

# Stock News Indexer

This application is a stock news discovery engine that consists of 3 components: the indexer, a worker, and a web app for browsing indexed content and managing the system.

## Architecture

### Indexer

The indexer is a robots.txt-obeying recursive indexer that discovers news articles for stock tickers. The indexer takes as input a stock symbol, crawls several sources for news articles that contain said ticker, and when a match is found, it is written to a datastore.

The indexer contains two parts, the indexing portion and the worker. The worker monitors a queue and when it receives a job, calls the indexer on that job. The indexer begins by searching both preset sources and those it has discovered to be worthy source sites. It checks a potentially-cached robots.txt file to ensure it is 'friendly' and searches the page for the supplied argument, in this case a stock symbol.

### Web App

The web application is a control center for operating the indexer, browsing news articles for the various stock tickers, and managing the system as a whole. It consists of a simple Sinatra web application and leverages Twitter Bootstrap for consistent and easy styling. Using the web application one can add new tickers to be indexed, add new sources, view the articles indexed for a given stock symbol, view jobs currently in the queue, and clear all data.