

# Geschriebene Sprache inhaltlich verstehen

Herausforderungen, Anwendungsmöglichkeiten, Potential

Government Round Table, IBM Watson

Marcus Hassler, CTO  
econob GmbH

20.03.2014, Bern, CH



# IBM Watson – Informationen für sich arbeiten lassen



# Maschinelle Verarbeitung natürlicher Sprache

- Textanalyse ermöglicht
  - Facettenreichtum der Sprache abzubilden
  - Unstrukturierten Inhalt zu erfassen
  - Inhalte „richtig“ zu interpretieren



## WIE?

Low-Level

Mid-Level

High-Level

- Statistiken bilden, Wörter filtern, Muster finden

Natürliche  
Sprache

Ich habe meinen Geburtstag am 15.02.2014 mit meiner Familie in Bern verbracht und wir

besuchten das Hotel Sc

über Sonderwünsche ,

Zimmers , das perfekte

Direktors war alles einfa

Heimat der Schweiz , sc

Hotel Schweizerhof nur

besuchen . Bravo an de

### Summary

Clouds	Concepts	Statistics
Characters		681
Tokens		113
Words		99
Concepts		5
Events		0
Orientations		6



# Maschinelle Verarbeitung natürlicher Sprache

- Textanalyse ermöglicht
  - Facettenreichtum der Sprache abzubilden
  - Unstrukturierten Inhalt zu erfassen
  - Inhalte „richtig“ zu interpretieren



## WIE?

Low-Level

Mid-Level

High-Level

- Wortarten identifizieren, Stammformen bilden, Phrasen erkennen, Sätze abgrenzen



Ich habe meinen Geburtstag am 15 . 02 . 2014 mit meiner Familie in Bern verbracht und wir besuchten das Hotel Schweizerhof . Von der Reservierung , die freundliche Anfrage vor der Anreise über Sonderwünsche , Valat - Parking , sehr herzlicher Empfang an der Reception , das zeigen des Zimmers , das perfekte Zimmer liebevoll hergerichtet , die handgeschriebene Willkommenskarte des Direktors war alles einfach nur super – bis hin zum opulenten Frühstück . Ich habe in meiner Heimat der Schweiz , schon lange nicht mehr so einen perfekten Service erlebt . Wir können das Hotel Schweizerhof nur wärmstens empfehlen und werden dieses vorzügliche Haus gerne wieder besuchen . Bravo an den Direktor und sein Team .

# Maschinelle Verarbeitung natürlicher Sprache

- Textanalyse ermöglicht
  - Facettenreichtum der Sprache abzubilden
  - Unstrukturierten Inhalt zu erfassen
  - Inhalte „richtig“ zu interpretieren



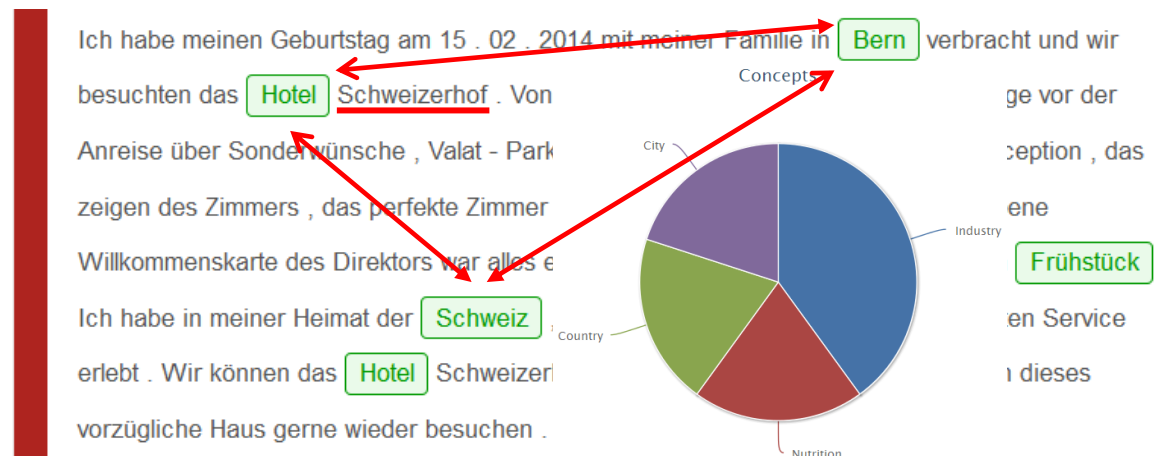
## WIE?

