

23-24

MÁSTER UNIVERSITARIO EN
INVESTIGACIÓN EN INTELIGENCIA
ARTIFICIAL

GUÍA DE ESTUDIO PÚBLICA



FUNDAMENTOS DEL PROCESAMIENTO LINGÜÍSTICO

CÓDIGO 31080027

UNED

23-24

FUNDAMENTOS DEL PROCESAMIENTO
LINGÜÍSTICO
CÓDIGO 31080027

ÍNDICE

PRESENTACIÓN Y CONTEXTUALIZACIÓN
REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA
ASIGNATURA
EQUIPO DOCENTE
HORARIO DE ATENCIÓN AL ESTUDIANTE
COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE
RESULTADOS DE APRENDIZAJE
CONTENIDOS
METODOLOGÍA
SISTEMA DE EVALUACIÓN
BIBLIOGRAFÍA BÁSICA
BIBLIOGRAFÍA COMPLEMENTARIA
RECURSOS DE APOYO Y WEBGRAFÍA

Nombre de la asignatura	FUNDAMENTOS DEL PROCESAMIENTO LINGÜÍSTICO
Código	31080027
Curso académico	2023/2024
Título en que se imparte	MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL
Tipo	CONTENIDOS
Nº ETCS	6
Horas	150.0
Periodo	ANUAL
Idiomas en que se imparte	CASTELLANO

PRESENTACIÓN Y CONTEXTUALIZACIÓN

En sentido estricto, cualquier actividad que implique un análisis o generación de la lengua utilizando el ordenador puede considerarse como procesamiento del lenguaje natural (PLN). No obstante, mientras los lingüistas prefieren hablar de la lingüística computacional como un área de conocimiento dentro de la lingüística aplicada, los informáticos consideran el procesamiento del lenguaje natural como una rama de la inteligencia artificial debido a la posibilidad de desarrollar sistemas de computación que simulen algún aspecto de la capacidad lingüística del ser humano.

En este contexto, Fundamentos del Procesamiento Lingüístico es una asignatura teórico-práctica dirigida principalmente a la parte del PLN que tiene más que ver con el procesamiento lingüístico automático: análisis morfológico, sintáctico y semántico, así como sus aplicaciones. En otras palabras, se centra en los fundamentos teóricos básicos del enfoque simbólico en la modelización del conocimiento lingüístico para la construcción de sistemas computacionales que analicen y/o generen textos en lenguaje natural.

En general, esta asignatura posibilita la formación tanto de profesionales que deseen organizar y coordinar proyectos lingüísticos dentro del marco de las tecnologías del lenguaje, como de investigadores que deseen realizar avances científicos en alguno de los aspectos teóricos del procesamiento del lenguaje natural.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Conocimientos previos necesarios:

- Lectura fluida en inglés
- Conocimientos de programación

Conocimientos previos recomendables:

- Conocimientos básicos de álgebra, cálculo y lógica
- Aprendizaje automático
- RDF y SPARQL

El alumno deberá ser capaz de leer y comprender con relativa facilidad textos en inglés, ya que muchas de las lecturas propuestas en esta asignatura están escritas en esta lengua.

Los alumnos deberán instalar, programar y ejecutar aplicaciones informáticas de procesamiento lingüístico, por lo que deberán tener facilidad en el uso de ordenadores y conocimiento de lenguajes de programación.

Esta asignatura puede ser cursada aisladamente.

EQUIPO DOCENTE

Nombre y Apellidos	ANSELMO PEÑAS PADILLA (Coordinador de asignatura)
Correo Electrónico	anselmo@lsi.uned.es
Teléfono	91 398 7750
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	LENGUAJES Y SISTEMAS INFORMÁTICOS
Nombre y Apellidos	ALBERTO PEREZ GARCIA-PLAZA
Correo Electrónico	alberto.perez@lsi.uned.es
Teléfono	91 398 8412
Facultad	ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
Departamento	LENGUAJES Y SISTEMAS INFORMÁTICOS

HORARIO DE ATENCIÓN AL ESTUDIANTE

Anselmo Peñas:

- Jueves de 9:30 a 13:30 horas.

Alberto Pérez:

- Jueves de 11:00 a 15:00 horas.

Medios de comunicación

- Correo electrónico:
- anselmo@lsi.uned.es (canal de comunicación preferente)
- alberto.perez@lsi.uned.es (canal de comunicación preferente)
- Número de teléfono:
- Anselmo Peñas: (+34) 91 398 7750
- Alberto Pérez: (+34) 91 398 8412
- Dirección postal: Departamento de Lenguajes y Sistemas Informáticos; ETSI Informática; C/Juan del Rosal, 16; 28040. Madrid

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

Competencias generales

Con la superación del curso se espera que el alumno complete las siguientes competencias generales especificadas en la memoria del máster:

CG1 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más

amplios (o multidisciplinares) relacionados con su área de estudio.

CG2 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios.

