



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

GUÍA DOCENTE

2026 - 2027

FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Raspado web (Web Scraping)
Código	DCIA-TITLE-721
Título	Graduado o Graduada en Traducción, Interpretación y Tecnologías del Lenguaje por la Universidad Pontificia Comillas [GRIRL 23]
Impartido en	Grado en Traducción, Interpretación y Tecnologías del Lenguaje [Cuarto Curso] Grado en Traducción, Interpretación y Tecnologías del Lenguaje [Tercer Curso]
Nivel	Reglada Grado Europeo
Cuatrimestre	Semestral
Créditos	3,0 ECTS
Carácter	Optativa
Departamento / Área	Departamento de Ciencias de la Computación e Inteligencia Artificial
Responsable	Luis Alarcon Massó
Horario de tutorías	previa cita

Datos del profesorado	
Profesor	
Nombre	Luis Alarcón Massó
Departamento / Área	Departamento de Organización Industrial
Correo electrónico	lalarcon@comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura
Aportación al perfil profesional de la titulación
En el perfil profesional del graduado en Grado en Traducción, Interpretación y Tecnologías del Lenguaje, esta asignatura proporciona al alumno las competencias necesarias para obtener y preparar de forma automatizada información textual disponible en la web para su posterior análisis y tratamiento en el ámbito de las tecnologías del lenguaje. Para ello, el alumno adquirirá una comprensión básica de la estructura de las páginas web y aprenderá a utilizar técnicas y herramientas de <i>web scraping</i> que le permitan localizar, extraer, almacenar y limpiar datos textuales procedentes de diferentes fuentes y formatos. El conocimiento de estas técnicas permitirá al futuro profesional trabajar con grandes volúmenes de información lingüística y desarrollar recursos y corpus que puedan ser utilizados en tareas de traducción, análisis lingüístico y procesamiento del lenguaje natural. Asimismo, la asignatura introduce los principios de uso responsable y ético de la información disponible en Internet, prestando especial atención al respeto de las condiciones de uso de los sitios web y a la gestión adecuada de los datos obtenidos.
Prerrequisitos
No hay.



Resultados del proceso de formación y aprendizaje

Conocimientos o contenidos

CN 10	Conoce los elementos básicos del desarrollo de la IA aplicada al lenguaje humano, tales como anotación de diverso tipo, ontologías y reconocimiento de voz.
CN 7	Conoce los fundamentos de la programación en su aplicación al procesamiento del lenguaje natural para la mediación interlingüística.
CN 8	Conoce los procedimientos de inteligencia artificial en aspectos centrales como aprendizaje automático, aprendizaje profundo o procesamiento del lenguaje

Habilidades o destrezas

HB 1	Puede aplicar teorías y conceptos lingüísticos a la comprensión y análisis de la lengua y evaluar la calidad de la comunicación en diferentes contextos.
HB 5	Puede trabajar en equipo con otros profesionales del lenguaje, la comunicación y el PLN.
HB 6	Puede adaptarse a los cambios tecnológicos y a las nuevas tendencias en el campo del multilingüismo, y desarrollar habilidades de aprendizaje autónomo y continuo en dicho ámbito.

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

Bloque 0. Presentación de la asignatura

- Introducción al *Web Scraping* y su aplicación en el ámbito de las tecnologías del lenguaje.
- Fuentes de información textual disponibles en la web.
- Casos de uso en traducción, lingüística y procesamiento del lenguaje natural.

Bloque 1. Estructura y tecnologías de la web

- Fundamentos de HTML y CSS.
- Estructura de una página web y modelo DOM.
- Identificación y selección de elementos para la extracción de información.

Bloque 2. Técnicas y herramientas de Web Scraping

- Introducción al *Web Scraping* con Python.
- Extracción de información mediante **BeautifulSoup** y **Scrapy**.
- Automatización de páginas web mediante **Selenium**.
- Selección de herramientas en función de las características de la fuente.

Bloque 3. Extracción y gestión de datos textuales

- Diseño de procesos automatizados de extracción.
- Extracción de textos y metadatos.



