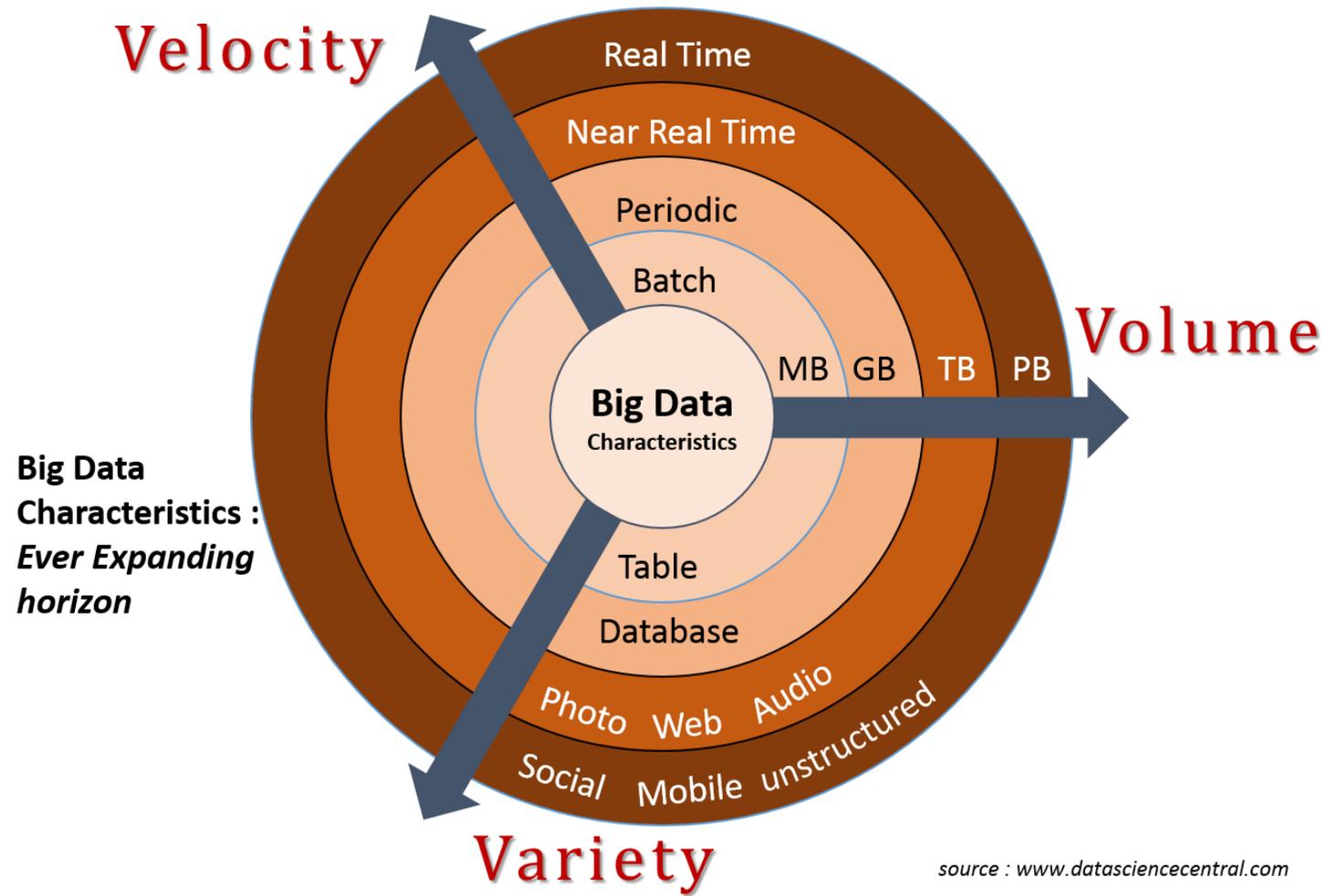


# Prozesse der Datenkonzentration

TOM ZIMMERLING

# Inhalt

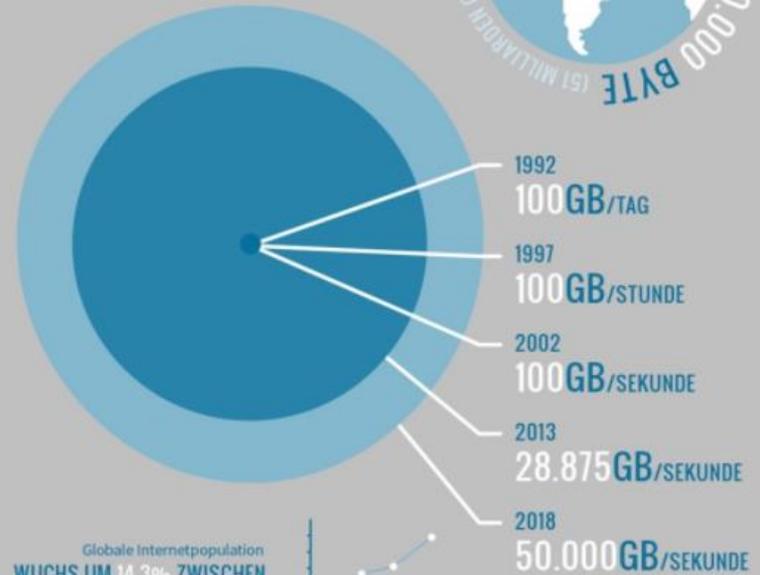
- Rückblick
- Data Mining
- Text Mining
- Wie verbessert Big Data unseren Alltag?
- Diskussion



DER GLOBALE DATENVERKEHR LAG 2013 BEI ETWA

5.000.000.000.000.000

## EIGENSCHAFTEN VON BIG DATA



Globale Internetpopulation  
WUCHS UM 14,3% ZWISCHEN  
2011 & 2013



3 MILLIARDEN  
Personen haben heute Zugang zum  
Internet, dies entspricht der  
gesamten Weltbevölkerung von 1960



# Data Mining

- automatisiertes Durchsuchen von Daten nach Mustern, Modellen oder Abweichungen
- Anwendung meist bei strukturierten Daten
- Hypothesen sind nicht notwendig
  
- Risiken entstehen bei Personenbezogenen Daten

# Data Mining

- **Umsatzdaten** WalMart führte eine Warenkorb-Analyse durch, 20 Millionen Transaktionen pro Tag erforderten eine Datenbank mit einer Kapazität von 24 Terabyte.
- **Bilder** Die NASA nimmt mit ihrem Earth Observing System Oberflächenbilder der Erde auf: 50 GByte pro Sekunde. Diese Erdbeobachtungssatelliten ermöglichen langfristige Erkundungen der Landoberfläche, der Biosphäre, der Atmosphäre und der Ozeane.
- **Textinformationen** Das World Wide Web enthält eine rapide wachsende Anzahl von Seiten und damit verbunden Daten und Informationen.

