

20-21

MÁSTER UNIVERSITARIO EN
INVESTIGACIÓN EN INTELIGENCIA
ARTIFICIAL

GUÍA DE ESTUDIO PÚBLICA



PROCESAMIENTO DEL LENGUAJE NATURAL

CÓDIGO 31101269

UNED

20-21

PROCESAMIENTO DEL LENGUAJE
NATURAL
CÓDIGO 31101269

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 | PROCESAMIENTO DEL LENGUAJE NATURAL |
| Código | 31101269 |
| Curso académico | 2020/2021 |
| 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

Este curso introductorio al procesamiento computacional del lenguaje natural aborda el diseño y la construcción de programas que pueden tratar, comprender y generar lenguaje natural. El temario y enfoque se centra en los modelos y técnicas simbólicas y estadísticas. Se estudiarán los problemas y soluciones básicas en los niveles lingüísticos morfológico, sintáctico, semántico y pragmático. Un capítulo de introducción y otro de áreas de aplicación, situarán la asignatura desde una perspectiva histórica, y permitirán conocer el estado actual de las realizaciones en este campo. La realización de un proyecto permitirá poner en práctica los conocimientos adquiridos. Se incluyen algunas técnicas de aprendizaje automático, pero no las basadas en deep-learning que serían objeto de otra asignatura por sí mismas. Si bien las técnicas de deep learning están suponiendo un salto espectacular en resultados para algunas aplicaciones de procesamiento del lenguaje, no deja de tener interés el conocer el enfoque clásico, que es el que ayudará a entender y analizar los resultados obtenidos por las técnicas de aprendizaje no supervisado.

Ficha técnica:

| | |
|--------------------------|----------|
| Tipo | Optativa |
| Créditos/horas totales | 6/150 |
| Horas de estudio teórico | 45 |
| Horas de prácticas | 45 |
| Proyecto | 60 |

Esta asignatura está encuadrada en los siguientes másteres:

- MÁSTER UNIVERSITARIO EN I.A. AVANZADA: FUNDAMENTOS, MÉTODOS Y APLICACIONES
- MÁSTER UNIVERSITARIO EN TECNOLOGÍAS DEL LENGUAJE

Permitirá por tanto al alumno poner en práctica los fundamentos y métodos adquiridos tanto en IA como en LSI para el procesamiento del Lenguaje Natural. A su vez le proporcionará conocimiento y tecnología para valorar la incorporación del lenguaje natural en diferentes aplicaciones interactivas, o que hagan uso de de información textual.

REQUISITOS Y/O RECOMENDACIONES PARA CURSAR ESTA ASIGNATURA

Es importante una lectura fluida del inglés y disponer de conexión a internet. En cuanto a contenidos, esta asignatura tiene relación estrecha con las asignaturas de la carrera de Ingeniería Informática relativas a Teoría de autómatas, Procesadores de Lenguaje, e Introducción a la Inteligencia Artificial, que proporcionan la base en cuanto a formalismos y técnicas computacionales. Así mismo las asignaturas de programación, y especialmente aquellas en que se estudian paradigmas declarativos, constituyen un complemento interesante para cursar Procesamiento de Lenguaje Natural. El lenguaje de programación que se utilizará es Python.

EQUIPO DOCENTE

Nombre y Apellidos
Correo Electrónico
Teléfono
Facultad
Departamento

MARIA FELISA VERDEJO MAILLO (Coordinador de asignatura)
felisa@lsi.uned.es
91398-6484
ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
LENGUAJES Y SISTEMAS INFORMÁTICOS

Nombre y Apellidos
Correo Electrónico
Teléfono
Facultad
Departamento

ENRIQUE AMIGO CABRERA
enrique@lsi.uned.es
91398-8651
ESCUELA TÉCN.SUP INGENIERÍA INFORMÁTICA
LENGUAJES Y SISTEMAS INFORMÁTICOS

HORARIO DE ATENCIÓN AL ESTUDIANTE

La tutorización de los alumnos se llevará a cabo a través de una plataforma de aprendizaje en internet, que incluye comunicación electrónica con el equipo docente, y áreas de compartición de trabajo. Los trabajos que se proponen se evalúan a lo largo del curso. Se puede contactar telefónicamente con el equipo docente:

Jueves lectivos:

de 15 a 19 h. Teléfono: 91 398 86 51

de 11 a 13. Teléfono: 913986484

Miércoles lectivos: de 11 a 13. Teléfono: 913986484

Se pueden acordar otras citas previa solicitud por correo electrónico

COMPETENCIAS QUE ADQUIERE EL ESTUDIANTE

MÁSTER UNIVERSITARIO EN LENGUAJES Y SISTEMAS INFORMÁTICOS

Competencias Básicas:

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 Generales:

CPG1 - Adquirir capacidad de abstracción, análisis, síntesis y relación de ideas.

