

Geschriebene Sprache inhaltlich verstehen

Herausforderungen, Anwendungsmöglichkeiten, Potential

IBM Developer Days 2014

Marcus Hassler, CEO & CTO
econob GmbH

21.05.2014, Wien, AUT



Informationssysteme – Informationen für sich arbeiten lassen



Maschinelle Verarbeitung natürlicher Sprache

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



WIE?

Technische Standard-Realisierungen

- **Dictionary Lookup:** Länder, Städte, Firmen
- **Pattern Lookup:** Date & Time, Credit Card Numbers
- **Parsing:** Firma + Stadt + Land, Titel + Person + Phone

➔ **Reicht (leider) oft nicht aus**

INPUT

Natürliche
Sprache

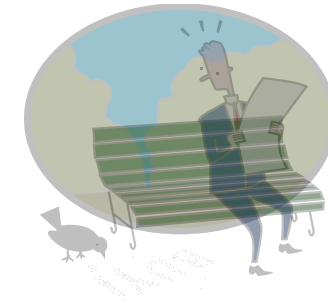
Herausforderungen für die maschinelle Sprachverarbeitung

- Obskürität 1

Meine Frau kaufte mir ein Schloss.



Ich sitze auf der Bank in der Bank.

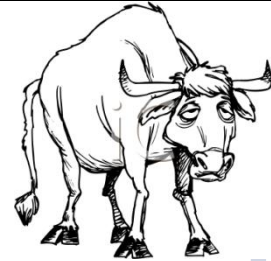


➔ **Doppeldeutigkeit (Ambiguität)**

Herausforderungen für die maschinelle Sprachverarbeitung

- Obskürität 2

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



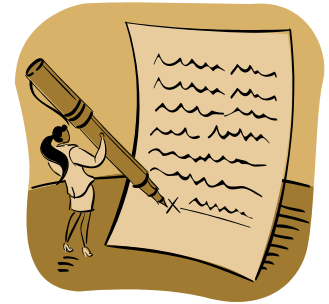
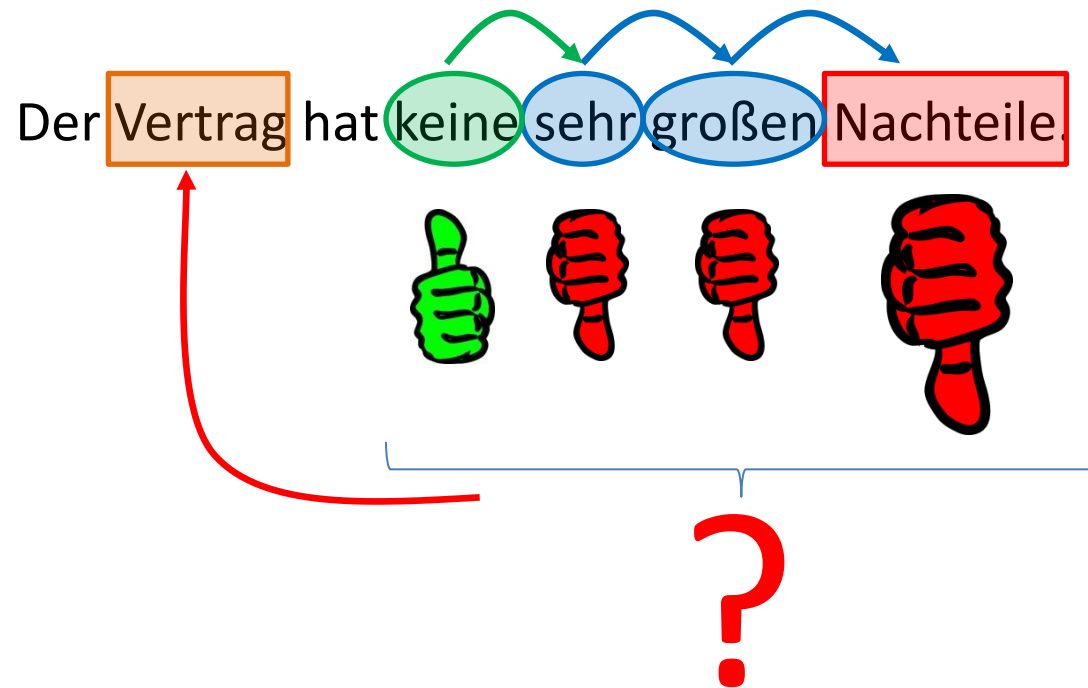
IBM kommt nach Bern. Das Unternehmen wird ... ?



