Universidad Nacional de Ingeniería (UNI)
Escuela Profesional de
Ciencia de la Computación
Sílabo 2027-I

1. CURSO

AI262. Procesamiento de Lenguaje Natural (Electivo)

2. INFORMACIÓN GENERAL

2.1 Curso	:	AI262. Procesamiento de Lenguaje Natural
2.2 Ciclo	:	CS300 Ciclo.
2.3 Créditos	:	4
2.4 Horas	:	2 HT; 4 HP;
2.5 Duración del periodo	:	16 semanas
2.6 Condición	:	Electivo
2.7 Modalidad de aprendizaje	:	Presencial
2.8 Prerrequisitos	:	STA251. Estadística y Probabilidades. (2^{do} Sem)

3. PROFESORES

Atención previa coordinación con el profesor

4. INTRODUCCIÓN AL CURSO

Este curso introduce los fundamentos teóricos y prácticos del Procesamiento de Lenguaje Natural (NLP), cubriendo desde técnicas clásicas hasta modelos modernos basados en deep learning. Los estudiantes aprenderán a implementar pipelines completos de NLP para aplicaciones como análisis de sentimientos, traducción automática y chatbots.

5. OBJETIVOS

- Comprender la representación computacional del lenguaje humano.
- Implementar algoritmos para procesamiento de texto en Python.
- Aplicar modelos de NLP a problemas reales usando frameworks modernos.

6. RESULTADOS DEL ESTUDIANTE

AG-C09) Diseño y Desarrollo de Soluciones: Diseña, implementa y evalúa soluciones para problemas complejos de computación. (Usage)

AG-C11) Uso de Herramientas: Aplica herramientas modernas de computación en la resolución de problemas. (Usage)

7. TEMAS

Unidad 1: Fundamentos de NLP (15 horas)	
Resultados esperados: AG-C09,AG-C11	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Tokenización y representación de texto ▪ Modelos de n-gramas y suavizado ▪ Extracción de características (TF-IDF, word embeddings)	▪ Preprocesar texto para modelos de NLP [Usar] ▪ Explicar limitaciones de modelos estadísticos clásicos [Familiarizarse]
Lecturas : [JM21], [MS99]	

Unidad 2: Modelado de Lenguaje (15 horas)	
Resultados esperados: AG-C09,AG-C11	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Word2Vec, GloVe y FastText ▪ Redes neuronales recurrentes (RNNs) para NLP ▪ Modelos de secuencia a secuencia (seq2seq)	▪ Entrenar embeddings de palabras [Usar] ▪ Implementar RNNs básicas para generación de texto [Evaluar]
Lecturas : [Gol17], [Mik+13]	

Unidad 3: Transformers y NLP Moderno (15 horas)	
Resultados esperados: AG-C09,AG-C11	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Arquitectura Transformer (atención auto-supervisada) ▪ BERT y modelos pre-entrenados ▪ Fine-tuning para tareas específicas	▪ Aplicar modelos pre-entrenados con HuggingFace [Usar] ▪ Evaluar métricas de desempeño en NLP [Evaluar]
Lecturas : [Vas+17], [Dev+19]	

Unidad 4: Aplicaciones Prácticas (15 horas)	
Resultados esperados: AG-C09,AG-C11	
Temas	**Aprendizaje esperado (*Learning Outcomes*)**
▪ Análisis de sentimientos con transformers ▪ Sistemas de pregunta-respuesta ▪ Despliegue de modelos con Flask/FastAPI	▪ Construir un pipeline completo de NLP [Usar] ▪ Optimizar modelos para producción [Evaluar]
Lecturas : [JM21], [Cho21]	

8. PLAN DE TRABAJO

8.1 Metodología

Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las diferentes etapas de la evaluación del curso.

8.2 Sesiones Teóricas

Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

8.3 Sesiones Prácticas

Las sesiones prácticas se llevan en clase donde se desarrollan una serie de ejercicios y/o conceptos prácticos mediante planteamiento de problemas, la resolución de problemas, ejercicios puntuales y/o en contextos aplicativos.

9. SISTEMA DE EVALUACIÓN

Sesiones Teóricas:
Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

Sesiones de Laboratorio:
Para verificar que los alumnos hayan alcanzado el logro planteado para cada una de las unidades de aprendizaje, realizarán actividades que les permita aplicar los conocimientos adquiridos durante las sesiones de teoría y se les propondrá retos que permitan evaluar el desempeño de los alumnos.

Exposiciones individuales o grupales:
Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las diferentes etapas de la evaluación del curso.

Lecturas:
A lo largo del curso se proporcionan diferentes lecturas, las cuales podrían ser evaluadas. El uso de herramientas Online permite a cada estudiante acceder a la información del curso, e interactuar fuera de aula con el profesor y con los otros estudiantes.

La nota final NF depende de de las siguientes notas:

- EP: Examen Parcial.
- EF: Examen Final.
- P: Evaluacion permanente (Promedio de prácticas y Trabajos calificados).

La nota final NF se calcula así:

$$NF = \frac{EP + EF + P}{3}$$

10. BIBLIOGRAFÍA BÁSICA

[MS99] Christopher D. Manning e Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[Mik+13] Tomas Mikolov et al. ■Efficient Estimation of Word Representations in Vector Space■. En: *arXiv:1301.3781* (2013).

[Gol17] Yoav Goldberg. *Neural Network Methods for Natural Language Processing*. Morgan & Claypool, 2017.

[Vas+17] Ashish Vaswani et al. ■Attention Is All You Need■. En: *Advances in Neural Information Processing Systems* 30 (2017).

[Dev+19] Jacob Devlin et al. ■BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding■. En: *arXiv:1810.04805* (2019).

[Cho21] François Chollet. *Deep Learning with Python*. 2nd. Manning, 2021.

[JM21] Daniel Jurafsky y James H. Martin. *Speech and Language Processing*. 3rd. Pearson, 2021.