

<p style="text-align: center;">UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA</p>  <p style="text-align: center;">FACULTAD DE INGENIERÍA</p> <p style="text-align: center;">PROGRAMA ANALÍTICO DE LA UNIDAD DE APRENDIZAJE:</p> <p style="text-align: center;">PROCESAMIENTO DE LENGUAJE NATURAL</p>	DES:	Ingeniería
	Programa académico	Doctorado en Ingeniería
	Tipo de materia (Obli/Opta):	Optativa
	Clave de la materia:	DI24OP25
	Semestre:	1, 2, 3
	Área en plan de estudios (B, P y E):	G, E
	Total de horas por semana:	4
	<i>Teoría: Presencial o Virtual</i>	2
	<i>Laboratorio o Taller:</i>	0
	<i>Prácticas:</i>	2
	<i>Trabajo extra-clase:</i>	6
	Créditos Totales:	10
	Total de horas semestre (x 16 sem):	160
	Fecha de actualización:	Marzo 2024
	<i>Prerrequisito (s):</i>	Programación, Cálculo, Álgebra lineal, Probabilidad y Estadística Básicas, Fundamentos de Machine Learning
DESCRIPCIÓN DEL CURSO		
<p>El curso aborda tanto los conceptos fundamentales como avanzados del procesamiento de lenguaje natural (PLN), capacitando a los alumnos para diseñar y desarrollar aplicaciones prácticas. Estas aplicaciones incluyen traducción de idiomas, reconocimiento de entidades nombradas, análisis de sentimientos y modelos de lenguaje, entre otras. Al finalizar el curso, los estudiantes estarán equipados con las competencias necesarias para crear soluciones efectivas en PLN que puedan aplicarse en una variedad de contextos y sectores industriales.</p>		
COMPETENCIAS A DESARROLLAR		
INVESTIGACIÓN		
<p>Difunde con responsabilidad ética y social el conocimiento científico, tecnológico, artístico y/o humanístico que produce de forma objetiva.</p>		
DISEÑO Y GESTIÓN DE INFRAESTRUCTURAS SOSTENIBLES PARA EL DESARROLLO		
<p>Diseña y gestiona infraestructuras seguras, eficientes y sostenibles que promueven el desarrollo socioeconómico y ambiental, integrando conocimientos de áreas como infraestructura para el transporte, estructura y materiales, computación e hidrología. Este diseño y gestión considera la sostenibilidad en todos sus aspectos y se rige por altos estándares éticos y profesionales.</p>		

DOMINIOS	OBJETOS DE ESTUDIO	RESULTADOS DE APRENDIZAJE	METODOLOGÍA	EVIDENCIAS
<p>Análisis Integral de Requerimientos de Infraestructura: Identifica los requerimientos de infraestructura considerando las condiciones del entorno, las necesidades socioeconómicas y los principios del desarrollo sostenible.</p>	<p>1. Representaciones textuales y espacios vectoriales.</p> <p>1.1 Probabilidades en Natural Language Processing (NLP)</p> <p>1.1.1 Probabilidades condicionales (Regla de Bayes)</p> <p>1.1.2 Naïve Bayes para Sentiment Analysis.</p> <p>1.2 Modelos espacio-vectoriales</p> <p>1.2.1 Representaciones vectoriales de palabras y documentos</p> <p>1.2.2 Distancia euclidiana y similitud de coseno.</p> <p>1.2.3 Visualización de vectores n-dimensionales: Principal Component Analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE).</p>	<p>El estudiante será capaz de comunicar de manera ética y socialmente responsable los conceptos fundamentales y aplicaciones prácticas relacionadas con las representaciones textuales y los espacios vectoriales en el procesamiento de lenguaje natural.</p>	<p>Estrategias:</p> <p>Clases magistrales para la presentación de conceptos fundamentales.</p> <p>Ejercicios prácticos de implementación de representaciones vectoriales.</p> <p>Análisis de casos de estudio para comprender aplicaciones prácticas.</p> <p>Recursos didácticos:</p> <p>Presentaciones multimedia.</p> <p>Herramientas de programación para la implementación práctica.</p> <p>Material de lectura complementario sobre representaciones textuales.</p>	<p>Implementación de un modelo de representación vectorial de palabras.</p> <p>Informe técnico que describe la aplicación de representaciones vectoriales en un problema específico de procesamiento de lenguaje natural.</p> <p>Presentación oral de un caso de estudio que ilustra la aplicación práctica de representaciones textuales en NLP.</p>
<p>Aplicación de Criterios y Reglamentos de Diseño Sostenible; Utiliza criterios y reglamentos de diseño sostenible vigentes y apropiados para el tipo de</p>	<p>2. Modelos probabilísticos.</p> <p>2.1 Creando un modelo para autocorrección de palabras</p> <p>2.1.1 Probabilidades de palabras</p>	<p>El estudiante demostrará habilidades para comunicar de forma objetiva los principios y aplicaciones de los modelos probabilísticos en el procesamiento de lenguaje natural,</p>	<p>Estrategias:</p> <p>Sesiones de discusión para profundizar en los fundamentos teóricos.</p> <p>Prácticas guiadas de implementación de modelos probabilísticos.</p>	<p>Desarrollo de un modelo de autocorrección de palabras basado en probabilidades.</p> <p>Evaluación de un modelo de Part of Speech Tagging mediante métricas de desempeño.</p>