Low-Level

Mid-Level

High-Level

- Konzepte finden, Beziehungen zwischen Konzepten finden, Konzepte bewerten



# Maschinelle Verarbeitung natürlicher Sprache

- Textanalyse ermöglicht
  - Facettenreichtum der Sprache abzubilden
  - Unstrukturierten Inhalt zu erfassen
  - Inhalte „richtig“ zu interpretieren



## WIE?

Low-Level

Mid-Level

High-Level

- Konzepte finden, Beziehungen zwischen Konzepten finden, Konzepte bewerten

Ich habe **meinen** Geburtstag am 15 . 02 . 2014 **mit** meiner Familie **in** **Bern** verbracht und wir besuchten das **Hotel** Schweizerhof . Von der Reservierung , die **freundliche** Anfrage vor der Anreise über Sonderwünsche , Valat - Parking , **sehr** **herzlicher** Empfang an der Reception , das **zeigen** des Zimmers , das **perfekte** Zimmer **liebepoll** hergerichtet , die handgeschriebene Willkommenskarte des Direktors war alles **einfach** **nur** **super** – **bis** hin zum opulenten **Frühstück** . Ich habe **in** meiner Heimat der **Schweiz** , schon lange **nicht** mehr so einen **perfekten** Service erlebt . Wir können das **Hotel** Schweizerhof **nur** wärmstens **empfehlen** und werden dieses vorzügliche Haus **gerne** wieder besuchen . Bravo an den Direktor und sein Team .

Natürliche  
Sprache

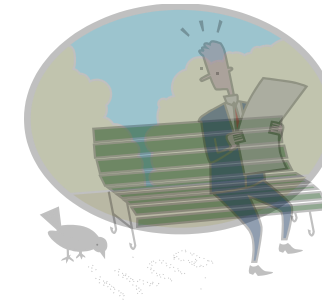
# Herausforderungen für die maschinelle Sprachverarbeitung

- Obskürität 1

Meine Frau kaufte mir ein Schloss.



Ich sitze auf der Bank in der Bank.

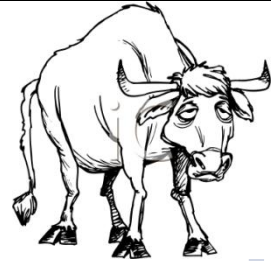


➔ **Doppeldeutigkeit (Ambiguität)**

# Herausforderungen für die maschinelle Sprachverarbeitung

- Obskürität 2

Die **Bäuerin** verkaufte die **Kuh**, weil **sie** alt und krank war. ?