- Almacenamiento y organización de los datos obtenidos.
- Tratamiento y limpieza de datos textuales.
- Preparación de corpus para su posterior análisis y procesamiento lingüístico.

Bloque 4. Uso responsable y ético del Web Scraping

- Condiciones de uso y restricciones de acceso a la información disponible en la web.
- Respeto de términos de servicio y buenas prácticas de extracción.
- Consideraciones éticas en la recopilación y utilización de datos.
- Uso responsable de la información obtenida mediante técnicas automatizadas.

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

Metodologías docentes:

1. Lecciones de carácter expositivo
2. Estudios de casos
3. Metodología de aprendizaje activo y colaborativo
4. Gamificación

Supervisión tutorial

Metodología Presencial: Actividades

- **Clases magistrales expositivas y participativas:**
 - El profesor combinará exposición de los contenidos teóricos y con ejemplos prácticos, tanto matemáticos como de programación.
 - El alumno dispondrá de algunos ejemplos prácticos de código, generado dentro y fuera del aula por el profesor.
- **Ejercicios prácticos y resolución de problemas:**
 - El alumno resolverá problemas planteados por el profesor de forma presencial durante la segunda hora de clase, fomentando las dinámicas de trabajo cooperativo.
- **Sesiones prácticas con uso de software:**
 - Se dedicarán las sesiones prácticas a resolver dudas de la práctica semanal y a que los alumnos finalicen su práctica.
 - Las prácticas se graduarán por dificultad que los alumnos irán implementando conforme vayan finalizando cada hito.
- **Actividades de evaluación continua del rendimiento:** se realizarán pruebas, desarrollarán prácticas complementarias a las semanales y retos gamificados.

CN 10, HB 5, HB 6



Metodología No presencial: Actividades

Las actividades formativas serán:

- Ejercicios prácticos y resolución de problemas:**

- El alumno dispondrá de problemas concretos enfocados a asimilar los conceptos explicados teóricos en la sesión anterior de teoría para desarrollar de forma no presencial.
- La solución de estos problemas será subida a la plataforma la semana siguiente, antes de empezar el nuevo tema.

- Sesiones prácticas con uso de software:**

- Una vez liberada la práctica semanal después de la sesión de teoría correspondiente, el alumno trabajará sobre ella de forma no presencial. El alumno deberá llegar a la sesión presencial de prácticas con los objetivos propuestos en el enunciado al 80%
- En el aula se extenderá el enunciado de manera incremental y se cubrirán los hitos planteados de forma progresiva

- Estudio personal:**

- El objetivo principal del trabajo no presencial es llegar a entender y comprender los conceptos teóricos de la asignatura, así como ser capaz de poner en práctica estos conocimientos para resolver los diferentes tipos de problemas.
- Después de cada explicación teórica el profesor subirá a la web todos los códigos desarrollados y el alumno deberá revisarlos y plantearse cuestiones "Whatif" para asimilar mejor los conceptos teóricos.

CN 10,
HB 6,
CN 7

RESUMEN HORAS DE TRABAJO DEL ALUMNO

HORAS PRESENCIALES			
AF1 - Lecciones de carácter expositivo - Exposición en que el/a docente explica una serie de nociones de forma clara, estructurada y motivadora con mayor o menor grado de participación activa y colaborativa del alumnado. Incluirá presentaciones dinámicas y la participación reglada o espontánea del estudiantado por medio de actividades diversas.	AF2 - Ejercicios prácticos/resolución de problemas - Ejercicios y resolución de problemas planteados por el/a docente a partir de una breve lectura, un material preparado para la ocasión, o cualquier otro tipo de datos o informaciones que supongan un desafío intelectual para el/ la estudiante.	AF3 - Trabajos individuales/grupales - Trabajos de carácter teórico o práctico, individuales o grupales, que implican la lectura de artículos, revistas, informes de investigación, capítulos de libros, informaciones en Internet, etc. y la redacción de una reflexión personal (de diverso calado y extensión) que va más allá de la mera recopilación de la información proveniente de diversas fuentes.	AF4 - Exposiciones individuales/grupales - Presentación ante el/a docente, el resto del alumnado y cualquier otro público invitado o participante. Tiene lugar de forma individual o colectiva. Se valorará el dominio del asunto tratado, la organización conceptual, la claridad expositiva, el respeto y racionalidad de las distintas fases y, caso de ser colectivo el ejercicio, la colaboración activa de cada una de las personas pertenecientes al equipo.
12.00	10.00	15.00	12.00
HORAS NO PRESENCIALES			



