Real-Time News Analytics With Big Data Technologies

Volker Stümpflen
CEO
Clueda AG

Robert Feckl
CIO
Baader Bank AG

26.3.2014
Partner

Baader Bank AG
- Leading Investmentbank in Germany
- Market leader in trading of financial instruments

Clueda AG
- Spin-Off Institute for Bioinformatics and Systemsbiology of the Helmholtz Zentrums München
- Real-time software solutions for semantic and associative knowledge processing and analysis
Newsflood

740,000 financial instruments

From stocks to derivatives

Increasing

100 traders and analysts

Decreasing time for increasing information

Is constant

500,000 news p.d.

~4 bn sentences p.a.

From news agencies to social media channels (Blogs, Tweets)

Strongly increasing
Clueda analysis ready
trader is buying

Automated analysis

Commercial advantage

News published

News reading

News reading finished
trader is buying
User-Centric Decision Making

See
Concepts, relations and events as they happen in multiple information sources

Understand
Trends, mood and relationships using semantics and systems biology approaches

Answer
Questions that only specialists could answer before
Market Moving Influences

Events

Insider Knowl.

Information

Market Moving

Mood

Sentiment
Elementary Processing Steps

1. Recognizing Concepts (Companies, Persons, ...)
2. Recognizing Relations and Events
3. Generating Knowledge Networks
4. Advanced Analytics (e.g. Sentiment)
Concept Detection

- Recognizing the meaning of unknown words
- Self-learning capabilities based on machine learning approaches
- After initial training knowledge base is extended automatically
Real-Time Event Detection and Processing

… big launch celebrations at hardware stores with Galaxy Tab III were canceled. **Apple sues Samsung** in Australia. Following earlier legal disputes …

- Understands textual information and relations
- Generates a semantic knowledge network
- Identifies market moving news in real-time
Event Determination With Big Data Analytics

- Price vs. Time graph
- t0 and t1 markers
- Open, Low, and Close labels
- News Release event
- Close = High
- Move caused by news
- Measurement error
- Market move
Analysis Of News From One Year

Number of news

Optimal threshold

Meaningful news events

Clustering

Threshold market move

Event Type 2

Event Type 1
<table>
<thead>
<tr>
<th>Event</th>
<th>Rel Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDS Price Move</td>
<td>1</td>
</tr>
<tr>
<td>Analyst Forecast</td>
<td>1</td>
</tr>
<tr>
<td>Business Climate Change</td>
<td>1</td>
</tr>
<tr>
<td>CEO Search</td>
<td>1</td>
</tr>
<tr>
<td>Company Forecast</td>
<td>1</td>
</tr>
<tr>
<td>Customer Problems</td>
<td>1</td>
</tr>
<tr>
<td>Debt Financing</td>
<td>1</td>
</tr>
<tr>
<td>Equity Financing</td>
<td>1</td>
</tr>
<tr>
<td>ErrorSymbolAssignment</td>
<td>1</td>
</tr>
<tr>
<td>Fraud Investigation</td>
<td>1</td>
</tr>
<tr>
<td>Government Decision (no bailout)</td>
<td>1</td>
</tr>
<tr>
<td>Incorporation Change</td>
<td>1</td>
</tr>
<tr>
<td>Legal Settlement</td>
<td>1</td>
</tr>
<tr>
<td>M&amp;A</td>
<td>1</td>
</tr>
<tr>
<td>Restructuring</td>
<td>1</td>
</tr>
<tr>
<td>Supply Chain</td>
<td>1</td>
</tr>
<tr>
<td>Trading Halt</td>
<td>1</td>
</tr>
<tr>
<td>Asset Liquidation</td>
<td>2</td>
</tr>
<tr>
<td>Stocks Fall (Peers)</td>
<td>2</td>
</tr>
<tr>
<td>Dividend Change</td>
<td>3</td>
</tr>
<tr>
<td>Broker Rating</td>
<td>9</td>
</tr>
<tr>
<td>Quarterly Results</td>
<td>10</td>
</tr>
</tbody>
</table>
Statement-Centric Information Compression and Detection

- Approximately 30-40% of all news contain redundant information

- Only one out of 500 news is market moving
Trading: clueda.trader
Systemic Interrelations / Systemic Mood

- Sentiment influences with systems biological methods
- Mood propagation in networks
- Identification of indirect mood drivers
Research: clueda.analyzer
Integration Into Baader IT Infrastructure

- Concerted Action for Implementing Clueda.Trader into Baader Infrastructure
- **JSON** and **RabbitMQ** as tools of choice for data integration
- 3rd-party application support for I-Trader (Dion Systems)
- Standard x86 server hardware components
- Operational costs pleasant due to commodity hardware and open source software
Costs

- Agile development process / SCRUM
- End user involved in the complete development cycle
- Project in time and costs
- Low total cost of ownership
- Horizontal scalable with commodity hardware
Thank You!

Volker Stümpflen
Robert Feckl