Procesado del Lenguaje Natural

Curso Académico: (2022 / 2023) **Fecha de revisión: 20-05-2022**

Departamento asignado a la asignatura:

Coordinador/a: ARENAS GARCIA, JERONIMO

Tipo: Optativa Créditos ECTS : 3.0

Curso : 1 Cuatrimestre : 2

REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

* Se recomienda haber superado la asignatura de Aprendizaje Automático
* La asignatura de Aprendizaje Profundo también proporciona competencias de interés, si bien no resulta imprescindible haberla cursado. Durante las primeras sesiones del curso se revisarán los conceptos necesarios para el cálculo de representaciones de palabras y documentos basadas en redes neuronales

OBJETIVOS

* Familiarizar a los estudiantes con algunos métodos de uso habitual para el procesamiento de lenguaje natural, tanto para el preprocesado de texto sin estructura, como para la construcción de modelos basados en aprendizaje automático
* Conocer diversas aproximaciones para el cálculo de la similitud semántica entre documentos y su uso para construir y analizar grafos semánticos
* Presentación de algunas herramientas para la visualización interactiva de modelos de aprendizaje automático y procesamiento de lenguaje natural basadas en grafos y dashboards interactivos
* Familiarizar a los estudiantes con algunas aplicaciones relevantes del procesamiento del lenguaje natural
* Favorecer la madurez en el conocimiento de estas tecnologías, y la autonomía para profundizar los conceptos explicados en clase, mediante el trabajo en un proyecto final grupal

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

1. Introducción al Procesamiento de Lenguaje Natural
2. Representación vectorial de palabras y documentos
   2.1. One-hot encoding
   2.2. Word Embeddings. Word2Vec. GloVe
   2.3. Otras representaciones
3. Preprocesado de Textos
   3.1. Adquisición de corpus documentales y parseado de documentos
   3.2. Pipelines de preprocesado
   3.3. Homogeneización y limpieza de textos
   3.4. Reconocimiento de entidades nombradas
   3.5. Spacy y Spark NLP
4. Modelado de tópicos
   4.1. Latent Semantic Indexing
   4.2. Latent Dirichlet Allocation
   4.3. Visualización. Creación de Dashboards BI Enriquecidos
5. Transformers
   5.1. Introducción a los Transformers. Hugging Face
   5.2. Clasificación de Textos: Análisis de sentimiento
   5.3. Clasificación Zero-shot
   5.4. Generation de texto
   5.5. Traducción Automática Neuronal
   5.6. Question & Answering
6. Análisis semántico
   6.1. Métricas de similitud semántica
   6.2. Grafos Semánticos
   6.3. Análisis de Grafos
   6.4. Visualización de Grafos
   6.5. Sistemas Semánticos para recuperación de la información

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

Se emplearán las siguientes actividades y metodologías de aprendizaje:

- Clases combinadas magistrales-laboratorio: las clases magistrales brindan una descripción general de los principales conceptos teóricos y matemáticos del procesamiento del lenguaje natural junto con las herramientas analíticas. En estas clases, se introducirán ejemplos de laboratorio como parte de las exposiciones teóricas: todas las sesiones formativas se llevarán a cabo en el laboratorio (siempre que haya disponibilidad) para imbricar ejemplos prácticos dentro de las explicaciones y agregar dinamismo a la clase.

- Proyecto final: El alumno trabajará en un proyecto en el que programará un sistema modular completo de una de las herramientas explicadas en clase. Los estudiantes recibirán algunas pautas y algunas sesiones preparatorias mediante el uso del aprendizaje basado en problemas.

Los profesores estarán disponibles durante 2 horas a la semana para tutorías

SISTEMA DE EVALUACIÓN

Convocatoria ordinaria:
- Participación en clase y Tareas (30%)
- Tests de evaluación intermedia (30%)
- Proyecto final (40%)

Convocatoria extraordinaria:
- Examen Final (60%)
- Proyecto Final (40%)

Peso porcentual del Examen Final:	0
Peso porcentual del resto de la evaluación:	100

BIBLIOGRAFÍA BÁSICA

- Aston Zhang, Zachary C. Lipton, Mu Li, Alexander J. Smola Dive into Deep Learning, https://d2l.ai, 2020

- Christopher D. Manning, Hinrich Schütze Foundations of Statistical Natural Language Processing, MIT Press, 1999

- Dan Jurafsky and James H. Martin Speech and Language Processing, Prentice Hall, 2018

- Li Deng (Editor), Yang Liu (Editor) Deep Learning in Natural Language Processing, Springer, 2018