

Taller Semana 3 - Métodos Avanzados

Estrategias y herramientas para la captura, estructuración y análisis de datos.

1. Media Cloud (30 min)

- Crear una cuenta de usuario ir a: <https://mediacloud.org/>
- [Guía básica](#): source manager, explorer y topic mapper.
- Ejercicios con Explorer, siguiendo indicaciones de guía básica.

2. Databasic.io (20 min)

- <https://databasic.io/es/>
- Ejercicio con WTFCSV utilizando el archivo descargado de Media Cloud : All story URLs.
- Ejercicio con una de las noticias encontradas en Media Cloud. Análisis de frecuencias de esa noticia con WORD COUNTER. Seleccionar dos noticias de las “sample stories” arrojadas en la búsqueda en Explorer. Visualizar la frecuencia de n-gramas en las dos noticias.
- Ejercicio con uno de los csv files descargados de DMI-Tcat.

3. DMI-TCAT (30 min)

- [Guía básica](#)
- Ejercicios de guía básica : queries con bin, filtrar de datos y crear un data-set para análisis, descargar archivos CSV.

4. OpenRefine (10 min) (pueden utilizar alternativamente excel o google spreadsheets)

- Abrir los datos (CSV) descargados durante los ejercicios en Media Cloud o DMI-TCAT.
- Limpiar uno de los conjuntos de datos que descargaron. Se borran los datos de las filas que ocasionen ruido y hayan sido incluidos por error. Por ejemplo noticias que no son de Colombia, o que están fuera de los límites temporales de la investigación. Algunos tweets también pueden ocasionar ruido si son generados en otro país.

Tarea

- Crear un post en el blog de la clase con visualizaciones de los resultados de los ejercicios realizados en Media Cloud.
- Las imágenes deben ir acompañadas con una breve descripción del query realizado, el conjunto de datos utilizado (media collections en Media Cloud), y las preguntas que se intentaron resolver.

- Crear un nuevo bin en TCAT para captura de tweets que sea útil para su proyecto de investigación de doctorado o para su mini-proyecto del módulo.