

Fall 12-1-2012

An All-In-One NLP Stock Market Backtester

Shaochen Zhong
Case Western Reserve University

Jiaqi Yu
Case Western Reserve University

Mocun Ye
Case Western Reserve University

Follow this and additional works at: <https://commons.case.edu/intersections-fa20>

Recommended Citation

Zhong, Shaochen; Yu, Jiaqi; and Ye, Mocun, "An All-In-One NLP Stock Market Backtester" (2012).
Intersections Fall 2020. 23.
<https://commons.case.edu/intersections-fa20/23>

This Book is brought to you for free and open access by the Intersections: CWRU Undergraduate Poster Session at Scholarly Commons @ Case Western Reserve University. It has been accepted for inclusion in Intersections Fall 2020 by an authorized administrator of Scholarly Commons @ Case Western Reserve University. For more information, please contact digitalcommons@case.edu.

An All-In-One NLP Stock Market Backtester

**– we handled all the engineering,
so that you can focus on your model.**

**Shaochen (Henry) Zhong, Jiaqi Yu, Mocun Ye
Supervised by Dr. Shuai Xu**

Demands

of doing NLP backtesting

- NLP/text-mining is popular for stock market prediction
- Yet it is engineeringly costly to do backtesting
 - Gather plain-text data
 - Gather market data
 - Categorize them in a way that is communicable with each other
 - Build a model
 - Plot visualization for evaluation

Required Skills

rather deviated from the nature of research

- Website scraping
- Retrieving metadata of the mentioned companies
- Fundamental understanding of stock market: mechanism of exchange, ticker, etc
- Transaction logging
- Interactive visualization plotting

Backtesting

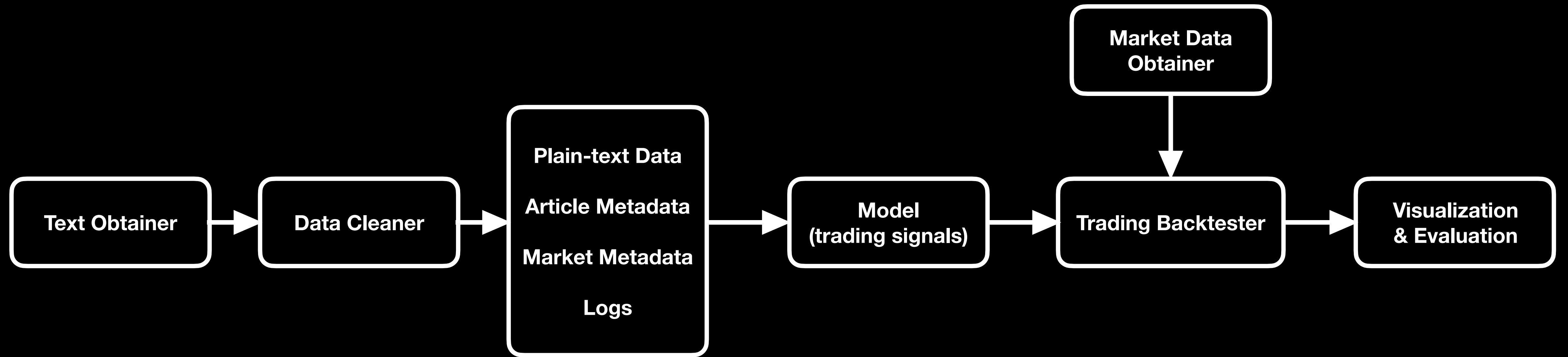
Problems with the existing platforms:

- Only one stock per trading strategy
- Unintuitive interface for strategy implementation
 - The strategy has to be implemented as a callback function
 - No active control over the iteration and the emulation

