Week 2

Heuristics

Finding out an appropriate way to extract specific information from the article:

- URL
- Headline
- Article text
- Author
- Published date
- References to other sites inside the article text (like references validating the information/claim)

- [Newspaper Library](#) & source code

Defining and creating database of trustful / non trustable sites.

"On this step we filter with pandas the articles we want to extract per their credibility category, we only take: Fake News, Satire News and Reliable News. For this we used a technique learned in the Master; i.e.-chunksize which took a lot of time. This can be replicated and revised on the following notebook: [FNC Big Data file](#)

The result of this operation was a CSV file named news_csv_filtered. This file is note at the data folder due to its size, so we included its name at the .gitignore list."

From [Spanish Fake News Detecto](#)r(TFM by pipe11 on github)

- List of sites:

    For country: oldest, most traffic.

Identify a valid source and contrast it with the news claim.

- Leverage google search: extracting relevant sentences or keywords from the article o directly searching the headline through google api, then process the top results and report a summary of this to the user.

  Custom Search JSON API

  SerpApi

  Google Knowledge Graph Search API

- Wikipedia has been used for automatic fact checking or contextualization, so it could be another option

  WikiGraphs

  MediaWiki API:All search modules