

# Big Data & Data Science Einführung / Begriffe

WS2025/26

Prof. Dr. Klemens Waldhör

## Modulbeschreibung:

Die Studierenden können nach erfolgreichem Abschluss des Moduls:

Management-Entscheidungen durch die Beschaffung, Analyse und Präsentation von internen und externen Informationen durch Techniken und Tools vorbereiten und unterstützen,

die Wechselwirkungen der Management-Entscheidungen zwischen den unterschiedlichen Bereichen eines Unternehmens / einer Organisation bestimmen und deren Auswirkungen insbesondere hinsichtlich des Marketings, der Performance- Messung oder der Unterstützung strategischer Entscheidungen einschätzen,

"Big Data" erklären und Lösungsvorschläge zur Bearbeitung der Daten erarbeiten, bestimmen, welche Informationen für welche Entscheidungen relevant sein können, sowohl Zahlen und Daten, aber auch unstrukturierte Texte als Quellen untersuchen, grundsätzliche Methoden aus dem Bereich der "Data Science" bewerten und anwenden,

durch "Data Science" ermittelte Erkenntnisse durch statistische und mathematische Grundlagen untermauern.

## Klausur

- Klausur 90 Minuten (100% der Modulnote)
- Transferaufgabe in der Klausur (etwa 10% des Klausurumfangs):
- In der Veranstaltung werden Arbeits- und/oder Rechercheaufgaben mit explizitem Transferbezug zum praktischen Umfeld der Studierenden gestellt.
- Im Rahmen der Klausur wird dann durch angelehnte (nicht zwingend identische) Fragestellungen die Übertragung wissenschaftlicher Inhalte und Methoden auf konkrete betriebliche oder gesellschaftliche Probleme reflektiert.

## Studium

Hochschule		1. Semester	2. Semester	3. Semester	4. Semester			
Management Basics	10 CP	Datenbankmanagement	7 CP	IT- & Medienrecht	5 CP	Projektmanagement	5 CP	Bet...
K	72 4	K	56 4	K	36 4	K	40 4	K
Wirtschaftsinformatik Basics	5 CP	E-Business	5 CP	Big Data & Data Science	5 CP	Geschäftsprozess-modellierung	5 CP	Sof...
K	36 4	K	36 4	K	36 4	K	36 4	K
Mathematische Grundlagen der Informatik	5 CP	Quantitative Methoden der Informatik	5 CP	IT-Infrastruktur	8 CP	Verteilte Systeme	5 CP	IT-...
K	36 4	K	36 4	S, K	56 4	K	36 4	P
Konzepte des prozeduralen Programmierens	7 CP	Konzepte des objektorientierten Programmierens	6 CP	Konzepte des skriptsprachenorientierten Programmierens	7 CP	Algorithmen & Datenstrukturen	8 CP	We...
K	56 4	K	48 4	K	56 4	K	52 4	S
		Fallstudie/Wissenschaftliches Arbeiten	5 CP	IT-Management	5 CP			
		S, P	20 0	K	36 4			
UE	216		212		200		220	
CP	27		28		25		28	
<b>Legend:</b>		<b>E:</b> Exam	<b>KC:</b> Kompetenz-Checks	<b>D:</b> Präsentation	<b>S:</b> Seminar			

1	Begriffe / Einführung / Wissensrepräsentation
2	Data Science Prozess
3	Data Exploration
4	Data Preprocessing
5	Maschinelles Lernen: Einführung, Allgemeine Ansätze, Model Evaluation
6	Maschinelles Lernen: Clusterverfahren, K-Means, k-NN, Dimensionsreduktion
7	Maschinelles Lernen: Entscheidungsbäume
8	Maschinelles Lernen: Neuronale Netze, Transformer Modelle
9	Maschinelles Lernen: Reinforcement Learning
10	Big Data Architekturen
11	Ethische Aspekte
12	<b>Klausur</b>

Anzahl der Gliederungspunkte entspricht nicht der Anzahl der Lehrveranstaltungseinheiten.

## Google Graphcast

- Google DeepMind (2024) GraphCast: AI model for faster and more accurate global weather forecasting. <https://deepmind.google/discover/blog/graphcast-ai-model-for-faster-and-more-accurate-global-weather-forecasting/>. Abruf am 2024-02-21.
- „Unserer künstliche Graphcast-Intelligenz erwartet zu Beginn der nächsten Woche ein von Italien über Österreich nach Tschechien ziehendes Tief, welches in ganz Bayern kräftige Schneefälle bringen würde. Bei den konventionellen Modellen hat GFS eine ähnliche Entwicklung, allerdings auf einem höheren Temperaturniveau mit Regen in tieferen Lagen. Bei den anderen konventionellen Wettermodellen gibt es dieses Tief aber gar nicht. Ich weiß aber nicht, ob man daraus etwas über die Zuverlässigkeit von Graphcast wird ableiten können. Vielmehr wird wohl nach dem Orkantief am Donnerstag die Entwicklung zu Beginn der nächsten Woche noch außerhalb der Reichweite aller Modelle liegen.“ Wetterochs, 19.02.2024

Wetterochs Wettermail  
Das Wetter im Einzugsgebiet der Regnitz  
<http://www.wettermail.de>

**Google Graphcast**

**Wetterrochs Wettermail**  
**Das Wetter im Einzugsgebiet der Regnitz**  
<http://www.wettermail.de>

Neues von der künstlichen Intelligenz. Hier ein Vergleich des derzeit besten konventionellen Wettermodells IFS (ECMWF) mit neuartigen Wettermodellen, die auf maschinellem Lernen (d.h. auf künstlicher Intelligenz) beruhen. Neu dazugekommen ist mit FuXi ein KI-Wettermodell aus China (Fudan University and Shanghai Qi Zhi Institute):

[https://charts.ecmwf.int/products/plwww\\_3m\\_fc\\_aimodels\\_wp\\_mean?area=Northern%20Extra-tropics&parameter=Temperature%20850hPa&score=Anomaly%20correlation](https://charts.ecmwf.int/products/plwww_3m_fc_aimodels_wp_mean?area=Northern%20Extra-tropics&parameter=Temperature%20850hPa&score=Anomaly%20correlation)

Hier sieht man, wie die Vorhersagegüte in Bezug auf die Temperatur in der freien Atmosphäre mit der zeitlichen Länge der Vorhersage abnimmt. Je weniger stark die Kurve absinkt, um so besser also. Und da schmiert das konventionelle IFS im Vergleich zur künstlichen Intelligenz geradezu ab. Bis Tag 5 schlägt sich Google Graphcast am besten, bevor dann FuXi eine (scheinbar) größere Zuverlässigkeit hat. Nach zehn Tagen sind FuXi und Graphcast demnach so gut wie IFS nach achteinhalb bzw. neun Tagen.

