

	<b>ENSENYAMENT d' ENGINYERIA INFORMÀTICA</b>		
	<b>ASSIGNATURA: Anàlisi de la Web: mètodes matemàtics i algorítmics</b>		
	<b>PROFESSORS: Josep Conde Colom i Joan Gimbert Quintilla</b>		
	<b>CURS: 2n</b>	<b>CRÈDITS: 6</b>	<b>TIPUS: Optativa</b>

## 1. OBJECTIUS

Els objectius específics de l'assignatura són:

- Conèixer l'arquitectura d'un cercador (ex. *Google*).
- Comprendre el funcionament dels principals algorismes d'ordenació de pàgines web (ex. *PageRank*).
- Conèixer les tècniques estadístiques que hi ha darrere dels procediments actuals per fer recomanacions (ex. *Amazon*).
- Aplicar mètodes i algorismes de l'Anàlisi Cluster per descobrir grups (ex. Xarxes Socials).
- Conèixer l'ús de la teoria de probabilitats en el filtratge de documents (ex. filtres de *spams*).

## 2. ESTRUCTURA

L'assignatura, que s'imparteix durant el segon quadrimestre del segon curs, consta de 6 crèdits, dels quals 4.5 són de docència presencial. Hi haurà classes teòriques, sessions pràctiques i exposicions de treballs.

## 3. PROGRAMA

### 1. Fonaments i tècniques per al disseny d'un cercador.

- La cerca d'informació en la Web.
  - Modelització de la Web.
- Arquitectura d'un cercador.
  - Components d'un cercador.
- Algorismes d'ordenació de pàgines web.
  - Algorisme *PageRank* emprat per *Google*.
  - Algorisme *HITS*.

### 2. Tècniques estadístiques per fer recomanacions.

- Recollida i gestió de les preferències.

- Mesures de similaritat per cercar usuaris amb preferències paregudes.
  - Distància euclídea.
  - Índex de correlació de Pearson.
- Aplicació a la recomanació de productes (ex. *Amazon*).

### 3. Ús de l'anàlisi clúster per descobrir grups.

- Preparació de les dades.
- Tècniques de l'anàlisi clúster.
  - Classificació piramidal.
  - L'algoritme de les  $k$ -mitjanes.
- Representació de dades multivariants en dos dimensions.
  - L'algoritme d'escalatge multidimensional.
- Aplicació a l'estudi de xarxes socials.

### 4. Mètodes probabilístics per filtrar documents.

- Anàlisi dels documents.
- Tècniques probabilístiques de filtratge.
  - El mètode bayesià.
  - El mètode de Fisher.
- Aplicació als filtres de *spam*.

### 5. Optimització estocàstica per a la resolució de problemes col.laboratius.

- Modelització: representació de les solucions i funció de cost.
- Heurístiques.
  - *Hill-climbing*.
  - *Simulated annealing*.
  - Algorismes genètics.
- Aplicació a la visualització d'una xarxa.

## 4. MATERIALS DE L'ASSIGNATURA I PROGRAMARI

Per a les classes teòriques farem servir, com a guió, transparències. Així mateix, ens recolzarem en la lectura i anàlisi d'articles divulgatius i d'altres de caràcter científic-tècnic. Tot aquest material el penjarem al campus virtual.

## 5. BIBLIOGRAFIA

- A. Arasu, J. Cho, H. Garcia-Molina, A. Paepcke and S. Raghavan, *Searching the Web* (manuscript), ACM Transactions on Internet Technology, 1 (1) (2001), 2-43.
- R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, Addison Wesley, 1999.
- M. Berry and M. Browne, *Understanding Search Engines. Mathematical Modeling and Text Retrieval*, SIAM, 1999.
- S. Brin and L. Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Computer Networks and ISDN Systems, 30 (1-7) (1998), 107-117.

- P. Fernández, *El secreto de Google y el Álgebra lineal*, Boletín de la Sociedad Española de Matemática Aplicada, **30** (2004), 114–141.
- J. Kleinberg, *Authoritative sources in a hyperlinked environment*, Journal of the ACM, **46** (5) (1999), 604–632.
- L. Page, S. Brin, R. Montwani and T. Winograd, *The PageRank Citation Ranking: Bringing Order to the Web*, Stanford Digital Library Technologies Project (1998).
- T. Segaran, *Programming Collective Intelligence*, O'Reilly, 2007.

## 6. AVALUACIÓ

Hi haurà dues modalitats d'avaluació:

### *I. Avaluació continuada:*

- Realització d'una activitat (10%).
- Realització d'una pràctica (35%).
- Realització i exposició d'un treball (35%).
- Prova de validació de la pràctica i dels continguts bàsics del programa (20%).

Es tindrà en compte la participació en les classes.

### *II. Avaluació no continuada:*

- Realització d'una pràctica de síntesi (50%).
- Realització d'un examen sobre els continguts de tot el programa (50%).