

FICHA DE LA ASIGNATURA:



Enfoques Avanzados para la Lingüística Computacional Optativa-6 créditos-

Equipo docente

Alberto Díaz Esteban

Área de conocimiento: Lenguajes y Sistemas Informáticos

Tif: 91 394 7561 albertodiaz@fdi.ucm.es

<https://orcid.org/0000-0003-1966-3421>

Elena del Olmo Suárez

Área de conocimiento: Lingüística General, Lingüística Computacional

Tif: 91 394 7727 elenadelolmo@ucm.es

<https://produccioncientifica.ucm.es/investigadores/144321/detalle>

Objetivos didácticos

El objetivo básico es el estudio de la Lingüística Computacional utilizando enfoques avanzados.

Este objetivo se desgrana en los siguientes objetivos específicos:

1. Conocer y saber utilizar los principales modelos y métodos de representación y procesamiento del lenguaje natural en sus niveles fonético-fonológico, léxico, morfológico, sintáctico, semántico, pragmático y discursivo.
2. Conocer y saber seleccionar críticamente el modelo de representación del conocimiento más adecuado a un problema o aplicación de la Lingüística computacional.
3. Conocer y saber realizar tareas comunes en el procesamiento del lenguaje natural: segmentación, etiquetado morfo-sintáctico, lematización y análisis sintáctico.
4. Conocer y saber utilizar soluciones (bibliotecas o repositorios) consolidadas para el procesamiento del lenguaje natural, la generación de representaciones gráficas, acceso y extracción de datos de la web y el almacenamiento y gestión de versiones de proyectos software.
5. Conocer y saber definir y describir los fundamentos matemáticos de las representaciones vectoriales y de las redes neuronales.
6. Saber enumerar, definir y aplicar a tareas sencillas de procesamiento del lenguaje los principales modelos de redes neuronales para el Procesamiento del Lenguaje Natural (PLN).
7. Saber explicar cómo pueden contribuir las aplicaciones de la Lingüística computacional para mejorar la vida de las personas y para facilitar el progreso científico y tecnológico en beneficio de la humanidad.

Conocimientos y/o destrezas previas recomendadas

- Conocimiento básico de Python

- capacidad de leer en inglés
- capacidad de autoaprendizaje y planificación
- capacidad crítica: no dar nada por supuesto
- capacidad de aprender del error

Posibles asignaturas del Máster directamente relacionadas con ésta:

Aplicaciones de la Lingüística Computacional

Carga de trabajo/estudio prevista por semana para el alumno

10-12 horas de dedicación, repartidas entre el estudio del material docente (35%), el análisis de material adicional (10%), la resolución de casos prácticos (35%), la comunicación con el tutor y los compañeros (10%), la asistencia a seminarios presenciales o en línea (10%).

Descripción general de la asignatura

Para responder a la cuestión de la comunicación hombre-máquina en lenguaje natural, la idea es, en primer lugar, conseguir representar computacionalmente (en una máquina) el conocimiento que tienen las personas en su cerebro y, en segundo lugar, conseguir que con ese conocimiento las máquinas puedan procesar, es decir entender y producir, el lenguaje natural. Esta asignatura aborda el estudio de la Lingüística Computacional utilizando enfoques avanzados. Se centra en el estudio de los fundamentos matemáticos y de los modelos y técnicas de programación más recientes aplicados a la Lingüística computacional.

De forma más concreta, se tratará la extracción, estructuración y explotación de datos lingüísticos, su tratamiento automático con bibliotecas de Procesamiento del Lenguaje Natural, los fundamentos matemáticos de las representaciones vectoriales y de las redes neuronales, y los modelos de redes neuronales utilizados para el procesamiento del lenguaje natural.

Cronograma del curso

Semana	Tema	Objetivos didácticos	Profesor
Semana 1	Presentación e introducción a la asignatura. Planteamiento del primer caso práctico.	Objetivos 1, 4, 6 y 7	Elena del Olmo Suárez Alberto Díaz Esteban
Semana 2	Gestión web	Objetivos 1, 2 y 3	Elena del Olmo Suárez
Semana 3	Manipulación de datos estructurados	Objetivos 1, 2 y 3	Elena del Olmo Suárez
Semana 4	Generación de representaciones gráficas	Objetivos 1, 2 y 3	Elena del Olmo Suárez
Semana 5	Bibliotecas consolidadas para el PLN	Objetivos 4 y 7	Elena del Olmo Suárez
Semana 6	Bibliotecas consolidadas para el PLN	Objetivos 4 y 7	Elena del Olmo Suárez
Semana 7	Control de versiones	Objetivo 4	Elena del Olmo Suárez
Semana 8	Introducción a las redes neuronales	Objetivos 4, 5 y 6	Alberto Díaz Esteban
Semana 9	Representaciones vectoriales	Objetivos 4, 5 y 6	Alberto Díaz Esteban
Semana 10	Redes recurrentes	Objetivos 4, 5 y 6	Alberto Díaz Esteban
Semana 11	Redes secuencia-secuencia	Objetivo 4, 5 y 6	Alberto Díaz Esteban

Semana 12	Atención y <i>transformers</i>	Objetivos 4, 5 y 6	Alberto Díaz Esteban
Semana 13	Redes convolucionales	Objetivos 4, 5 y 6	Alberto Díaz Esteban
Semana 14	<i>Transfer learning</i>	Objetivos 4, 5 y 6	Alberto Díaz Esteban
Examen final		Objetivos 1 a 7	

Breve descripción de la Metodología(s) de aprendizaje(s) que se prevé utilizar:

El curso se impartirá en el Campus Virtual de la UCM. Se organizará en semanas. Cada semana se hará una propuesta temática de trabajo y se trabajará en dos partes: **1era. parte: presentación de la propuesta temática** para que sirva de guía y punto de partida para la reflexión individual sobre su aplicación en cada asignatura concreta. Se realizará el lunes por parte de los profesores.

