

1. Datos Generales de la asignatura

Nombre de la asignatura:	Lenguaje Natural
Clave de la asignatura:	MPSCO-0131
SATCA¹:	2 – 4 - 6
Carrera:	Maestría en Sistemas Computacionales

2. Presentación

Caracterización de la asignatura
La asignatura Lenguaje Natural aporta al perfil de egreso del estudiante de la Maestría en Sistemas Computacionales los conocimientos, habilidades y competencias necesarios para diseñar, desarrollar e implementar soluciones capaces de analizar, comprender y generar información textual de forma automatizada, optimizando la comunicación hombre-máquina y el manejo de grandes volúmenes de datos. La importancia de esta asignatura radica en su capacidad para habilitar el desarrollo de aplicaciones avanzadas como chatbots, sistemas de búsqueda semántica, análisis de sentimientos, extracción automática de información y modelos de lenguaje basados en redes neuronales. Lenguaje Natural se relaciona con la asignatura de Aprendizaje Profundo, ya que esta proporciona las técnicas y modelos necesarios para implementar redes neuronales y transformadores aplicados al procesamiento de texto. Asimismo, se vincula con la asignatura Bases de Datos y Bases de Conocimiento, que ofrece los fundamentos para la gestión, almacenamiento y consulta eficiente de información textual y semántica, permitiendo integrar corpus, ontologías y bases de conocimiento como insumos para las tareas de PLN. Para abordar la asignatura de Lenguaje Natural, el temario se organiza en seis temas principales. El primer tema, Introducción, presenta los fundamentos, aplicaciones y herramientas para el trabajo en PLN. El segundo tema, Expresiones regulares y autómatas, se enfoca en la detección y extracción de patrones textuales. El tercer tema, Análisis sintáctico, aborda el uso de técnicas y herramientas para obtener estructuras gramaticales. El cuarto tema, Análisis de dependencias, introduce métodos para identificar relaciones gramaticales y semánticas entre palabras. El quinto tema, Ontologías, se centra en la representación estructurada del conocimiento y su integración en aplicaciones de PLN. Finalmente, el sexto tema, Modelo de Lenguaje Neuronal y N-gramas, explora la construcción de modelos estadísticos y neuronales para la predicción, clasificación y generación de texto.
Intención didáctica
La extensión de los contenidos antes descritos debe ser adecuada de tal forma que sea posible cubrir los conceptos fundamentales de cada tema, así como proporcionar ejemplos prácticos y casos de estudio que permitan una comprensión más profunda. El enfoque de la asignatura debe ser práctico y orientado al desarrollo de proyectos que permitan a los estudiantes aplicar los conceptos aprendidos bajo un paradigma colaborativo y dar solución a problemas de la vida real.

¹ Sistema de Asignación y Transferencia de Créditos Académicos

El enfoque de la asignatura debe permitir a los estudiantes desarrollar competencias tales como resolución de problemas complejos, trabajo colaborativo, pensamiento crítico y analítico, comunicación efectiva, aprendizaje autónomo y la adaptabilidad y flexibilidad en el uso de diferentes tecnologías y herramientas de desarrollo.

Por su parte, el docente debe actuar como facilitador de aprendizaje, proporcionando orientación y apoyo durante el proceso de aprendizaje, ejemplos claros y casos de estudio que permitan a los estudiantes comprender los conceptos teóricos y su aplicación práctica y retroalimentación constructiva que permita un desarrollo continuo de habilidades y competencias. Debe fomentar la participación, discusión e intercambio de ideas de los estudiantes en el desarrollo de proyectos para la solución de problemas de la vida real. Además, debe estar actualizado con las últimas tendencias y tecnologías en el campo de procesamiento de lenguaje natural.

3. Participantes en el diseño y seguimiento curricular del programa

Lugar y fecha de elaboración o revisión	Participantes	Observaciones
Instituto Tecnológico Superior de Teziutlán. Enero 2025.	Instituto Tecnológico Superior de Teziutlán.	Elaboración de la especialidad de Inteligencia Artificial para la Maestría en Sistemas Computacionales.