**IBM** kommt nach Bern. **Das Unternehmen** wird ... ?

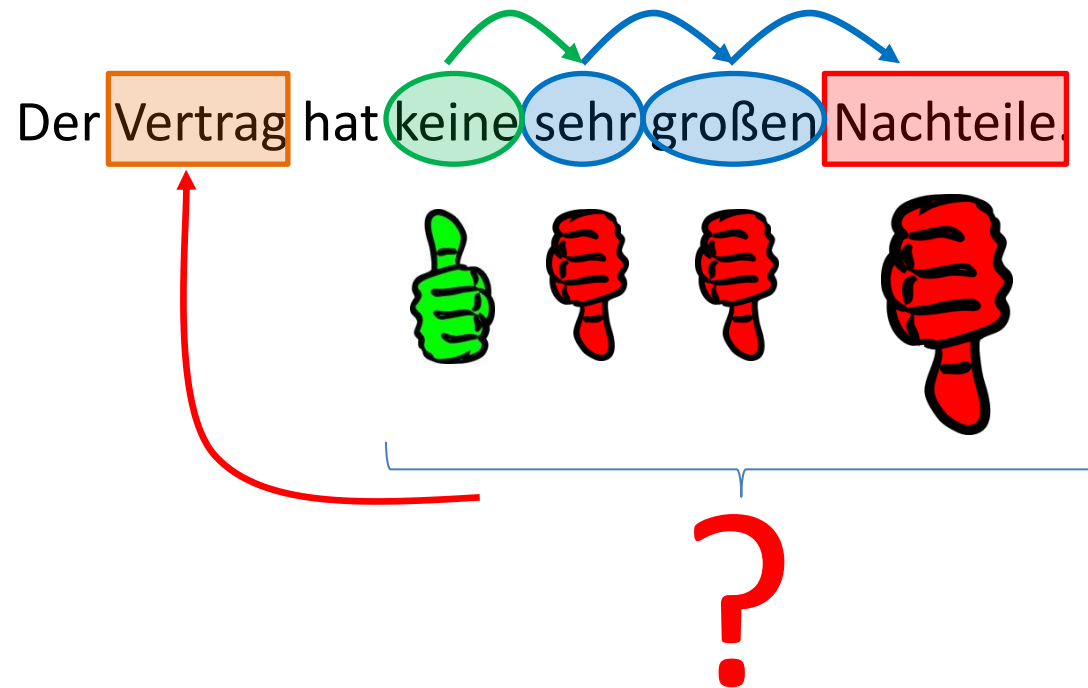


➔ **Koreferenz**



# Herausforderungen für die maschinelle Sprachverarbeitung

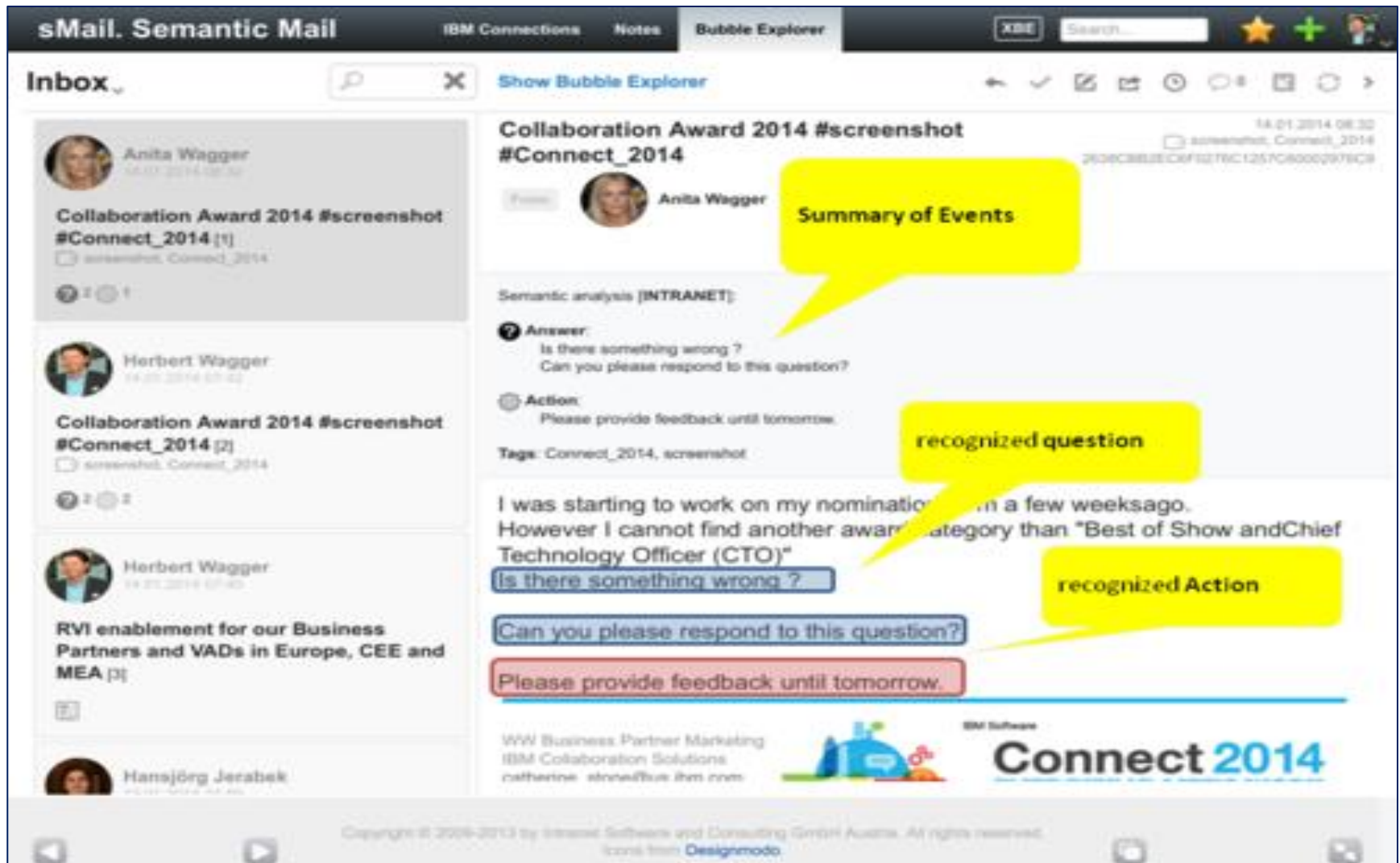
## ▪ Obskürität 3



## ➔ Unklare Bewertung

- Komplexe Kombinationen
- Nicht eindeutige Aussage (Interpretation)

- Automatische eMail – Klassifikation



The screenshot displays the 'sMail. Semantic Mail' interface. The main email content is titled 'Collaboration Award 2014 #screenshot #Connect\_2014' and is from Anita Wagner. The email body includes a 'Semantic analysis [INTRANET]' section with the following items:

- Answer:** Is there something wrong? Can you please respond to this question? (Annotated with a yellow speech bubble: 'Summary of Events')
- Action:** Please provide feedback until tomorrow. (Annotated with a yellow speech bubble: 'recognized question')
- Tags:** Connect\_2014, screenshot

The main text of the email reads: 'I was starting to work on my nomination in a few weeks ago. However I cannot find another award category than "Best of Show and Chief Technology Officer (CTO)"'. Below this, the question 'Is there something wrong?' is highlighted in blue and annotated with a yellow speech bubble: 'recognized Action'. The response 'Can you please respond to this question?' is also highlighted in blue. The action 'Please provide feedback until tomorrow.' is highlighted in red.

The footer of the email includes 'WW Business Partner Marketing', 'IBM Collaboration Solutions', and the 'Connect 2014' logo.



## ■ Automatisierter Aktienhandel

A-TRAP (Absolute Trading Application) - News Trader 2.7.1

File View Help