CPG2 - Adquirir capacidad crítica y de decisión

CPG3 - Adquirir capacidad de estudio y autoaprendizaje

CPG4 - Adquirir capacidad creativa y de investigación

CPG5 - Adquirir habilidades sociales para el trabajo en equipo

Competencias Específicas:

CE1 - Adquirir capacidad de comprender y manejar de forma básica los aspectos más importantes relacionados con los lenguajes y sistemas informáticos en general y, de manera especial, en los siguientes ámbitos: Tecnologías del lenguaje y de acceso a la información en web

CE3 - Adquirir capacidad de estudio de los sistemas y aproximaciones existentes y para distinguir las aproximaciones más efectivas.

CE4 - Adquirir capacidad para detectar carencias en el estado actual de la ciencia y la tecnología

CE5 - Adquirir capacidad para proponer nuevas aproximaciones que den solución a las carencias detectadas.

CE6 - Adquirir capacidad de especificar, diseñar, implementar y evaluar tanto cualitativa como cuantitativamente los modelos y sistemas propuestos.

CE7 - Adquirir capacidad para proponer y llevar a cabo experimentos con la metodología adecuada como para poder extraer conclusiones y determinar nuevas líneas de actuación e investigación.

**MÁSTER UNIVERSITARIO EN INTELIGENCIA ARTIFICIAL AVANZADA:
FUNDAMENTOS, MÉTODOS Y APLICACIONES****Competencias Básicas:**

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 Generales:

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 Específicas:

CE1 - Conocer los fundamentos de la Inteligencia Artificial y las fronteras actuales en investigación.

CE2 - Conocer un conjunto de métodos y técnicas tanto simbólicas como conexionistas y probabilistas, para resolver problemas propios de la Inteligencia Artificial.

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

En la primera parte del curso, mediante el estudio de la bibliografía el alumno adquirirá una visión amplia de las técnicas de procesamiento de lenguaje natural en los niveles léxico, sintáctico y semántico y sus aplicaciones. Los conocimientos adquiridos a nivel teórico se pondrán en práctica mediante la realización de ejercicios con una herramienta, NLTK, de libre disposición que los alumnos utilizarán para la elaboración de un analizador morfológico y un analizador sintáctico y semántico sobre un subdominio abordable del lenguaje. Paralelamente, los conocimientos adquiridos a nivel global y la capacidad de síntesis se pondrán en práctica mediante el desarrollo de una serie de resúmenes (el primero de ellos guiado por preguntas) .

En la segunda parte del curso el alumno adquirirá la destreza necesaria para elaborar un sistema de procesamiento de lenguaje orientado a una tarea específica. Con este curso el alumno asimilará tanto el potencial de las técnicas existentes de procesamiento de lenguaje como de sus limitaciones, siendo capaz de analizar en qué casos es factible aplicar estas técnicas en la resolución de un problema.

CONTENIDOS

Tema 0- Python y NLTK-

Una introducción al lenguaje de programación Python para adquirir el nivel necesario para el uso de las herramientas disponibles en NLTK.

Tema 1. Introducción

Se identifican algunos de los problemas más importantes que se plantean en el estudio y tratamiento computacional del lenguaje natural, y se da una breve descripción histórica del desarrollo de esta disciplina.

Tema 2: Autómatas finitos, procesamiento de unidades morfológico-léxicas

Se fijan los conceptos de expresiones regulares y los operadores asociados además de autómatas finitos y lenguajes regulares. Se introduce además el concepto de morfología en inglés y, mediante lecturas complementarias, morfología castellana. El tema aborda a continuación las técnicas de procesamiento morfológico basadas en lexicones, transductores y la aproximación de stemming. Finalmente se estudian los N-gramas. Se proponen unos ejercicios prácticos a realizar con NLTK.

