



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA ACADÉMICA
DIRECCIÓN DE EDUCACIÓN SUPERIOR



PROGRAMA SINTÉTICO

UNIDAD ACADÉMICA: UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA CAMPUS COAHUILA (UPIIC), ESCUELA SUPERIOR DE CÓMPUTO (ESCOM), UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA CAMPUS TLAXCALA (UPIIT)	
PROGRAMA ACADÉMICO: Licenciatura en Ciencias de Datos	
UNIDAD DE APRENDIZAJE: Procesamiento de lenguaje natural	SEMESTRE: VI

PROPÓSITO DE LA UNIDAD DE APRENDIZAJE:				
Descubre aspectos relevantes de textos con base la normalización, procesamiento semántico, clasificación y agrupamiento, análisis de personalidad, opinión y relación con otros tipos de datos y extracción de información.				
CONTENIDOS:	I. Contextualización y normalización de textos II. Procesamiento semántico de texto III. Clasificación y agrupamiento de textos IV. Análisis de personalidad y opinión V. Análisis de textos en relación con otros tipos de datos			
ORIENTACIÓN DIDÁCTICA:	Métodos de enseñanza		Estrategias de aprendizaje	
	a) Inductivo	X	a) Estudio de casos	
	b) Deductivo	X	b) Aprendizaje basado en problemas	
	c) Analógico		c) Aprendizaje orientado a proyectos	X
	d) Heurístico	X		
EVALUACIÓN Y ACREDITACIÓN:	Diagnóstica	X	Saberes Previamente Adquiridos	X
	Solución de casos	X	Organizadores gráficos	
	Problemas resueltos		Problemarios	
	Reporte de proyectos	X	Reporte de seminarios	
	Reportes de indagación	X	Otras evidencias a evaluar: presentación electrónica de resultados de indagación, reporte de uso de software, presentación del proyecto.	
	Reportes de prácticas	X		
	Evaluaciones escritas			
BIBLIOGRAFÍA BÁSICA:	Autor(es)	Año	Título del documento	Editorial / ISBN
	Bird, S., Klein, E., Loper, E.	2009	Natural Language Processing with Python	O'Reilly/ 9780596516499
	Eisenstein, J.	2019	Introduction to Natural Language Processing	MIT press/ 9780262042840
	Jurafsky, D., & Martin, J.	2008	Speech and Language Processing	Pearson Prentice Hall/ 9780131873216
	Vajjala, S., Majumder, B., Gupta, A., & Surana, H.	2020	Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems	O'Reilly Media/ 9781492054054
	Zhai, C., Massung, S.	2016	Text Data Management and Analysis	ACM and Morgan & Claypool Publishers/ 9781970001167

*Bibliografía básica



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA ACADÉMICA
DIRECCIÓN DE EDUCACIÓN SUPERIOR



PROGRAMA DE ESTUDIOS

UNIDAD DE APRENDIZAJE: Procesamiento de lenguaje natural

HOJA 2 DE 9

UNIDAD ACADÉMICA: UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA CAMPUS COAHUILA (UPIIC), ESCUELA SUPERIOR DE CÓMPUTO (ESCOM), UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA CAMPUS TLAXCALA (UPIIT)

PROGRAMA ACADÉMICO: Licenciatura en Ciencia de Datos

SEMESTRE:
VI

ÁREA DE FORMACIÓN:
Profesional

MODALIDAD:
Escolarizada

TIPO DE UNIDAD DE APRENDIZAJE:
Teórica-Práctica/Obligatoria

VIGENTE A PARTIR DE:
Agosto 2022

CRÉDITOS:

TEPIC: 7.5

SATCA: 6.7

INTENCIÓN EDUCATIVA

La unidad contribuye al perfil de egreso de la Licenciatura en Ciencia de Datos con el desarrollo de habilidades de procesamiento de lenguaje natural a partir de los procesos de normalización de textos, procesamiento semántico de textos, extracción de información, clasificación y agrupamiento de textos, análisis de personalidad y opinión y análisis de textos en relación con otros tipos de datos. Además, fomenta el desarrollo de habilidades transversales como la toma de decisiones basada en estadística y el trabajo en equipo con un alto sentido ético y de responsabilidad.

