

Curso Académico: (2022 / 2023)

Fecha de revisión: 13-05-2022

Departamento asignado a la asignatura: Departamento de Informática

Coordinador/a: MORATO LARA, JORGE LUIS

Tipo: Optativa Créditos ECTS : 6.0

Curso : 4 Cuatrimestre : 2

REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

- Ficheros y Bases de Datos (Obligatoria, 2º Curso, 2º Semestre)
- Programación Orientada a Objetos (Obligatoria, 1er Curso, 2º Semestre)

OBJETIVOS

El propósito de la asignatura es mejorar el conocimiento de varios aspectos:

1. Conocimiento de los modelos de recuperación
2. Conocimientos de las técnicas de procesamiento del lenguaje natural
3. Conocimiento de sistemas para formalizar, sintetizar y estructurar la información
4. Sistemas de seguimiento de trazas
5. Conocimiento una presentación de resultados óptima
6. Conocimiento de los sistemas de mejora en la recuperación y reutilización de conocimiento en la Web y en la Ingeniería del Software

Para lo cual se realizará:

1. Diseño de sistemas de recuperación
2. Diseño de analizadores del lenguaje natural
3. Utilización de minería de textos para mejorar la representación y ordenación de resultados

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

Se presentan los descriptores asociados a la asignatura:

Modelos de recuperación, Procesamiento del Lenguaje Natural, Análisis Semántico, metadatos, linked data, Recuperación de información, reutilización de conocimiento, minería de datos

La asignatura muestra los fundamentos para crear sistemas de recuperación de información. Esto incluye la utilización de sistemas de organización del conocimiento, procesamiento del lenguaje natural, algoritmos de posicionamiento y evaluación mediante métricas. La asignatura también incluye la creación de páginas Web posicionables y accesibles.

Contenido del curso, 3 unidades:

Unidad 1: Recuperación de información

- Tema 1. Fundamentos de la búsqueda en los distintos tipos de web: web semántica, web social, web de datos, dark y deep web, pregunta-respuesta y web comercial.
- Tema 2. Optimización en motores de búsqueda web (SEO/SEM)
- Tema 3. Modelos de recuperación
- Tema 4. Acceso, adquisición y limpieza de datos procedentes de la web semántica y bigdata
- Tema 5. Crawlers y scrapers

Unidad 2. Evaluación de la recuperación

- Tema 6. Métricas para evaluar la recuperación

Unidad 3. Técnicas de recuperación de información avanzadas

- Tema 7. Procesamiento del lenguaje (PLN)
- Tema 8. Técnicas de extracción de información
- Tema 9. Realimentación y expansión de consultas

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

Clases Teóricas: 1.6 ECTS. Tienen por objetivo alcanzar las competencias específicas cognitivas de la asignatura

Clases Prácticas: Desarrollan las competencias específicas instrumentales y la mayor parte de las transversales, como son la de trabajo en equipo, capacidad de aplicar los conocimientos a la práctica, de planificar y organizar y de análisis y síntesis. También tienen por objetivo desarrollar las capacidades específicas actitudinales. Consisten en el diseño y desarrollo de un sistema de recuperación de información en un entorno Web. Además deben elaborar una memoria con ejercicios propuestos en clase sobre algoritmos de posicionamiento, métricas de recuperación y tecnologías para recuperar información. Talleres y prácticas 0.2 ECTS, ejercicios individuales o en grupo 3 ECTS
Las tutorías para solventar dudas prácticas y teóricas serán 1 ECTS

Ejercicios y Examen: 0.2 ECTS. Tienen por objeto incidir y complementar en el desarrollo de las capacidades específicas cognitivas y procedimentales. Las clases prácticas y ejercicios se expondrán oralmente para su posterior discusión.

SISTEMA DE EVALUACIÓN

Los ejercicios y exámenes además de servir como actividad formativa tienen el doble objetivo de ser medida para el sistema de evaluación.

Los criterios de evaluación son:

- Ejercicios de clase: Calidad, organización y corrección de la memoria escrita.
- Trabajo en grupo: presentación oral, corrección de los juicios de relevancia, y precisión y fiabilidad de la aplicación
- Página Web: calidad del texto en inglés y castellano; criterios de posicionamiento aplicados; posición alcanzada en el motor Google; novedad y calidad de los contenidos

El sistema de evaluación incluye la valoración de las actividades académicas prácticas según la siguiente ponderación.

Examen: examen es 40%, nota mínima 4.0

Prácticas, incluida la defensa de las mismas: 60%

Peso porcentual del Examen Final: 40

Peso porcentual del resto de la evaluación: 60

BIBLIOGRAFÍA BÁSICA

- Aurelien Geron Hands-On Machine Learning with Scikit-Learn and TensorFlow, OReilly, 2017
- Baeza-Yates, Ricardo and Berthier Ribeiro-Neto Modern information retrieval, ACM Press, Addison-Wesley, 2011
- Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda Applied Text Analysis with Python, OReilly, 2018
- Verborgh, R., De Wilde, M., & Sawant, A. Using OpenRefine: The essential OpenRefine guide that takes you from data analysis and error fixing to linking your dataset to the web, Packt Publishing, 2013

BIBLIOGRAFÍA COMPLEMENTARIA

- Anne Ahola Ward The SEO battlefield: winning strategies for search marketing programs, OReilly, 2017
- Dale R Handbook of Natural Language Processing, Marcel Dekker, 2000
- Dean Allemang, James Hendler Semantic Web for the Working Ontologists: Effective Modeling in RDFS and OWL, Elsevier, 2011
- Gábor László Hajba Website Scraping with Python: Using BeautifulSoup, Google Books, 2018
- J. Urbano, M. Marrero, D. Martín y J. Morato Bringing Undergraduate Students Closer to a Real-World Information Retrieval Setting: Methodology and Resources, ACM SIGCSE ITICSE, 2011
- Moens Marie-Francine Information Extraction: algorithms and prospects in a retrieval context (Chps. 1, 2 & 4), Springer, 2006
- Morato, J, Sánchez-Cuadrado, S, Moreno, V, Moreira JA Evolución de los factores de posicionamiento web y adaptación de las herramientas de optimización, Revista española de Documentación Científica, Vol 36, No 3, 2013
- Nadeau D. and Sekine S. A survey of named entity recognition and classification, Linguisticae Investigationes vol. 30 n.1, 2007
- Stuart Russell, Peter Norvig Artificial Intelligence: A Modern Approach, Pearson, 2018

RECURSOS ELECTRÓNICOS BÁSICOS

- . How to Sparql. My experiment.: <http://rdf.myexperiment.org/howtosparql>
- . Google. Search Engine Optimization (SEO) Starter Guide : <https://support.google.com/webmasters/answer/7451184?hl=en>
- . MOZ. The beginner's guide to SEO: <https://moz.com/beginners-guide-to-seo>
- . W3schools tutorials (HTML, CSS, JS): <https://www.w3schools.com/>
- Manning, Christopher D; Raghavan, Prabhakar; Schütze, Hinrich . Introduction to Information Retrieval. Cambridge University Press: <https://nlp.stanford.edu/IR-book/information-retrieval-book.html>