➔ Koreferenz

Herausforderungen für die maschinelle Sprachverarbeitung

▪ Obskürität 3

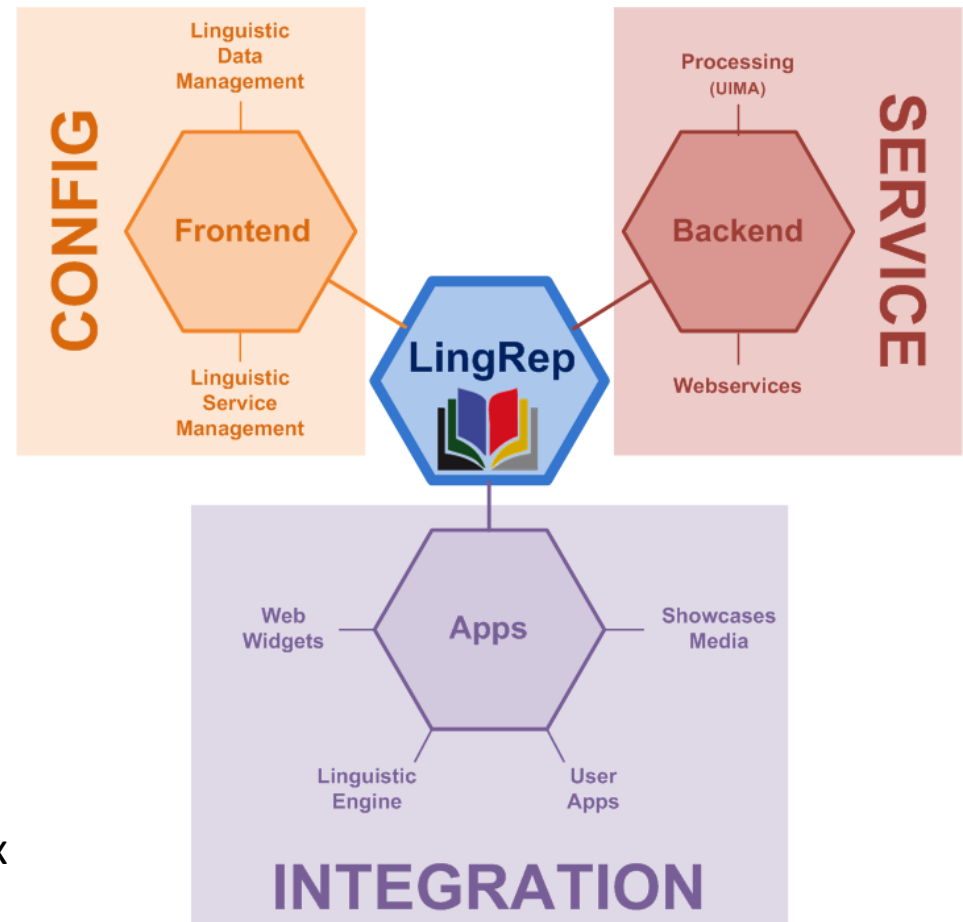


➔ Unklare Bewertung

- Komplexe Kombinationen
- Nicht eindeutige Aussage (Interpretation)



- **LingRep is a**
 - Comprehensive Linguistic Framework
 - NLP Outsourcing Platform
 - Technology
- **LingRep comes with**
 - Data Management
 - Webservice API
 - Extensive Know-How
 - Consulting & Support
- **LingRep is not a**
 - Ready-to-use Application
 - Software working out-of-the-box





- 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 'Summary of Events' section, a 'Semantic analysis [INTRANET]' section with an 'Answer' and an 'Action', and a paragraph of text. Several yellow callout boxes highlight specific elements: 'Summary of Events' points to the top of the email body; 'recognized question' points to the question 'Is there something wrong?'; 'recognized Action' points to the action 'Please provide feedback until tomorrow.'; and another 'recognized question' points to the question 'Can you please respond to this question?'. The interface also shows a sidebar with other emails and a footer with copyright information.



