



Procesamiento de Lenguaje Natural (NLP)



Dr. Gaddiel Desirena López

Maestría en Inteligencia Artificial 2021



Presentación

El **procesamiento de lenguaje natural**, abreviado **PLN**—en inglés, *natural language processing*, NLP—es un campo de las ciencias de la computación, de la Inteligencia Artificial y de la lingüística que estudia las interacciones entre las computadoras y el lenguaje humano. Se ocupa de la formulación e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas y máquinas por medio del lenguaje natural, es decir, de las lenguas del mundo. No trata de la comunicación por medio de lenguas naturales de una forma abstracta, sino de diseñar mecanismos para comunicarse que sean eficaces computacionalmente —que se puedan realizar por medio de programas que ejecuten o simulen la comunicación—.

A background image showing a person's hands typing on a laptop keyboard. The person is wearing a plaid shirt. The laptop screen displays a blurred image of a person. The overall scene is dimly lit, with a warm, orange-toned light source visible in the upper left corner, creating a soft glow.

Propósitos Transversales

- Colaborar con los compañeros de clase.
- Generar el hábito de resolver problemas con herramientas computacionales.
- Usar de manera efectiva y eficaz las tecnologías de información para representar e interpretar los conceptos en diferentes formas: numérica, geométrica, algebraica y algorítmica.
- Implementar en Python los conocimientos adquiridos en este curso.

Actitudes y Valores

Se espera que al cursar esta asignatura se desarrolle responsabilidad ante la actividad académica, manifiesta en al menos los siguientes aspectos:

- Participación activa, con compromiso, perseverancia y actitud positiva.
- El cumplimiento de las normas de disciplina establecidas.
- El cumplimiento en tiempo y forma de las actividades que se encomienden como trabajo independiente.
- El desarrollo de espíritu crítico y autocrítico (constructivo) en el análisis del desempeño propio y de los compañeros.



Actitudes y Valores

- El sentido de la ética, evitando, en particular, cometer actos deshonestos en la realización de las actividades evaluativas.
- El desarrollo de la capacidad para identificar características personales al afrontar procesos de aprendizaje y, como consecuencia, para aprender con mayor independencia.
- Diálogo abierto, directo y respetuoso tanto con el profesor como con los compañeros.
- Tolerancia y respeto