Date and Time	Source	Title	Items	Indicators/Wordings
2014-03-12 11:17:07	Local	[TEST SAP 1] DGA...	SAP	Software Revenue, Sales, SSRS Revenue, Bea...
2014-03-12 11:17:06	Local	[MUV2 2011] Muni...	MUV2	MUV2 Profit, Dividend, Net income GAAP
2014-03-12 11:17:05	Local	[MUV2 9M 2012] M...	MUV2	MUV2 Profit, Operating Profit, Net income ...
2014-03-12 11:17:05	Local	[MUV2 6M H1 2012]...	MUV2	ORD Earnings per share, MUV2 Profit, Opera...
2014-03-12 11:17:04	Local	[MUV2 3M Q1 2012]...	MUV2, MMM	MUV2 Profit
2014-03-12 11:17:03	Local	[TEST LHA WARNIN...	LHA	Sales
2014-03-12 11:17:03	Local	[TEST LHA Q1 201...	LHA	Net Profit, Net income GAAP, Sales
2014-03-12 11:17:02	Local	[TEST LHA Q1 201...	LHA	Operating Profit, Sales, Net income GAAP
2014-03-12 11:17:02	Local	[TEST LHA FY 201...	LHA	Operating Profit, MUV2 Profit, Dividend, S...
2014-03-12 11:17:01	Local	[TEST LHA FY 201...	LHA	Operating Profit, MUV2 Profit, Net Profit, ...
2014-03-12 11:17:01	Local	[TEST LHA 9M 201...	LHA	Operating Profit, MUV2 Profit, Sales, Net ...
2014-03-12 11:17:00	Local	[TEST LHA 6M 201...	LHA	Operating Profit, MUV2 Profit, Operating P...
2014-03-12 11:16:59	Local	[TEST LHA 6M 201...	LHA	Operating Profit, MUV2 Profit, Operating P...
2014-03-12 11:16:59	Local	[TEST FPE3 Q1 20...	FPE3	EBIT, Sales, MUV2 Profit, ORD Earnings per...
2014-03-12 11:16:58	Local	[TEST FPE3 H1 20...	FPE3	EBIT, Sales, MUV2 Profit, ORD Earnings per...
2014-03-12 11:16:58	Local	[TEST FPE3 FY 20...	FPE3	Sales, ORD Earnings per share, EBIT, Divid...
2014-03-12 11:16:57	Local	[TEST FPE3 9M 20...	FPE3	EBIT, Sales, MUV2 Profit, ORD Earnings per...
2014-03-12 11:16:56	Local	[TEST DAT FY 201...	DAT	EBIT, Net Profit, Net income GAAP, MUV2 Pr...