Esta unidad se relaciona de manera antecedente con Estadística, Aprendizaje de máquina e inteligencia artificial y Matemáticas avanzadas para ciencia de datos, de manera lateral con Análisis de series de tiempo, Temas selectos de procesamiento de lenguaje natural y Sistemas de información geográfica. y de manera consecuente con Big data

PROPOSITO DE LA UNIDAD DE APRENDIZAJE

Descubre aspectos relevantes de textos con base la normalización, procesamiento semántico, clasificación y agrupamiento, análisis de personalidad, opinión y relación con otros tipos de datos y extracción de información

TIEMPOS ASIGNADOS

HORAS TEORÍA/SEMANA: 3.0

HORAS PRÁCTICA/SEMANA: 1.5

HORAS TEORÍA/SEMESTRE: 54.0

HORAS PRÁCTICA/SEMESTRE:
27.0

**HORAS APRENDIZAJE
AUTÓNOMO:** 32.0

HORAS TOTALES/SEMESTRE: 81.0

**UNIDAD DE APRENDIZAJE
DISEÑADA POR:**

Comisión de Diseño del Programa
Académico

APROBADO POR:

Comisión de Programas
Académicos del H. Consejo
General Consultivo del IPN

dd/mm/aaaa

**AUTORIZADO Y VALIDADO
POR:**

Mtro. Mauricio Igor Jasso
Zaranda
Director de Educación Superior



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA ACADÉMICA
DIRECCIÓN DE EDUCACIÓN SUPERIOR



UNIDAD DE APRENDIZAJE: Procesamiento de lenguaje natural

HOJA 3 DE 9

UNIDAD TEMÁTICA I Contextualización y normalización de texto	CONTENIDO	HORAS CON DOCENTE		HRS AA
		T	P	
UNIDAD DE COMPETENCIA Analiza la normalización a textos en lenguaje natural con base en fundamentos matemáticos, lingüísticos y normalización.	1.1 Contextualización 1.1.1 Sistemas y aplicaciones de lenguaje natural 1.1.2 Fundamentos lingüísticos para procesamiento de lenguaje 1.1.3 Fundamentos matemáticos para procesamiento de lenguaje 1.1.4 Lenguajes de programación, librerías y software para procesamiento de lenguaje	3.0		3.0
	1.2 Normalización de textos 1.2.1 Segmentación en palabras y oraciones 1.2.2 Eliminación de elementos no relevantes 1.2.3 Etiquetado con categorías gramaticales	6.0	3.0	1.5
	Subtotal	9.0	3.0	4.5

UNIDAD TEMÁTICA II Procesamiento semántico de texto	CONTENIDO	HORAS CON DOCENTE		HRS AA
		T	P	
UNIDAD DE COMPETENCIA Extrae aspectos semánticos e información a partir de la representación vectorial de textos, minería de tópicos, análisis sintáctico y expresiones regulares.	2.1 Representación vectorial de textos para su procesamiento semántico 2.1.1 Modelo de bolsa de palabras 2.1.2 Modelo de espacio vectorial para textos	3.0		2.0
	2.2 Selección de características en vectores 2.2.1 Frecuencia y probabilidad de palabras 2.2.2 Funciones de frecuencia de término (TF) 2.2.3 Frecuencia inversa del documento (IDF) 2.2.4 Técnicas de mapeo de palabras a vectores numéricos (word embeddings)	4.5	1.5	2.0
	2.3 Extracción de aspectos semánticos a partir de textos 2.3.1 Palabras similares 2.3.2 Asociaciones entre palabras 2.3.3 Terminología 2.3.4 Palabras clave	4.5	1.5	1.5
	2.4 Minería de tópicos 2.4.1 Representación de tópicos mediante palabras clave 2.4.2 Modelo generativo de tópicos 2.4.3 Algoritmo de Asignación latente de Dirichlet (LDA)	4.5	1.5	2.0
	2.5 Extracción de información a partir de textos 2.5.1 Análisis sintáctico superficial y profundo para extracción de entidades e información. 2.5.2 Expresiones regulares para extracción de información 2.5.3 Resumen automático	3.0	1.5	1.5
	Subtotal	19.5	6.0	9.0



