurà d'acabar-les correctament. Això donarà la oportunitat a l'estudiant de corregir els seus errors i defectes de disseny abans de continuar amb la pràctica. El treball addicional requerit fora de l'aula per part de l'estudiant variarà depenent del temps requerit per corregir els errors.

8. Fonts d'informació i recursos didàctics

9.1. Fonts d'informació per a l'aprenentatge. Bibliografia bàsica (suport paper i electrònic)

- Modern Information Retrieval, de R. Baeza-Yates i B. Riberio-Nieto (Nova York: Addison-Wesley, 1999).
- Transparències utilitzades durant les classes teòriques.

9.2. Fonts d'informació per a l'aprenentatge. Bibliografia complementària (suport paper i electrònic)

- Managing Gigabytes, (2a. ed.), de I. Witten, A. Moffat i T. Bell (Nova York: Morgan Kaufmann Publishers, 1999).

Recursos electronics en anglès:

Introduction to Information Retrieval

<http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html>

Recursos electronics en castellà:

Introducción:

- <http://www.tejedoresdelweb.com/307/article-1056.html>
- <http://www.tejedoresdelweb.com/307/propertyvalue-23553.html>

Motores de búsqueda (obtenido en wikipedia):

- http://www.buscarportal.com/articulos/motores_busqueda.php
- <http://microasist.com.mx/noticias/tp/ksotp2207.shtml>
- <http://www.monografias.com/trabajos/buscadores/buscadores.shtml>