2ª parte: estudio y actividades de consolidación, se realizarán a lo largo de la semana por parte de los alumnos y con la guía de los profesores.

3ª parte: resolución de casos prácticos, se realizarán a lo largo de la semana por parte de los alumnos.

Las tareas a realizar cada semana son:

- 1.- Presentación de los objetivos y contenidos del módulo. Guía para su estudio (lunes)
- 2.- Estudio, por parte del alumno, de los contenidos básicos. Incluirán lecturas (en modo texto, audio y/o vídeo) y casos prácticos resueltos.
- 3.- Realización de un caso práctico no resuelto evaluable.
- 4.- Ampliación opcional de conocimientos con el material complementario
- 5.- Reflexión entre iguales: uso del foro para preguntar y aclarar cuestiones de los contenidos básicos. El profesor hace el seguimiento sin intervenir excepto que sea necesario porque se observan errores o no se resuelven las dudas.

Se podrán realizar tutorías síncronas en la que el profesor atenderá las dudas que se planteen en foro.

Enumeración de las actividades de aprendizaje que se prevén utilizar para las sesiones virtuales:

Se utilizarán todas las actividades previstas en la memoria para aplicar los conocimientos adquiridos en la asignatura:

- Visionado de video/audiolecciones y sesiones síncronas o asíncronas
- Estudio individual del material básico
- Lectura y análisis de material complementario
- Resolución de casos prácticos
- Comunicación virtual con el profesor
- Foros y comunicación colaborativa
- Seminarios presenciales

Enumeración de las actividades de aprendizaje que se prevén utilizar para las sesiones presenciales:

- Resolución de casos teóricos o prácticos
- Resolución de dudas de los contenidos propuestos por el profesor

Procedimiento de evaluación

La asignatura se evalúa a partir de las actividades siguientes:

- Actividades evaluables de la semana: 60% de la nota final
- Examen presencial: 40% nota final

Competencias y destrezas que se desarrollarán

CE8. Conocer los modelos y métodos más relevantes (simbólicos, estadísticos y biológicos) para el procesamiento automático de las lenguas y los fenómenos lingüísticos en sus diferentes niveles lingüísticos: fonético-fonológico, morfológico, sintáctico, semántico, pragmático y de discurso.

CE10. Saber utilizar con suficiente destreza los principales paquetes de programación para resolver tareas de procesamiento del lenguaje natural en formato texto.

CE11. Conocer el manejo de las herramientas software existentes para el procesamiento de las producciones lingüísticas en diferentes lenguas (ej. segmentadores, analizadores morfológicos, sintácticos, semánticos).

Procedimiento para mostrar el progreso del alumno:

Boletín de calificaciones de la asignatura virtual

Mecanismos de comunicación docente

- Foro, correo electrónico de la asignatura virtual y videoconferencia

Mecanismos de tutorización virtual:

- Foro, correo electrónico en la asignatura virtual y videoconferencia al menos una vez en semana

Mecanismos de contacto

- Foro de la asignatura virtual que podrá complementarse cuando sea necesario con sesiones de videoconferencia. Lo atenderán los profesores de la asignatura.
- Correo electrónico institucional de los profesores en caso de que no tenga acceso al campus virtual
- Además, el alumno dispone de un servicio de ayuda para las incidencias informáticas de la Universidad en <https://ssii.ucm.es/estudiante>

Mecanismos de contacto para quejas y sugerencias de la asignatura:

- El alumno debe presentar su queja, en primer lugar, al profesor, y hacerle cuantas sugerencias considere oportunas sobre la asignatura, tanto por vía de correo electrónico como por un buzón anónimo de "quejas y sugerencias" en la página de Presentación de la asignatura virtual.
- El alumno se podrá dirigir también al Coordinador del Máster, así como al representante de alumnos en caso de que su queja o sugerencia no sea atendida.
- Además, el máster dispone de un buzón de quejas y sugerencias en su página web atendido por el Coordinador del máster

Mecanismos para recoger la opinión de los alumnos sobre la asignatura:

Participación en el programa Docencia de la UCM complementado con una encuesta anónima preparada por los profesores en la asignatura virtual

Requisitos técnicos especiales (no de Campus Virtual):

Ordenador portátil con conexión a internet

Bibliografía básica

- Bratko, I. Programming for Artificial Intelligence. Pearson. 2001.
- Eisenstein, J. Introduction to Natural Language Processing. 2019. Borrador disponible en: <https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf>
- Jurafsky, D. and Martin, J.H. Speech and Language Processing. Pearson Prentice Hall. 2008.
- Listerri, J, Martí, M. A. (Eds). Tratamiento del Lenguaje Natural. Edicions Universitat de Barcelona. 2003.
- Manning, C., Raghavan, P. and Schütze, H. Introduction to Information Retrieval. Cambridge University Press. 2008. Disponible en: <https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>
- Deep Learning with Python. Francois Chollet. Second edition; Manning, 2021
- Deep Learning with Python. Francois Chollet. First edition; Manning, 2018