Large scale information extraction and integration infrastructure for supporting financial decision making.
Why FIRST? - Motivations

The most reliable data sources today...

...also have their weakness!
They do not consider unstructured data, rumors, market sentiments, etc.
**Why FIRST? - Motivations**

**Example: Apple iPhone 1 Announcement on 2007-01-09**

Stock prices were skyrocketing after the announcement. However, the announcement could be sensed before...

- 2006-05-12: Apple said to be working on co-branded handset
- 2006-07-29: Peter Oppenheimer: ‘We’re not sitting around doing nothing’
- 2006-12-15: iPhone may include JAJAH voip software
- 2007-01-09: iPhone Announcement
Why FIRST? - Motivations

Example: Market surveillance via FIRST (the Google news case)

- September 2008: Google news announced “United Airlines bankruptcy”.
- Within 12 minutes → stock price decreased 75% → wiped out US $1bn.

The “news” was actually 6 years old...

Plausibility checking will help in identifying hoaxes: consistence with regulatory news and other sources.
Why FIRST? – Motivations
A growing universe of unstructured data
Why FIRST? – Motivations
A growing universe of unstructured data
Why FIRST? – Motivations
A growing universe of unstructured data
Why FIRST? – Motivations

A growing universe of unstructured data
Why FIRST? – Motivations
A growing universe of unstructured data
Why FIRST? – Motivations
A growing universe of unstructured data

... how to separate the wheat from the chaff?
Motivation
Vision
Innovation
Tools
FIRST Project

European-funded research project

Project facts

Running from October 2010 until September 2013
9 partners
More than 30 people
Preliminary results available
More to come...

Stay tuned (http://project-first.eu/)
Who is behind FIRST?
is to make available the relevant information of the entire financial information space (including unreliable, unstructured, sentiment sources) to the decision maker in near-real time in an automated way
FIRST Vision

Financial Resources

Structured

Unstructured

Blog, analysis, bulletin boards...
Unreliable, poor quality, noisy...

AUTOMATION

Acquisition
Processing
Analysis
Decision support
Mining the Web for financial texts

Data Acquisition pipeline: Web mining

- Social media
- Blogs
- News

Streaming

Natural Language preprocessing and entity extraction

Cleaning

HTML

Content

Annotated Content

Financial terms, Companies, Instruments ...
Data acquisition after one year

- Some numbers
  - 176 Web sites
  - 2,671 RSS sources
  - ~40,000 documents per day
  - >5,000,000 documents by end of 2011
    - And growing

Essential for future evaluation and analysis
Analysing sentiments in Web texts

The Analytical Pipeline: Identify, extract, classify, aggregate

Document with basic annotations

SENTIMENT CLASSIFICATION per object and feature

Document with sentiment sentences

SENTIMENT AGGREGATION per object and feature

Aggregated sentiments

Object

Indicators

Positive sentiment

Sentiment Sentences

DocumentData

Indicator

Original markups

Sentiment

SentimentSentencePositive

SentimentPhrase

SentimentPhrasePositive

Sentiment_Object

50

Positive sentiment

When Apple (Nasdaq: AAPL) reports fourth-quarter and fiscal 2011 results on October 18th, the company will show that earnings per share grew 84 percent year-over-year. For a company the size of Apple this is an astonishing feat.

However, despite this amazing growth and still limitless opportunities with the iPhone and iPad upgrades, next year analysts are only predicting growth of a modest 18 percent.

While the law of large numbers will most certainly catch up to Apple one day, they haven't yet. This morning, the company announced it had record iPhone 4S pre-orders of 1 million in the first 24 hours. Analysts said the entire weekend could have seen a total of 3 million. It wasn't even the iPhone 5!

With numbers like this from a product many said was a disappointment, it is easy to see that 2012 could be another Year of the Apple. Think iPad 3 and iPhone 5!

What if the company grew earnings again at 84 percent again? If this was the case you'd have a company earning $51 per share. That's right $51!

Slap a 20x multiple on this and you have $1020 stock. Slap a 16x multiple on this and you have a $768 stock. Even slap a 10x multiple and you have a $510 stock.

Shares of Apple are up 4 percent today to $384.57.
Supporting the decision making process

The Decision Support techniques: Analysis and visualization

First Acquisition & Analytical Pipelines

Knowledge Base

Machine Learning Techniques

Qualitative Modeling

Visualization Techniques

Forecasting Models

Outputs:
Forecasts of volatility or returns, Alert on pump and dump, Reputation change of a counterpart Signals, Charts, Topic Spaces, Topic Trends, Reports ...

...
Glassbox model

- Objects
- Sentiment
  - low
  - high
- Features

Drill down

Document

sentences
The three FIRST use cases & their relevance for the industry

Market Surveillance
- **Capital markets compliance** can be automated today using structured data, but the automation **does not take unstructured data into account**
- **FIRST will**
  - make use of **large volumes of unstructured data** into financial compliance;
  - develop automated techniques to **better detect market abuse/insider trading**.

Reputational Risk Management
- **No off-the-shelf solutions or methodologies** for reputational risk management.
- **FIRST will**
  - provide a **sustainable tool for reputational risk monitoring**;
  - contribute to break new ground in this field of dramatically **high impact in FSI**.

Retail Brokerage
- **Today**, mainly based on **quantitative analysis and key figures**.
- **FIRST will**
  - use **unstructured data to leverage** both **information** for private investors and sophisticated **tools** for professional users.
Stay tuned (http://project-first.eu/)
Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°257928.

THANKS