**News Scraper Comparison:**

Newspaper: <https://github.com/codelucas/newspaper/>

News-please: <https://github.com/fhamborg/news-please/>

**Possible Metrics of Comparison:**

* Number of sources (test on all sources found, identify any errors in parsing, controls e.g. not a news site)
* Speed (python timeit on all sources found, plot averages or histograms of times or something similar, along with table, also test auto detection of languages vs manual)
* Number of languages (test our 10, accuracy per language statistics, how do I test this in foreign languages?)
* Features checklist (big table similar to e.g. subscription sites?)
  + News-please has only CLI compatibility with extracting full websites, whereas newspaper can provide useful objects for entire websites.
* Compatibility checklist (e.g. with different databases and pipelines)
* Frequency (I don’t think this is applicable as scraping isn’t self-scheduled afaik)

**Newspaper:**

**Newsplease:**