

 <p>UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS</p>	<p align="center">UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS</p> <p align="center">FACULTAD DE INGENIERÍA</p> <p align="center">SYLLABUS</p> <p align="center">Página 1 de 9</p>	
---	---	--

Maestría en Ciencias de la Información y las Comunicaciones

-Énfasis (Cuando aplique)

ESPACIO ACADÉMICO (ASIGNATURA): Procesamiento de Lenguaje Natural					
Código del espacio académico:					
Obligatorio		Básico		Complementario	
Electivo	X	Intrínseco		Extrínseco	X
Fecha última actualización		Grupo:			
Número de créditos:	4				
TIPO DE CURSO					
Teórico		Práctico		Teórico-Práctico	X
				Virtual	
ALTERNATIVAS METODOLÓGICAS					
Clase magistral	X	Seminario		Seminario-Taller	X
Taller	X	Prácticas	X	Proyectos tutorados	X
Otro:					
HORARIO					
Día	Horas		Salón		
JUSTIFICACIÓN ESPACIO ACADÉMICO					
<p><i>El procesamiento de lenguaje natural (o conocido mejor por sus siglas en inglés NLP Natural Language Processing) es una rama de la inteligencia artificial, las ciencias de la computación y la lingüística que se ocupa de la construcción de algoritmos y software que sean capaces de procesar y comprender el lenguaje natural (también conocido como texto) de forma similar y en algunos casos mejor, que como lo hacemos los humanos.</i></p> <p><i>El NLP es fundamental en la Inteligencia Artificial ya que los seres humanos pensamos, hablamos, escribimos e incluso soñamos en lenguaje natural. Mas del 80% del conocimiento de la humanidad está en lenguaje natural y cada 18 meses se duplica, lo que hace fundamenta contar con el apoyo del software para su proceso y comprensión.</i></p> <p><i>Finalmente, el NLP es fundamental para la AI ya que la interacción hombre maquina será</i></p>					



UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

FACULTAD DE INGENIERÍA
SYLLABUS

Página 2 de 9

en lenguaje natura, para lo cual es fundamental avanzar en el NLP para que sea lo más cercano a como interactuamos los seres humanos.

CONOCIMIENTOS PREVIOS: Ciencias de la Computación (Programación en Python), Matemáticas (Algebra lineal básica-Vectores y matrices y sus operaciones), Probabilidades y Estadística Básica

PROGRAMACIÓN DE CONTENIDO

OBJETIVO GENERAL

El objetivo de este curso es estudiar y aplicar de forma práctica todos los conceptos relacionados con el NLP en su intersección entre las ciencias de la computación con el Machine Learning y el procesamiento de texto, como también la lingüística en el estudio del lenguaje en los aspectos de: sintaxis, semántica y pragmática y discurso.

OBJETIVOS ESPECÍFICOS

Al finalizar el curso el estudiante estará en capacidad de:

- *Conocer las características básicas de la lingüística del lenguaje en los aspectos: sintaxis, semántica y pragmática y discurso*
- *Conocer que es el procesamiento y análisis de texto desde una perspectiva conexionista (estadística y machine learning), también conocido como text mining o NLP*
- *Conocer las diferentes etapas del text mining o NLP*
- *Conocer diferentes técnicas, herramientas y APIs de text mining desde modelos estadísticos descriptivos básicos, pasando por el machine learning hasta llegar al Deep Learning con los modelos de lenguaje (transformadores)*
- *Conocer las técnicas de machine learning y Deep learning para el análisis predictivo.*
- *Aplicar el NLP en un dominio de negocio utilizando diferentes librerías y modelos de NLP en Python.*

COMPETENCIAS DE FORMACIÓN

Las competencias que se desarrollarán en esta materia son:

contexto: *desarrollar la competencia para identificar las necesidades de la sociedad en lo relacionado con el NLP*

básicas *entender todos los conceptos de algoritmos y matemática relaciona con el NLP y*



UNIVERSIDAD DISTRICTAL
FRANCISCO JOSÉ DE CALDAS

UNIVERSIDAD DISTRICTAL
FRANCISCO JOSÉ DE CALDAS

FACULTAD DE INGENIERÍA
SYLLABUS

Página 3 de 9

su desarrollo en la Inteligencia Artificial y como se conecta el NLP con la lingüística del lenguaje.

laborales el estudiante tendrá la capacidad de desarrollar software y sistemas de NLP para las empresas.