4. Competencia(s) a desarrollar

Competencia(s) específica(s) de la asignatura
Diseña e implementa soluciones basadas en técnicas y modelos de Procesamiento de Lenguaje Natural que permiten analizar, comprender y generar información textual, optimizando la comunicación hombre-máquina y el manejo de grandes volúmenes de datos mediante herramientas y tecnologías avanzadas.

5. Competencias previas

• Utiliza herramientas de desarrollo de software y entornos de programación, demostrando habilidades para escribir, probar y depurar código en Python u otros lenguajes aplicables al Procesamiento de Lenguaje Natural. • Identifica los conceptos básicos de manejo y preprocesamiento de datos textuales, tales como limpieza, tokenización y normalización, para preparar corpus que serán utilizados en tareas de PLN. • Comprende los fundamentos de la teoría de lenguajes formales y autómatas, reconociendo su aplicación en la modelación y procesamiento de datos lingüísticos. • Conoce los principios básicos de probabilidad, estadística y álgebra lineal, identificando su relevancia en la construcción y evaluación de modelos de lenguaje estadísticos y neuronales.
--

6. Temario

No.	Temas	Subtemas
1	Introducción	1.1 Historia y evolución del Procesamiento del Lenguaje Natural (PLN) 1.2 Aplicaciones actuales del PLN en la industria y la investigación 1.3 Tipos de datos textuales y su preprocesamiento inicial 1.4 Herramientas y entornos para el desarrollo en PLN (Python, NLTK, SpaCy, Hugging Face)
2	Expresiones regulares y autómatas	2.1 Sintaxis y operadores básicos de expresiones regulares 2.2 Uso de expresiones regulares para extracción y limpieza de datos 2.3 Implementación de autómatas finitos para reconocimiento de patrones 2.4 Integración de expresiones regulares en scripts y pipelines de PLN
3	Análisis sintáctico	3.1 Tipos de gramáticas y su representación formal 3.2 Técnicas de parsing: descendente, ascendente y dependencias 3.3 Uso de librerías para análisis sintáctico (NLTK, SpaCy, Stanford Parser) 3.4 Extracción de información estructurada a partir de árboles sintácticos
4	Análisis de dependencias	4.1 Conceptos básicos de dependencias gramaticales 4.2 Algoritmos y modelos para análisis de dependencias 4.3 Implementación práctica con SpaCy y UDPipe 4.4 Aplicaciones en extracción de relaciones y minería de información
5	Ontologías	5.1 Fundamentos y componentes de una ontología 5.2 Herramientas para modelado y edición de ontologías (Protégé, OWL) 5.3 Integración de ontologías con aplicaciones de PLN 5.4 Casos de uso: búsqueda semántica, chatbots y sistemas de recomendación
6	Modelos de Lenguaje Neuronal y N-gramas	6.1 Fundamentos de los modelos de lenguaje estadísticos y neuronales 6.2 Construcción e implementación de modelos N-grama

		6.3 Introducción a word embeddings y representaciones vectoriales (Word2Vec, GloVe, FastText) 6.4 Uso de modelos preentrenados y transformadores (BERT, GPT) para tareas de PLN
--	--	---

7. Actividades de aprendizaje de los temas

1. Introducción	
Competencias	Actividades de aprendizaje
Específica(s): - Analiza los fundamentos y aplicaciones del Procesamiento de Lenguaje Natural para identificar su potencial en la optimización de la comunicación hombre-máquina y el manejo de grandes volúmenes de datos. Genéricas: - Capacidad de análisis y síntesis. - Habilidad para buscar, procesar y analizar información procedente de fuentes diversas. - Capacidad de investigación.	- Investigar y presentar casos de uso de PLN en diferentes sectores. - Identificar y clasificar distintos tipos de datos textuales presentes en repositorios abiertos. - Instalar y configurar entornos de trabajo en Python para PLN (NLTK, SpaCy, Hugging Face). - Realizar un script de ejemplo que cargue y procese un corpus de texto.
2. Expresiones regulares y autómatas	
Competencias	Actividades de aprendizaje
Específica(s): - Implementa expresiones regulares y autómatas finitos para la detección y extracción de patrones en textos. Genéricas: - Capacidad de análisis y síntesis. - Capacidad de investigación. - Habilidad para buscar, procesar y analizar información procedente de fuentes diversas. - Capacidad de trabajo en equipo.	- Construir expresiones regulares para localizar patrones comunes en textos (fechas, correos, direcciones, etc.). - Integrar expresiones regulares en scripts de Python para limpieza y normalización de datos textuales. - Implementar un autómata finito simple para validar cadenas de texto. - Desarrollar un ejercicio de extracción de información

