

Introducción al Procesamiento de Lenguaje Natural

Guía de asignatura

Última actualización: agosto de 2022

1. Información general

Nombre de la asignatura	Introducción al procesamiento de Lenguaje Natural
Código	
Tipo de asignatura	Electivo
Número de créditos	3
Tipo de crédito	3A
Horas de trabajo con acompañamiento directo del profesor	64
Horas de trabajo independiente del estudiante	132
Total, horas por periodo académico	196
Prerrequisitos	Ninguno
Correquisitos	
Horario	Miércoles y viernes de 11:00 a.m. a 1:00 p.m.
Líder de área	Edgar Andrade

2. Información del profesor

Nombre del profesor	Fabián Sánchez Salazar
Perfil profesional	Doctor, magister y profesional en matemáticas de la Universidad Nacional con más de 17 años de experiencia docente en cursos de matemáticas y estadística en diferentes niveles. Además, profesor e investigador en el campo de la ciencia de los datos.
Correo electrónico institucional	fabian.sanchez@urosario.edu.co

3. Resumen y propósitos del curso

En este curso los estudiantes aprenderán los elementos básicos que conforman un sistema de procesamiento de lenguaje natural, así mismo como algunos métodos utilizados en este campo.

4. Conceptos fundamentales

- 1) Introducción al PLN. Conceptos generales. Librerías
- 2) Preparación de textos: Python-NLTK. Normalización
- 3) Procesamiento descriptivo de textos. Bolsa de palabras. Embeddings
- 4) Machine Learning y PLN. Clasificación con textos.
- 5) WebScraping. Análisis de sentimientos
- 6) Machine Learning y PLN. Clustering con documentos
- 7) LSI, LDA, Topic Modeling
- 8) Similitud, generación de textos, traducción.
- 9) Word2vec, Glove, Fast Text. Machine Translation
- 10) Modelos Auto-atención
- 11) Modelos de lenguaje: RNN, LSTM.
- 12) Transformers- BERT.

5. Resultados de aprendizaje esperados (RAE)

Se espera que el estudiante desarrolle los siguientes resultados de aprendizaje a medida que avance en el curso:

- Definir las principales características de un sistema de procesamiento de lenguaje natural
- Conocer algunas técnicas básicas utilizadas en este campo.
- Describir lo que son sistemas de dialogo, y entender su implementación.
- Conocer las bases para el desarrollo de un sistema de procesamiento de lenguaje natural.

6. Modalidad del curso

El curso se desarrollará de manera presencial.

7. Estrategias de aprendizaje

El desarrollo del curso se efectuará por medio de la construcción del fundamento teórico a través de la acción facilitadora del docente y la interacción con los estudiantes, para lo cual el estudiante debe desarrollar una lectura previa del contenido a partir de la segunda sesión, de esta manera podrá aportar a la discusión y a la consolidación del fundamento teórico. Una vez conseguido el fundamento teórico, este será reforzado y contextualizado por medio de laboratorios y casos de estudio. También se estimulará la participación del estudiante mediante trabajos de indagación con la debida sustentación ante el profesor y el curso.

Se trabajarán Notebook de Google-Colab o Jupiter para el desarrollo de las sesiones prácticas en la sala de cómputo. Además, se realizarán actividades evaluativas, tales como:

- Talleres de ejercicios
- Elaboración de código (Python)
- Estudio de caso

8. Actividades de evaluación

La evaluación del curso estará dada de la siguiente forma:

- Talleres y actividades en clase 20%.
- Dos Estudios de caso cada uno de 20%. Semanas 5 y 11
- Estudio de caso final 40%.

9. Programación de actividades

Fecha	Tema	Trabajo acompañado por el docente	Trabajo independiente del estudiante
Semana 1	Presentación del programa y guía de asignatura. Reglas del juego del curso Introducción al PLN Conceptos fundamentales Librerías de trabajo: NLTK, Spacy, Gensim. Expresiones regulares Tabla de frecuencias de palabras	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Desarrollar las tareas Hacer los ejercicios propuestos.
Semana 2	Nube de palabras Librería BeautifulSoup Procesamiento básico de texto en Python. Introducción a Normalización de Textos	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Desarrollar las tareas Hacer los ejercicios propuestos.
Semana 3	Preparación de textos <ul style="list-style-type: none"> • Tokenización • Lematización • Palabras de Parada • Etiquetado e identificación de estructura • Bolsa de palabras: TF , TF+ID 	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Desarrollar las tareas Hacer los ejercicios propuestos.

Semana 4	Procesamiento descriptivo de textos. Machine Learning en textos 1 Clasificación	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Hacer los ejercicios propuestos.
Semana 5	Machine Learning en textos 1 Clasificación PRESENTACIÓN ESTUDIO DE CASO 1. (20%)	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Presentar el estudio de caso.
Semana 6	Web Scraping. Análisis de Sentimientos.	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Hacer los ejercicios propuestos.
Semana 7	Web Scraping. Análisis de Sentimientos. Taller Machine Learning en textos 2 Agrupación	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Hacer los ejercicios propuestos.
Semana 8	Machine Learning en textos 2 Agrupación LSI: Latent Semantic Indexing	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Hacer los ejercicios propuestos.
Semana 9	LSI: Latent Semantic Indexing LDA: Latent Dirichlet Allocations Topic Modeling	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Hacer los ejercicios propuestos.
Semana 10	Word embeddings Word2vec Glove Fast Text	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Hacer los ejercicios propuestos.
Semana 11	PRESENTACIÓN ESTUDIO DE CASO 2. (20%) Redes Neuronales y PLN	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Presentar el estudio de caso.
Semana 12	Redes Neuronales RNN y LSTM y PLN Generación de Texto Taller en clase	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Hacer los ejercicios propuestos.
Semana 13	Modelos de auto-atención Seq2seq Machine Translation	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Hacer los ejercicios propuestos.
Semana 14	Taller Transformers BERT	Exposición por parte del docente.	Preparar la sesión Revisar la bibliografía

