



Formato para registro de Unidades de aprendizaje 2021

INSTRUCTIVO para el correcto llenado del formato SIP-30

- El formato SIP-30 es un formato digital el cual puede ser completado con un procesador de texto y guardarse como archivo PDF para su envío.
- Adicionalmente será necesario anexar la solicitud firmada por el director de la Unidad Académica respectiva y el acuerdo de Colegio donde se avaló su registro; tenga listos los archivos al momento de ingresar su solicitud en el formulario en línea.
- El enlace de atención única para esta y otras gestiones es: <https://forms.office.com/r/c8DLS6VBv1> (copie y pegue en un navegador web si el enlace no funciona)
- Tome en cuenta los criterios establecidos en el Reglamento de Estudios de Posgrado ([REP 2017](#)) para el llenado de este formato, a continuación, se presentan algunas definiciones útiles:
 - *Número de semanas por semestre del programa*: Es el número de semanas lectivas efectivas al semestre, indicadas en el acuerdo de creación del programa académico o en alguna actualización posterior del programa. En caso de haber tenido una actualización en este sentido, la misma deberá haber sido presentada y avalada en reunión del Colegio de Profesores de la Unidad Académica, además de haber sido aprobada por la SIP. El rango de semanas lectivas al semestre es mínimo 15 y máximo 18.
 - *Tipo de horas*: Las unidades de aprendizaje, en cuanto a las horas asignadas, están clasificadas como: Teóricas, Prácticas y Teórico-prácticas. Estas denominaciones son excluyentes, es decir, las unidades de aprendizaje solo pueden ser de un solo tipo, no pueden tener horas combinadas.
 - *Número de horas – semana*: Es el número de horas asignadas para ser impartida la Unidad de Aprendizaje a la semana.
 - *Total de horas al semestre*: Es el número de horas totales a impartir de la Unidad de Aprendizaje al semestre. Se calcula multiplicando Número de semanas por número de horas-semana.
 - *Créditos* (Reglamento de Estudios de Posgrado 2017): FÓRMULA DE CÁLCULO: $16 \text{ hrs.} = 1 \text{ crédito}$ (horas totales / 16), no deben asignarse fracciones, los créditos deben redondearse a número entero.
- Para el registro de unidades de aprendizaje de modalidad no escolarizada o mixta incluya adicionalmente los campos marcados con el color azul
- En todos los campos existen comentarios en forma de  globo que sirven de ayuda para el requisitado correspondiente, en caso de duda solicite apoyo del asesor didáctico de la UTEyCV de su Unidad Académica.



Formato para registro de Unidades de aprendizaje 2021

I.- Datos de identificación de la unidad de aprendizaje

Unidad académica:	Multisede (CIC, CIDETEC, ESCOM, ESFM, UPIITA)									
Programa académico:	Maestría en Ciencia y Tecnología de Inteligencia Artificial y Ciencia de Datos									
		Doctorado					Orientación profesional			
	x	Maestría				x	Orientado a la investigación			
		Especialidad					Con la industria			
							Especialidad médica			
Nombre de unidad de aprendizaje:	Sesión de colegio donde se propuso:						Fecha de propuesta:			
	Natural Language Processing									
Tipo de unidad de aprendizaje:	Clave de la unidad de aprendizaje:						Créditos:		5 REP 2017	
	Semanas del semestre		18		Horas a la semana:		4		Horas totales: 74	
	Obligatoria:				Optativa:		x		Observaciones:	
	Semestre:		1-3							
	Teórica (%):				Práctica (%):				Teórico-prácticas (%): 100	
Área del conocimiento:	Ingeniería y Ciencias Fisicomatemáticas		x		Ciencias Sociales y Administrativas				Ciencias Médico Biológicas	
									Interdisciplinario	
Modalidad no escolarizada:	No escolarizada				Nombre de la Plataforma:					
	Mixta				Presencial (%):				En plataforma (%):	
Horas establecidas en el programa de estudios:	Presenciales (si procede) (horas x semana)						En plataforma (horas x semana):			