Sieht man sich aber nun die Ergebnisse von FuXi im Detail an, erkennt man Erstaunliches. Je weiter die Vorhersage fortschreitet, um so detailärmer werden die Prognosen. Bei Tag 10 sieht das wie eine Klimakarte aus und nicht wie eine echte Wettervorhersage. Insbesondere sind die Luftdruck- und Temperaturgegensätze dann viel zu gering. D.h. alle extremeren Ereignisse wie Sturm oder Starkregen werden weggebügelt, der Regen wird auf riesige Gebiete verschmiert, statt konzentriert an Wetterfronten aufzutreten. In der folgenden Grafik sieht man auch, dass bei FuXi die erwartete Wetteraktivität ab Tag 5 viel zu gering ausfällt.

[https://charts.ecmwf.int/products/plwww\\_3m\\_fc\\_aimodels\\_wp\\_mean?area=Northern%20Extra-tropics&parameter=Temperature%20850hPa&score=Forecast%20activity](https://charts.ecmwf.int/products/plwww_3m_fc_aimodels_wp_mean?area=Northern%20Extra-tropics&parameter=Temperature%20850hPa&score=Forecast%20activity)

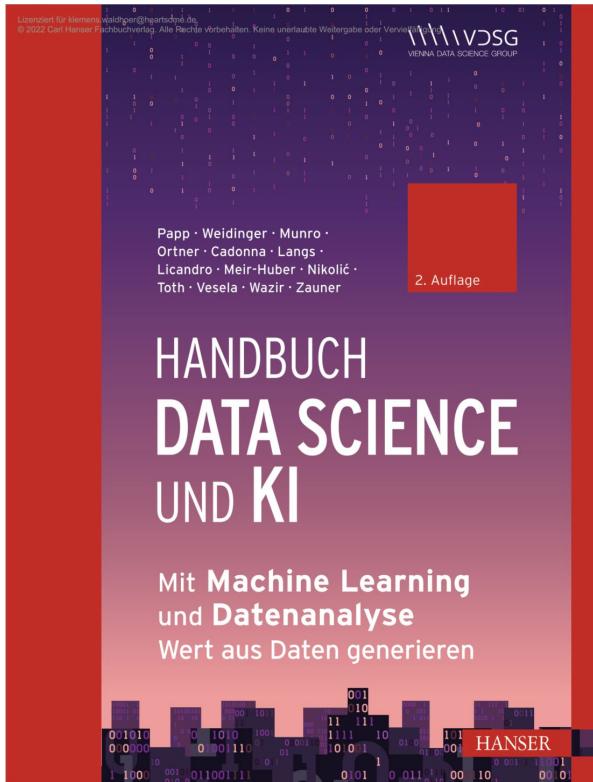
FuXi trickst also sozusagen die eine Statistik aus und fällt bei der anderen durch. Daher haben die FuXi-Prognosen sicherlich einen akademischen, aber keinen praktischen Wert. Somit geht der Titel des derzeit besten Mittelfristmodells (4-10 Tage) an Google Graphcast, welches auch am Tag 10 noch echte Wettervorhersagekarten liefert und keinen konturlosen Brei. Nur minimal schlechter als Graphcast ist übrigens AIFS, die hauseigene KI-Vorhersage des ECMWF, die ebenfalls bis zum Ende meteorologisch sinnvolle und detailreiche Ergebnisse liefert.

Der qualitative Abstand zwischen Graphcast und IFS ist übrigens ungefähr so groß wie der Abstand zwischen IFS und den übrigen konventionellen Wettermodellen (GFS, ICON, etc). D.h. der Vorsprung von Graphcast gegenüber z.B. GFS ist dann schon enorm. Ich verfolge ja die Ergebnisse des Graphcast-Modells regelmäßig und die Überlegenheit macht sich tatsächlich auch in der täglichen Praxis bemerkbar. Insbesondere sind mir bis jetzt bei Graphcast nicht diese spezifischen Schwächen aufgefallen, die bei allen konventionellen Modellen gelegentlich auftreten und zu teilweise skurrilen Entwicklungsvarianten führen, die dann wieder fallengelassen werden.

17.02.204



## Literatur



## Literatur – Data Science

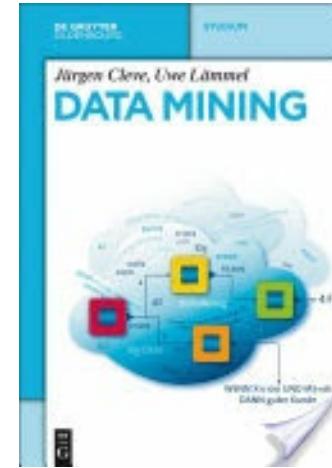
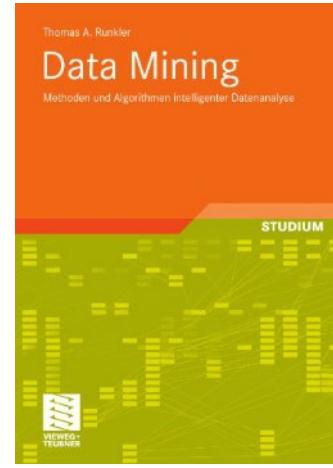
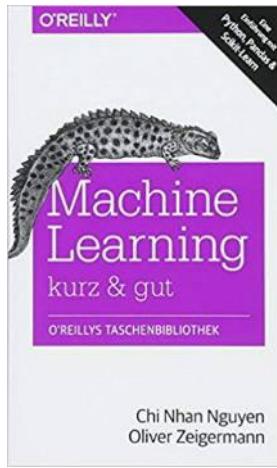
- bitkom (2018) Digitalisierung gestalten mit dem Periodensystem der Künstlichen Intelligenz. Ein Navigationssystem für Entscheider.
- Carstensen K-U, Ebert C, Ebert C, Jekat S, Klabunde R, Langer H (2010) Computerlinguistik und Sprachtechnologie. Eine Einführung. Spektrum Akad. Verl., Heidelberg.
- Cleve J, Lämmel U (2014) Data Mining. Oldenbourg Wissenschaftsverl., München. [CL]
- Géron A (2018) Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow. Konzepte, Tools und Techniken für intelligente Systeme. O'Reilly, Heidelberg.
- Hausser R (2000) Grundlagen der Computerlinguistik. Mensch-Maschine-Kommunikation in natürlicher Sprache ; mit 772 Übungen. Springer, Berlin [u.a.].
- Kemper H-G, Baars H, Mehanna W (2010) Business Intelligence - Grundlagen und praktische Anwendungen. Eine Einführung in die IT-basierte Managementunterstützung. Vieweg+Teubner Verlag / GWV Fachverlage GmbH Wiesbaden, Wiesbaden. **Online bei Springer verfügbar [KE]**
- Nguyen CN, Zeigermann O (2018) Machine Learning – kurz & gut. Eine Einführung mit Python, Pandas und Scikit-Learn. O'Reilly, Heidelberg.
- Runkler TA (2010) Data Mining. Methoden und Algorithmen intelligenter Datenanalyse. Vieweg + Teubner, Wiesbaden.

