

<p style="text-align: center;">UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA</p>  <p style="text-align: center;">UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA</p> <p style="text-align: center;">UNIDAD ACADÉMICA INGENIERÍA</p> <p style="text-align: center;">PROGRAMA DE LA UNIDAD DE APRENDIZAJE:</p> <p style="text-align: center;"><u>PROCESAMIENTO DE LENGUAJE NATURAL</u></p>	DES:	Ingeniería
	Programa académico	Maestría en Ingeniería en Computación
	Tipo de materia (Obli/Opta):	Optativa
	Clave de la materia:	MICOP2315
	Semestre:	2,3,4
	Área en plan de estudios (G y E):	G, E
	Total de horas por semana:	7
	Teoría: Presencial o Virtual	2
	Laboratorio o Taller:	0
	Prácticas:	2
	Trabajo extra-clase:	3
	Créditos Totales:	7
	Total de horas semestre (x 16 sem):	112
	Fecha de actualización:	Agosto 2023
	Prerrequisito (s):	Ninguno

DESCRIPCIÓN DEL CURSO:

El curso describe conceptos fundamentales y avanzados del procesamiento de lenguaje natural, así como también aporta al alumno las competencias necesarias para que pueda diseñar y desarrollar aplicaciones útiles como traducción de idiomas, reconocimiento de entidades nombradas, análisis de sentimientos y modelos de lenguaje entre otras.

COMPETENCIAS A DESARROLLAR:

INTELIGENCIA ARTIFICIAL. Aplica la Inteligencia Artificial para resolver problemas en los sectores industrial, gubernamental, académico y social bajo esquemas de colaboración ética y multidisciplinaria.

GESTIÓN DE PROYECTOS. Coordina y administra de forma responsable, proyectos que atienden criterios de sustentabilidad para contribuir a la mejora de la calidad de vida.

GESTIÓN DEL CONOCIMIENTO. Demuestra conocimientos y habilidades para la búsqueda, análisis crítico, síntesis y procesamiento de información para su transformación en conocimiento, con actitud ética.

COMUNICACIÓN CIENTÍFICA. Difunde con responsabilidad ética y social el conocimiento científico, tecnológico, artístico y/o humanístico que produce de forma objetiva para aportar ideas y hallazgos científicos.

INVESTIGACIÓN. Desarrolla investigación original, tecnología y/o innovaciones en procesos, servicios o productos que contribuyan a la solución de problemas, mejoren la convivencia, generen oportunidades para el desarrollo sustentable y propicien una mejor calidad de vida.

DOMINIOS	OBJETOS DE ESTUDIO	RESULTADOS DE APRENDIZAJE	METODOLOGÍA	EVIDENCIAS
Construye prototipos innovadores utilizando herramientas de software y hardware adecuadas a la solución de	<ol style="list-style-type: none"> 1. Representaciones textuales y espacios vectoriales. <ol style="list-style-type: none"> 1.1. Representaciones dispersas. <ol style="list-style-type: none"> 1.1.1. Extracción de características. 1.1.2. Preprocesamiento 1.1.3. Sentiment Analysis: implementación de un ejemplo usando Logistic 	<ul style="list-style-type: none"> • Aplica los fundamentos de ingeniería y de ingeniería especializada en la identificación, formulación, análisis y 	<ul style="list-style-type: none"> • Para cada unidad, se presenta una introducción por parte del maestro. • Para cada unidad, el maestro deja una tarea donde se aplican los conceptos vistos en clase para 	<ul style="list-style-type: none"> • Reportes de investigación • Proyectos

<p>problemas en cualquier ámbito de desempeño.</p> <p>Comprueba los resultados obtenidos de un prototipo contra las investigaciones recientes a fin de identificar nuevas contribuciones</p> <p>Accede a diferentes fuentes de información (journals o revistas científicas, bases de datos, índices, etc.) de calidad.</p> <p>Analiza y recupera información pertinente mediante diversas estrategias de búsqueda de datos científicos.</p> <p>Evalúa de manera crítica la información, considerando su calidad y pertinencia.</p> <p>Se comunica en forma oral y escrita con propiedad, relevancia, oportunidad y ética.</p> <p>Aplica los elementos fundamentales de la redacción científica.</p> <p>Interpreta y expresa ideas, sentimientos, teorías y corrientes de pensamiento con un enfoque ecuménico (universal).</p> <p>Aplica procesos metodológicos para el desarrollo de investigación o intervención, en</p>	<p>regression.</p> <p>1.2. Probabilidades en Natural Language Processing (NLP)</p> <p>1.2.1. Probabilidades condicionales (Regla de Bayes)</p> <p>1.2.2. Naïve Bayes para Sentiment Analysis.</p> <p>1.2.3. Laplacian Smoothing para evitar probabilidades 0</p> <p>1.2.4. Función Log-likelihood para evitar valores pequeños de probabilidades (underflow)</p> <p>1.2.5. Ejemplos de aplicaciones: ataques adversarios, detección de sarcasmo, ironía y eufemismos.</p> <p>1.3. Modelos espacio-vectoriales</p> <p>1.3.1. Representaciones vectoriales de palabras y documentos</p> <p>1.3.2. Distancia euclidiana y similitud de coseno como medidas de significado.</p> <p>1.3.3. Visualización de vectores n-dimensionales: Principal Component Analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE)</p> <p>1.4. Traducción con vectores de palabras.</p> <p>1.4.1. La norma Frobenius.</p> <p>1.4.2. Hash tables.</p> <p>1.4.3. Local Sensitive Hashing (LSH).</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>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>2.2.3. El algoritmo de Viterbi.</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>resolución de problemas complejos con el fin de alcanzar conclusiones fundamentadas.</p> <ul style="list-style-type: none"> Realiza investigaciones de problemas complejos por métodos que incluyen experimentos apropiados, análisis e interpretación de datos y síntesis de la información con el fin de llegar a conclusiones válidas. 	<p>la resolución de problemas. La tarea requiere que el alumno revise las técnicas y concepto vistos en clase, aclare dudas y aplique las técnicas ya sea manualmente o las implemente utilizando un lenguaje de programación.</p> <ul style="list-style-type: none"> La discusión y el análisis se propician a partir del planteamiento de una situación problemática, donde el estudiante aporte alternativas de solución o resolver un ejercicio en el que aplique conceptos ya analizados. En algunas unidades el maestro muestra directamente en una computadora, posiblemente con la ayuda de un proyector, cómo se implementan las técnicas vistas en clase usando un lenguaje de programación. Material de Apoyo didáctico: Recursos Talleres para realizar ejercicios Materiales gráficos: artículos, libros, diccionarios, etc. Cañón Rotafolio Pizarrón, pintarrones Proyector de acetatos * Plataforma 	
---	---	---	--	--

