

Processament Automàtic de Textos (21542)

Titulació/estudi: Grau en Llengües Aplicades

Curs: quart

Trimestre: primer

Nombre de crèdits ECTS: 4

Hores de dedicació de l'estudiant: 100

Tipus d'assignatura: Optativa

Professor/s GG: Toni Badia i Carme Colominas

Llengua de docència: Català

1. Presentació de l'assignatura

En aquesta assignatura s'aprendrà a utilitzar elements simples de programació per preparar un text de cara al seu processament automàtic. Els estudiants adquiriran la capacitat de processar textos per poder-ne extreure fragments, freqüències i grups de paraules.

2. Competències que s'han d'assolir

- Coneixements d'informàtica (G.6): utilitzar rudiments de programació (dades, variables, estructures de control, creació i ús de funcions)
- Aplicació de coneixements a la pràctica (G.17): confegir un petit programa que faci el que interessa
- Pràctica en la utilització d'eines informàtiques aplicades a la mediació lingüística (E.11): accedir a textos, segmentar-los, indexar-los, calcular-hi freqüències, usar-hi expressions regulars

3. Continguts

- Representació dels textos: caràcters, cadenes i llistes
- Segmentació de paraules i frases
- Indexació i variables
- Càlcul de freqüències
- Expressions regulars
- Estructures de control.

4. Avaluació i recuperació

Avaluació		Recuperació			
Activitat d'avaluació	Ponderació sobre la nota final	Recuperable/ No recuperable	Ponderació sobre la nota final	Forma de recuperació	Requisits i observacions
Pràctiques setmanals de programació (una per tema)	30%	Recuperable	30%	Revisió de pràctiques, completant o corregint les fetes	
Examen	70%	Recuperable	70%	Examen	

5. Metodologia: activitats formatives

Per a l'estudiant, aquesta assignatura suposa 7/8 hores de treball setmanal. D'aquestes 7/8 hores, 2:30 h seran presencials i la resta (4:30 h) seran de treball autònom (preparació de les sessions teòriques a través de lectures, finalització de les pràctiques començades a l'hora de pràctiques sota el guiatge del professor i realització d'una traducció cooperativa d'una pàgina web). Els materials utilitzats en les classes es penjaran al Moodle de l'assignatura.

6. Bibliografia bàsica de l'assignatura

Bird, Steven, Klein, Ewan & Loper, Edward (2009) *Natural Language Processing with Python*, O'Reilly

<https://sites.google.com/site/naturallanguagetoolkit/book>

Gauld, Alan (2011) *Learning to program*

<http://www.alan-g.me.uk/tutor/index.htm>

Jeffrey Elkner, Allen B. Downey and Chris Meyers. C (2010) *Learning with Python* 2nd Edition (Using Python 2.x)

<http://openbookproject.net/thinkcs/python/english2e/>

Jurafsky, Daniel, and James H. Martin. 2009. [Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics](#). 2nd edition. Prentice-Hall.

Perkins, Jacob (2010) *Python Text Processing with NLTK 2.0 Cookbook*