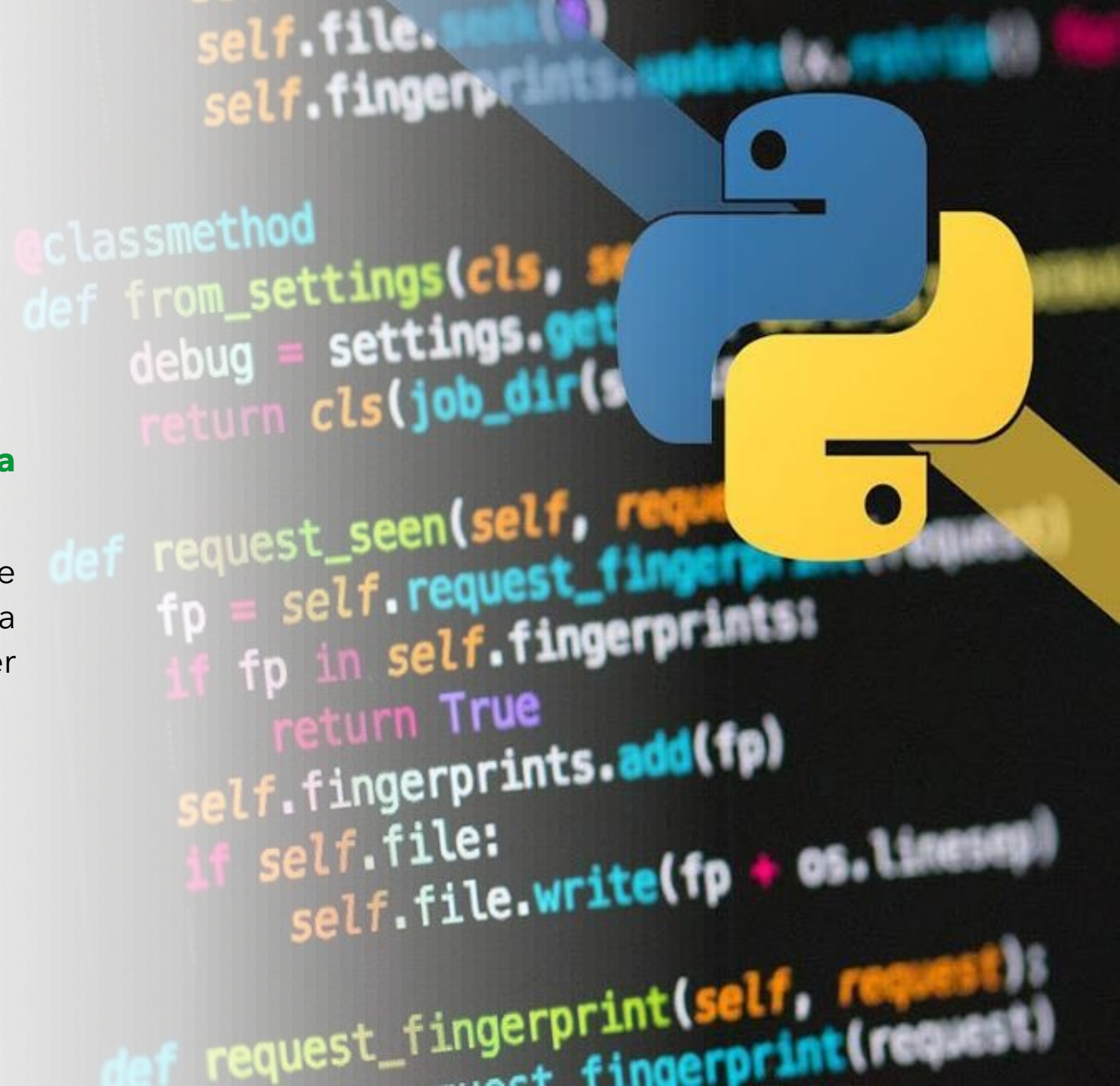
Recursos

- Para tener un desempeño satisfactorio de esta asignatura es indispensable contar con conocimientos de **programación, probabilidad y estadística, aprendizaje automático** . Es deseable manejar algún lenguaje científico de programación, por ejemplo, **Python**.



Recursos

- **Tener instalado Suite de Anaconda**
(<https://www.anaconda.com/>)
- Anaconda viene con una suite de herramientas gráficas llamada "Anaconda Navigator". Trabajaremos con Spyder y Jupyter Notebook.



The background of the slide is a dark, blurred image. On the left, a portion of a black calculator is visible, showing buttons with a plus sign, a minus sign, and a division sign. In the center, there is a line graph with a fluctuating line and numerical values like 6,000, 6,500, and 6,750. On the right side, several coins are stacked. A small orange horizontal bar is located in the top left corner.

¿Qué se espera al terminar el curso?

- Al finalizar con éxito la asignatura se tendrán los conceptos básicos de Procesamiento Natural del Lenguaje. Se tendrán los conocimientos y habilidades sobre clasificación de documentos, traducción automática, extracción de información.

Programa del Curso

4 Módulos

1.- Introducción

- Análisis básico de texto
- Archivos de texto con Python
- Expresiones Regulares
- Librerías de Python (spacy, NLTK)

- Modelo Vectorial de texto
- Preprocesamiento
- Clasificación de textos
- Word **embedding**

2.- Fundamentos de Modelos de Lenguaje

3.- Etiquetado de Secuencia

- Procesos de Markov
- Cadenas de Markov ocultas
- Redes Neuronales Convolucionales para la detección de n-gramas
- Redes Neuronales Recurrentes

- Modelo secuencia-secuencia
- Arquitectura codificador-decodificador
- Capas de atención

4.- NLP con Modelos de secuencia-secuencia

Evaluación del Aprendizaje



Productos

% de Calificación

Proyecto Final

40

Examen

30

Actividades

30

TOTAL:

100%



Políticas y Lineamientos del curso



Clases en línea (Google meet).



Evaluación por proyecto y tareas.



Se realizarán prácticas durante las sesiones de clase



Todas las prácticas serán realizadas en Python.



Políticas y Lineamientos del curso



Entrega de tareas en plataforma ALINCO.



No se recibirán tareas tiempo después de la fecha límite.



Micrófonos apagados excepto si se les avisa lo contrario.