

What is it?

Problem: Many financial tools aggregate news about companies, but don't analyze the sentiment of the articles or sources

Solution: By scraping information off the internet and using natural language processing, **DaBuzz** will be able to gauge the overall public perception of companies in the stock market

How it Works

Building the Classifier:

- 1000 articles are downloaded and loaded into training set
- Every article is read and manually classified
- After all are manually classified, a naïve bayesian classifier is compiled

Processing Incoming Articles:

- URLs from finance RSS feeds are scraped
- Main article text is extracted from URL
- Text is run against classifier to obtain sentiment
- Article is mapped to a company using keywords
- Processed data saved in database for future use

See it Live!

stonelinks.org/dabuzz/

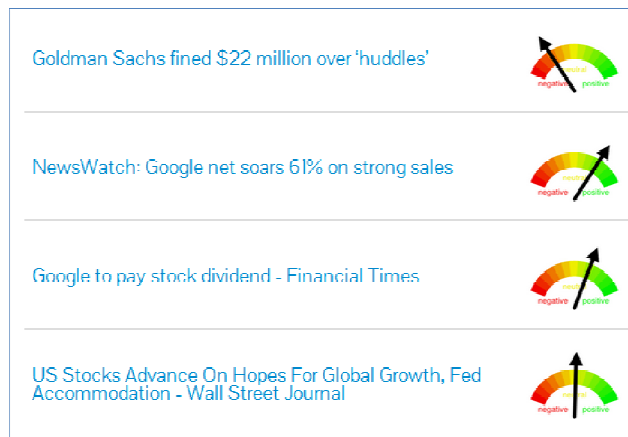


Figure 1: Sample of article titles and their perceived sentiment

About RCOS

- A group of RPI students who help society by creating useful software
- Students participate in semester long projects for credit or funding
- A creative, intellectual and entrepreneurial outlet for students
- Great practice for the real world:
 - Write code that actually ships!
 - Practice giving and receiving feedback
 - Discuss technical and non-technical parts of projects

Technologies

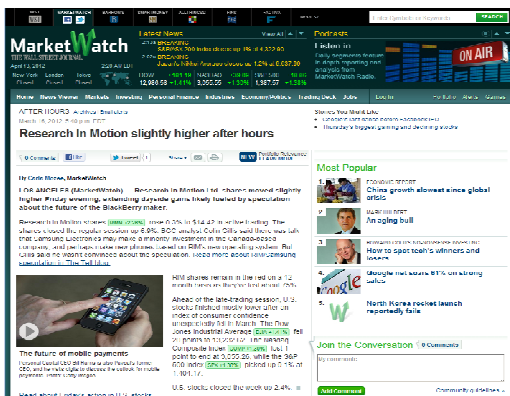


Figure 2: Web page before being processed

Plain Text
Extraction

Training:

LOS ANGELES (MarketWatch) — Research In Motion Ltd. shares moved slightly higher Friday evening, extending day's gains likely fueled by speculation about the future of the BlackBerry maker. Personal Capital CEO Bill Harris is also PayPal's former CEO, and he visits digits to discuss the outlook for mobile payments. Photo: Getty Images. China's manufacturing got a mixed assessment on Sunday, as the official survey for March improved to its... Following a big market rally in the year's first quarter, investors are scrutinizing data in search of its... Charts of week's data show consumers looking past gasoline prices, in part due to improving jobs picture. A security breach of more than 10 million Visa and MasterCard accounts could leave you exposed to fraud, so be on guard. Here's what to do. Mutual fund juggernaut Franklin Resources has recovered smartly from the dark days of 2008. Why the good times could continue. Redesigning coupe now a legitimate challenger to Mustang and Camaro. The action you requested requires a MarketWatch Community display name. MarketWatch Community is a free service that lets you discover, organize and share MarketWatch stories with other readers. The action you requested is only available to MarketWatch members.

Positive Irrelevant Negative

Figure 3: Text extracted and loaded into training

Future Goals and Applications

- Allow articles to represent information for more than one company
- Use price data to dynamically evaluate reliability of sources
- Scrape more sources:
 - Facebook
 - Twitter
 - Newspapers
- Detect mass-propagation of news that is above average importance
- Autonomous or semi-autonomous trading (LFT Algorithm)