<p>infraestructura a desarrollar.</p>	<p>2.1.2 Planteando el problema como alineación de secuencias.</p> <p>2.1.3 Solución como programación dinámica.</p> <p>2.2 Part of Speech Tagging (POS).</p> <p>2.2.1 Cadenas de Markov.</p> <p>2.2.2 Modelos ocultos de Markov</p> <p>2.3 Modelos de Lenguaje</p> <p>2.3.1 N-gramas</p> <p>2.3.2 Probabilidad condicional de secuencias de palabras.</p> <p>2.3.3 Métricas para evaluación de modelos de lenguaje.</p> <p>2.4 Word Embeddings.</p> <p>2.4.1 One-Hot vectors.</p> <p>2.4.2 Significado n-vectorial embebido.</p> <p>2.4.3 Métodos de Word Embeddings.</p> <p>2.4.3.1 No-contextuales.</p> <p>2.4.3.1.1 Word2vec, GloVe, fastText.</p>	<p>asegurando su comprensión ética y socialmente responsable.</p>	<p>Análisis de casos reales de aplicación de modelos en el procesamiento de lenguaje natural.</p> <p>Recursos didácticos:</p> <p>Ejemplos de código para la implementación de modelos.</p> <p>Artículos y estudios de investigación sobre modelos probabilísticos en NLP.</p> <p>Herramientas de software para experimentación y evaluación de modelos.</p>	<p>Presentación de resultados de un proyecto de investigación que utiliza modelos de lenguaje para generar texto de manera automática.</p>
---------------------------------------	--	---	---	--

	<p>2.4.3.2 Contextuales.</p> <p>2.4.3.2.1 BERT, GPT-X, etc.</p> <p>2.4.3.3 EL método Skipgram.</p>			
<p>Modelado Multidisciplinario para la Sostenibilidad: Aplica modelos multidisciplinares para representar las condiciones actuales y futuras de las infraestructuras, considerando aspectos ambientales, sociales y económicos.</p>	<p>3. Modelos secuenciales.</p> <p>3.1 Redes Neuronales en NLP.</p> <p>3.2 Sentiment Analysis con redes neuronales.</p> <p>3.3 Redes Neuronales Recurrentes (RNNs).</p> <p>3.3.1 Arquitecturas vector-to-sequence, sequence-to-vector-sequence-to-sequence.</p> <p>3.3.2 RNNs y el vanishing gradient.</p> <p>3.3.3 Gated Recurrent Unit (GRU) RNNs.</p> <p>3.3.4 Long Short-Term Memory (LSTM) RNNs.</p> <p>3.3.5 Named Entity Recognition (NER) con LSTMs.</p>	<p>El estudiante será capaz de difundir de manera ética y socialmente responsable los conceptos clave y aplicaciones de los modelos secuenciales en el procesamiento de lenguaje natural, promoviendo una comprensión objetiva y responsable.</p>	<p>Estrategias:</p> <p>Talleres prácticos de implementación de redes neuronales recurrentes.</p> <p>Estudio de casos de éxito en el uso de modelos secuenciales en NLP.</p> <p>Sesiones de debate sobre las ventajas y limitaciones de los diferentes enfoques de modelos secuenciales.</p> <p>Recursos didácticos:</p> <p>Notebooks interactivos con ejemplos de implementación.</p> <p>Documentación técnica de frameworks de aprendizaje profundo.</p> <p>Artículos académicos sobre el estado del arte en modelos secuenciales.</p>	<p>Implementación de un modelo de Named Entity Recognition (NER) utilizando redes neuronales recurrentes.</p> <p>Análisis comparativo de diferentes arquitecturas de RNNs para tareas específicas en NLP.</p> <p>Informe de investigación que presenta resultados experimentales sobre el desempeño de modelos secuenciales en tareas de procesamiento de lenguaje natural.</p>
<p>Uso Eficiente de Tecnologías y Herramientas Especializadas:</p>	<p>4. Modelos de Atención.</p>	<p>El estudiante demostrará la capacidad de comunicar de manera</p>	<p>Estrategias:</p> <p>Conferencias invitadas de expertos</p>	<p>Desarrollo de un modelo de Neural Machine</p>