## Literatur – Expertensysteme u.ä.

- Dengel A (Hrsg) (2012) Semantische Technologien. Grundlagen - Konzepte - Anwendungen. Spektrum Akademischer Verlag, Heidelberg.
- Jackson P (1987) Expertensysteme. E. Einf. Addison-Wesley Bonn, Reading, Massachusetts, Menlo Park, California, Don Mills, Ontario, Wokingham, England, Amsterdam, Sydney, Singapore, Tokyo, Madrid, Bogotá, Santiago, San Juan
- Puppe F (1988) Einführung in Expertensysteme. Springer Berlin [u.a.]
- Reichenberger K (2010) Kompendium semantische Netze. Konzepte, Technologie, Modellierung. Springer Heidelberg
- Schnabel M (1996) Expertensysteme in der Medizin. Eine Einführung mit Beispielen ; 15 Tabellen. Fischer Stuttgart
- Schnupp P, Nguyen Huu CT (1987) Expertensystem-Praktikum. Springer-Verlag Berlin Heidelberg
- Spreckelsen C, Spitzer K (2009) Wissensbasen und Expertensysteme in der Medizin. KI-Ansätze zwischen klinischer Entscheidungsunterstützung und medizinischem Wissensmanagement, 1. Aufl. Vieweg+Teubner Verlag / GWV Fachverlage GmbH Wiesbaden Wiesbaden
- STYCZYNSKI ZA, Rudion K, Naumann A (2017) Einführung in Expertensysteme. Grundlagen, Anwendungen und Beispiele aus der elektrischen Energieversorgung. Springer Berlin Heidelberg Berlin, Heidelberg



## Literatur



## Anwendungen - Literatur

- Aichele C, Herrmann J (2021) Betriebswirtschaftliche KI-Anwendungen. Springer Fachmedien Wiesbaden Wiesbaden
- Barthelmeß U (2019) Künstliche Intelligenz Aus Ungewohnten Perspektiven. Ein Rundgang Mit Bergson, Proust und Nabokov. Springer Fachmedien GmbH Wiesbaden
- Bostrom N (2016) Superintelligenz. Szenarien einer kommenden Revolution. Suhrkamp Berlin
- Buchkremer R, Heupel T, Koch O (Hrsg.) (2020) Künstliche Intelligenz in Wirtschaft & Gesellschaft. Springer Fachmedien Wiesbaden Wiesbaden
- Dreyfus HL (1985) Die Grenzen künstlicher Intelligenz. Was Computer nicht können. Athenäum Königstein/Ts
- Gläß R (2018) Künstliche Intelligenz im Handel 1 - Überblick. Digitale Komplexität managen und Entscheidungen unterstützen. Springer Vieweg Wiesbaden
- Kurzweil R, Göttler S, Koch V, Heinemann E (op. 1993) Das Zeitalter der künstlichen Intelligenz. Carl Hanser München, Wien
- Mainzer K (2016) Künstliche Intelligenz - Wann übernehmen die Maschinen?, 1. Aufl. Springer Berlin Heidelberg Berlin, Heidelberg
- Paaß G, Hecker D (2020) Künstliche Intelligenz. Springer Fachmedien Wiesbaden Wiesbaden
- Shapiro SC (Hrsg.) (1987) Encyclopedia of artificial intelligence. Wiley New York, NY
- Stoyan H (1988) Programmiermethoden der künstlichen Intelligenz. Springer Berlin, Heidelberg



Sr

Si

Ar

Ai

Pi

Pl

Fr

Fi

Ei

Ps

Lr

Ir

Ii

Sy

Dm

Lg

Lc

Ml

Cm

Gr

Gi

Da

Te

Lu

Lt

Ms

Ma

Cn

<https://periodensystem-ki.de/Mit-Legosteinen-die-Kuenstliche-Intelligenz-bauen>

### Erstellen Sie einen zeitlichen Überblick über die Entwicklung folgender Begriffe:

Gruppe 1

- Künstliche Intelligenz
- Machine Learning
- Expertensysteme
- Problemlösen

Gruppe 2

- Neuronale Netze
- Deep Learning

Gruppe 3

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Gruppe 4

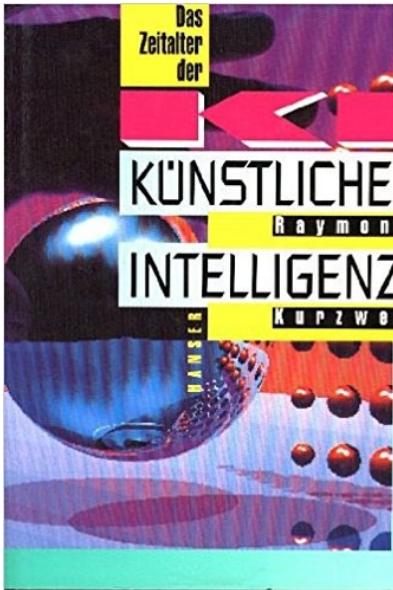
- Transformer Modelle
- CNN

## Wie alles begann

“We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.”

