

Marcatge de Textos (21543)

Titulació/estudi: Grau en Llengües Aplicades

Curs: quart

Trimestre: segon

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 anotar un text de cara al seu processament automàtic. Els estudiants adquiriran la capacitat de processar i anotar textos per poder-ne extreure fragments, freqüències i grups de paraules de diferents categories.

2. Competències que s'han d'assolir

- Coneixements d'informàtica (G.6):

- Consultar i extreure informació de corpus lematitzats i marcats morfosintàcticament
- Marcar textos morfosintàcticament de forma automàtica
- Calcular la cobertura i precisió dels programes de marcatge

- Aplicació de coneixements a la pràctica (G.17): fer servir eines informàtiques per al processament de text

- Capacitat per poder reflexionar sobre el funcionament de la llengua (E.9): extreure conclusions lingüístiques a partir de les operacions de processament de textos i parla

3. Continguts

- Identificació de frases
- Etiquetatge morfosintàctic
- Mètodes d'anotació: per regles
- Mètodes d'anotació: estadístics
- Cobertura i precisió
- Elements bàsics de programació: manipulació de dades, funcions, recursivitat etc.

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 i 1/2 seran presencials i la resta (5-6) 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). 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*