	específica a partir de documentos no estructurados.
3. Análisis sintáctico	
Competencias	Actividades de aprendizaje
Específica(s): • Aplica técnicas de análisis sintáctico para extraer estructuras gramaticales y relaciones jerárquicas entre los elementos de un texto. Genéricas: • Capacidad de análisis y síntesis. • Capacidad de aplicar los conocimientos en la práctica. • Habilidad para buscar, procesar y analizar información procedente de fuentes diversas. • Capacidad de comunicación oral y escrita.	1. Implementar un parser descendente y uno ascendente sobre un conjunto de oraciones. 2. Usar NLTK o SpaCy para generar árboles sintácticos a partir de textos. 3. Analizar y comparar los resultados de diferentes técnicas de parsing. 4. Extraer frases nominales y verbales para construir un resumen estructurado de un texto.
4. Análisis de dependencias	
Competencias	Actividades de aprendizaje
Específica(s): • Implementa modelos de análisis de dependencias para la identificación de relaciones gramaticales entre palabras en un texto. Genéricas: • Capacidad de análisis y síntesis. • Capacidad de aplicar los conocimientos en la práctica. • Habilidad para buscar, procesar y analizar información procedente de fuentes diversas. • Capacidad de comunicación oral y escrita.	1. Utilizar SpaCy para realizar análisis de dependencias sobre un conjunto de oraciones. 2. Visualizar gráficamente las dependencias gramaticales. 3. Implementar un script que extraiga relaciones sujeto-verbo-objeto a partir de un corpus. 4. Integrar el análisis de dependencias en un pipeline para extracción de información.
5. Ontologías	
Competencias	Actividades de aprendizaje

Específica(s): • Diseña e integra ontologías para la representación y consulta de conocimiento en aplicaciones de PLN. Genéricas: • Capacidad de análisis y síntesis. • Capacidad de aplicar los conocimientos en la práctica. • Habilidad para buscar, procesar y analizar información procedente de fuentes diversas. • Capacidad de comunicación oral y escrita.	1. Modelar una ontología sencilla en Protégé utilizando OWL. 2. Definir clases, propiedades y relaciones para un dominio específico. 3. Consultar una ontología mediante SPARQL. 4. Integrar la ontología en un script que realice búsquedas semánticas sobre un conjunto de datos.
6. Modelo de Lenguaje Neuronal y N-gramas	
Competencias	Actividades de aprendizaje
Específica(s): • Desarrolla e implementa modelos estadísticos y neuronales para la predicción y representación de texto. Genéricas: • Capacidad de análisis y síntesis. • Capacidad de aplicar los conocimientos en la práctica. • Habilidad para buscar, procesar y analizar información procedente de fuentes diversas. • Capacidad de comunicación oral y escrita.	1. Implementar un modelo N-grama sobre un corpus y evaluar su desempeño. 2. Entrenar un modelo de embeddings con Word2Vec o FastText. 3. Utilizar un modelo preentrenado como BERT para realizar clasificación de texto. 4. Comparar el desempeño entre modelos estadísticos y neuronales en una tarea de PLN.

8. Práctica(s)

Con el objetivo de vincular y fortalecer el aprendizaje del saber con el saber hacer se recomienda la elaboración y desarrollo de las siguientes prácticas:

- Tema 1: Introducción
 - Instalación y configuración de entornos para PLN en Python (NLTK, SpaCy, Hugging Face) y carga de un corpus para exploración inicial.
 - Implementación de un script que realice conteo de palabras, frecuencia de términos y visualización de las 10 palabras más comunes de un corpus.
- Tema 2: Expresiones regulares y autómatas