John McCarthy, Marvin Minsky, Nathaniel Rochester und Claude Shannon im Jahr 1955

- Alan Turing 1950:  
„ . . . um das Jahr 2000 herum Computer in der Lage sein werden, Fragen auf die gleiche Weise zu beantworten wie Menschen“
- Herbert Simon (Nobelpreisträger) 1958  
„innerhalb von zehn Jahren ein digitaler Computer der Schachweltmeister sein wird“
- Dito 1965  
„Maschinen innerhalb von zwanzig Jahren in der Lage sein werden, jede Arbeit zu verrichten, die ein Mensch tun kann“
- Marvin Minsky 1970  
„In drei bis acht Jahren werden wir eine Maschine mit der allgemeinen Intelligenz eines durchschnittlichen Menschen haben.“



**Raymond Kurzweil**

Exponentielle Zunahm  
Entwicklung

2045 Singularität - KI €  
Menschheit Unsterblich

**Transhumanismus:** Er  
menschlichen Fähigke  
kybernetischen Masch  
insbesondere der **Sing**

„Verpflichtung zum Fortschritt“

## EU-Plan: 20 Milliarden Euro für Künstliche Intelligenz bis 2020

25.04.2018 15:46 Uhr – Stefan Krempl

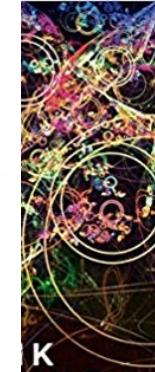
vorlesen



(Bild: Tatiana Shepeleva/Shutterstock.com)

Die EU-Kommission will mit einem Maßnahmenpaket die öffentlichen und privaten Investitionen in Künstliche Intelligenz deutlich steigern. Die Entwicklung soll mit Daten aus den Bereichen Verkehr, Gesundheit und Forschung angefüttert werden.

Krempel S (2018) EU-Plan: 20 Milliarden Euro für Künstliche Intelligenz bis 2020. <https://www.heise.de/newsticker/meldung/EU-Plan-20-Milliarden-Euro-fuer-Kuenstliche-Intelligenz-bis-2020-4034360.html>. Abruf am 2018-06-05.



SUPERCOMPUTER WATSON

Im Krankenhaus fällt die Wahrheit durch

VON SEBASTIAN BALZTER - AKTUALISIERT AM 03.06.2018 - 16:11



- Eine Gruppe hochrangiger Experten warnt Gesellschaft, die durch künstliche Intelligenz gefährdet wird.
- Zu der Gruppe gehören unter anderen Forscher aus Yale, Oxford und Tohoku sowie Entwickler aus dem Bereich der KI.
- Die Entwicklung der KI sei an einem Punkt eingreifen können - wenn Politiker, Forscher zusammenarbeiten.

Andrian Kreye (2018) Führende künstlicher Intelligenz.  
<http://www.sueddeutsche.de/digital/fuehrende-forscher-warnten-vor-intelligenz-1.3878669>. Abruf am 2018-06-05.

## Stephen Hawking warns artificial intelligence could end mankind

By Rory Cellan-Jones  
Technology correspondent

2 December 2014

[f](#) [Twitter](#) [Facebook](#) [Email](#) [Share](#)


Prof Stephen Hawking, one of Britain's pre-eminent scientists, has said that efforts to create thinking machines pose a threat to our very existence.

He told the BBC: "The development of full artificial intelligence could spell the end of the human race."

BBC (2014) Hawking: AI could end human race. <http://www.bbc.com/news/technology-30290540>. Abruf am 2018-06-05.

's Elon Musk

vorlesen



ägt, aber SpaceX-Gründer Elon Musk ist zurück. Einmal mehr erinnerte dringende

"en": Elon Musk erneut seine Warnung: "Viel gefährlicher als Atomwaffen-Elon"

3000AC

100AC 500

1000 1500

1800

1900

2000

2020

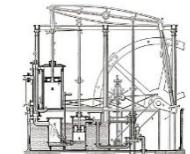
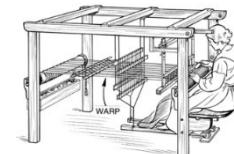
Himmels-  
scheibe  
von BebraAntikythera:  
Tragbarer  
ComputerWasserrad, Windmühlen (aber  
schon AC verwendet)

Neuzeit: Rationales, wissenschaftlich geprägtes Denken  
Messen, Bewerten, systematische Vorgehensweise



Rom – Sklaven billiger als  
Maschinen, obwohl  
durchaus bekannt, Problem  
Energieerzeugung

1. Industrielle Revolution  
Mechanisierung der menschlichen Arbeit  
Dampfmaschine  
Webstuhl



2. Industrielle Revolution  
Elektrizität  
Fließband  
Massenproduktion



3. Industrielle Revolution  
Computer, Informatik,  
Automatisierung, Robotik



Digitalisierung:  
Umwandlung analoger  
Daten in digitale Form

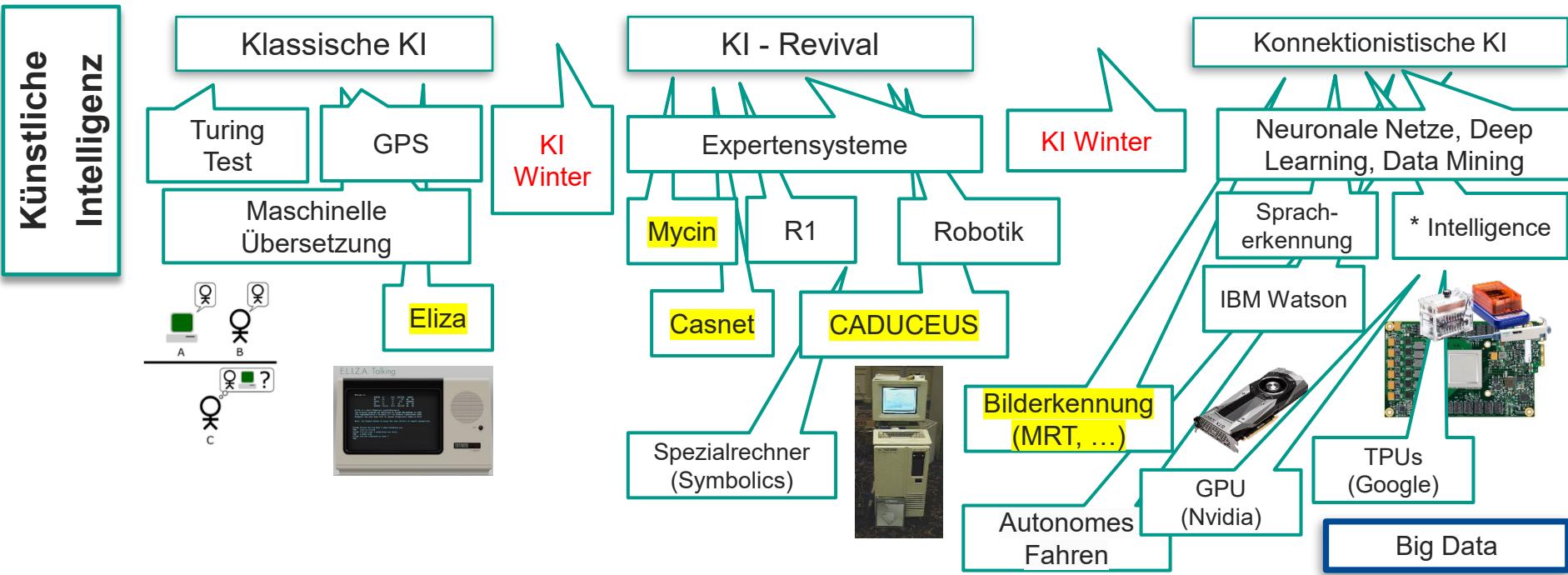
4. Industrielle Revolution  
Digitale Revolution, Industrie 4.0, CPS  
Informatisierung, IoT, VR, AR, ...