Temas 3: Etiquetado

En este tema se establece un puente entre los niveles léxico y sintáctico. Se describe la taxonomía de palabras aplicables a diferentes lenguas, y las diferentes técnicas de etiquetado sintáctico existentes.

Temas 4 y 5 : Gramáticas de contexto libre y parsing

En el tema 4 se introducen las estructuras de la oración, incluyendo los conceptos de constituyente, sintagmas nominales y verbales, oraciones coordinadas, y su representación mediante gramáticas de contexto libre. Se verá también la gramática de dependencias.

El tema 5 se centra en las técnicas fundamentales de análisis sintáctico: descendente ("top-down") y ascendente ("bottom-up").

Tema 6: Estructuras de rasgos y unificación

Se describe el análisis sintáctico mediante la unificación de rasgos, su implementación y el diseño de restricciones de unificación.

Tema 7: Semántica y análisis semántico

Este tema aborda en general las diferentes técnicas de procesamiento a nivel semántico del lenguaje. Se introducen los conceptos de nivel semántico, predicados de primer orden y análisis semántico dirigido por sintaxis, entre otros. Incluye además el nivel léxico semántico en el que se describen relaciones semánticas entre palabras, y bases de datos léxico semánticas. Se propone un ejercicio práctico con la herramienta NLTK.

Tema 8: Nivel de discurso y aplicaciones de extracción, sistemas pregunta-respuesta y resumen

Este tema incluye el estudio de conceptos básicos de nivel de discurso como son la segmentación y resolución de correferencias. Finalmente nos centraremos en tres tipos de aplicaciones que son hoy día muy utilizadas: la extracción de información, los sistemas pregunta-respuesta y los resúmenes, para estudiar la clase de problemas que se plantean y el alcance de las técnicas para tratarlos. A partir de esta base, se propone un trabajo personal de carácter teórico y práctico, que pone en juego los conocimientos adquiridos en la asignatura.

METODOLOGÍA

El curso consta de un conjunto de temas cuyo estudio se realiza con la siguiente metodología dentro de un paradigma de construcción de conocimiento:

Para cada tema, el alumno debe acceder al material propuesto por el equipo docente. Este material consta de:

- Bibliografía básica común a todos los temas. Se trata un libro con un conocimiento ya estructurado facilitando la introducción del alumno en la materia.
- Artículos científicos. Se propone la lectura de algunos artículos de carácter científico. Su contenido es más específico. Aparte de conocer su contenido, el alumno se familiarizará con la estructura y formato que deben seguir los textos de estas características.
- Enlaces web que apuntan a recursos y herramientas relacionados con el tema.

A partir de este material (y con la guía de unas preguntas para el primero), el alumno debe realizar tres síntesis (de 10-30 páginas), en forma de artículos, correspondiendo cada una a un bloque de temas, con el objetivo de sintetizar el conocimiento que ha adquirido. La elaboración del artículo se dirige a:

- Estimular la lectura detenida del material propuesto.
- Provocar la necesidad de buscar información que complete el material propuesto inicialmente. Esta búsqueda es un entrenamiento necesario en la formación del alumno como investigador. Con cada trabajo tendrá mayor capacidad para encontrar y discriminar fuentes de información relevantes, requisito para desarrollar cualquier trabajo de investigación posterior.
- Estimular una reflexión sobre el material estudiado, elaborando ejemplos propios, necesaria para poder realizar una síntesis de calidad.
- Adquirir destreza para la escritura de documentos técnicos.

Junto con la elaboración de artículos de síntesis, el alumno deberá realizar una serie de ejercicios para los que dispondrá de la herramienta NLTK. De esta forma, el alumno podrá centrarse en aspectos del lenguaje y diseño de gramáticas sin necesidad de implementar desde un principio autómatas y mecanismos de unificación. Para manejar esta herramienta es conveniente tener conocimientos del lenguaje de programación Python, por lo que la primera semana del curso se dedicará a este tema.

Los últimos meses del curso se dirigen a afianzar los conocimientos adquiridos mediante la elaboración de un proyecto en el que se pondrán en práctica las técnicas aprendidas con datos y herramientas similares a los utilizados en la práctica profesional

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

Proyecto: se planteará un problema de tratamiento de información textual para una aplicación, que conlleve experimentación: modelización, procesado automático de datos y análisis de resultados, en donde se pongan en juego los conocimientos y técnicas de la asignatura.