<p>Maneja software especializado para la resolución de problemas de ingeniería con un enfoque en el desarrollo sostenible, reconociendo y abordando sus limitaciones de manera responsable y eficiente.</p>	<p>4.1 Neural Machine Translation (NMT).</p> <p>4.2 Modelos sequence-to-sequence.</p> <p>4.2.1 Encoders-Decoders y el cuello de botella de información.</p> <p>4.2.2 Alineación de palabras y atención.</p> <p>4.2.3 Keys, Queries y Values para recuperación de información en la capa de atención.</p> <p>4.2.4 Beam search para mejores decoders.</p> <p>4.3 Bilingual Evaluation Understudy (BLEU score) como medida de evaluación de calidad en traducción de textos.</p> <p>4.4 Recall-Oriented Understudy for Gisting Evaluation (ROUGE score) como medida de evaluación de calidad en textos generados automáticamente.</p> <p>4.5 Los Transformers.</p> <p>4.5.1 Aplicaciones: Sumarización de textos, auto-complete, NER, question answering,</p>	<p>objetiva y ética los principios y aplicaciones de los modelos de atención en el procesamiento de lenguaje natural, promoviendo una difusión responsable del conocimiento científico.</p>	<p>en atención y modelos de traducción automática.</p> <p>Ejercicios prácticos de implementación de modelos de atención.</p> <p>Estudio de casos de aplicaciones prácticas de modelos de atención en NLP.</p> <p>Recursos didácticos:</p> <p>Tutoriales en video sobre la implementación de modelos de atención.</p> <p>Documentación técnica de modelos de traducción automática.</p> <p>Conjuntos de datos y benchmarks para evaluación de modelos de atención.</p>	<p>Translation (NMT) con atención.</p> <p>Evaluación del desempeño del modelo utilizando métricas de calidad de traducción como BLEU score y ROUGE score.</p> <p>Presentación de un proyecto de aplicación de modelos de atención en NLP, junto con resultados y conclusiones obtenidas.</p>
---	---	---	---	--

	<p>traducción, chatbots, sentiment analysis, text classification, spell-checking.</p> <p>4.5.2 Arquitectura básica.</p> <p>4.5.3 Modelos de lenguaje pre-entrenados basados en Transformers: BERT, GPT-X, T5.</p> <p>4.5.4 Self attention y las matemáticas detrás.</p> <p>4.5.5 Multi-head attention.</p>			
<p>Identifica las necesidades del contexto global en congruencia con los retos de la sociedad del conocimiento.</p>	<p>5. Transfer learning</p> <p>5.1 Embeddings pre-entrenados.</p> <p>5.2 Feature extractors.</p> <p>5.3 Full Fine-tuning.</p> <p>5.4 Adapters.</p> <p>5.4.1 Sequential adapters.</p> <p>5.4.2 Parallel adapters.</p> <p>5.4.3 Adapters for multi-task learning.</p> <p>5.4.4 Adapters for Transformers.</p> <p>5.5 Algunas aplicaciones.</p> <p>5.5.1 Question answering.</p>	<p>El estudiante será capaz de difundir de forma ética y socialmente responsable los conceptos y aplicaciones del aprendizaje por transferencia en el procesamiento de lenguaje natural, asegurando una comunicación objetiva y responsable del conocimiento científico.</p>	<p>Estrategias:</p> <p>Sesiones de revisión de literatura sobre métodos de aprendizaje por transferencia en NLP.</p> <p>Prácticas de adaptación de modelos pre-entrenados a diferentes tareas.</p> <p>Proyectos de investigación sobre aplicaciones innovadoras de transfer learning en NLP.</p> <p>Recursos didácticos:</p> <p>Artículos científicos sobre técnicas de transfer learning en NLP.</p> <p>Repositorios de modelos pre-entrenados y</p>	<p>Adaptación de un modelo pre-entrenado de BERT a una tarea específica de clasificación de texto.</p> <p>Reporte técnico que describe el proceso de fine-tuning y evaluación del modelo adaptado.</p> <p>Presentación de un proyecto de investigación que utiliza técnicas de transfer learning para mejorar el desempeño de modelos de procesamiento de lenguaje natural en un dominio específico.</p>