5. Industrielle Revolution  
**Künstliche Intelligenz,**  
**BigData, IoT**



1950 1960 1970 1980 1990 2000 2010 2020



## 2017

„Attention Is All You Need“ – Transformer eingeführt (NMT, Self-Attention).

## 2018

GPT (Vortrainieren + Feintuning).

BERT (Masked LM) hebt Benchmarks.

## 2019

Transformer-XL (lange Kontexte), XLNet.

RoBERTa (stärkeres Training), T5 (Text-to-Text).

## 2020

GPT-3 (Few-Shot).

ViT (Transformer fürs Sehen).

Longformer/BigBird (lange Kontexte).

RAG (Retrieval-augmentierte Generierung).

## 2021

Switch Transformer (MoE, Skalierung).

CLIP & DALL·E (Text↔Bild).

Codex (Code-Generierung).

## 2022

Chinchilla (compute-optimal). PaLM/LaMDA.

InstructGPT/RLHF (Alignment).

Whisper (ASR), Stable Diffusion (LDM).

Flamingo (multimodale Few-Shot).

## 2023

GPT-4 (multimodal).

LLaMA/Llama 2 (offene Gewichte).

## 2024–2025

GPT-4o (Echtzeit multimodal), Claude 3/3.5, Gemini 1.5 (1–2M Token).

Llama 3/3.1, Mixtral (offene MoE), Qwen2/2.5, Phi-3 (Edge).

DiT-Backbones & Text-zu-Video (z. B. Sora).

„Reasoning“-Modelle ( $o_1 \rightarrow o_3$ ).

Das erste Jahrhundert im dritten Jahrtausend ist nach einhelliger Ansicht von IT-Experten – dabei muss nicht einmal das Marktforschungsinstitut Gartner bemüht werden – gekennzeichnet von

- Zunehmenden, internen und vor allem externen Datenvolumina, die zum einen durch verbesserte Mess- und Sensortechnik (siehe z. B. RFID Technik) anfallen, und zum anderen als unstrukturierte oder semi-strukturierte Daten aus Webseiten im Web 2.0 und sozialen Netzwerken extrahiert und passend in die betrieblichen Informationssysteme integriert werden,
- Orts- und zeitunabhängiger (mobiler) Kommunikation von Mitarbeitern untereinander und mit Dritten mittels Emails, Twitter, Facebook unter Nutzung von PCs, Smartphones oder Tablets,
- Virtuellen Arbeitsplätzen und damit einhergehender Änderung der Arbeitsgewohnheiten und -organisation, wobei – wenn nötig in Echtzeit – Daten und Informationen personalisiert für Entscheidung, Planung und Controlling bereitgestellt oder abgerufen werden müssen, sowie
- Verkürzten Planungs- und Entscheidungsdauern durch die zunehmende Globalisierung, Dynamik und Interaktion von Wirtschaft und Gesellschaft.

Lit. 1

- Gestiegenes Datenvolumen und erhöhte Dimensionalität der Datenräume (Stichwort Big Data),
- Erhöhte Erfassungs-, Mess- und Veränderungsgeschwindigkeit der Daten,
- Zunehmende Heterogenität insbesondere der externen Daten aus dem Web,
- Stärkere Verteilung der Daten in sog. Clouds,
- Hohe Erwartungen der Endanwender an (schnelle) Informationsbedarfsdeckung in Echtzeit und
- Gestiegene Verarbeitungskomplexität der explorativen Datenanalyse (Data Analytics) hinsichtlich hochdimensionaler (> 100 Dimensionen) und massiver Datenbestände in der Größenordnung von Petabytes (~ 1 Mio. GB). Dies liegt an der Nutzung un- bzw. semistrukturierter Daten in (sozialen) Netzwerken und an Anforderungen an Aktualität, Informationsbedarf und Datenqualität.

Lit. 1

### Data Science

- Mathematisches Wissen
- Statistisches Wissen
- Spezielle DM-Modelle (Neuronale Netze)
- KI-Wissen
- Programmiersprachen (Python, ...)

### Big Data

- Architekturen
- Informationstechnisches Wissen
- Datenbankenwissen

### Anwendungswissen

- Anwendungsspezifisches Wissen

## HOW MUCH DATA IS NEEDED TO CLASSIFY 1000 DIFFERENT OBJECTS?

**1**

< 1 Mio. Pictures

**2**

1 Mio. Pictures

**3**

10 Mio. Pictures

**4**

100 Mio. Pictures

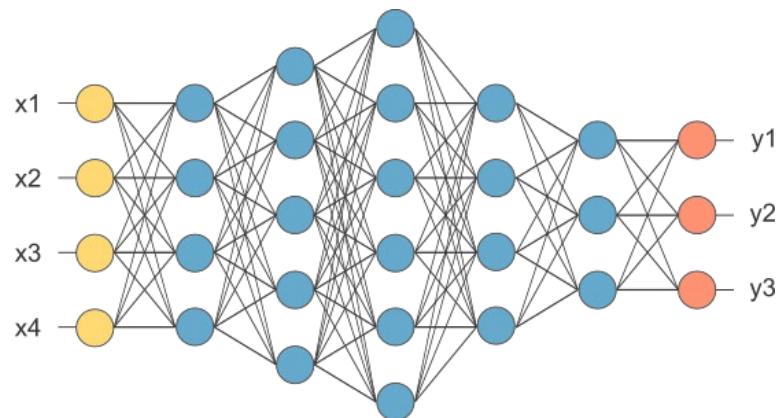




### Wieviele Daten zum Trainieren?

#### 10 Layer Neural Network

- 1 Milliarde Parameter
- 10 Mio. Bilder @ 8MB = 80TB



#### 1000 Kategorien identifizieren

- 100 Mio. Images @ 2MB per image = 200TB

### Wie lange braucht man, um ein Modell zu trainieren?

**1**

1h  
of computation

**2**

8h  
of computation

**3**

1 week  
of computation

**4**

4 weeks  
of computation



Search...