CG3 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

CG4 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Competencias básicas

Asimismo, se espera que el alumno complete las competencias básicas establecidas para el título:

CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

Competencias específicas

Con la superación del curso se espera que el alumno complete las siguientes competencias específicas indicadas en la memoria del máster:

CE3 - Conocer los procedimientos específicos de aplicación de estos métodos a un conjunto relevante de dominio (educación, medicina, ingeniería, sistemas de seguridad y vigilancia, etc.), que representan las áreas más activas de investigación en IA.

RESULTADOS DE APRENDIZAJE

El objetivo principal de la asignatura consiste en conocer y aplicar las técnicas de procesamiento lingüístico automático de los textos como paso previo a su conversión en representaciones simbólicas manipulables por los ordenadores.

De acuerdo con esto, los resultados del aprendizaje son:

1. Conocer y aplicar las técnicas de análisis morfosintáctico de las palabras en contexto.
2. Conocer y aplicar las técnicas de análisis sintáctico de las oraciones, tanto el análisis de constituyentes como de dependencias.
3. Conocer y aplicar las técnicas para obtener representaciones semánticas de las oraciones.
4. Saber utilizar algunas herramientas existentes de procesamiento lingüístico de los textos.

CONTENIDOS

Tema 1: Introducción al procesamiento lingüístico de los textos

Tema 2: Análisis morfosintáctico y léxico de las palabras

Tema 3: Análisis sintáctico de las oraciones

Tema 4: Análisis semántico

METODOLOGÍA

Hemos asignado a cada tema de la asignatura una serie de lecturas obligatorias que presentan con claridad los contenidos teórico-prácticos sobre el procesamiento lingüístico y sus aplicaciones informáticas. A partir de estas lecturas, los alumnos demostrarán la asimilación de los contenidos mediante la elaboración de varias prácticas y un proyecto personal.

La asignatura adoptará una metodología de enseñanza a distancia que favorezca el trabajo autónomo. Con el fin de consolidar el proceso de aprendizaje, esta metodología se centrará tanto en los contenidos teóricos mediante la lectura de artículos científicos, como en el desarrollo de actividades prácticas basadas en dichas lecturas:

- Sigue las directrices del Espacio Europeo de Educación Superior adaptadas a la enseñanza a distancia.
- No existen clases presenciales. Los contenidos se imparten a distancia, de acuerdo con las normas y estructuras con soporte telemático de enseñanza en la UNED.
- El material docente incluye un resumen de los contenidos de cada tema y distintos tipos de actividades relacionadas con la consulta bibliográfica y el modelado, simulación e implementación de diversos ejemplos de los distintos mecanismos descritos en la teoría.
- El alumno deberá estudiar el material docente (escrito y/o audiovisual) y realizar distintas actividades según el plan de trabajo propuesto por el Equipo Docente relacionadas con la consulta bibliográfica, con el modelado y con la implementación de distintas técnicas

relativas a cada una de las fases implicadas en el proceso de extracción del conocimiento, así como la realización de prácticas con software especializado.

- Se organizarán foros de discusión para dudas y debates.

SISTEMA DE EVALUACIÓN

TIPO DE PRIMERA PRUEBA PRESENCIAL

Tipo de examen No hay prueba presencial

TIPO DE SEGUNDA PRUEBA PRESENCIAL

Tipo de examen2 No hay prueba presencial

CARACTERÍSTICAS DE LA PRUEBA PRESENCIAL Y/O LOS TRABAJOS

Requiere Presencialidad No

Descripción

La **evaluación global** de la asignatura está basada en la superación de prácticas no presenciales.

Criterios de evaluación

Ponderación de la prueba presencial y/o los trabajos en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC? Si,PEC no presencial

Descripción

La **evaluación global** de la asignatura está basada en la superación de cuatro prácticas (no presenciales) con el siguiente peso en el cómputo final de la asignatura:
Práctica 1: Conceptos en Procesamiento del Lenguaje Natural (20%) [Octubre - Noviembre]

Práctica 2: Análisis léxico y sintáctico (20%) [Noviembre - Enero]

Practica 3: Análisis semántico (20%) [Enero - Marzo]

Práctica 4: Trabajo individual a propuesta del estudiante (40%) [Marzo - Mayo]

La fecha concreta de entrega de cada práctica se especificará en el entorno virtual.

Criterios de evaluación

Los principales criterios de evaluación:

Aportación de contenidos correctos según lo estudiado durante el curso

Aportación de información nueva.

Originalidad e innovación.

Corrección de formato: adaptación a las instrucciones de formato que se indiquen.

Extensión adecuada del trabajo según los márgenes establecidos.

Ponderación de la PEC en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

OTRAS ACTIVIDADES EVALUABLES

¿Hay otra/s actividad/es evaluable/s? No

Descripción

Criterios de evaluación

Ponderación en la nota final

Fecha aproximada de entrega

Comentarios y observaciones

¿CÓMO SE OBTIENE LA NOTA FINAL?