Criterios de evaluación

Para el proyecto se valorará:

El análisis del problema y el modelo propuesto

La descripción del sistema desarrollado, tecnologías empleadas. modularización, etc.

Descripción de los problemas encontrados durante el desarrollo y soluciones empleadas.

Análisis comparativo y justificación de los resultados obtenidos en relación al estado del arte

Análisis comparativo en relación a los trabajos realizados por el resto de compañeros.

Ponderación de la prueba presencial y/o los trabajos en la nota final 50%

Fecha aproximada de entrega proyecto: principios de mayo

Comentarios y observaciones

MUY IMPORTANTE : No hay mas que un UNICO plazo de entrega para cada uno de los trabajos y para el proyecto en el curso académico. No hay examen final.

PRUEBAS DE EVALUACIÓN CONTINUA (PEC)

¿Hay PEC? Si, PEC no presencial

Descripción

Los trabajos (PEC) constan de: (A) tres síntesis en forma de artículos, sobre el temario de la asignatura y (B) tres ejercicios prácticos con las herramientas de NLTK. Cada PEC definida como una tarea en el entorno de la asignatura, tiene **un unico plazo de entrega**. Estos plazos están también especificados en el **cronograma** del curso donde pueden visualizarse conjuntamente. Es necesario entregar en su plazo correspondiente tanto las síntesis como los ejercicios prácticos.

Criterios de evaluación

¿Cómo se obtiene la nota final?

Requisito 1- Todos los trabajos deben entregarse.

Requisito 2- Cada uno de los dos tipos (A,B) tiene que estar aprobado.

La nota final es ponderada.

PEC

A) - 3 trabajos de síntesis (25%)

B)- 3 practicas (25%)

Proyecto - (50%)

Ponderación de la PEC en la nota final 50%

Fecha aproximada de entrega 2 semana diciembre, mediados febrero, principio marzo

Comentarios y observaciones

Es obligatorio entregar todos los trabajos. Para cada entrega hay una fecha única. Conviene consultar el plan de entregas para organizarse adecuadamente el estudio de la asignatura.

En caso de suspender alguno de los trabajos, hay una opción de recuperación, con un plazo de entrega para la convocatoria extraordinaria.

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?

Cada entrega puntúa de 1 a 10. La nota final es una ponderación de los tres tipos de entregas:

PEC

A-síntesis (8,9, 8 % respectivamente)

B-trabajos practicos (7, 8 y 10% respectivamente)

Proyecto (50%)

Cada tipo de entrega (A,B, proyecto) debe estar aprobada.

BIBLIOGRAFÍA BÁSICA

ISBN(13):9780135041963

Título:SPEECH AND LANGUAGE PROCESSING (segunda)

Autor/es:Jurafsky, Daniel ; Martin, James H. ;

Editorial:PEARSON EDUCATION

El libro base (segunda edición) es imprescindible. La primera edición del libro de bibliografía básica no incluye algunos contenidos del curso.

Para los ejercicios prácticos:

Libro: Natural language processing with Python, disponible en <https://www.nltk.org/book/>

Herramientas

Python (lenguaje de programación)

<https://docs.python.org/3/>

NLTK 3.0

<http://www.nltk.org/>

BIBLIOGRAFÍA COMPLEMENTARIA

En el entorno virtual de la asignatura se pone a disposición de los alumnos material de estudio complementario (artículos, recopilaciones y referencias a otro material disponible en la web)

RECURSOS DE APOYO Y WEBGRAFÍA

La plataforma de aprendizaje en internet, proporcionará interfaz de interacción entre el alumno y sus profesores, a través del curso virtual. Ofrece la gestión de entregables en forma de tareas, compartir documentos, participar en foros temáticos, así como la comunicación con el equipo docente.

Además, los alumnos tendrán que descargarse la herramienta NLTK, para la realización de ejercicios prácticos

¿Qué es NLTK? Una presentación breve

<http://desilinguist.org/pdf/crossroads.pdf>

Para el proyecto se facilitará el acceso a los datos y herramientas necesarias para llevarlo a cabo.

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.