Home + Products + Markets + Demos Licensing + Resources + Clients + News + Company

Adjust your content based on emotional responses

FACE ANALYTICS  
GENDER MALE  
AGE ± 32  
MOOD HAPPY

FACE ANALYTICS  
GENDER MALE  
AGE ± 32  
MOOD MAD

FACE ANALYTICS  
GENDER MALE  
AGE ± 32  
MOOD SAD

• • • • •



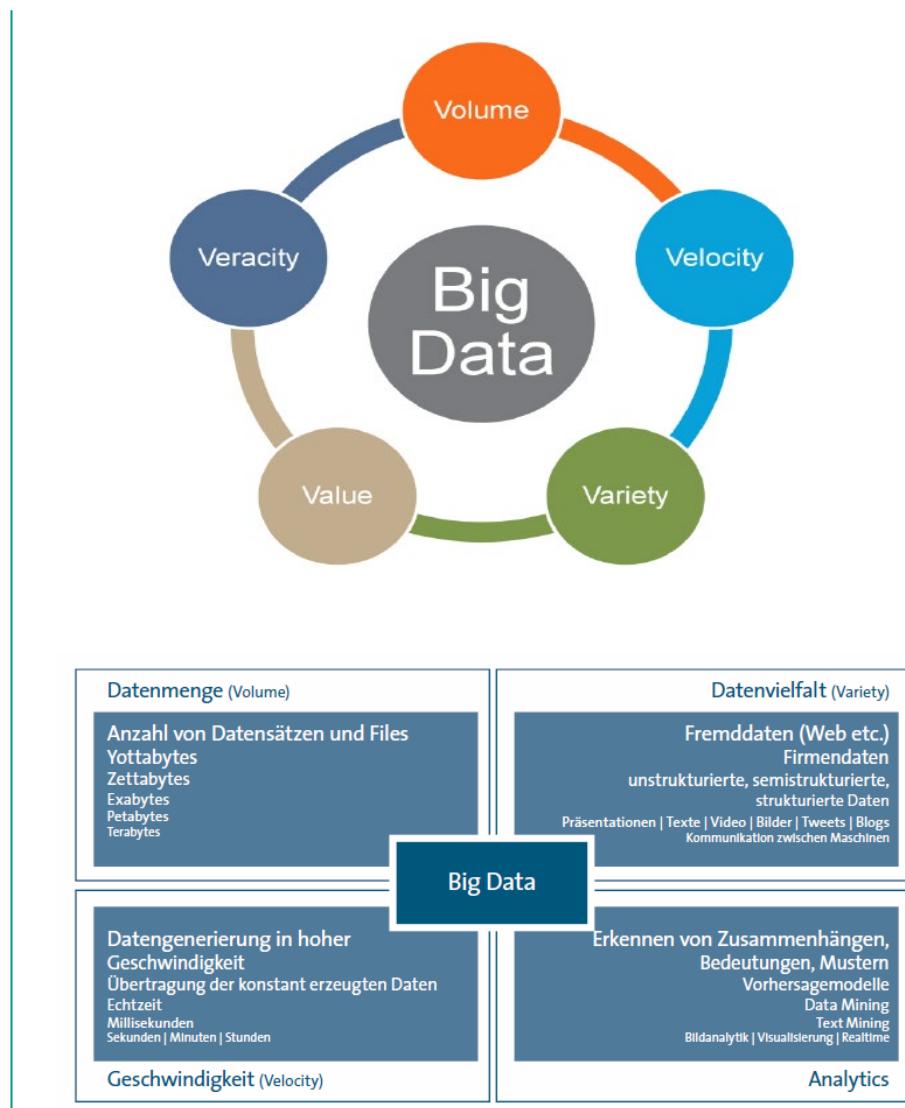
Heidborn T (2017) Daten zum Stapeln. MaxPlank Forschung(1):26-33.

## Begriffe und Definitionen

## Big Data

Big Data bezeichnet den Einsatz großer Datenmengen aus vielfältigen Quellen mit einer hohen Verarbeitungsgeschwindigkeit zur Erzeugung wirtschaftlichen Nutzens.

- **Volume (Menge)**
  - Datenbestand: umfangreich; liegt im Ter ( $10^{12}$ ) bis Zetta ( $10^{21}$ ) bytebereich; Skalierungsfähigkeit
- **Variety (Vielfalt)**
  - Speicherung von strukturierten, semi-strukturierten und unstrukturierten Multimedia-Daten (Text, Grafik, Bilder, Audio und Video)
- **Velocity (Geschwindigkeit)**
  - Datenströme (Data Streams) sollen in Echtzeit ausgewertet und analysiert werden können
- **Value (Wert) – Viability (Brauchbarkeit)**
  - Big Data Anwendung sollen den Unternehmenswert steigern - Investitionen in Personal und technischer Infrastruktur werden dort vorgenommen, wo eine Hebelwirkung besteht und einen Mehrwert generieren kann
- **Veracity (Richtigkeit / Validität )**
  - da viele Daten vage oder ungenau sind, müssen spezifische Algorithmen zur Bewertung der Aussagekraft und zur Qualitätseinschätzung der Resultate eingesetzt werden.



BITKOM (2017) Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte.

