Pipeline:

1. Cominciare da facta.news
2. Usare scraper (tool) per scaricare tutte le pagine di facta [usare HTTrack] [https://facta.news]
3. Python: estrazione fact-checking json [quante pagine lo hanno]
   1. “Generated by Full Fact Claim Review Schema”
   2. "claimReviewed"
   3. "itemReviewed"
   4. "reviewRating"
4. Identificazione lingua titolo [https://pypi.org/project/langdetect/]
5. Analisi dati
   1. keywords → word cloud
      1. remove stopwords
   2. topic detection

* Aggiungere altri news outlets
* Recuperare pagina notizia originale → download html

Google Fact Check Tools: <https://toolbox.google.com/factcheck/explorer/search/covid;hl=it>

Lista siti fact-checking:

https://www.agcom.it/siti-di-fact-checking