Note:

Semantic Vector for every sentences.

Baseline model semantic classification.

Store headline, store publish headline and store original source.

Scrappy library

Proxy for scrappy

Tor python

Find correcaltion of the semantic

Second look up a mapping semantic

Third. Different news have different weight. Depending on the source. Financial news can have a bigger impact

Homoginies relevant stocks

Focuse on just one industrie of stocks

How far can I go back in time to retrieve the data.

Download it hourly the data of the courses

Hourly Html + csv containing publish date source my own retrievate. The name of the file. Name of the html

Loader object

Needs cordonation of timestamp

Build object. Download from the source with time stamp. Should have a presentation method.

Timeseries of the price and from the news.

Know if sources are good and reliable

Dataframe with timeseries with ticker.