PROGRAMA (UNIDADES TEMÁTICAS Y CONTENIDO DETALLADO)

- *Introducción al NLP*
 - *Definiciones*
 - *Tipos de procesamiento de Texto*
 - *Simbólico*
 - *Conexionista: estadístico y machine learning*
 - *Algo de lingüística*
 - *Sintaxis, Semántica y Pragmática y Discurso*
 - *Organización del texto: corpus, documentos, párrafos, tokens, caracteres*
 - *Procesamiento básico de texto (Python)*
 - *Expresiones regulares*
 - *Distancia entre textos*
- *Etapas del proceso de NLP o Text Mining*
 - *Preparación de textos (Python NLTK y spaCy)*
 - *Tokenización*
 - *Lematización*
 - *Palabras de Parada*
 - *Etiquetado e identificación de estructura*
 - *Modelos computables de sintaxis*
 - *Bag of Words*
 - *Term Matrix Frequence*
 - *TF-IDF*
 - *Procesamiento estadístico descriptivo de textos*
 - *Wordclouds*
- *Machine Learning clásico con textos (Python Pycaret, scikit-learn Tensor Flow)*
 - *Clasificación de textos*
 - *Agrupamiento de textos*
 - *Tópicos seleccionados*
- *Deep Learning y el procesamiento de lenguaje natural*
 - *Introducción a las redes neuronales*
 - *Palabras Imbricadas (Python gensim)*
 - *Word2Vec*



UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

**UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS**

**FACULTAD DE INGENIERÍA
SYLLABUS**

Página 4 de 9

- *Glove*
- *Fasttext*
- *Modelos de lenguaje (Python tensorflow y Hugging Face)*
 - *Rede Neuronales Recurrentes*
 - *LSTM*
 - *Attention is all you need.*
 - *Transformers*
 - *Modelos preentrenados (<http://huggingface.co/>)*
 - *Aplicaciones*
 - *Operaciones clásicas de NLP con Transformadores*
 - *Resúmenes*
 - *Generación de Textos*
 - *Sistemas de QA*

Todos los trabajos prácticos se realizan con Python utilizando librerías OpenSource como se referencian en el contenido temático

ESTRATEGIAS

Metodología Pedagógica y Didáctica:

El desarrollo de la materia está basado en la definición de un caso de aplicación de NLP para su empresa o entorno y con base en lo estudiado en clase, construir el software para su implementación y sustentación final. Paralelo a esto en la clase el profesor presentara ejemplos concretos para estudio y discusión

Tipo de Curso	Horas			Horas profesor/semana	Horas Estudiante/semana	Total, Horas Estudiante/semestre	Créditos
	TD	TC	TA	(TD + TC)	(TD + TC +TA)	X 16 semanas	
	3	1	12	4	16	256	4

Trabajo Presencial Directo (TD): trabajo de aula con plenaria de todos los estudiantes.

Trabajo Mediado Cooperativo (TC): Trabajo de tutoría del docente a pequeños grupos o de forma individual a los estudiantes.

Trabajo Autónomo (TA): Trabajo del estudiante sin presencia del docente, que se puede realizar en distintas instancias: en grupos de trabajo o en forma individual, en casa o en biblioteca, laboratorio, etc.

RECURSOS

Portal Web Institucional (PWI): es el espacio virtual donde se puede compartir de manera pública la información de un área específica. Puede crearlo un profesor para subir



UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

FACULTAD DE INGENIERÍA
SYLLABUS

Página 5 de 9

información de sus espacios académicos o proyectos, pueden crearlas los estudiantes de manera individual o colectiva y los semilleros y grupos de investigación, así como las distintas figuras académicas o administrativas que requieran el uso de este recurso.

Aulas virtuales: son un espacio de aprendizaje virtual donde se comparte información específica de cada área de estudio y está organizada por proyectos curriculares.

Correo Institucional: es el correo electrónico institucional por medio del cual se recibe información propia de la Universidad Distrital y puede ser usado con diferentes fines sin que afecte la seguridad individual o colectiva de las personas o instituciones. Los estudiantes y profesores pueden solicitar su correo institucional y clave en la coordinación a la cual esté adscrito o en la oficina asesora de sistemas en la sede de la Universidad de la Calle 40.