	<p>5.5.2 Chat-bots</p> <p>5.6 Limitantes de los Transformers</p> <p>5.6.1 Secuencias largas.</p> <p>5.6.2 Complejidad computacional cuadrática de la capa de atención.</p> <p>5.6.3 LSH attention.</p> <p>5.6.4 Reversible Layers.</p> <p>5.6.5 El Reformer, un Transformer más eficiente.</p>		<p>conjuntos de datos públicos.</p> <p>Herramientas de software para fine-tuning y evaluación de modelos pre-entrenados.</p>	
--	--	--	--	--

FUENTES DE INFORMACIÓN	EVALUACIÓN DE LOS APRENDIZAJES
<ul style="list-style-type: none"> • NLTK Documentation.: Documentación oficial de Natural Language Toolkit (NLTK), una biblioteca popular de Python para procesamiento de lenguaje natural. https://www.nltk.org/ • spaCy Documentation. Documentación oficial de spaCy, otra biblioteca de procesamiento de lenguaje natural en Python, conocida por su rapidez y eficiencia. https://spacy.io/ • TensorFlow Text Documentation. Documentación oficial de TensorFlow Text, una biblioteca de TensorFlow específicamente diseñada para tareas de procesamiento de texto. https://www.tensorflow.org/text • Hugging Face Transformers . Repositorio de modelos preentrenados y herramientas para trabajar con Transformers en NLP, incluyendo documentación detallada y ejemplos. https://huggingface.co/transformers/ • Stanford NLP Group. Página oficial del Grupo de Procesamiento de Lenguaje Natural de la Universidad de Stanford, que incluye recursos, investigaciones y herramientas desarrolladas por el grupo. https://nlp.stanford.edu/ • Deep Learning Book. Página oficial del libro "Deep Learning" de Ian Goodfellow, Yoshua Bengio y Aaron Courville, que proporciona recursos adicionales, códigos y enlaces relevantes. http://www.deeplearningbook.org/ • Natural Language Processing with Python. La página oficial del libro "Natural Language 	<p>Promedio de todos los proyectos de cada tema.</p>

<p>Processing with Python" de Steven Bird, Ewan Klein y Edward Loper proporciona recursos complementarios, ejemplos de código y datos de práctica. https://www.nltk.org/book/</p> <ul style="list-style-type: none"> • Hugging Face Blog. El blog de Hugging Face ofrece artículos, tutoriales y actualizaciones sobre los últimos avances en procesamiento de lenguaje natural, incluyendo el uso de Transformers y modelos de lenguaje preentrenados. https://huggingface.co/blog/ • The Allen Institute for AI (AI2). AI2 es una organización líder en investigación en inteligencia artificial que proporciona acceso a herramientas, modelos y recursos en el campo del procesamiento de lenguaje natural. https://allenai.org/ • The Linguistic Data Consortium (LDC). LDC es un consorcio que ofrece una amplia gama de recursos lingüísticos, incluyendo corpus de texto, datos de audio y herramientas para la investigación en procesamiento de lenguaje natural. https://www ldc.upenn.edu/ 	
---	--

Cronograma del avance programático

Objetos de aprendizaje	Semanas															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Representaciones textuales y espacios vectoriales.																
2. Modelos probabilísticos.																
3. Modelos secuenciales.																
4. Modelos de Atención.																
5. Transfer learning.																