**But with our project,
we will handle all the engineering,
so that you may focus on your research**

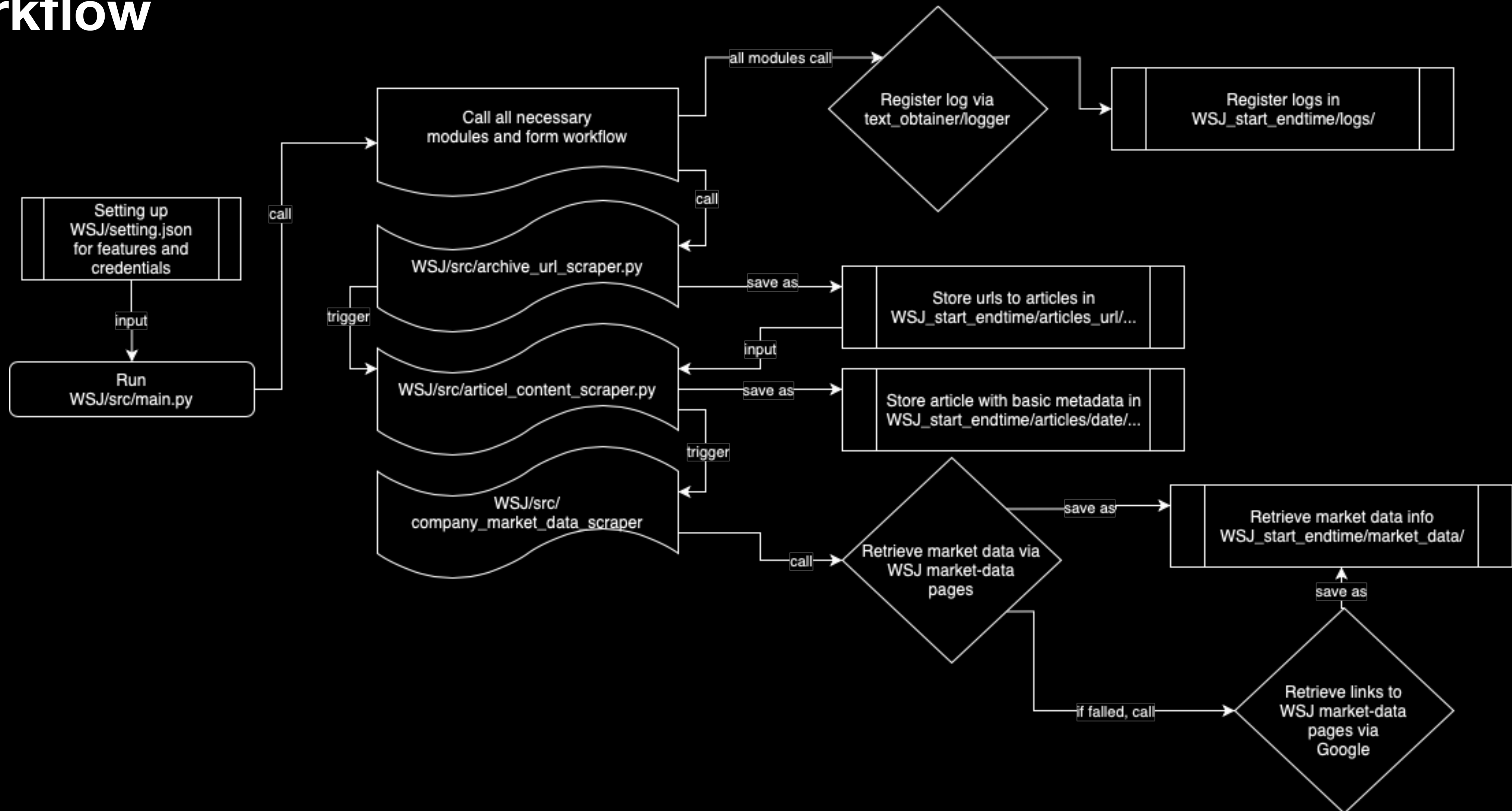
General Workflow

when using our project



Text Obtainer

Workflow



Text Obtainer

End Result

```

text_obtainer
├── __pycache__
├── channel
│   ├── WSJ
│   │   ├── __pycache__
│   │   ├── scheme
│   │   │   └── unit_schema.json
│   │   └── src
│   │       ├── __pycache__
│   │       ├── archive_url_scraper.py
│   │       ├── article_content_scraper.py
│   │       ├── company_market_data_scraper.py
│   │       ├── main.py
│   │       └── uuid_generator.py
│   └── setting.json
├── logger.py
├── text_obtainer_output
│   ├── WSJ_20080101_20080101
│   └── WSJ_20120102_20120102
│       ├── article_urls
│       │   └── 20120102.txt
│       ├── articles
│       │   └── 20120102
│       ├── logs
│       │   ├── article_urls_log.txt
│       │   ├── articles_log.txt
│       │   └── WSJ_market_data_log.txt
│       └── market_data
│           ├── company_market_LUT.json
│           └── raw_company_market_LUT.json

```

\$ python3 text_obtainer/channel/WSJ/src/main.py

```

{
  "Leap Wireless International": {
    "market_data_url": "https://www.wsj.com/
market-data/quotes/LEAP.UT",
    "ticker": "LEAP.UT",
    "exchange": "U.S.: NYSE",
    "legal_full_name": "Ribbit LEAP Ltd. Un",
    "quoted_in": [
      "6d2cb1fa-a88a-407a-a71d-1c3ca7407447"
    ]
  },
  "Bayer": {
    "market_data_url": "https://www.wsj.com/
market-data/quotes/BAYRY",
    "ticker": "BAYRY",
    "exchange": "U.S.: OTC",
    "legal_full_name": "Bayer AG ADR",
    "quoted_in": [
      "090faba5-1f06-462f-b685-16e1f8e95768"
    ]
  },
  "BASF": {
    "market_data_url": "https://www.wsj.com/
market-data/quotes/BASFY",
    "ticker": "BASFY",
    "exchange": "U.S.: OTC",
    "legal_full_name": "BASF SE ADR",
    "quoted_in": [
      "090faba5-1f06-462f-b685-16e1f8e95768"
    ]
  },
  ...
}