## Herausforderungen

- Heterogene, verteilte Daten
  - Mit unterschiedlichen Datenstrukturen
- Großes Datenvolumen
  - Große Datenmengen ergeben noch keinen Mehrwert
- Schutz personenbezogener Daten
- Analyseprobleme
  - Scheinkausalität zwischen Daten und Attributen
  - Interpretation der Ergebnisse (Data Mining)

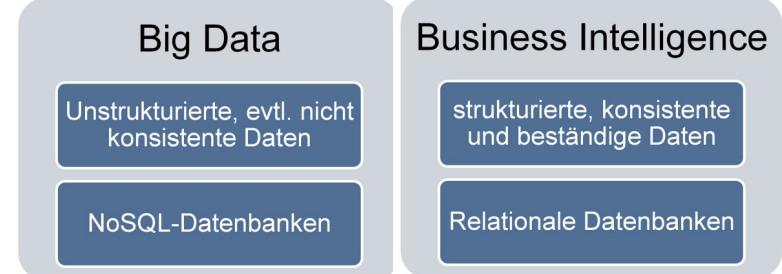
## Beispiele

- Verbrechensvorhersage
- Überwachung (NSA)
- Verkehrssteuerung
- Produkt/Prozesssteuerung
- Analyse von Gesundheitsdaten
- Kreditkartenüberprüfung (Fraud Detection)
- Meist im Zusammenhang mit Data Mining und KI

- „Der Begriff Business Intelligence [...] wird als betriebliche Entscheidungsunterstützung durch einen integrierten, aufs Unternehmen bezogenen IT-basierten Gesamtansatz definiert.
- Dabei werden die Daten aus unterschiedlichen Abteilungen extrahiert, transformiert und in einem zentralen Datenlager (engl. Data Warehouse (DW)) abgelegt.
- Selektionen auf den gesammelten Daten bezüglich eines bestimmten Kriteriums werden Data-Marts genannt und stellen einen nicht persistenten Zwischenspeicher zum Data Warehouse dar.
- Über definierte Schnittstellen können Business Intelligence Anwendungen, wie Leistungsanalysen oder Berichterstattungen, darauf zugreifen und die Daten weiterverarbeiten.“
- Daten meist aus dem internen Unternehmensumfeld verwendet.

## Beispiel

Aufgabe; Suchen / Nennen Sie einige Beispiel für eine sinnvolle BI Anwendung (in Ihrem Unternehmen)!



## Competitive Intelligence

- Konkurrenzbeobachtung, Marktforschung
- Systematische Prozess der Informationserhebung und –analyse des Marktes
  - Legale (?) Datensammlung, Daten stammen meist außerhalb des eigenen Unternehmens (im Ggs. zu Business Intelligence)
  - Keine Wirtschaftsspionage
  - Datenquellen: Internet, Unternehmensberichte, Messen, ...
  - durch den aus mehr oder minder strukturierten Informationen über Märkte, Wettbewerber und Technologien ein Verständnis über ihr Unternehmensumfeld entsteht.
  - Darauf basierend Entscheidungen treffen zu können

## Beispiel

- Patentbeobachtung und Auswertung



<https://www.questel.com/2016/03/14/questel-webinar-competitive-intelligence/>

## OLAP (Online Analytical Processing)

- Daten als **mehrdimensionaler Würfel**, auf den je nach Bedarf bestimmte Operationen ausgeführt werden
- Client Server Orientierung
- Kontrast zu **OLTP** – Online Transactional Processing

## Grundoperationen

- Slicing, Dicing, Pivoting (Rotieren), Drilling (Down, Up)

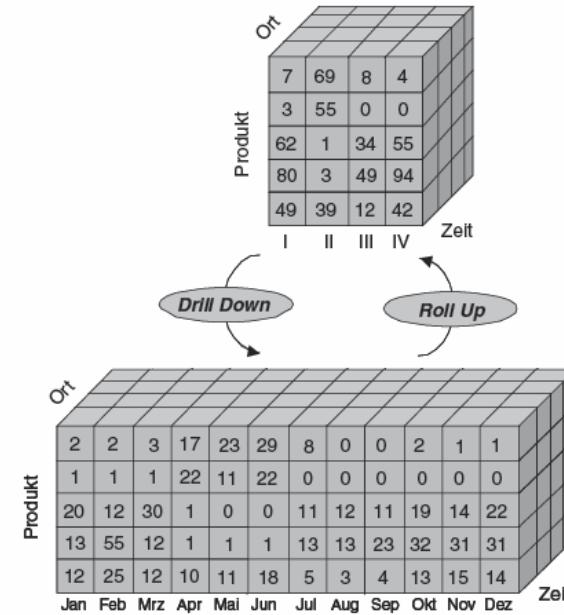


Bild 14.7 Roll Up und Drill Down

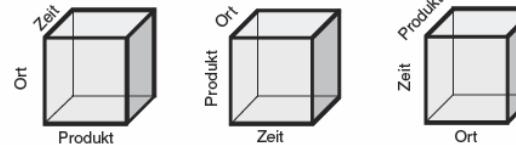
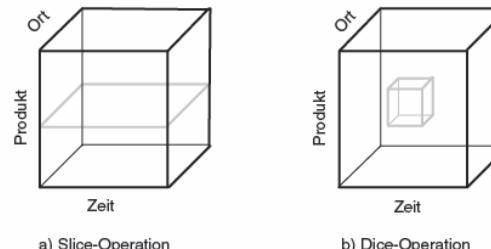


Bild 14.8 Rotation



Kudrass T (2015) Taschenbuch Datenbanken. Carl Hanser Verlag, Leipzig, S 445

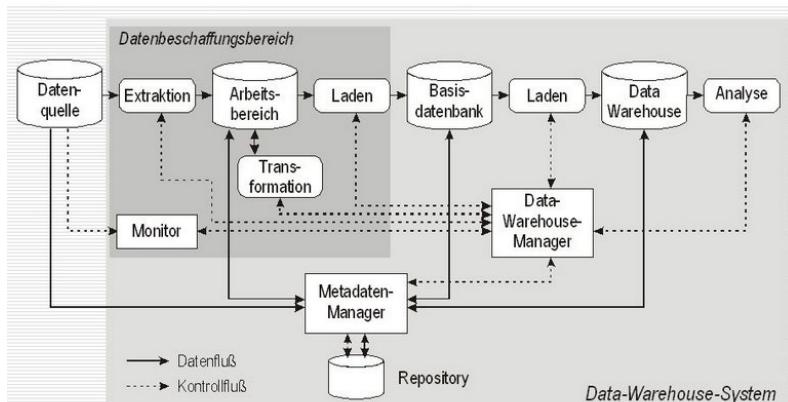
Bild 14.9  
Slice und Dice

# Data Warehouse (DWH)

- Eine physische Datenbank, die eine integrierte Sicht auf beliebige Daten (eines Unternehmens) ermöglicht und der Analyse und Unterscheidungsunterstützung dient

## Eigenschaften

- Physisch realisiert (non-volatile = nicht flüchtig)
- Schema und Daten sind integriert
- Schema ist analyseorientiert (zeitbezogen!)
- Keine Datenmodifikationen
- Verwendet historische Daten



Saake G, Sattler K-U, Heuer A (2008) Datenbanken. Konzepte und Sprachen. mitp, Heidelberg.