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA ACADÉMICA
DIRECCIÓN DE EDUCACIÓN SUPERIOR



UNIDAD DE APRENDIZAJE: Procesamiento de lenguaje natural

HOJA 4 DE 9

UNIDAD TEMÁTICA III Clasificación y agrupamiento de textos	CONTENIDO	HORAS CON DOCENTE		HRS AA
		T	P	
UNIDAD DE COMPETENCIA Aplica el algoritmo de clasificación o agrupamiento de textos con base a los clasificadores, aprendizaje de máquina supervisado y no supervisado	3.1 Uso práctico de clasificación de textos	0.5		1.5
	3.2 Tipos de aprendizaje de máquina y clasificadores	1.0		1.5
	3.2.1 Clasificadores generativos			
	3.2.2 Clasificadores discriminativos			
	3.3 Aprendizaje de máquina supervisado para clasificación de textos	3.0	1.5	1.5
	3.3.1 Clasificación de textos mediante Bayes ingenuo			
	3.3.2 Clasificación de textos mediante Regresión logística			
	3.3.3 Clasificación de textos mediante K vecinos más cercanos			
	3.4 Aprendizaje de máquina no supervisado para agrupamiento de textos	3.0	1.5	1.5
	3.4.1 Métodos generativos para agrupamiento de textos			
3.4.2 Agrupamiento jerárquico de textos				
3.4.3 Agrupamiento no jerárquico de textos				
3.4.4 Agrupamiento aglomerante de textos				
3.4.5 Agrupamiento de textos mediante Algoritmo de K promedios				
	Subtotal	7.5	3.0	6.0

UNIDAD TEMÁTICA IV Análisis de personalidad y opinión	CONTENIDO	HORAS CON DOCENTE		HRS AA
		T	P	
UNIDAD DE COMPETENCIA Examina aspectos de personalidad y sentimientos en textos a partir de rasgos y detección de polaridad.	4.1 Detección y análisis de aspectos de personalidad en textos	3.0	3.0	1.5
	4.1.1 Edad			
	4.1.2 Sexo			
	4.1.3 Rasgos personales: extraversión, neurosis, simpatía, diligencia, franqueza			
	4.2 Análisis de opinión sobre entidades	3.0	1.5	1.5
	4.2.1 Regresión logística ordinal para opinión			
	4.2.2 Detección de polaridad de opinión con base en diccionarios de polaridad			
	4.2.3 Detección de polaridad de opinión usando algoritmos de aprendizaje de máquina			
	4.3 Análisis de opinión sobre aspectos y características específicas de entidades	3.0	1.5	1.5
	4.3.1 Extracción de aspectos y características			
4.3.2 Minería de opinión sobre aspectos y características				
4.4. Análisis de sentimientos				
4.4.1 Análisis basado en diccionarios de sentimientos	3.0	3.0	1.5	
4.4.2 Análisis basado en aprendizaje de máquina				
	Subtotal	12.0	9.0	6.0



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA ACADÉMICA
DIRECCIÓN DE EDUCACIÓN SUPERIOR



UNIDAD DE APRENDIZAJE: Procesamiento de lenguaje natural

HOJA 5 DE 9

UNIDAD TEMÁTICA V Análisis de textos en relación con otros tipos de datos	CONTENIDO	HORAS CON DOCENTE		HRS AA
		T	P	
UNIDAD DE COMPETENCIA Detecta relaciones entre textos con otro tipo de datos a partir de su análisis de: geolocalización, datos temporales, series de tiempo y redes sociales	5.1 Textos como datos no estructurados y datos estructurados	0.5		1.5
	5.2 Análisis de textos en relación con datos de geolocalización	1.0	1.5	1.0
	5.3 Análisis de textos en relación con datos temporales	1.5	1.5	1.0
	5.4 Análisis de textos en series de tiempo	1.5	1.5	1.0
	5.5 Análisis de mensajes en redes sociales	1.5	1.5	2.0
	Subtotal		6.0	6.0

ESTRATEGIAS DE APRENDIZAJE	EVALUACIÓN DE LOS APRENDIZAJES
Estrategia de Aprendizaje Orientado a Proyectos El alumno desarrollará las siguientes actividades: <ol style="list-style-type: none">1. Desarrollo de conceptos teóricos e indagación documental2. Presentación oral de los resultados de indagación documental.3. Resolución de casos de estudio en equipo e individual de los temas que sean vistos en clase que le permita el análisis de los temas4. Desarrollo de un proyecto donde se implemente los procesos de lenguaje natural vistos en clase.5. Presentación del proyecto6. Uso de software para procesamiento de textos en lenguaje natural.7. Realización de prácticas	Evaluación diagnóstica. Portafolio de evidencias: <ol style="list-style-type: none">1. Reporte de indagación2. Presentación electrónica de resultados de indagación3. Solución de casos4. Reporte de proyecto5. Presentación y reporte de proyecto6. Reporte de uso de software7. Reporte de prácticas