```

\$ python3 data_cleaner/main.py

```

{
  "channel": "WSJ",
  "date": "20120102",
  "url": "https://www.wsj.com/articles/
SB10001424052970203462304577136061852705298",
  "author": "Barbara Kollmeyer",
  "headline": "Europe Stocks Start 2012 Strong",
  "quotes": [
    "Bayer",
    "BASF",
    "Siemens",
    "Allianz",
    "Intesa Sanpaolo",
    "Sunways",
    "Vestas Wind Systems"
  ],
  "content": "MADRID\u2014better-than-expected
purchasing-managers index out of Germany lifted
European stock markets on the first day of trading
for the new year.\nThe Stoxx Europe 600 index rose
1.1% to 247.15, its highest close since Oct. 28.
U.K. and U.S. markets were closed, making for thin
trading conditions.\n The push was led by the
German DAX 30 index, which jumped 3% to 6075.52,
its best one-day gain since Dec. 20.
...
as investors focused on claims it and other
Norwegian insurers will face from a storm that
swept over Sweden, Finland and Norway with
hurricane-strength winds on Dec. 26 and Dec.
27.\nWrite to
...
"article_id": "8534f652-
eb8c-47e7-9204-3f7cd43d7de2",
"mention": {
  "Total": 1,
  "Bayer": 1,
  "BASF": 1,
  "Siemens": 1,
  "Allianz": 1,
  "Intesa Sanpaolo": 1,
  "Sunways": 1,
  "Vestas Wind Systems": 1
}
}

```

\$ python3 model/dummy_model.py

```

{
  "Ford Motor": {
    "market_data_url": "https://www.wsj.com/market-data/
quotes/F",
    "ticker": "F",
    "exchange": "U.S.: NYSE",
    "legal_full_name": "Ford Motor Co.",
    "quoted_in": [
      "bc72c575-dc74-4588-9809-ad141b16dd41"
    ],
    "mentioned_in": {
      "bc72c575-dc74-4588-9809-ad141b16dd41": {
        "mentioned_time": 1,
        "date": "20120103"
      },
      "00b5ee0b-36dd-4076-872a-d790fc1371af": {
        "mentioned_time": 1,
        "date": "20120102"
      }
    },
    "sentiment_indicator": {
      "20120102": {
        "negative": 10,
        "positive": 4,
        "uncertainty": 2,
        "litigious": 0,
        "strongmodal": 0,
        "weakmodal": 2,
        "constraining": 2,
        "trade_info": {
          "trade_indicator": [
            0.44,
            [
              5.5,
              12.5
            ]
          ],
          "trade_signal": "sell"
        }
      },
      ...
    },
    "total_actionable_days": 2,
    "total_mentioned_times": 2
  },
  ...
}

```

Trading Backtester

End Result (partially powered by backtesting.py)

```
jupyter trader Last Checkpoint: 12 minutes ago (autosaved) Python 3.8

File Edit View Insert Cell Kernel Windows Help

In [ ]:
import os, time
import datetime as dt
import threading
import numpy as np
import pandas as pd
import backtesting as bt
import json

class dummystrg_active(bt.Strategy):
    def init(self):
        self.runflag = True
        self.iteration_holder = threading.Lock()
        self.iteration_holder.acquire()
        #print("iteration init")

    def next(self):
        if self.runflag:
            self.iteration_holder.acquire()
            #print("iteration issued")

    def rc_buy(self):
        self.buy()

    def rc_sell(self):
        self.sell()

    def rc_iterate(self):
        if self.iteration_holder.locked():
            self.iteration_holder.release()
        else:
            #print("Info: Iteration ended!")

class dummystrg_passive(bt.Strategy):
    def init(self):
        pass

    def next(self):
        pass

'''
Workflow:
import trader as TD
t=TD.Trader(time_start = "YYYY-MM-DD", time_end = "YYYY-MM-DD")
t.import_stocks('AAPL') #import as much as one want as long as the data for such symbol exists.
#optionally:
t.trade_on_close = True #So as to let trade happen on closing
t.trade_mode = 'passive' #If the passive mode is desired (and the user would then have to define the strategy callb
#Then:
t.trade_init(balance = 100, preowned_stocks = []) #Arguments are optional. Balance default to 100. Preowned_stocks :
#Then trades:
t.trade_buy('some_symbol')
t.trade_sell('some_symbol')
```

Visualization (under development)

Expected End Result

Interactive overlay/
parallel display of NLP/
text-mining related
indicator(s) on top of
the trading signal
visualization.