<p>congruencia con el planteamiento y objetivos del proyecto a abordar</p> <p>Manifiesta capacidad de innovar y creatividad al producir soluciones apropiadas para los contextos en los que se desenvuelve.</p>	<ul style="list-style-type: none"> 2.3.3. Probabilidad logarítmica. 2.3.4. Un modelo de lenguaje generativo. 2.3.5. Perplexity y su variante logarítmica como medida de evaluación de modelos de lenguaje. 2.3.6. Palabras Out of Vocabulary (OOV). 2.4. Word Embeddings. <ul style="list-style-type: none"> 2.4.1. Representaciones por medio de enteros 2.4.2. One-Hot vectors. 2.4.3. Significado n-vectorial embebido. 2.4.4. Métodos de Word Embeddings. <ul style="list-style-type: none"> 2.4.4.1. No-contextuales. <ul style="list-style-type: none"> 2.4.4.1.1. Word2vec, GloVe, fastText. 2.4.4.2. Contextuales. <ul style="list-style-type: none"> 2.4.4.2.1. ELMo, BERT. 2.4.4.3. EL método Skipgram a fondo. 3. Modelos secuenciales. <ul style="list-style-type: none"> 3.1. Redes Neuronales en NLP. 3.2. Frameworks para redes neuronales profundas: Tensorflow, Pythorch, TRAX. 3.3. Sentiment Analysis con redes neuronales. 3.4. Redes Neuronales Recurrentes (RNNs). <ul style="list-style-type: none"> 3.4.1. Arquitecturas one-to-one, one-to-many, many-to-many y many-to-one. 3.4.2. Vanilla RNN 3.4.3. Gated Recurrent Unit GRU RNNs. 3.4.4. Bi-Directional RNNs. 3.4.5. Deep RNNs. 3.4.6. RNNs y el vanishing gradient. 3.4.7. Long Short-Term Memory LSTM RNNs. 3.4.8. Named Entity Recognition (NER) con LSTMs. 3.5. Redes Siamesas 4. Modelos de Atención. <ul style="list-style-type: none"> 4.1. Neural Machine Translation (NMT). 4.2. Modelos sequence-to-sequence. <ul style="list-style-type: none"> 4.2.1. Encoders-Decoders y el cuello de botella de información. 			
---	---	--	--	--

	<ul style="list-style-type: none"> 4.2.2. Alineación de palabras y atención. 4.2.3. Keys, Queries y Values para recuperación de información en la capa de atención. 4.2.4. Beam search para mejores decoders. 4.3. Bilingual Evaluation Understudy (BLEU score) como medida de evaluación de calidad en traducción de textos. 4.4. Recall-Oriented Understudy for Gisting Evaluation (ROUGE score) como medida de evaluación de calidad en textos generados automáticamente. 4.5. Los Transformers. <ul style="list-style-type: none"> 4.5.1. Aplicaciones: Sumarización de textos, auto-complete, NER, question answering, traducción, chatbots, sentiment analysis, text classification, spell-checking. 4.5.2. Arquitectura básica. 4.5.3. Modelos de lenguaje pre-entrenados basados en Transformers: BERT, GPT-3, T5. 4.5.4. Self attention y las matemáticas detrás. 4.5.5. Multi-head attention. 4.5.6. Transfer learning. 4.5.7. Fine-tuning. 4.5.8. Algunas aplicaciones. <ul style="list-style-type: none"> 4.5.8.1. Question answering. 4.5.8.2. Chat-bots 4.5.9. Limitantes de los Transformers <ul style="list-style-type: none"> 4.5.9.1. Secuencias largas. 4.5.9.2. Complejidad computacional cuadrática de la capa de atención. 4.5.9.3. LSH attention. 4.5.9.4. Reversible Layers. 4.5.9.5. El Reformer, un Transformer más eficiente. 			
--	---	--	--	--

FUENTES DE INFORMACIÓN	EVALUACIÓN DE LOS APRENDIZAJES
<p>Deep Learning for NLP and Speech Recognition. Uday Kamath, John Liu, James Whitaker. Springer.</p> <p>Natural Language Processing with Python. Steven Bird, Ewan Klein, Edward Loper. O'reilly.</p> <p>Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow. Aurélien Géron. O'reilly.</p> <p>Natural Language Processing with PyTorch. Delip Rao, Brian McMahan. O'reilly.</p>	<p>Proyectos 100%</p>

CRONOGRAMA

Objetos de aprendizaje	Semanas															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Introducción al Diseño y Análisis de Algoritmos	■	■	■	■												
2. Modelos probabilísticos.					■	■	■	■								
3. Modelos secuenciales.									■	■	■	■				
4. Modelos de Atención.													■	■	■	■