Medios y Ayudas: Estos se refieren tanto a los físicos como humanos necesarios para la actividad pedagógica y didáctica. No sólo se hace referencia a las ayudas audiovisuales: presentación de imágenes de computador, programas o software, sino también a la posibilidad de recursos para salidas de campo trabajo práctico de laboratorio, requerimientos para la logística y el trabajo con invitados o colaborativos con otros docentes en el aula.

Textos Guía y complementarios: hace referencia a los libros que soportan teóricamente el desarrollo de formación. Los complementarios son aquellos libros que no son de carácter prioritario pero que contienen información relevante para complementar el proceso de aprendizaje.

Revistas: Se recomienda para los espacios académicos de las áreas de profundización y/o investigación centralizarse más en artículos de revistas y de bases de datos.

Direcciones de internet: Escribir las direcciones de internet que usted como docente considera relevantes para el buen desarrollo de la formación del estudiante. Adicional, le solicitamos respetuosamente agregar aquí las bases de datos a las que la Universidad Distrital tiene acceso y que contengan información relacionada con el área de estudio.

BIBLIOGRAFÍA

- Daniel Jurafsky and James H. Martin. **Speech and Language Processing**. 2nd Edition. Prentice-Hall, Inc. 2009
- Steven Bird, Ewan Klein, Edward Loper. **Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit**. O'Reilly. 2009
- Christopher D. Manning , Hinrich Schütze. **Foundations of Statistical Natural Language Processing**. MIT Press. Cambridge. 1999
- Nitin Indurkha, Fred J. Damerau. **Handbook of Natural Language Processing (Chapman & Hall/CRC Machine Learning & Pattern Recognition)**. Second



UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

FACULTAD DE INGENIERÍA
SYLLABUS

Página 6 de 9

- Edition. CRC Press. 2010
- Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta (Author), Harshit Surana. **Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems.** O'Reilly. 2020
- Denis Rothman. **Transformers for Natural Language Processing.** Packtpub. 2021
- Savaş Yıldırım, Meysam Asgari-Chenaghlu. **Mastering Transformers.** PacktPub. 2021

REVISTAS

ACM Transactions on Speech and Language Processing

DIRECCIONES DE INTERNET

URLs de las librerías con su documentación

- <https://www.nltk.org/>
- <https://spacy.io/>
- <https://pycaret.org/>
- <https://scikit-learn.org/stable/>
- <https://www.tensorflow.org/>
- <https://huggingface.co/>

ORGANIZACIÓN/TIEMPOS

Espacios, Tiempos, Agrupamientos:

Se recomienda trabajar una unidad cada cuatro semanas, trabajar en pequeños grupos de estudiantes, utilizar Internet para comunicarse con los estudiantes para revisiones de avances y solución de preguntas (esto considerarlo entre las horas de trabajo cooperativo).

Semana/ unidad temática	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Introducción al NLP	X	X	X	X												
2. Text Mining					X	X	X	X								
3. Machine Learning en NLP									X	X	X	X				
4. Deep Learning en NLP													X	X	X	X

EVALUACIÓN

Es importante tener en cuenta las diferencias entre evaluar y calificar. El primero es un



UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

**UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS**

**FACULTAD DE INGENIERÍA
SYLLABUS**

Página 7 de 9

proceso cualitativo y el segundo un estado terminal cuantitativo que se obtiene producto de la evaluación. Para la obtención de la información necesaria para los procesos de evaluación se requiere diseñar distintos formatos específicos de autoevaluación, coevaluación y heteroevaluación.

	TIPO DE EVALUACIÓN	FECHA	PORCENTAJE
PRIMER CORTE	Teórico-Practica		30%
SEGUNDO CORTE	Teórico-Práctica		30%
EXAMÉN FINAL	Proyecto de clase		40%

ASPECTOS A EVALUAR DEL CURSO

Para la evaluación se realizarán 2 exámenes teóricos y para la parte práctica se dejarán cada semana talleres para desarrollar, de esta manera se realizarán las evaluaciones Teórico-Practica, estas evaluaciones serán individuales. El proyecto de clase será grupal para lo cual los estudiantes deberán hacer una sustentación de su resultado.