

Marcado de textos (21543)

Titulación/estudio: Grado en Lenguas Aplicadas

Curso: cuarto

Trimestre: segundo

Número de créditos ECTS: 4

Horas de dedicación del estudiante: 100

Tipo de asignatura: optativa

Profesor/es GG: Toni Badia

Lengua de docencia: Catalán

1. Presentación de la asignatura

En esta asignatura se aprenderá a utilizar elementos simples de programación para anotar un texto de cara a su procesamiento automático. Los estudiantes adquirirán la capacidad de procesar textos para poder extraer de ellos fragmentos, frecuencias y grupos de palabras de diferentes categorías.

2. Competencias que se deben alcanzar

G.6. Conocimientos de informática

- Consultar y extraer información de corpus lematizados y marcados morfosintácticamente
- Marcar textos morfosintácticamente de forma automática
- Calcular la cobertura y precisión de los programas de marcado

G.17. Aplicación de conocimientos a la práctica

- Utilizar herramientas informáticas para el procesamiento de texto

E.9 Capacidad de reflexión sobre el funcionamiento de la lengua

- Extraer conclusiones lingüísticas a partir de las operaciones de procesamiento de textos y habla

3. Contenidos

- identificación de frases
- etiquetado morfosintáctico
- métodos de anotación: por reglas

- métodos de anotación: estadísticos
- cobertura y precisión
- elementos básicos de programación: manipulación de datos, funciones, recursividad, etc.

4. Evaluación y recuperación

Evaluación		Recuperación			
Actividad de evaluación	Ponderación sobre la nota final	Recuperable/ No recuperable	Ponderación sobre la nota final	Forma de recuperación	Requisitos y observaciones
Prácticas semanales de programación (una por tema)	30 %	sí	30 %	Revisión de prácticas, completando o corrigiendo las que ya se han hecho	
Examen	70 %	sí	70 %	Examen	

5. Metodología: actividades formativas

Para el estudiante, esta asignatura supone 7/8 horas de trabajo semanal. De estas 7/8 horas, 2 y 1/2 serán presenciales y el resto (5-6) serán de trabajo autónomo (preparación de las sesiones teóricas a través de lecturas, finalización de las prácticas comenzadas en la hora de prácticas bajo la guía del profesor). Los materiales utilizados en las clases se colgarán en el Moodle de la asignatura.

6. Bibliografía básica de la asignatura

Bird, Steven, Klein, Ewan & Loper, Edward (2009) *Natural Language Processing with Python*, O'Reilly
<http://www.nltk.org/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*