Formato para registro de Unidades de aprendizaje 2021

I. Aprendizajes que el estudiante deberá demostrar al finalizar

Conocimientos	Habilidades y destrezas	Actitudes y valores
• The student will be able to perform language analysis at various levels: lexical and syntactic. • Handling ambiguity at different levels. • The ability to classify texts using simple algorithms.	• Clearly identify the advantages and disadvantages of each acquired knowledge. • Ability to extrapolate the acquired knowledge.	• Desire to independently research the topics covered in class. • Ability to understand and reflect on the topics of the course.

Resolución que aborda la propuesta con su enfoque disciplinar

The goal of Computational Linguistics is for computers to perform useful tasks involving human language, such as enabling human-machine communication, improving human-human communication, or simply processing text and speech. In the era of online information, individuals and institutions are increasingly confronted with large amounts of text that are critical for performing relevant tasks. At the same time, phenomena that were previously undetectable can be studied thanks to the automatic analysis of these large amounts of text. In this course, students will learn the necessary statistical tools that encompass quantitative approaches to natural language processing, including probabilistic modeling, information theory, and linear algebra.

II. Proximidad formativa

Áreas multi, inter y transdisciplinarias	Líneas de Generación y Aplicación de Conocimiento	Sectores sociales
• Computational Cognitive Science • Natural Language Processing	• Aprendizaje automático • Redes neuronales y aprendizaje profundo • Reconocimiento de patrones • Minería de datos, descubrimiento de conocimiento y analítica avanzada • Minería de texto y procesamiento de lenguaje natural	• Improvement of skills in implementing solutions that take into account the diverse sectors of the population, considering the implications that these solutions may have.
Estrategia de asociación:		



Formato para registro de Unidades de aprendizaje 2021

The techniques learned will be applied in solving problems faced by governmental organizations and non-governmental associations with the activity associated with natural language processing.

III Metodología de enseñanza – aprendizaje

Descripción

The units are complemented with audiovisual material to help students acquire the concepts presented in each unit. Throughout the course, the techniques covered in each unit will be implemented. These implementations will be taken into consideration for evaluation, with a weightage of 60%.

Additionally, at the end of each unit, a knowledge exam will be conducted, accounting for 20% of the overall grade. Attendance and student presentations will contribute to the remaining 20% of the grade.

Evidencias como proceso de aprendizaje

The implemented algorithms, summaries, and mind maps of the topics. Projects presentation related to the topics covered in class, as well as their own areas of interest.

Evidencias integradoras (resultados que contribuyen al curriculum)

Ponderación

IV. Descripción de la participación esperada en el estudiante

Receptiva

Resolutiva

Autónoma

Estratégica



Formato para registro de Unidades de aprendizaje 2021

--	--	--	--

Contenido temático

1. Introduction to Natural Language Processing (NLP) (2H)

- 1.1 Definition
- 1.2 Applications
- 1.3 Tools and Resources

2. Preparation of Textual Data (4H)

- 2.1 Data Acquisition
- 2.2 Text Normalization
 - 2.2.1 Tokenization
 - 2.2.2 Lemmatization and Stemming
 - 2.2.3 Removal of Irrelevant Elements
 - 2.2.4 Part-of-Speech Tagging