AF2 - Ejercicios prácticos/resolución de problemas - Ejercicios y resolución de problemas planteados por el/a docente a partir de una breve lectura, un material preparado para la ocasión, o cualquier otro tipo de datos o informaciones que supongan un desafío intelectual para el/ la estudiante.	AF3 - Trabajos individuales/grupales - Trabajos de carácter teórico o práctico, individuales o grupales, que impliquen la lectura de artículos, revistas, informes de investigación, capítulos de libros, informaciones en Internet, etc. y la redacción de una reflexión personal (de diverso calado y extensión) que va más allá de la mera recopilación de la información proveniente de diversas fuentes.	AF5 - Estudio personal y documentación - Estudio individual que el/la estudiante realiza para comprender, reelaborar y retener un contenido científico con vistas a una posible aplicación en el ámbito de su profesión. Lectura individual de textos de diferente tipo (libros, revistas, artículos sueltos, prensa, publicaciones en Internet, informes sobre experiencias prácticas, etc.) relacionados con las materias de estudio.
10.00	10.00	10.00
CRÉDITOS ECTS: 3,0 (79,00 horas)		

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

El uso de IA para crear trabajos completos o partes relevantes, sin citar la fuente o la herramienta o sin estar permitido expresamente en la descripción del trabajo, será considerado plagio y regulado conforme al Reglamento General de la Universidad.

Actividades de evaluación	Criterios de evaluación	Peso
Exámenes: - Prueba Intersemestral. - Examen Final.	Prueba Intersemestral (20%): comprensión de los conceptos fundamentales de IA y sus aplicaciones. Examen Final (50%): se evaluará el pensamiento computacional y abstracto para la resolución de problemas.	70 %
Sesiones prácticas: - Retos colaborativos. - Trabajos no presenciales. - Prácticas. - Proyectos.	La actitud, participación y realización de las prácticas semanales y los retos planteados en sesiones colaborativas o individuales.	20 %
Proyecto final.	Proyecto final de la asignatura que el alumno expondrá al final del curso.	10 %

Calificaciones

La calificación final en **convocatoria ordinaria** y **extraordinaria** de la asignatura dependerá de la evaluación de las siguientes actividades:



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

GUÍA DOCENTE

2026 - 2027

Nota Final convocatoria ordinaria = 20% Prueba Intersemestral + 50% Examen Final + 20% Prácticas semanales + 10% Proyecto final

- Para aprobar la asignatura los alumnos tienen que obtener al menos 5 puntos sobre 10 en el examen final de la asignatura y en el proyecto final, tanto en la convocatoria ordinaria como en la extraordinaria.
- En caso de que un alumno suspenda el proyecto final, pero apruebe el examen final, se planteará la opción de ofrecer una entrega extra del proyecto en convocatoria extraordinaria.
- Falta de asistencia al 15% o más de las horas presenciales de esta asignatura puede tener como consecuencia la imposibilidad de presentarse a las convocatorias ordinaria y extraordinaria.
- El proyecto final será **INDIVIDUAL**.
Las prácticas semanales se entregan individualmente, aunque en clase se pueda/deba trabajar en equipo.

Convocatoria Extraordinaria

1. **Examen (80% del total):** La nota mínima del examen de extraordinario es de 5. Si la nota es inferior a 5, ésta será la nota final de la asignatura.
2. **Proyecto final (20%):** La nota mínima para superar el ejercicio es de 5.

BIBLIOGRAFÍA Y RECURSOS

En cumplimiento de la normativa vigente en materia de **protección de datos de carácter personal**, le informamos y recordamos que puede consultar los aspectos relativos a privacidad y protección de datos que ha aceptado en su matrícula entrando en esta web y pulsando "descargar"

<https://servicios.upcomillas.es/sedeelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792>