

EX-2024-00605830- -UNC-ME#FAMAF

PROGRAMA DE ASIGNATURA	
ASIGNATURA: Minería de Datos para Texto	AÑO: 2024
CARACTER: Optativa	UBICACIÓN EN LA CARRERA: 5° año 2° cuatrimestre
CARRERA: Licenciatura en Ciencias de la Computación	
REGIMEN: Cuatrimestral	CARGA HORARIA: 120 horas.

ASIGNATURA: Minería de Datos para Texto	AÑO: 2024
CARACTER: Optativa	UBICACIÓN EN LA CARRERA: 4° año 2° cuatrimestre
CARRERA: Licenciatura en Matemática Aplicada	
REGIMEN: Cuatrimestral	CARGA HORARIA: 120 horas.

FUNDAMENTACIÓN Y OBJETIVOS

En esta materia se presentan métodos y técnicas de análisis exploratorio de datos y aprendizaje no supervisado con especial atención a sus aplicaciones en procesamiento del lenguaje natural. Se proveen los fundamentos teóricos y metodológicos de aproximaciones como las reglas de asociación, clustering o embeddings, y se desarrollan aplicaciones prácticas de estas técnicas en diferentes problemas de procesamiento del lenguaje natural. Se aborda la cuestión de la evaluación en aprendizaje no supervisado, y se proveen diferentes métodos para ello.

CONTENIDO

Introducción a la minería de datos, análisis exploratorio de datos, aprendizaje no supervisado

Inteligencia artificial, aprendizaje automático, aprendizaje supervisado y no supervisado. Minería de datos. Aprendizaje semi-supervisado.

Introducción al procesamiento del lenguaje natural

El lenguaje natural como objeto de estudio. Aplicaciones del tratamiento automático del lenguaje natural. Análisis por niveles del lenguaje natural. Generación. Métodos no supervisados para delimitar palabras, crear y enriquecer lexicones, análisis morfológico, análisis sintáctico y semántico.

Evaluación

Métricas. Limitaciones y fortalezas de las métricas. Concursos abiertos. Testbeds. Complemento entre análisis cualitativo y cuantitativo.

Reglas de asociación y Clustering

Correlación, significatividad. Diferentes formas de clustering: aglomerativo, jerárquico. Distancias. Combinaciones. Clustering como embedding.

Modelos de Lenguaje y Embeddings

Reducción de dimensionalidad y acercamiento a causas latentes mediante métodos proyectivos. PCA. ICA. Autoencoders. Embeddings neuronales. Embeddings de modelos de lenguaje. Modelos de lenguaje

El entorno de los aprendizajes automáticos

Representation learning. Transfer learning. Weak supervision. Agentes.

Cuestiones éticas



UNC

Universidad
Nacional
de Córdoba

FAMAF

Facultad de Matemática,
Astronomía, Física y
Computación

EX-2024-00605830- -UNC-ME#FAMAF

Inteligencia artificial responsable. Ética de la inteligencia artificial. Impactos sociales y ambientales. Equidad y sesgo. Métricas de equidad. Exploraciones.

BIBLIOGRAFÍA

BIBLIOGRAFÍA BÁSICA

C. Manning, H. Schütze. 1999. Foundations of Statistical Natural Language Processing. MIT Press.

Daniel Jurafsky and James H. Martin. 2000. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (1st. ed.). Prentice Hall PTR, USA.

BIBLIOGRAFÍA COMPLEMENTARIA

R. Barzilay, K. McKeown. 2001. Extracting Paraphrases from a Parallel Corpus. {it Proceedings of the Meeting of the Association for Computational Linguistics 2001} D. Brown et al. 1993. The Mathematics of Statistical Machine Translation. Computational Linguistics, 1993.

K. Church, P. Hanks. 1990. Word Association Norms, Mutual Information, and Lexicography. Computational Linguistics Vol. 16 (1), pp.22-29-

I. Goodfellow, Y. Bengio y A. Courville (2016). Deep Learning. MIT Press T.K. Landauer, S.T. Dumais. 1997. A Solution to Plato's Problem: The Latent Semantic Analysis: Theory of Acquisition, Induction and Representation of Knowledge. Psychological Review

NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook> D. Yarowsky. 1997. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. Proceedings of the Meeting of the Association for Computational Linguistics 1997

EVALUACIÓN

FORMAS DE EVALUACIÓN

La evaluación consistirá en la entrega de dos informes de trabajos prácticos, un informe de proyecto y defensa oral del proyecto, que incluye examen oral con preguntas teórico-prácticas sobre el programa de la materia.

REGULARIDAD

Aprobar al menos dos evaluaciones parciales o sus correspondientes recuperatorios.

PROMOCIÓN

Aprobar todas las evaluaciones parciales con una nota no menor a 6 (seis), y obteniendo un promedio no menor a 7 (siete) y aprobar el coloquio.

CORRELATIVIDADES

Para Licenciatura en Matemática Aplicada: (aprobadas)

Algoritmos y Estructuras de Datos

Probabilidad y Estadística

Para Licenciatura en Ciencias de la Computación:

Para cursar:

Modelos y Simulación (regular)

Algoritmos y Estructuras de Datos II; Probabilidad y Estadística (aprobadas)

Para rendir: (aprobadas)

Algoritmos y Estructuras de Datos II; Modelos y Simulación