		Desarrollo de casos y ejercicios guiados en clase.	Hacer los ejercicios propuestos.
Semana 15	Transformers BERT: Bidirectional Encoder Representations from Transformers	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Preparar la sesión Revisar la bibliografía Hacer los ejercicios propuestos.
Semana 16	PRESENTACIÓN ESTUDIO DE CASO FINAL (40%)	Exposición por parte del docente. Desarrollo de casos y ejercicios guiados en clase.	Presentar el estudio de caso.

- Esta es una aproximación al desarrollo semanal, pero puede variar dependiendo del desarrollo de las sesiones y el componente de Software.

10. Factores de éxito para este curso

A continuación, se sugieren una serie de acciones que pueden contribuir, de manera significativa, con el logro de metas y consecuentemente propiciar una experiencia exitosa en este curso:

1. Planificar y organizar el tiempo de trabajo individual que le dedicará al curso
2. Organizar el sitio y los materiales de estudios
3. Tener un grupo de estudio, procurar el apoyo de compañeros
4. Cultivar la disciplina y la constancia, trabajar semanalmente, no permitir que se acumulen temas ni trabajos
5. Realizar constantemente una autoevaluación, determinar si las acciones realizadas son productivas o si por el contrario se debe cambiar de estrategias
6. Asistir a las horas de consulta del profesor, participar en clase, no quedarse nunca con la duda
7. Utilizar los espacios destinados para consultas y resolución de dudas, tales como Sala Gauss y Sala Knuth
8. Propiciar espacios para el descanso y la higiene mental, procurar tener buenos hábitos de sueño
9. Tener presente en todo momento valores como la honestidad y la sinceridad, al final no se trata solo de aprobar un examen, se trata de aprender y adquirir conocimientos. El fraude es un autoengaño.

11. Bibliografía y recursos

[1] Jurafsky, D. & Martin J. (2000) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Upper Saddle River, NJ: Prentice-Hall.

[2] Manning, C. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

[3] Bird, S. & Klein, E. & Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly.

12. Bibliografía y recursos complementarios

[4] Perkins, H. & Joshi, C. (2016). *Natural Language Processing: Python and NLTK*. Packt Publishing Ltd.

[5] Deng and Liu. (2018) *Deep Learning in Natural Language Processing*. Springer Nature Singapore Pte Ltd.

[6] Eisenstein, J. (2019). *Natural Language Processing*. MIT Press.

13. Acuerdos para el desarrollo del curso

No está permitido comer o usar dispositivos móviles dentro de clase. No se realizará aproximación de notas al final del semestre. Las notas solo serán cambiadas con base en reclamos OPORTUNOS dentro de los límites de tiempo determinados por el Reglamento Académico. Si por motivos de fuerza mayor el estudiante falta a algún parcial, entrega de proyecto o quiz, deberá seguir el procedimiento regular determinado por el Reglamento Académico para presentar supletorios. No habrá acuerdos informales al respecto. No se eximirá a ningún estudiante de ningún examen.

Asistencia al Curso

Con el propósito de afianzar el modelo pedagógico contemplado en el Proyecto Educativo Institucional y promover un rendimiento académico óptimo, es necesario asegurar un espacio de interacción entre estudiantes y profesores que facilite la reflexión y el debate académico en torno al conocimiento. En este sentido, se valora la participación en las actividades académicas y esta se considera como un deber y un derecho del estudiante. (Artículo 48 Reglamento Académico).

Procesos Disciplinarios-Fraude En Evaluaciones

Teniendo en cuenta el reglamento formativo-preventivo y disciplinario de la Universidad del Rosario, y la certeza de que las acciones fraudulentas van en contra de los procesos de enseñanza y aprendizaje, cualquier acto corrupto vinculado a esta asignatura será

notificado a la secretaría académica correspondiente de manera que se inicie el debido proceso disciplinario. Se recomienda a los estudiantes leer el reglamento para conocer las razones, procedimientos y consecuencias que este tipo de acciones pueden ocasionar, así como sus derechos y deberes asociados a este tipo de procedimientos.

14. Respeto y no discriminación

A continuación, encontrará unas orientaciones institucionales básicas que sugerimos mantener en su guía de asignatura. Puede ampliar esta información si lo considera pertinente:

Si tiene alguna discapacidad, sea este visible o no, y requiere algún tipo de apoyo para estar en igualdad de condiciones con los(as) demás estudiantes, por favor informar a su profesor(a) para que puedan realizarse ajustes razonables al curso a la mayor brevedad posible. De igual forma, si no cuenta con los recursos tecnológicos requeridos para el desarrollo del curso, por favor informe de manera oportuna a la Secretaría Académica de su programa o a la Dirección de Estudiantes, de manera que se pueda atender a tiempo su requerimiento.

Recuerde que es deber de todas las personas respetar los derechos de quienes hacen parte de la comunidad Rosarista. Cualquier situación de acoso, acoso sexual, discriminación o matoneo, sea presencial o virtual, es inaceptable. Quien se sienta en alguna de estas situaciones puede denunciar su ocurrencia contactando al equipo de la Coordinación de Psicología y Calidad de Vida de la Decanatura del Medio Universitario (Teléfono o WhatsApp 322 2485756).