



Outline of course on:

## “Análisis computacional de textos y redes en Ciencias Sociales”

### Instructores:

Ernesto Calvo  
[ecalvo@umd.edu](mailto:ecalvo@umd.edu)  
Universidad de Maryland

Iñaki Sagarzazu  
[inaki.sagarzazu@ttu.edu](mailto:inaki.sagarzazu@ttu.edu)  
Texas Tech University

Este curso está orientado a profesionales y estudiantes de Ciencias Sociales, interesados en el uso de métodos cuantitativos para el análisis de textos y redes sociales. El objetivo es ofrecer una exposición sistemática de los conceptos fundamentales para quienes quieran participar del vínculo cada vez más estrecho entre el Análisis de Datos y las Ciencias Sociales.

El seminario está estructurado en 10 clases de 4 horas. Se complementarán las clases con un laboratorio/práctico durante la tarde. Será dictado en R/R-Studio y se concentrará en los paquetes Quantero y Igraph.

### Programa:

- Introducción a R: Interfaz de R-Studio, operaciones básicas, vectores. Carga y manipulación de datos, *Data Frames*.
- Análisis de texto: Bolsa-de-palabras vs. Análisis semántico. Codificación de Diccionario.
- Análisis semántico: extracción de espacio latente de un corpus.
- Extracción de conocimiento: tópicos de aprendizaje automático.
- Análisis de redes sociales: conceptos de redes, detección de comunidades y visualización de redes.
- Análisis de corpus de textos en redes sociales: hashtags, análisis de clusters, redes semánticas, visualización de contenidos.



### Requisitos:

- Concurrir con una computadora con R y RStudio instalado. (<https://cran.r-project.org/> y <https://www.rstudio.com/>).
- Tener un conocimiento básico de R, como el disponible en el [siguiente tutorial](#).

### Para instalar los paquetes requeridos para la clase correr en R o R-Studio el código:

```
sapply(c("rvest","RCurl","XML","httr","xml2","RCurl","tm","igraph","foreign","twitter","httpuv","lmer","streamR","quantda","readtext"),install.packages)
```

### Selección de Textos de Apoyo

<https://www.dropbox.com/sh/skwrakeilmcqn9j/AADrdgsFqWq8CeXzp-rXVIwba?dl=0>

## Programa

### Semana 1: Iñaki Sagarzazu

#### Módulo 1: Lunes 4 de Junio, Teórico - Introducción al análisis de Texto, 2 horas.

- 9 a 10:45hs: Introducción a analisis de Texto. Codificacion manual, *validity*, *reliability*.

#### Módulo 2, Lunes 4 de Junio, Taller I : Introducción a R, 1 ½ horas

- 11:15 a 13:00hs: Introducción a R. Tipos de datos, vectores, matrices, arrays, listas, loops, apply, tapply/sapply. Programación de funciones.

#### Módulos 3 y 4, Martes 5 de Junio: Textos como Datos, Procesamiento de Lenguaje Natural

- 9 a 10:45hs / 11:15 a 13:00hs: Quanteda, TM y expresiones regulares.

#### Módulos 5 y 6, Miércoles 6 de Junio: Análisis de Texto: Análisis de Tópicos

- 9 a 10:45hs / 11:15 a 13:00hs: Analisis de Topicos mediante Dictionarios y modelos sin supervision.

#### Módulos 7 y 8, Jueves 7 de Junio: Análisis de Texto: Análisis de Sentimiento

- 9 a 10:45hs / 11:15 a 13:00hs: Análisis de Texto III: Análisis de Sentimiento mediante el uso de diccionarios

#### Módulo 9 y 10, Viernes 8 de Junio: Análisis de Texto: *Scaling* de Textos

- 9 a 10:45hs / 11:15 a 13:00hs: Organización de documentos en escalas con metodos semi-supervisados.



## Semana 2: Ernesto Calvo

### Módulo 11, Lunes 11 de Junio: Web-Scrapping en R I.

- 9 a 10:45hs: Ernesto Calvo
- Web Scrapping. Some basics. ReadLines, Rvest.

Codigo 3: Ejemplo de codigo para google scholar

<https://www.dropbox.com/s/n7notu537dcpyji/Sample%20code%20for%20google%20scholar-function.R?dl=0>

### Módulo 12, Lunes 11 de Junio: Web-Scrapping en R II.

- 11:15 a 13:00hs: Ernesto Calvo
- Otro ejemplo de web Scrapping.

Codigo 4: Ejemplo de C IMDB

<https://www.dropbox.com/s/zrmg67h6moce7a1/Example%20with%20Kevin%20Bacon%20en%20IMDB.R?dl=0>

### Módulo 13, Martes 12 de Junio: Introducción al análisis de Redes.

- 9 a 10:45hs: Ernesto Calvo
- Las Redes Sociales como Objeto. Edges, Nodes. Types of networks.

<https://www.dropbox.com/s/tc00tchtua91ciz/Twitter%20in%20R%20using%20streamR-Remplazar%20key%20and%20Secret.R?dl=0>

Toy JSON File: <https://www.dropbox.com/s/3e1n3pf9eu2mcxn/macri.json?dl=0>

### Módulo 14, Martes 12 de Junio: Filas JSON, Vertices y Aristas, Manipulacion de Redes en R.

- 11:15 a 13:00hs: Ernesto Calvo
- Una introducción al paquete Igraph. Cargando distintos tipos de redes. Variables, atributos.

### Módulo 15, Miércoles 13 de junio: Detección de Comunidades y Layouts.

- 9 a 10:45hs: Ernesto Calvo
- Community detection y Layouts. Densidad, centralidad, importancia.

### Módulo 16, Miércoles 13 de Junio: Visualización de Actividad en Twitter.

- 11:15 a 13:00hs: Ernesto Calvo
- Visualizacion de atributos en la red (Nodos, Aristas, tipos de datos).



**Módulo 17, Jueves 14 de Junio: Propagación de información en Redes**

- 9 a 10:45hs: Ernesto Calvo
- Paradoja de los amigos, difusión, propagación.

**Módulo 18, Jueves 14 de Junio: Propagación de Información en Redes II.**

- 11:15 a 13:00hs: Ernesto Calvo.  
Jerarquias, Nodalidad y Autoridad.

**Módulo 19, Viernes 15 de Junio: Análisis de Sentimiento en Redes Sociales.**

- 9 a 10:45hs: Ernesto Calvo

**Módulo 20, Viernes 15 de Junio: Cierre.**

11:15 a 13:00hs: Ernesto Calvo