Jump to selection  
  Summarize I/W  
  Simplify  
  Hide uncompered  
 Hide implausible  
  Hide irrelevant

...	R	Item	Indicator/Wording	Period	Value	Comparison
0		LHA	Operating Profi	H1 2011	NONE 2,011 k	
0	!	LHA	Operating Profi	H1 2011	€ 3,00 m	≥ € 2,70 m
1		LHA	Operating Profi	H1 2011	NONE 2,011 k	
1		LHA	Operating Profi	H1 2011	NONE 2,011 k	
1		LHA	Operating Profi	H1 2011	NONE 2,011 k	
1	!	LHA	Operating Profi	H1 2011	€ 3,00 m	≤ € 2,20 m
2		LHA	Operating Profi	H1 2011	NONE 2,011 k	
2		LHA	Operating Profi	H1 2011	NONE 2,011 k	

Search   Reset

A-TRAP Markup

Select None   Select All

Bracket  
 Item  
 KeyIndicator  
 Matching Rule  
 Period  
 Value

[TEST LHA 6M 2011] Lufthansa generates operating profit in the first half year of 2011 and confirms the outlook to improve the operating result 2011 on previous year

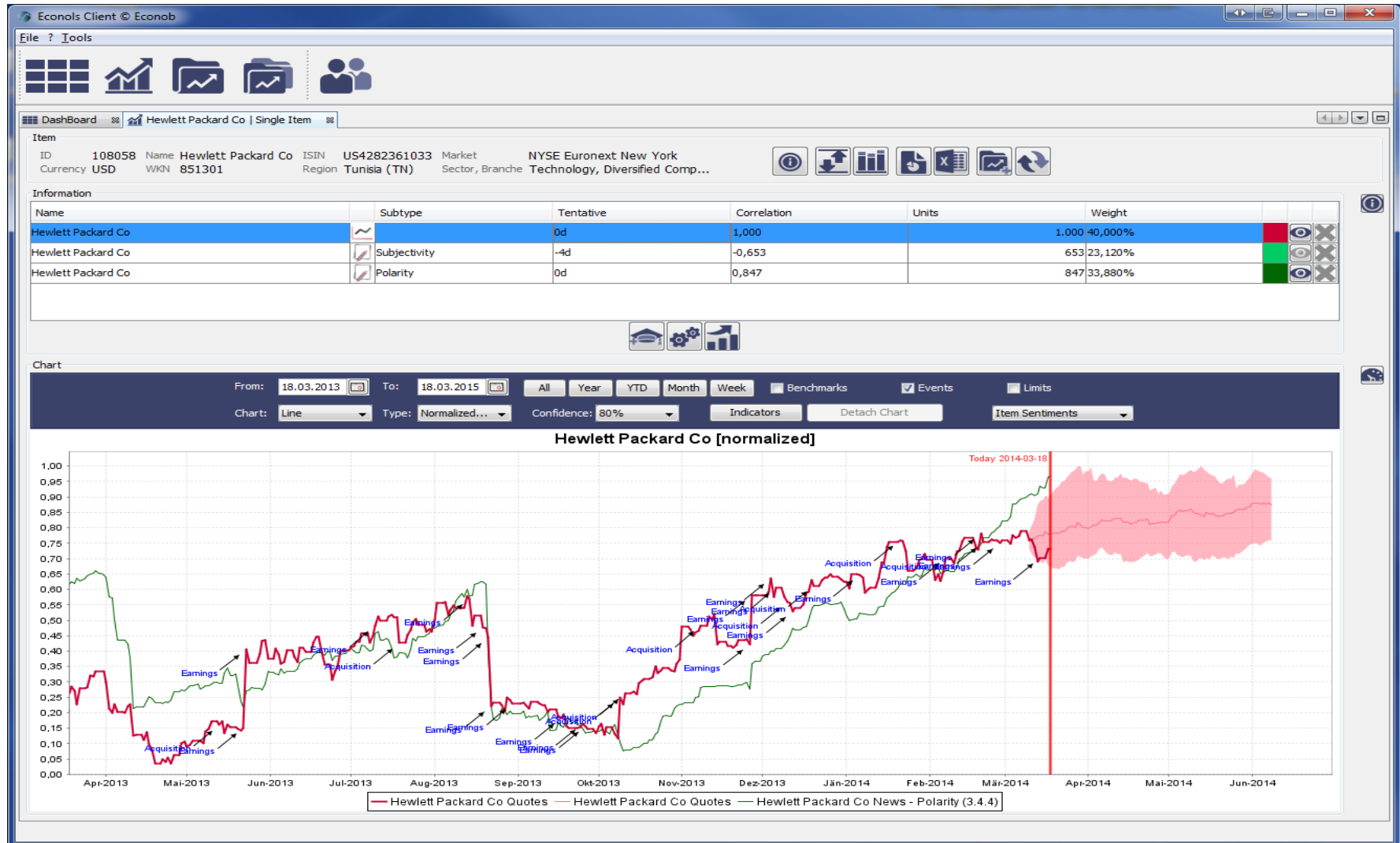
Lufthansa generates operating profit in the first half year of 2011 and confirms the outlook to improve the operating result 2011 on previous year.

