# Indexing Articles

- Use scrapy to crawl websites for new articles
- Use beautifulSoup to parse websites and extract data to be placed into databases
    - Write Custom bs4 parse code to get desired content from page:
        - Urls
        - Article Title
        - Img-urls & captions
        - (Top 10) Keywords w/NLP
            - Sklearn/NLTK/TensorFlow
        - Article Theme w/NLP
        - Authors
        - Created Date, Last Modified Date, Last Accessed Date

# Search Interface

- Use Django/Flask – create web app to search backend
- Provide clients with a programmable rest-API (either using flask – better for microservices, or Django-rest – quicker to deploy)
- Provide clients with a user-friendly front-end interface
    - Uses rest-API to access data

# Database Structure

- See attached PDF