

Equipo docente



Ana María Fernández-Pampillón Cesteros (2 créditos)
Área de conocimiento: Lingüística General, Lingüística Computacional
Tlf: 91 394 7730 apampi@filol.ucm.es
Página personal: <http://www.ucm.es/afpc>



Javier Ferreiros López (1 crédito)
Grupo de Tecnología del Habla y Aprendizaje Automático - UPM
javier.ferreiros@upm.es
<https://blogs.upm.es/gthau/javier-ferreiros-lopez/>



Juan Manuel Montero Martínez (1 crédito)
Grupo de Tecnología del Habla y Aprendizaje Automático - UPM
juanmanuel.montero@upm.es
<https://blogs.upm.es/gthau/juan-manuel-montero/>

Objetivos didácticos

- O1. Saber definir formalmente estructuras morfológicas o sintácticas de los lenguajes mediante gramáticas sintagmáticas
- O2. Saber definir expresiones regulares para especificar secuencias textuales
- O3. Conocer los tipos de autómatas y saber definir y aplicar autómatas de estados finitos para modelar lenguajes regulares
- O4. Aplicación de las gramáticas independientes del contexto al modelado de estructuras sintácticas de los lenguajes naturales
- O5. Conocer las estrategias básicas del análisis sintáctico basado en gramáticas sintagmáticas
- O6. Saber definir y aplicar el formalismo de las Gramáticas de atributos al modelado de fenómenos lingüísticos morfológicos o sintácticos.
- O7. Conocer los fundamentos de los modelos de lenguaje estadísticos basados en n-gramas.
- O8. Conocer y describir los fundamentos matemáticos de las redes neuronales.
- O9. Conocer y describir los fundamentos matemáticos de las representaciones vectoriales para las palabras.
- O10. Saber definir y describir los principales modelos de etiquetado morfológico-sintáctico automático del lenguaje natural.
- O11. Saber definir y describir modelos generativos basados en reglas con probabilidad.

O12. Conocer y describir cómo aplicar arquitecturas neuronales avanzadas para el Procesamiento del Lenguaje Natural (PLN).

O13. Experimentar con sistemas que resuelven tareas comunes en el procesamiento del lenguaje natural: etiquetado morfo-sintáctico, lematización y análisis sintáctico.

Conocimientos y/o destrezas previas recomendadas

- Conocimientos a nivel de usuario de informática e internet.
- Programación en Python.
- Lectura de textos en inglés.
- Autoaprendizaje y planificación.
- Trabajo colaborativo.
- Pensamiento crítico.
- Aprender del error.

Posibles asignaturas del Máster directamente relacionadas con ésta

- Métodos cualitativos y cuantitativos en Lingüística
- Tecnologías del habla y del texto
- Enfoques Avanzados en Lingüística Computacional (Redes Neuronales).

Carga de trabajo/estudio prevista por semana para el alumno

8 horas de dedicación semanal durante 12 semanas virtuales y 1 semana presencial. De las semanas virtuales, 2 horas serán sesiones virtuales con intervención directa del profesorado en que se presentarán los conceptos fundamentales y las pautas y punteros necesarios para el estudio personal durante las otras 6 horas. Adicionalmente habrá una sesión presencial de 4 horas al final del curso de carácter práctico con el fin de aplicar todos los conocimientos adquiridos en la asignatura.

Descripción general de la asignatura

En esta asignatura se estudiarán los principales modelos, simbólicos, estadísticos y de inspiración biológica, para la descripción y procesamiento (análisis y generación) de las lenguas naturales en sus niveles morfológico y sintáctico. La asignatura comenzará tratando los modelos simbólicos basados en reglas (gramáticas sintagmáticas) de la teoría formal de lenguajes, y basados en autómatas.

A continuación, se tratan los modelos automáticos de análisis morfológico y sintáctico en sus versiones: estadísticas y generativas basados en reglas con probabilidad. Finalmente, se trabajarán los modelos neuronales, repasando las nociones básicas que dan lugar a representaciones vectoriales de las palabras y evolucionando hasta arquitecturas avanzadas. Se incluye una componente práctica experimentando con tareas concretas a las que se aplicarán algunas de las herramientas automáticas estudiadas.

