

ASIGNATURA : PROCESADORES DE LENGUAJE

CURSO: 4º OPCION:

Créditos: 9

Semestre: 1º

Castellano

Euskera

1.-Objetivos:

- Aprender y aplicar técnicas de representación digital del lenguaje natural escrito, es decir, estructurar, ordenar y relacionar documentos digitales (o corpus documentales) que contienen lenguaje natural escrito: Taxonomías y ontologías.
- Aprender y aplicar técnicas de procesamiento automático sobre una representación digital del lenguaje natural escrito o corpus documental: traducción automática, resúmenes automáticos, buscadores, vistas parciales de documentos o corpus, etc. En general, Natural Language Processing (NLP): Natural Language Generation (NLG) y Natural Language Understanding (NLU).

2.-Programa:

TEORICA

1. INTRODUCCIÓN: PROCESADORES DEL LENGUAJE Y SUS APLICACIONES EN INTERNET.
2. ETIQUETADO DE DOCUMENTOS EN INTERNET: HTML Y EDITORES.
3. NUEVAS TÉCNICAS DE ETIQUETADO DE DOCUMENTOS EN INTERNET: XML, DTD, XSL (CSS) Y EDITORES.
4. LA WEB SEMÁNTICA Y SUS ESTÁNDARES.
5. ESTRUCTURACIÓN DE DOCUMENTOS.
6. TAXONOMÍAS O RELACIONES ENTRE DOCUMENTOS.
7. ONTOLOGÍAS O SIGNIFICADO DEL ETIQUETADO.
8. INTRODUCCIÓN A NATURAL LANGUAGE PROCESSING (NLP): NATURAL LANGUAGE GENERATION (NLG) Y NATURAL LANGUAGE UNDERSTANDING (NLU): CORPUS, CORPUS MULTILÍNGÜES PARALELOS, ALINEAMIENTO DE DOCUMENTOS, GENERACIÓN DE RESÚMENES, VISTAS PARCIALES DE DOCUMENTOS,

PRACTICA

1. PROYECTO CONCRETO DE APLICACIÓN QUE SE DIVIDE EN LA SIGUIENTES PARTES:
2. ESTUDIO DE LA ESTRUCTURA INTERNA DE LOS DOCUMENTOS
3. ESTUDIO TAXONÓMICO
4. ESTUDIO ONTOLÓGICO
5. CREACIÓN DEL CORPUS
6. PROCESAMIENTO DEL HÁBEAS PARA LA GENERACIÓN DE VISTAS DE DOCUMENTOS

3.- Bibliografía:

- GOLDFARB, CHARLES F.; PRESCOD, PAUL. *Manual de XML..* Ed: PRENTICE HALL.
- JURAFSKY, DANIEL; MARTIN, JAMES H. *Speech and language processing.* Ed: PRENTICE HALL.
- REITER, EHUD; DALE, ROBERT. *Building natural language generation systems..* Ed: CAMBRIDGE UNIVERSITY PRESS.