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA ACADÉMICA
DIRECCIÓN DE EDUCACIÓN SUPERIOR



UNIDAD DE APRENDIZAJE: Procesamiento de lenguaje natural

HOJA 6 DE 9

RELACIÓN DE PRÁCTICAS			
PRÁCTICA No.	NOMBRE DE LA PRÁCTICA	UNIDADES TEMÁTICAS	LUGAR DE REALIZACIÓN
1	Normalización de textos	I	Sala de cómputo
2	Generación de representación vectorial de textos	II	
3	Extracción de palabras similares	II	
4	Extracción de asociaciones entre palabras	II	
5	Extracción de terminología y palabras clave	II	
6	Minería de tópicos	II	
7	Extracción de información	II	
8	Generación de resumen	II	
9	Clasificación de textos	III	
10	Agrupamiento de textos	III	
11	Análisis de aspectos de personalidad	IV	
12	Análisis de polaridad en opinión	IV	
13	Análisis de opinión sobre entidades	IV	
14	Análisis de opinión sobre aspectos y características específicas de entidades	IV	
15	Análisis de textos en relación con otros tipos de datos	V	
		TOTAL DE HORAS 27.0	



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA ACADÉMICA
DIRECCIÓN DE EDUCACIÓN SUPERIOR



UNIDAD DE APRENDIZAJE: Procesamiento de lenguaje natural

HOJA 7 DE 9

Bibliografía							
Tipo	Autor(es)	Año	Título del documento	Editorial	Documento		
					Libro	Antología	Otros
C	Bengfort, B., Bilbro, R., & Ojeda, T.	2018	<i>Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning</i>	O'Reilly Media, Inc./ 9781491963043	X		
B	Bird, S., Klein, E., Loper, E.	2009	<i>Natural Language Processing with Python</i>	O'Reilly/ 9780596516499	X		
B	Eisenstein, J.	2019	<i>Introduction to Natural Language Processing</i>	MIT press/ 9780262042840	X		
C	Ganegedara, T.	2018	<i>Natural Language Processing with TensorFlow: Teach Language to Machines Using Python's Deep Learning Library</i>	Packt Publishing Ltd./ 9781788478311	X		
C	Ghosh, S., & Gunning, D.	2019	<i>Natural Language Processing Fundamentals: Build Intelligent Applications that can Interpret the Human Language to Deliver Impactful Results</i>	Packt Publishing Ltd./ 9781789954043	X		
C	Gomez-Perez, J. M., Denaux, R., & Garcia-Silva, A.	2020	<i>A Practical Guide to Hybrid Natural Language Processing: Combining Neural Models and Knowledge Graphs for NLP</i>	Springer Nature/ 9783030448295	X		
B	Jurafsky, D. & Martin, J.	2008	<i>Speech and Language Processing</i>	Pearson Prentice Hall/ 9780131873216	X		
C	Liu, B.	2020	<i>Sentiment analysis: Mining opinions, sentiments, and emotions</i>	Cambridge university press/ 9781107017894	X		
C	Silge, J., & Robinson, D.	2017	<i>Text Mining with R: A Tidy Approach</i>	O'Reilly Media, Inc./ 9781491981658	X		
B	Vajjala, S., Majumder, B., Gupta, A. & Surana, H.	2020	<i>Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems</i>	O'Reilly Media/ 9781492054054	X		
B	Zhai, C., Massung, S.	2016	<i>Text Data Management and Analysis</i>	ACM and Morgan & Claypool Publishers/ 9781970001167	X		