Contenido temático

1. Introducción a los autómatas y los lenguajes formales
2. Lenguajes regulares: autómatas finitos y expresiones regulares. Aplicación al procesamiento morfológico del lenguaje
3. Lenguajes independientes del contexto: autómatas con pila y gramáticas independientes del contexto. Aplicación al procesamiento sintáctico
4. Gramáticas aumentadas: gramáticas de atributos. Aplicación al procesamiento morfo-sintáctico
5. Modelos estadísticos basados en n-gramas
6. Fundamentos de redes neuronales para PLN
7. Representación vectorial de palabras
8. Técnicas de etiquetado morfo-sintáctico automático (1)

9. Técnicas de etiquetado morfo-sintáctico automático (2) y modelos generativos con probabilidades

10. Arquitecturas neuronales avanzadas para PLN

Cronograma del curso

Semana	Tema	Objetivos didácticos	Profesor
S1	Introducción a los autómatas y los lenguajes formales	O1, O3	Ana María Fernández-Pampillón
S2	Lenguajes regulares: autómatas finitos y expresiones regulares	O1, O2, O3	Ana María Fernández-Pampillón
S3	Lenguajes regulares: aplicación al procesamiento morfológico	O1, O2, O3, O13	Ana María Fernández-Pampillón
S4	Lenguajes independientes del contexto: gramáticas independientes del contexto y análisis sintáctico	O1, O4, O5	Ana María Fernández-Pampillón
S5	Gramáticas aumentadas: gramáticas de atributos	O1, O5, O6	Ana María Fernández-Pampillón
S6	Aplicación de las gramáticas de atributos a la descripción morfosintáctica de las lenguas	O1, O6, O13	Ana María Fernández-Pampillón
S7	Modelos estadísticos basados en n-gramas	O7	Javier Ferreiros
S8	Fundamentos de redes neuronales para PLN	O8	Javier Ferreiros
S9	Representación vectorial de palabras	O9	Javier Ferreiros
S10	Técnicas de etiquetado morfo-sintáctico automático (1)	O10	Juan Manuel Montero
S11	Técnicas de etiquetado morfo-sintáctico automático (2) y modelos generativos con probabilidades	O11	Juan Manuel Montero
S12	Arquitecturas neuronales avanzadas para PLN	O12	Juan Manuel Montero
	Prácticas con herramientas automáticas	Todos los objetivos	Todos
	Examen final	Todos los objetivos	Todos
	Examen extraordinario	Todos los objetivos	Todos

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:

1ª parte: presentación de la propuesta temática para que sirva de guía y punto de partida para la reflexión individual y la planificación personal del estudio. Esta presentación será teórica o práctica según la semana correspondiente. En el cronograma del curso se indica la secuenciación de los temas de estudio. Este cronograma, en todo caso, podrá modificarse para adaptarse al ritmo de aprendizaje de los alumnos.

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

Las tareas a realizar cada semana son:

- 1.- Presentación del profesor sobre los objetivos y contenidos del módulo. Guía para su estudio.
- 2.- Estudio, por parte del alumno, de los contenidos básicos. Incluirán lecturas (en modo texto, audio y/o vídeo) y/o ejercicios.

3.- 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.

4.- Ampliación opcional de conocimientos con el material complementario.

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

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

Se utilizarán las siguientes actividades:

- Visionado de video/audio lecciones y sesiones sincronas o asincronas.
- Estudio individual del material básico.
- Lectura y análisis de material complementario.
- Resolución de ejercicios prácticos.
- Comunicación virtual con el profesor.
- Foros y comunicación colaborativa.

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

- Aclaración y profundización de conceptos teóricos, en caso de que sea necesario.
- Planteamiento y resolución colaborativa de ejercicios teóricos o prácticos.
- Presentación de los alumnos de los resultados de las actividades de aprendizaje que se seleccionen

Procedimiento de evaluación

- Evaluación continua: 30% de la nota final.
- Examen final teórico-práctico 70 % de la nota final.

La nota de la evaluación continua será la media aritmética de los ejercicios evaluables de cada semana.

Para aprobar la asignatura es imprescindible aprobar el examen 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.
- CE9. Saber describir las características del lenguaje humano que hacen difícil su tratamiento automático.
- 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 a la Coordinación 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 la Coordinación del Máster

Requisitos técnicos especiales (no de campus virtual)

Ordenador con conexión a internet

Bibliografía

JURAFSKY, D.; MARTIN J.H. (2009) *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd. Edition (borrador de la 3ª edición 2020 en: <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>)

FERNÁNDEZ FERNÁNDEZ, G.; SAEZ VACAS, F. (1987). *Fundamentos de la Informática*. Alianza Editorial / 1995 Anaya Multimedia. Disponible en: https://oa.upm.es/22325/1/Fundamentos_de_informatica.pdf

KELLEY, Dean (2001). *Teoría de autómatas y lenguajes formales*. Prentice Hall, Madrid, 2001

KNUTH, D. 1968. *Semantics of Context-Free Languages. Mathematical Systems Theory*. Vol. 8 No.2

PAAKI, J. (1995) Attribute Grammar Paradigms. A High-Level Methodology in Language Implementation. *ACM Computing Surveys*. Num. 2. June 1995. <https://www.csee.umbc.edu/courses/graduate/631/Fall2002/p196-paakki.pdf>