- Implementación de expresiones regulares en Python para extraer correos electrónicos, fechas y números de teléfono de un texto.
- Construcción de un autómata finito simple en Python para validar cadenas de texto según reglas definidas (por ejemplo, formatos de identificadores).
- Tema 3: Análisis sintáctico
 - Uso de NLTK y SpaCy para generar y visualizar árboles sintácticos de un conjunto de oraciones.
 - Implementación de un script que identifique y extraiga frases nominales y verbales de un texto para su posterior análisis.
- Tema 4: Análisis de dependencias
 - Implementación de análisis de dependencias con SpaCy para identificar relaciones gramaticales sujeto-verbo-objeto en un corpus.
 - Visualización de dependencias gramaticales y extracción automática de relaciones para construir una red semántica básica.
- Tema 5: Ontologías
 - Creación de una ontología en Protégé para un dominio específico (por ejemplo, turismo, salud o educación) con clases, propiedades y relaciones.
 - Integración de una ontología en Python usando RDFLib y ejecución de consultas SPARQL para obtener información específica.
- Tema 6: Modelo de Lenguaje Neuronal y N-gramas
 - Implementación de un modelo N-grama en Python para predecir la siguiente palabra en una secuencia.
 - Uso de un modelo preentrenado (por ejemplo, BERT o GPT) para realizar clasificación de texto o análisis de sentimientos sobre un conjunto de datos.

9. Proyecto de asignatura

El objetivo del proyecto que planteé el docente que imparta esta asignatura, es demostrar el desarrollo y alcance de la(s) competencia(s) de la asignatura, considerando las siguientes fases:

- **Fundamentación:** marco referencial (teórico, conceptual, contextual, legal) en el cual se fundamenta el proyecto de acuerdo con un diagnóstico realizado, mismo que permite a los estudiantes lograr la comprensión de la realidad o situación objeto de estudio para definir un proceso de intervención o hacer el diseño de un modelo.
- **Planeación:** con base en el diagnóstico en esta fase se realiza el diseño del proyecto por parte de los estudiantes con asesoría del docente; implica planificar un proceso: de intervención empresarial, social o comunitario, el diseño de un modelo, entre otros, según el tipo de proyecto, las actividades a realizar los recursos requeridos y el cronograma de trabajo.
- **Ejecución:** consiste en el desarrollo de la planeación del proyecto realizada por parte de los estudiantes con asesoría del docente, es decir en la intervención (social, empresarial), o construcción del modelo propuesto según el tipo de proyecto, es la fase de mayor duración que implica el desempeño de las competencias genéricas y específicas a desarrollar.
- **Evaluación:** es la fase final que aplica un juicio de valor en el contexto laboral-profesión, social e investigativo, ésta se debe realizar a través del reconocimiento de

logros y aspectos a mejorar se estará promoviendo el concepto de “evaluación para la mejora continua”, la metacognición, el desarrollo del pensamiento crítico y reflexivo en los estudiantes.

10. Evaluación por competencias

Para evaluar las actividades de aprendizaje se recomienda solicitar:

- Mapas conceptuales o mentales
- Reporte de investigación
- Reportes de prácticas
- Prototipos
- Tablas comparativas
- Estudio de casos
- Exposiciones en clase
- Portafolio de evidencias, entre otros.

Para verificar el nivel del logro de las competencias del estudiante se recomienda utilizar:

- Listas de cotejo.
- Guías de observación
- Rúbricas, entre otros.

11. Fuentes de información

- Lee, R. S. T. (2023). Natural language processing: A textbook with Python implementation. Springer. <https://doi.org/10.1007/978-981-99-1999-8>
- Ara, H. (2024). Mastering natural language processing. Independently Published.
- Paaß, G., & Giesselbach, S. (2023). Foundation models for natural language processing. Springer. <https://doi.org/10.1007/978-3-031-40852-1>
- Vajjala, S., Majumder, B., Gupta, A., & Surana, H. (2024). Practical natural language processing: A comprehensive guide to building real-world NLP systems. O'Reilly Media.
- Tunstall, L., von Werra, L., & Wolf, T. (2024). Natural language processing with transformers: Building language applications with Hugging Face. O'Reilly Media.
- Jurafsky, D., & Martin, J. H. (2023). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (3rd ed., Draft). Prentice Hall.
- Goldberg, Y., & Hirst, G. (2023). Neural network methods in natural language processing. Morgan & Claypool Publishers. <https://doi.org/10.2200/S00979ED2V01Y202008HLT050>