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA ACADÉMICA
DIRECCIÓN DE EDUCACIÓN SUPERIOR



UNIDAD DE APRENDIZAJE: Procesamiento de lenguaje natural

HOJA 8 DE 9

Recursos digitales								
Autor, año, título y Dirección Electrónica	Texto	Simulador	Imagen	Tutorial	Video	Presentación	Diccionario	Otro
Proyecto NLTK. (2022). NLTK: Kit de herramientas de lenguaje natural. Nltk.org. Recuperado el 1 de junio de 2022, de https://www.nltk.org/ .								X
Universidad Politécnica de Cataluña. (2022). MCR - Repositorio central multilingüe TALP: Tecnologías y Aplicaciones del Lenguaje y el Habla. Talp.upc.edu. Recuperado el 1 de junio de 2022, de http://www.talp.upc.edu/demo-detail/440/MRC .							X	
Universidad de Princeton. (2022). WordNet Una base de datos léxica para inglés. Wordnet.princeton.edu. Recuperado el 1 de junio de 2022, de https://wordnet.princeton.edu/ .							X	
Autentia. (2022). Youtube.com. Recuperado el 1 de junio de 2022, de https://www.youtube.com/watch?v=7faeFE6whF0 .					X			
Estudiantes de IA. (2022). Youtube.com. Recuperado el 1 de junio de 2022, de https://www.youtube.com/watch?v=kNC8V7MjKp8					X			
Explosión. (2022). Spacy.io. Recuperado el 1 de junio de 2022, de https://spacy.io/models/es .								X
Kaggle. (2022). Buscar conjuntos de datos abiertos y proyectos de aprendizaje automático Kaggle . Kaggle.com. Recuperado el 1 de junio de 2022, de https://www.kaggle.com/datasets .								X
Google. (2022). TensorFlow. TensorFlow. Recuperado el 1 de junio de 2022, de https://www.tensorflow.org/ .								X
Moreno Sandoval, A. (2022). Llf.uam.es. Recuperado el 1 de junio de 2022, de http://www.llf.uam.es/ESP/pdf/Moreno-Sandoval_Analisis_opinion.pdf .	X							
Sanzón, Y., Vilariño, D., Somodevilla, M., & Zepeda, C. (2022). Modelos para detectar la polaridad de los mensajes en redes sociales . Rcs.cic.ipn.mx. de https://www.rcs.cic.ipn.mx/2015_99/Modelos%20para%20detectar%20la%20polaridad%20de%20los%20mensajes%20en%20redes%20sociales.pdf . Recuperado el 1 de junio de 2022,	X							
Agudo Praena, J. (2018). Clasificación automática de textos Reflexiones sobre el software. Joragupra.com. Recuperado el 1 de junio de 2022, de https://www.joragupra.com/2016/03/clasificacion-automatica-de-textos.html .	X							
Torres-Rondón, A., Hojas-Mazo, W., & Simón-Cuevas, A. (2022). MÉTODO DE DETECCIÓN DE TÓPICOS EN DOCUMENTOS BASADO EN ANÁLISIS CONTEXTUAL DEL CONTENIDO . Puerta de la investigación. Recuperado el 1 de junio de 2022, de https://www.researchgate.net/publication/333642843_METODO_DE_DETECCION_DE_TOPICOS_EN_DOCUMENTOS_BASADO_EN_ANALISIS_CONTEXTUAL_DEL_CONTENIDO .	X							



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA ACADÉMICA
DIRECCIÓN DE EDUCACIÓN SUPERIOR



UNIDAD DE APRENDIZAJE: Procesamiento de lenguaje natural

HOJA 9 DE 9

PERFIL DOCENTE: Ingeniería, Licenciatura y/o Maestría en Ciencias Físico Matemáticas o afines

EXPERIENCIA PROFESIONAL	CONOCIMIENTOS	HABILIDADES DIDÁCTICAS	ACTITUDES
Preferentemente dos años en la profesión en el área de Inteligencia Artificial y/o Procesamiento de lenguaje natural y dos años de docencia a nivel superior.	En inteligencia artificial y procesamiento de lenguaje natural. Del Modelo Educativo Institucional.	Discursivas Cognoscitivas Metodológicas De conducción del grupo Para evaluar	Responsabilidad Tolerancia Honestidad Respeto Paciencia Disciplina Constancia

ELABORÓ

REVISÓ

AUTORIZÓ

Dra. Olga Kolesnikova
Profesora Coordinadora

M. en C. Iván Giovanni Mosso
García
Subdirector Académico ESCOM

M. en C. Andrés Ortigoza Campos
Director ESCOM

Dr. Joel Omar Juárez Gambino
Profesor Colaborador

Ing. Enrique Lima Morales
Subdirector Académico UPIIT

Dr. Edgar Alfredo Portilla Flores
Director UPIIT

Ing. Carlos Alberto Paredes Treviño
Director Interino de UPIIC