

Código de asignatura: 47AA74

Nombre del programa académico	Maestría en Ingeniería Eléctrica		
Nombre completo de la asignatura	Técnicas de Procesamiento de Lenguaje Natural		
Número de créditos ECTS por categoría	Ciencias naturales y matemáticas	Módulos profesionales y especiales	Humanidades y ciencias sociales y económicas
	3	3	1
Semestre y año de actualización	Semestre 1 – 2021		
Semestre y año en que se imparte	Semestre 1 – Año 2		
Tipo de asignatura	[] Obligatoria [X] Electiva		
Director o contacto del programa	Andrés Escobar Mejía		
Coordinador o contacto de la	Julián David Echeverry Correa		

Descripción y contenidos**1. Breve descripción**

La asignatura de Técnicas de Procesamiento de Lenguaje Natural es de naturaleza teórica-práctica, el propósito que tiene es el estudio de las técnicas probabilísticas, estadísticas y de aprendizaje de máquina relacionadas con el desarrollo de sistemas de procesamiento de lenguaje natural. Se abordan los siguientes temas: procesamiento básico de texto, modelos de lenguaje probabilísticos, modelos de recuperación de información, sistemas de clasificación de textos, sistemas de diálogo (e.g. question-answering, chatbots), sistemas de sentiment analysis y sistemas automáticos de reconocimiento de habla y de traducción.

2. Objetivos. Se espera que al finalizar este curso el estudiante esté en la capacidad de:

- OA1: Introducir al estudiante a los conceptos del procesamiento de lenguaje natural desde la perspectiva de las ciencias computacionales. Se corresponde con los RAP: 9, 10, 11, 12 y 13.

- OA2: Formar al estudiante en habilidades prácticas en el preprocesamiento, preparación y análisis de textos. Se corresponde con los RAP: 11, 12 y 13.

- OA3: Diseñar y construir modelos de aprendizaje de máquina aplicados a tareas de procesamiento de lenguaje. Se corresponde con los RAP: 10, 11 y 13.

- OA4: Analizar y desarrollar sistemas de procesamiento de lenguaje natural. Se corresponde con los RAP: 6, 7, 11, 12 y 13

3. Resultados de aprendizaje. Los propósitos de formación en el estudiante de posgrado son:

- RAA-1. Definir los conceptos fundamentales de las técnicas de procesamiento y análisis de textos. Se corresponde con los RAP: 10, 11 y 12

- RAA-2. Distinguir entre modelos de lenguaje probabilísticos y modelos de lenguaje basados en reglas. Se corresponde con los RAP: 8 y 12

- RAA-3. Identificar los algoritmos óptimos para diversas tareas de clasificación de textos. Se corresponde con los RAP: 8, 10, 11, 12 y 13

- RAA-4. Aplicar modelos de recuperación de información clásicos y algebraicos. Se corresponde con los RAP: 10, 12 y 13

- RAA-5. Resolver problemas de análisis semántico mediante algoritmos computacionales. Se corresponde con los RAP: 10 y 13

- RAA-6. Diseñar sistemas de diálogo a partir del modelado secuencial de agentes conversacionales. Se corresponde con los RAP: 6, 7, 10 y 13

- RAA-7. Aplicar técnicas de modelado estadístico en el desarrollo de sistemas de reconocimiento automático de habla y sistemas de traducción. Se corresponde con los RAP: 6, 7, 8, 10, 11, 12 y 13

4. Contenido

- T-1. Introducción a las técnicas de procesamiento de lenguaje natural (3h)
- T-2. Procesamiento de texto (5h)
- T-3. Modelos de lenguaje probabilísticos (6h)
- T-4. Clasificación de textos (6h)
- T-5. Modelos de recuperación de información (6h)
- T-6: Análisis semántico (4h)
- T-7: Sistemas de diálogo (4h)
- T-8: Reconocimiento automático de habla (4h)
- T-9: Reconocimiento automático de locutor (3h)
- T-10: Alineamiento estadístico y traducción automática (3h)
- T-11: Aplicaciones de los sistemas de procesamiento de lenguaje (4h)

5. Requisitos. Los definidos en requisito de admisión de la IES.

Competencias: Capacidad de resolver problemas que involucren técnicas del álgebra lineal. Adecuada capacidad de lecto-escritura. Capacidad de llevar a cabo una adecuada metodología de investigación.

