

# Text Mining II

## Diploma en Big Data

Francisco Rangel  
Paolo Rosso



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

# Planificación

Viernes 12	Sábado 13	Viernes 19	Sábado 20
Presentación del módulo (5min)	Construcción del baseline (1h) Práctica semi-guiada	Optimización (0.25h) Francisco Rangel	Tarea de Author Profiling (1.5h)
Introducción al Author Profiling (2h) Paolo Rosso	Tarea de Author Profiling (3h)	Tarea de Author Profiling (4h)	Presentaciones PechaKucha (1h)
Exploración del dataset PAN-AP-2013 (2.5h) Práctica semi-guiada	EmoGraphs (0.5h) Francisco Rangel	Presentación de resultados (0.25h) Francisco Rangel	Plagiarism Detection (2h) Paolo Rosso
30 min descanso			

# Requisitos

- Conocimientos de Java (o el lenguaje que el alumno desee utilizar)
- Conocimientos de Latex (o en su defecto Word)
- Cuenta en GitHub

# Material

- Script Java + guión para exploración de los datos
- Script Java para generación de baseline
- Vídeos de optimización
- Esqueleto artículo

# Entregables

- En una cuenta de GitHub del alumno:
  - Artículo realizado individualmente contando su participación en la tarea. La longitud máxima del artículo será de 4 hojas adaptadas al formato proporcionado
  - Presentación PechaKucha (4 min) de resultados
  - Código fuente de los scripts utilizados para la experimentación más un guión para su reproducibilidad

# Criterios de evaluación

- Interacción en clase (10%)
- Presentación PechaKucha (40%)
  - Presentación resultados tarea\*
- Trabajo entregable (50%)
  - Artículo
  - Código fuente + guión de ejecución

\* Los resultados de la competición se evaluarán en dos aspectos confrontados: accuracy vs. time