■ 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

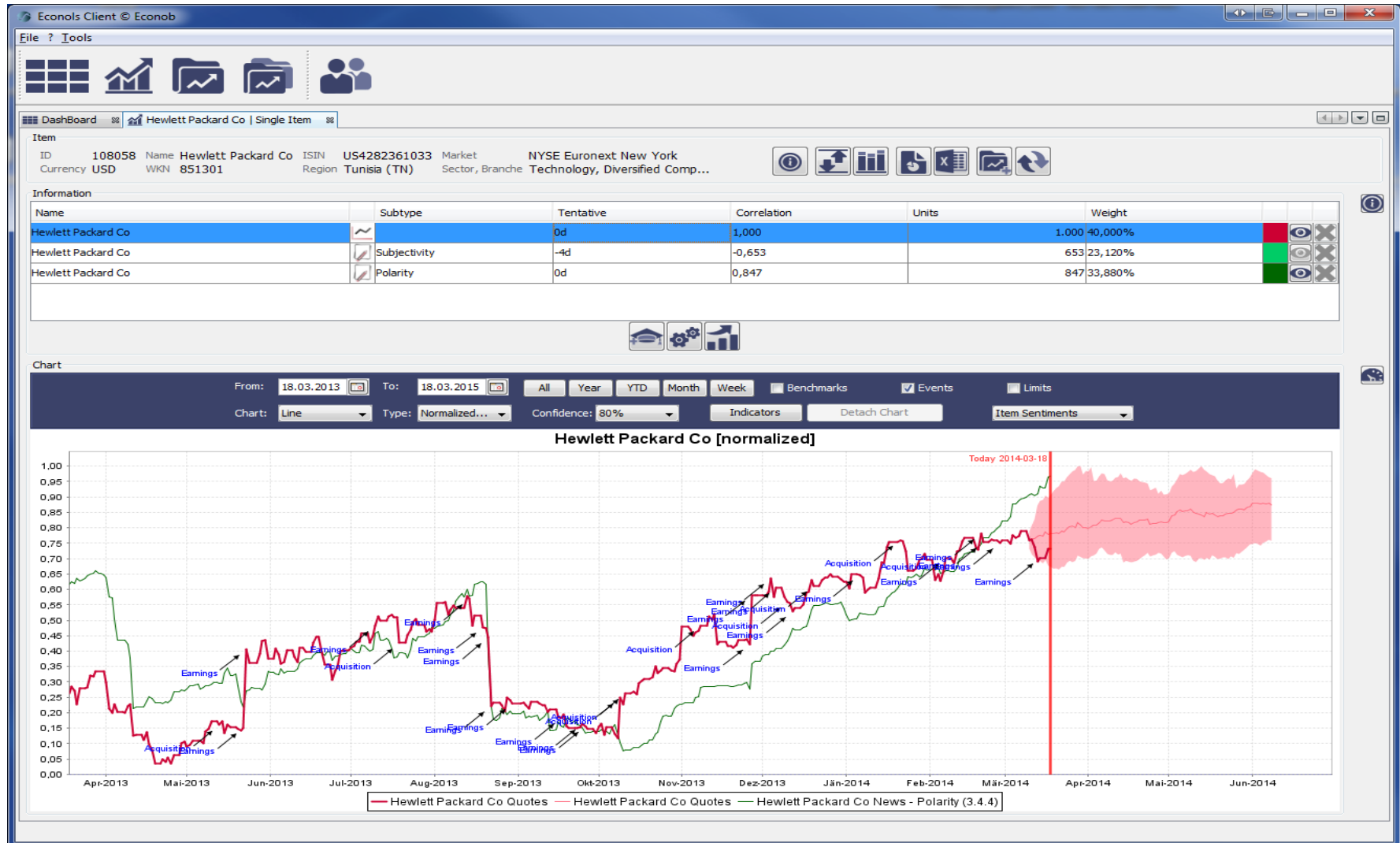
Lufthansa generates operating profit in the first half year of 2011 and confirms the outlook to improve the operating result 2011 on previous year.

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** um neben Stimmungen auch Wertschätzungen und daraus Bewertungen abzuleiten.
- **Medizinische Personalisierung** die Fachinformationen mit Patientenakten abgleicht um Best-Practices zu empfehlen.
- **Denkende Systeme** die den gesamten Inhalt aller Google Books kennen und daraus neues Wissen generieren.
- **Vollautomatisierte Kommunikation** bei der Computer Dialoge mit Menschen führen und komplette Sprachsteuerung erlauben.
- **Textforensik** zur Identifikation der Autorin bzw. des Autors anhand von Sprachstilanalysen.

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

**„Informations-Systeme sind maximal so gut,
wie deren darunterliegende Daten-Aufbereitung“**

Danke!



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

econob GmbH
A-9020 Klagenfurt, Lakeside B01a

+43 660 4890 448

marcus.hassler@econob.com

www.econob.com

Eine der zentralen Aufgaben von Informationssystemen 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.