# Ablauf einer Datenanalyse – Modell nach Fayyad

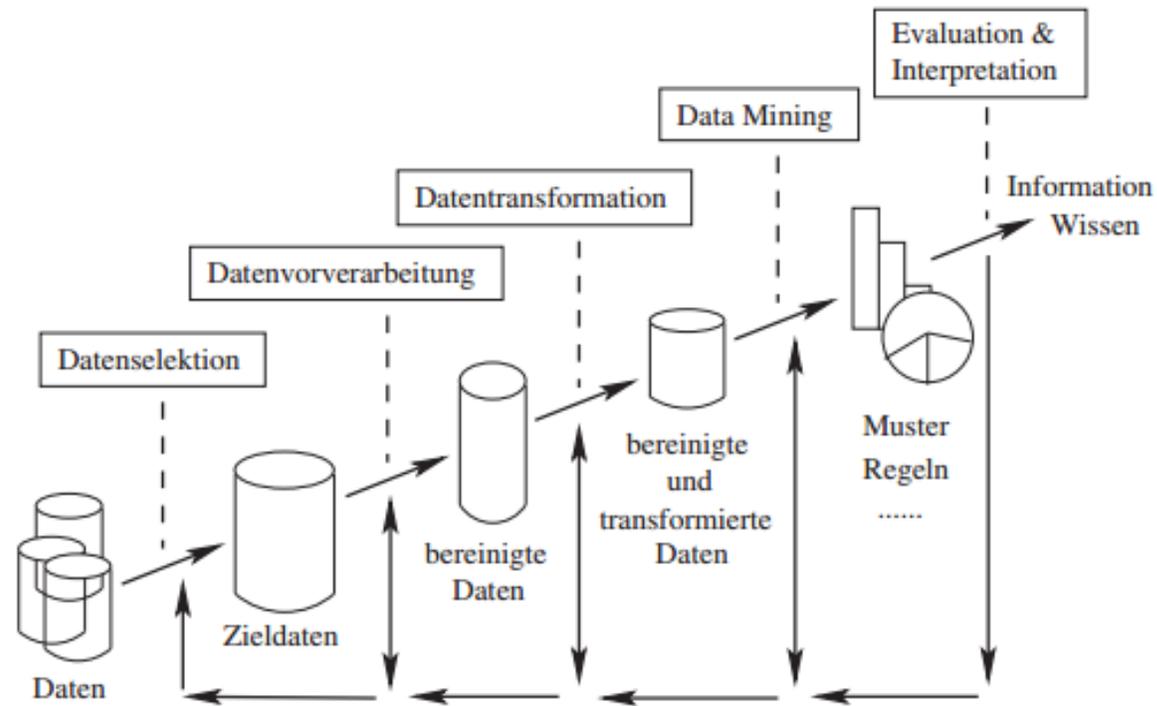


Abb. 1.2: Ablauf eines Data-Mining-Prozesses [FPSS96]