27.07.11

The Lufthansa Group generated an operating profit of EUR 3m in the first half year 2011, an improvement of EUR 174m on the previous year's period.

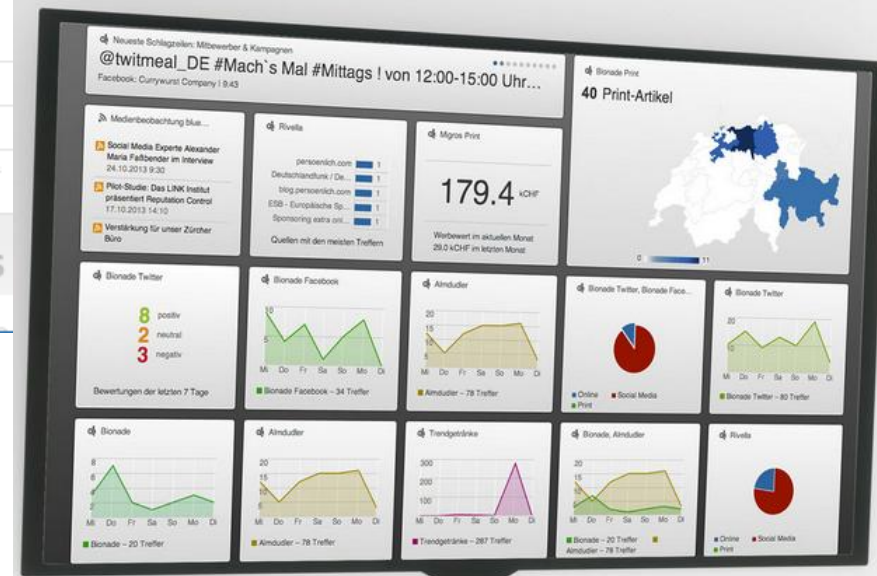
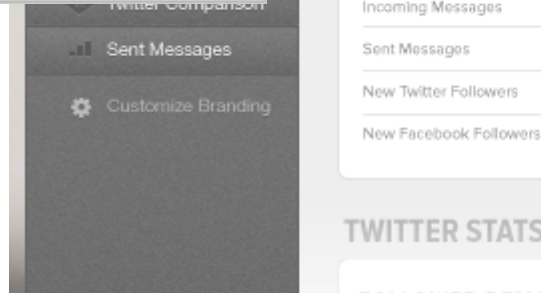
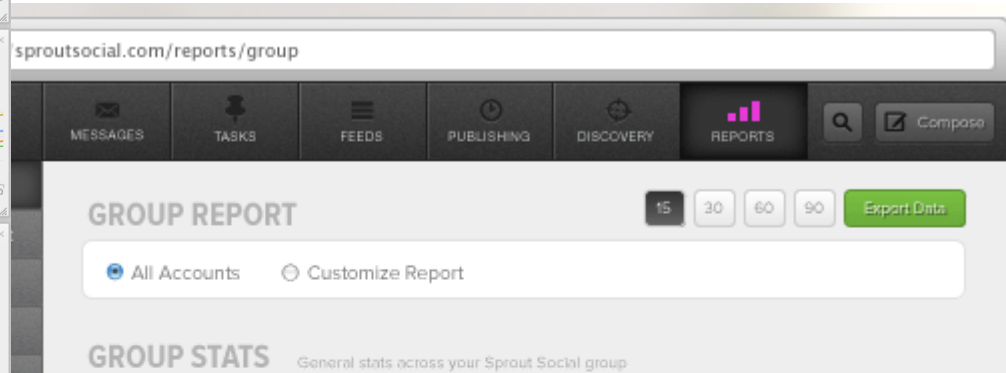
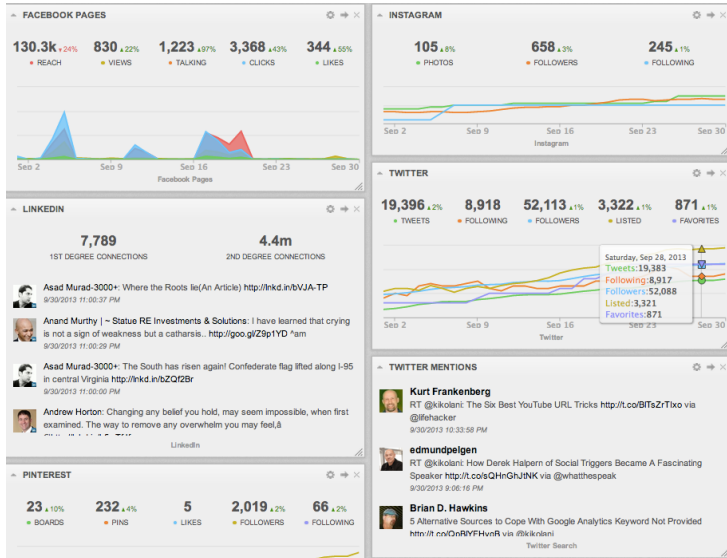
This development reflects a robust sales

