



Seminario del Programa de Doctorado “Formación en la Sociedad del Conocimiento” de la Universidad de Salamanca

Seminario Análisis de Big Data 1.

Análisis de sentimiento predictivo: Procesamiento del lenguaje natural con técnicas de aprendizaje automático en Python

Por Dr. Carlos Arcila Calderón

Profesor del Departamento de Sociología y Comunicación y miembro del Observatorio de Contenidos Audiovisuales (OCA)
Universidad de Salamanca

Fechas y horario

14 y 14 de junio de 2018, en horario e 09:00 a 14:00 horas (10 horas)

Aula 5, Facultad de Ciencias

Matricula gratuita. 15 plazas. Para inscripción escribir a carcila@usal.es

Contenidos del curso

1. Introducción a los métodos computacionales en la investigación en comunicación y al análisis de big data
2. Programación básica en Python
3. Recuperación de datos de Twitter (API Streaming y API REST)
4. Técnicas básicas de procesamiento del lenguaje natural con la librería NLTK
5. Técnicas básicas de aprendizaje automática (*machine learning*) con la librería Sci-Learn
6. Análisis de sentimiento predictivo

Requisitos para los asistentes

- (a) Se requiere conocimiento básico de informática en entornos Windows, Linux o MacOS.
- (b) Se requiere conocimiento básico de estadística y de la técnica de “análisis de contenido”
- (c) No se requiere ordenador personal ya que el seminario se llevará a cabo en un Laboratorio de informática

Lecturas recomendadas

- Arcila, C.; Barbosa, E. & Cabezuelo, F. (2016). Técnicas Big Data: Análisis de textos a gran escala para la investigación científica y periodística. *El Profesional de la Información*, 25 (4), 623-631.
<http://dx.doi.org/10.3145/epi.2016.jul.12>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.
- Arcila, C.; Ortega, F.; Jiménez, J. & Trulleque, S. (2017). Análisis supervisado de sentimientos políticos en español: Clasificación en tiempo real de tweets basada en aprendizaje automático. *El Profesional de la Información*, 26 (5), 978-987.