La **evaluación global** de la asignatura se hará según el siguiente plan de prácticas, todas ellas de tipo NO PRESENCIAL:

Práctica 1: 20%

Práctica 2: 20%

Practica 3: 20%

Práctica 4: 40%

IMPORTANTE (nota mínima): para superar la asignatura, la práctica final ("práctica 4") deberá superarse con una nota igual o superior a 5 sobre 10. Esta nota mínima será requisito indispensable para computar el resto de prácticas del curso, tanto en la convocatoria ordinaria como extraordinaria.

IMPORTANTE (fechas de entrega): la fecha máxima de entrega se indicará en la plataforma ALF desde la apertura del curso. Es importante recordar que es necesario ajustarse a las fechas marcadas desde el inicio para superar la asignatura.

IMPORTANTE (convocatoria extraordinaria): tanto el baremo como el sistema de evaluación serán los mismos en las convocatoria ordinaria y extraordinaria, es decir, cuatro tareas relacionadas con las lecturas del curso y otras lecturas que deberán recopilarse individualmente. Aquellas prácticas que se hayan superado correctamente durante el curso se guardarán para la convocatoria extraordinaria y se evaluarán según el baremo propuesto. Por tanto, sólo deberán completarse aquéllas que no hubieran sido entregadas a tiempo durante la convocatoria ordinaria.

IMPORTANTE (aclaración): no hay ninguna variación con respecto al planteamiento general en el caso de los estudiantes en centros penitenciarios o para las convocatorias extraordinarias.

BIBLIOGRAFÍA BÁSICA

TEXTO DE REFERENCIA

Daniel Jurafsky & James H. Martin, 2021. *Speech and Language Processing* (3rd edition)
<https://web.stanford.edu/~jurafsky/slp3/>

CONSULTA DE ARTÍCULOS (lista indicativa)

1. Cambria, E. y B. White (2014). "Jumping NLP curves: A review of Natural Language Processing research". *IEEE Computational Intelligence Magazine*, Mayo 2014: 48-57
2. Gómez Guinovart, X. (2000). "Perspectivas de la lingüística computacional". *Novática: Revista de la Asociación de Técnicos de Informática*, número especial del 25 aniversario (Horizonte 2025), 145: 85-87.
3. Llisterri, J. (2003). "Lingüística y tecnologías del lenguaje". *Lynx. Panorámica de Estudios Lingüísticos* 2: 9-71.
4. Mairal Usón, R. y F. Cortés Rodríguez (2006). "An overview of Role and Reference Grammar". En R. Mairal, M. A. Escobar Álvarez, M. S. Peña Cervel, E. Samaniego Fernández y F. Cortés Rodríguez (eds.) *Current Trends in Linguistic Theory*. Madrid: Universidad Nacional de Educación a Distancia, 97-176.
5. Peñalver-Martínez, I., R. Valencia-García y F. García-Sánchez (2011). "Minería de Opiniones basada en características guiada por Ontologías". *Procesamiento del Lenguaje Natural*, 46: 91-98. Sociedad Española para el Procesamiento del Lenguaje Natural, España.
6. Periñán Pascual, C. (2005). "Procesamiento del lenguaje natural: de lingüista a ingeniero del conocimiento". En I. K. Brady, M. Navarro Coy y C. Periñán Pascual (eds.) *Nuevas Tendencias en Lingüística Aplicada*. Murcia: Quaderna, 293-317.
7. Periñán Pascual, C. y R. Mairal Usón (2009). "Bringing Role and Reference Grammar to natural language understanding". *Procesamiento del Lenguaje Natural* 43: 265-273.
8. Rodríguez, H. (2000). "Técnicas básicas en el tratamiento informático de la lengua". *Quark. Ciencia, Medicina, Comunicación y Cultura* 19: 26-34.

BIBLIOGRAFÍA COMPLEMENTARIA

Dale, R., H. Moisl y H. Somers (eds.) (2000): *Handbook of Natural Language Processing*, Nueva York, Marcel Dekker.

Mitkov, R. (ed.) (2003): *The Oxford Handbook of Computational Linguistics*, Oxford, Oxford University Press.

RECURSOS DE APOYO Y WEBGRAFÍA

La plataforma de aprendizaje en Internet permitirá no solo la interacción entre los alumnos y el profesor, sino también el acceso a las lecturas obligatorias y a las instrucciones para la realización de los trabajos propuestos. Igualmente, el foro de debate y el correo electrónico se convierten en instrumentos imprescindibles para la consulta y la resolución de dudas relacionadas con el trabajo del alumno.

IGUALDAD DE GÉNERO

En coherencia con el valor asumido de la igualdad de género, todas las denominaciones que en esta Guía hacen referencia a órganos de gobierno unipersonales, de representación, o miembros de la comunidad universitaria y se efectúan en género masculino, cuando no se hayan sustituido por términos genéricos, se entenderán hechas indistintamente en género femenino o masculino, según el sexo del titular que los desempeñe.