3. Vector Representation (6H)

- 3.1 Bag of Words
- 3.2 One-Hot Encoding
- 3.3 TF-IDF



Formato para registro de Unidades de aprendizaje 2021

3.4 Word Embeddings

3.4.1 Word2Vec

3.4.2 GloVe

3.4.3 ELMo

4. Semantic Aspect Extraction (10H)

4.1 Language Modeling with n-grams

4.1.1 Language Model Evaluation

4.1.2 Smoothing

4.2 Word Association

4.3 Text Similarity

4.4 Keyword Extraction

4.5 Topic Modeling

5. Deep Learning for Natural Language Processing (18H)

5.1 Recurrent Neural Networks (RNN)

5.1.1 Applications of RNNs in NLP

5.1.2 RNN Architectures

5.1.3 LSTM and GRU Networks

5.1.4 Encoder-Decoder Model

5.2 Transformer Model

5.2.1 Attention Mechanisms

5.2.2 Pretraining

5.2.3 Bidirectional Encoder

5.2.4 Fine-Tuning

5.2.5 Transfer Learning

6. NLP Applications (30H)

6.1 Text Correction



Formato para registro de Unidades de aprendizaje 2021

6.1.1 Spelling 6.1.2 Grammar 6.1.3 Style 6.2 Text Classification 6.3 Sentiment Analysis 6.4 Question Answering 6.5 Chatbots and Dialogue Systems 6.6 Machine Translation 6.7 Automatic Summarization 7. Ethical Considerations for NLP (2H)

V. Secuencia programática

N o .		T e m a		Objetivo de aprendizaje / competencia específica		Tiempo/Horas/Semanas		
						8		
Actividad(es):		No. Descripción de las actividades: Contenidos:				Tipo de interacción(es):		
						Referencias (s):		
Evidencia(s):								



Formato para registro de Unidades de aprendizaje 2021

VI. Habilitadores tecnológicos

Disposiciones		Especificaciones / descripción de efectos
x	Conectividad	
x	Habilidades digitales	
x	Interoperabilidad	
x	Datos abiertos	
x	Servicios de nube	
x	Aprendizaje automático	
	Simulación	
	Realidad aumentada	
	Otro...	

VII. Referencias

Conferencias magistrales

1.
2.
3.

Notas complementarias

Documentales / electrónicas

1. Kumar, Ela. (2011) Natural language processing. IK International Pvt Ltd.
2. Calvo, Hiram. (2013) Procesamiento práctico de lenguaje natural, Editorial SMIA.
3. Gelbukh, A., & Calvo, H. (2018) Automatic syntactic analysis based on selectional preferences (Vol. 765). Springer.
4. Haro, S. N. G., & Gelbukh, A. (2007) Investigaciones en análisis sintáctico para el español. Instituto Politécnico Nacional, Dirección de Publicaciones.
5. Kamath, Uday (2019) Deep learning for NLP and speech recognition. Springer.



Formato para registro de Unidades de aprendizaje 2021

6. Mathangi Sri. (2021) Practical natural language processing with python. Apress.
7. L. Ashok Kumar, D. Karthika Renuka (2023) Deep learning approach for natural language processing, speech, and computer vision. CRC Press

VIII. Créditos y responsabilidades

Responsabilidad	Nombre completo	Clave de nombramiento /No. de empleado
Coordinador (Autor)	Francisco Hiram Calvo Castro	9016-EC-12
Participante (Coautor)	Idalia Maldonado Castillo	
Participante (Coautor)	Joel Omar Juárez Gambino	
Participante (Coautor)	Hind Taud	13748-EE-18
Participante (Coautor)		
Asesor didáctico / Diseñador Instruccional		
Tecnólogo educativo / Comunicólogo		
Corrector de estilo		
Programador multimedia / Diseñador gráfico		
Otro...		



Formato para registro de Unidades de aprendizaje 2021

Por la División de Operación y Promoción al Posgrado de la SIP Nombre _____ FIRMA _____	Por la Subdirección de Diseño y Desarrollo de la DEV Nombre _____ FIRMA _____
--	--

VERIFICACIÓN PARA SU PUESTA EN OPERACIÓN	REVISIÓN TÉCNICO-PEDAGÓGICA PARA LA MODALIDAD
Por la Dirección de Posgrado Nombre _____ FIRMA _____ SELLO DE VALIDACIÓN	Por la Dirección para la Educación Virtual Nombre _____ FIRMA _____