# CRISP – DM Cross-Industry Standard Process for Data Mining



# Text Mining

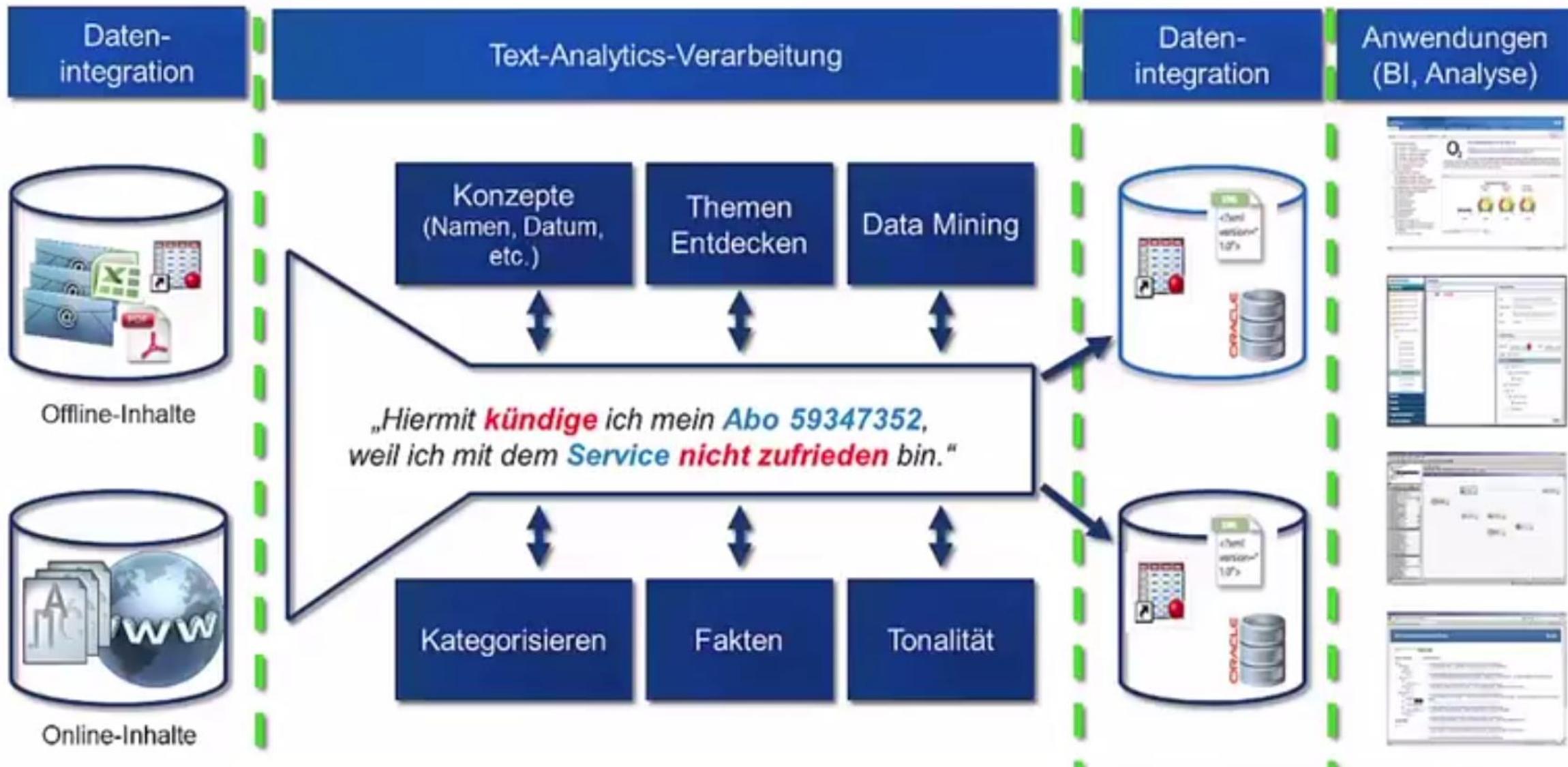
- Analyseverfahren von un- oder schwachstrukturierten Textdaten
- Szenario 1: „Wort des Tages“ Projekt der Uni Leipzig, welches Wörter anzeigt die im Netz häufig verwendet werden

(<http://wod.corpora.uni-leipzig.de/>)

- Szenario 2: Trump Tweets

# SAS TEXT ANALYTICS

# DIE INTELLIGENZ IM END-TO-END VERARBEITUNGSPROZESS



# Wie verbessert die Datenanalyse unseren Alltag?

- Verkehrsmeldesysteme
- Google flu trends
- Individuelle Kaufangebote
- Prognosen von Verbrechen

“Die Macht von Big Data

Digitale Freiheit oder  
Einschränkung?”

# Quellen

---

<https://www.youtube.com/watch?v=3Tus1SPoMWc>

**Data Mining (De Gruyter Studium)** Taschenbuch – 7. März 2016

[http://www-ai.cs.uni-dortmund.de/LEHRE/VORLESUNGEN/KDD/SS08/02\\_CRISP\\_4p.pdf](http://www-ai.cs.uni-dortmund.de/LEHRE/VORLESUNGEN/KDD/SS08/02_CRISP_4p.pdf)

<http://blog.wiwo.de/look-at-it/2015/04/22/big-data-25-trillionen-byte-daten-jeden-tag-wachst-vier-mal-schneller-als-weltwirtschaft/>

<https://www.youtube.com/watch?v=RAT0Cm5UEss>

<https://www.zdf.de/politik/frontal-21/die-macht-von-big-data-100.html>