6. Recursos

Libros de texto:

- [1] Daniel Jurafsky and James H. Martin. Speech and Language Processing. Prentice Hall. 2nd edition. 2009.
- [2] Christopher Manning and Hinrich Schütze. Foundations of Statistical Natural Language Processing. MIT Press. 1999.
- [3] Ricardo Baeza and Berthier Ribeiro-Neto. Modern Information Retrieval. Addison Wesley. 2nd edition. 2011.
- [4] Steven Bird et al. Natural Language Processing with Python. O'Reilly. 2009.
- [5] Witten et al. Data Mining. Morgan Kaufmann. 3rd edition. 2011.
- [6] Christopher Manning et al. Introduction to Information Retrieval. Cambridge University Press. 2008.

Herramientas informáticas:

- Natural Language Toolkit. <http://www.nltk.org/>
- Cloud Natural Language <https://cloud.google.com/natural-language?hl=es>
- Stanford Core NLP. <https://stanfordnlp.github.io/CoreNLP/>
- SpaCy. <https://spacy.io/usage/spacy-101>
- Gensim. <https://github.com/RaRe-Technologies/gensim>
- Mallet. MAchine Learning for Language Toolkit. <http://mallet.cs.umass.edu/>
- Speech Recognition Toolkit. <http://htk.eng.cam.ac.uk/>
- Statistical machine translation system. <http://www.statmt.org/moses/>

Recursos de internet:

- Página web de la asignatura en Google Classroom

7. Herramientas técnicas de soporte para la enseñanza

- Tareas enfocadas al desarrollo de algoritmos para la solución de problemas en el área del procesamiento de lenguaje.
- Desarrollo de sistemas automáticos de procesamiento de información.
- Ejercicios individuales, grupales y tareas propuestas.
- Otras herramientas técnicas presentadas previamente en el ítem 6 (Recursos).

8. Trabajos en laboratorio y proyectos

- Se proponen, en el desarrollo del curso, proyectos de laboratorio y trabajos que relacionen las temáticas tratadas en él.

9. Métodos de aprendizaje

- Clases magistrales apoyadas por tecnologías de la información para la presentación de las clases y como apoyo al proceso de aprendizaje del estudiante.
- Lecturas de artículos especializados y de material adicional propuesto por el profesor como parte del trabajo individual semanal del alumno.
- Talleres relacionados con los contenidos del curso que deben ser resueltos de forma grupal.

10. Métodos de evaluación

- Para la obtención de la nota final se realizan distintas entregas escritas individuales y grupales durante el semestre. Están previstas las siguientes entregas:

- Entrega 1: Taller propuesto para evaluar los objetivos de los temas Procesamiento básico de texto (T2) y Modelos de lenguaje probabilístico (T3). (20%) (RA1 y RA2). Entrega de carácter grupal.
- Entrega 2: Taller propuesto para evaluar los objetivos de los temas Clasificación de textos (T4), Modelos de recuperación de información (T5) y Análisis semántico (T6). (20%) (RA1, RA2, RA3, RA4 y RA5). Entrega de carácter grupal.
- Entrega 3: Anteproyecto en donde se presenten los objetivos del proyecto final, se identifique la problemática a resolver, se plantee la hipótesis de investigación y se enuncien las posibles soluciones metodológicas al problema. (10%) (RA1, RA2, RA3, RA4, RA5, RA6 y RA7). Entrega de carácter individual.
- Entrega 4: Proyecto propuesto para desarrollar un sistema de interacción persona-máquina basado en técnicas de procesamiento de lenguaje natural. Se evalúan los temas Sistemas de diálogo (T7), Reconocimiento automático de habla (T8), Reconocimiento automático de locutor (T9) y Alineamiento estadístico y traducción automática (T10). (50%) (RA1, RA2, RA3, RA4, RA5, RA6 y RA7). Entrega de carácter individual.