

3011176 - Procesamiento del Lenguaje Natural

JAIME ALBERTO GUZMAN LUNA, Ph.D

FACULTAD DE MINAS
Sede Medellín

SINTELWEB
Grupo de Investigación
Sistemas Inteligentes Web



1

Información de contacto

- Jaime Alberto Guzmán Luna
 - Doctor en Ingeniería de Sistemas e Informática
 - M.Sc. en Ingeniería de Sistemas e Informática
 - Especialista en comunicación educativa
 - Ingeniero Civil
- Áreas de trabajo
 - Inteligencia Artificial
 - Sistemas de recuperación de información
 - Planificación automática de procesos
 - Aprendizaje de máquinas
 - PLN con aprendizaje profundo
- Universidad Nacional de Colombia
 - Oficina M8A-306
 - Email: jaguzman@unal.edu.co
- Horario de Atención virtual (citas previas):
 - Lunes de 2:30 a 4:30 pm
- Apoyo:
 - Alejandro Jiménez Franco
 - Email: aljimenezfr@unal.edu.co



2



Objetivos



- **Objetivo general**
 - El objetivo de este curso es proporcionar a los estudiantes las habilidades y conocimientos fundamentales en el campo del Procesamiento de Lenguaje Natural, un campo de la inteligencia artificial que permite a las máquinas comprender y manipular el lenguaje humano. Los temas abarcan desde el procesamiento de texto básico hasta la comprensión del lenguaje, con aplicaciones en análisis de sentimientos, traducción automática, chatbots y más.
- **Objetivos específicos**
 - Comprender los conceptos clave en Procesamiento del Lenguaje Natural familiarizándose con las bibliotecas populares de PLN (como NLTK, SpaCy, Transformers de Hugging Face, etc.)
 - Familiarizarse con las técnicas y algoritmos utilizados para el procesamiento de texto, análisis de sentimiento, clasificación de texto y extracción de la información.
 - Entender los Modelos de Lenguaje y sus aplicaciones prácticas.
 - Aplicar modelos y herramientas de PLN avanzadas a problemas del mundo real (traducción automática, Preguntas&Respuestas- QA, análisis de conversaciones, etc)

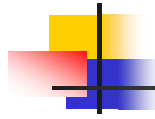
3



Contenido del Curso (1)

- **TEMA 1: Introducción al Procesamiento de Lenguaje Natural.**
 - Generalidades del Procesamiento de Lenguaje Natural y su importancia.
 - Limpieza y normalización de textos en Python
 - Manejo de errores ortográficos: Pyspellchecker
- **TEMA 2: Procesamiento de Texto.**
 - Tareas básicas de procesamiento de texto: tokenización, stopwords, stemming y lematización
 - Bibliotecas populares; NLTK, spaCy y TextBlob
- **TEMA 3: Manejo de corpus de texto.**
 - Modelos Estadísticos de Lenguaje (N-gramas y modelos probabilísticos; modelo TF-IDF; e introducción a la semántica latente-LSA).
 - Creación de corpus desde textos y la Web (text mining y web scraping)
 - Métricas básicas para corpus de texto

4



Contenido del Curso (2)

- **TEMA 4: Análisis de Sentimientos, clasificación y extracción de información en textos.**
 - Introducción al análisis de sentimientos; Métodos supervisados y no supervisados; y Aplicaciones prácticas y casos de estudio
 - Clasificación de Texto con algoritmos de aprendizaje automático (Modelos de clasificación populares: Naive Bayes, SVM, redes neuronales), Evaluación de modelos de clasificación de texto.
 - Extracción de Información, Reconocimiento de entidades nombradas (NER) y Extracción de relaciones entre entidades.
- **TEMA 5: Modelos Avanzados de Representación de Texto.**
 - Word embeddings: Word2Vec, GloVe.
 - Modelos basados en redes neuronales: embeddings de contexto (FastText, ELMo).
 - Transformadores y el modelo BERT.
- **TEMA 6: Generación de Texto y Modelos de Lenguaje.**
 - Redes neuronales recurrentes (RNN) y LSTM.
 - Transformadores para generación de texto (GPT).
 - Aplicaciones prácticas en PLN Traducción automática, chatbots, resumen de texto, QA (Question Answering), análisis de conversaciones.

5



Metodología

- **Actividades presenciales:**
 - Clases magistrales por parte del docente.
 - Talleres: Alumnos en compañía del docente y monitores
 - Exposiciones de los estudiantes
 - Presentación de prácticas por parte de los estudiantes.
- **Actividades Asíncronas:**
 - Videos de temas complementarios
 - Talleres en casa

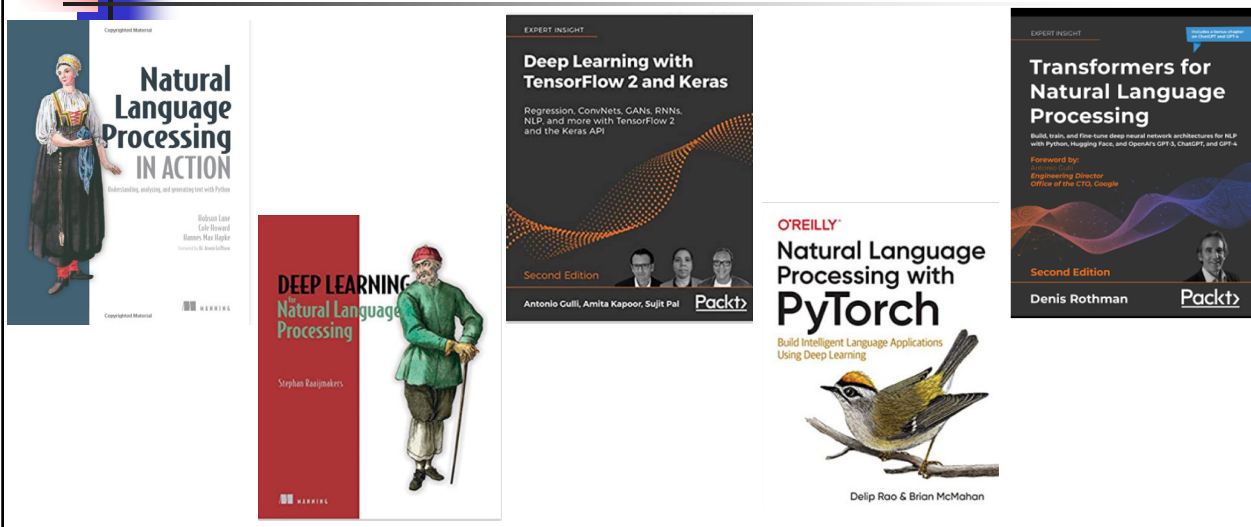
6

Evaluación

- 2 trabajos prácticos:
 - Práctica 1: Aplicativo para el análisis de sentimientos desde textos(25%)
 - Fecha: miércoles 27 de noviembre de 4 a 8 pm.
 - Práctica 2: Aplicativo practico basado en la Generación de Texto y Modelos de Lenguaje (25%)
 - Fecha: miércoles 26 de febrero de 4 a 8 pm.
- Exposiciones sobre temáticas (20%)
 - Fecha miércoles 15 de enero de 4 a 8 pm
- Seguimiento a Talleres (30%)
 - Miércoles 12 m

7

Lecturas y Recursos Recomendados



8