## ■ Nachrichten-basierte Prognose von Aktienkursen



# Anwendungen und Potential

## ■ Großflächige Meinungsanalyse




# Anwendungen und Potential

- **Medizinische Personalisierung** die auf Geninformationsbasis und Lebenssituation zugeschnittene Behandlungen und Medikamente vorschlagen und entwerfen.
- **Denkende Systeme** die den gesamten Inhalt aller Google Books kennen und daraus neues Wissen generieren.
- **Vollautomatisierte Kommunikation** bei der Computer korrekte Antworten geben, Dialoge mit Menschen führen und Sprachsteuerung (zB. in der Robotik) erlauben.
- **Textforensik** zur Identifikation der Autorin bzw. des Authors anhand von Sprachstilanalysen.

- Sprachliche Informationsflut bewältigen
- Wissen aus Texten extrahieren
- Ergebnisse direkt in Mehrwert umwandeln

**„Google findet, Watson erklärt“**

# Danke!



Mag. Dipl.-Ing. Dr. **Marcus Hassler**  
CTO

**econob GmbH**  
A-9020 Klagenfurt, Lakeside B01a

+43 660 4890 448

[marcus.hassler@econob.com](mailto:marcus.hassler@econob.com)

[www.econob.com](http://www.econob.com)



Eine der zentralen Aufgaben von Watson ist das **inhaltliche Verstehen natürlichsprachiger Texte**. Anhand kurzer Beispiele werden einige der **Herausforderungen** veranschaulicht, denen man sich gegenüber sieht, wenn man Sprache automatisch analysieren will. Darauf aufbauend sind weitere Schritte notwendig, um den analysierten Inhalt auch wirklich **richtig zu „verstehen“**. Es ist ein langer und mühsamer Weg, doch der Lohn ist gross: **Neue Applikationen und Systeme** werden möglich, die unsere Zukunft massgeblich mitgestalten werden.