## Komponenten

- Datenquellen** enthalten die relevanten Daten, oft heterogen, können internen oder externen Ursprungs (Internet) sein
- Back-End-Bereich** – Werkzeuge zwischen Datenquellen und DWH
  - Extraktoren, Transformatoren, Monitor (Datenänderungen erkennen und behandeln), Datenqualitätssicherung
  - ETL = Extract, Transform, Load
- Operational Data Store** – physischer Zwischenspeicher
- Datenmigration** – Transformatoren, die der Vereinheitlichung dienen

## Beispiel

- Analyse der Geschäftsdaten von global operierenden Unternehmen, z.B. mit vielen Tochtergesellschaften (Amazon, WalMart)

## Data Mart

- **Teildatenbestand eines DWH** („kleine DWHs“)
- Definition aber oft uneinheitlich verwendet, Übergang zu DWH fließend
- Meist für einen **bestimmten Bereich** eines Unternehmens eingesetzt
- Spezielle Anwendungsbereiche
- „**materialisierte Sicht**“ eines Ausschnitts des DWHs – meist aus Effizienzgründen
- Datenmodell wird oft in einer **OLAP** Datenstruktur reflektiert

- **Struktureller Extrakt:** Nur Teil des DHW-Schemas extrahiert für bestimmte Analysen
- **Inhaltlicher Extrakt:** Gesamtes Schema, aber nur Teile der Daten extrahiert (z.B. bestimmte Periode)
- **Aggregator Extrakt:** Höhere Abstraktionsstufe gespeichert (Tage -> Monate)

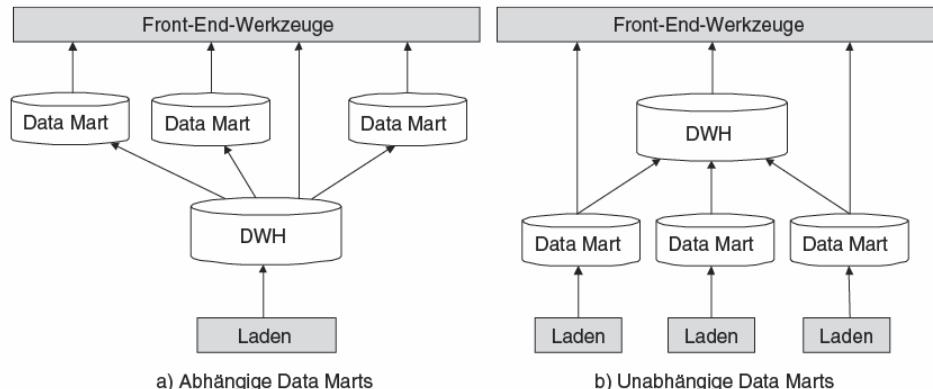


Bild 14.2 Abhängige und unabhängige Data Marts

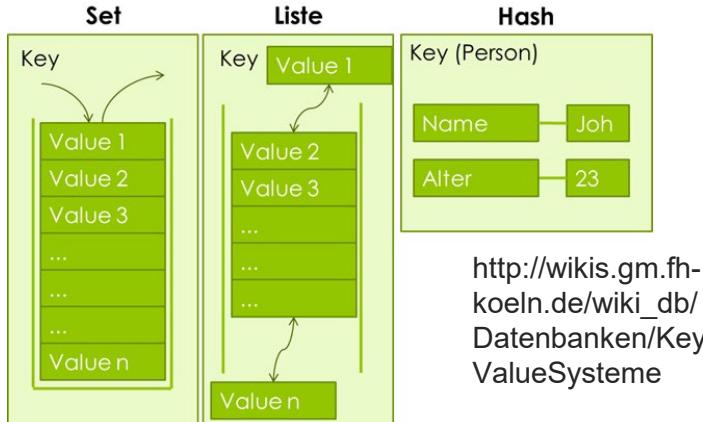
Kudrass T (2015) Taschenbuch Datenbanken. Carl Hanser Verlag, Leipzig. S 437

## Beispiel

- Webmining, insbesondere Nutzungsdaten

## NoSQL – Not Only SQL

- Traditionelle Datenbanken (SQL) erfordern ein Schema zur Daten/Strukturbeschreibung
- NoSQL:
  - Datenhaltung ohne spezielle Vorgaben („Schemas“)
  - Nicht relationales Datenmanagement
  - Hochverfügbarkeit
  - Massiv verteilte Webanwendungen
  - Horizontale Skalierung, dh. mehr Rechner (Nodes) hinzufügen



## CAP Theorem

- Consistency, Availability und Partition Tolerance, wobei nur zwei der drei parallel gewährleistet werden können.
  - Die **Konsistenz** der gespeicherten Daten. In verteilten Systemen mit replizierten Daten muss sichergestellt sein, dass nach Abschluss einer Transaktion auch alle Replikate des manipulierten Datensatzes aktualisiert werden. Diese Konsistenz sollte nicht verwechselt werden mit der Konsistenz aus der ACID-Transaktionen, die nur die innere Konsistenz eines Datenbestandes betrifft.
  - Die **Verfügbarkeit** im Sinne akzeptabler Antwortzeiten. Alle Anfragen an das System werden stets beantwortet.
  - Die **Ausfalltoleranz** der Rechner-/Servernetze. Das System arbeitet auch bei Verlust von Nachrichten, einzelner Netzknoten oder Partition des Netzes weiter.
  - Konsistenzgewährung: aufgrund des CAP-Theorems ist die Konsistenz lediglich verzögert gewährleistet (**weak consistency**), falls hohe Verfügbarkeit und Ausfalltoleranz angestrebt werden

# In-Memory Datenbanken

## ■ Prinzip

- Nutzung des RAM zur Datenspeicherung im Ggs. zu Festplatten
- Vorteil: Hohe Zugriffsgeschwindigkeit
- Nachteil: Preis

## ■ Problem

- Persistenz von Transaktionen (ACID)
- Lösungen  
Snapshot-Dateien  
Protokolldateien  
Replikation auf Festplatten

## ■ Varianten

- Hybride Systeme
- Tlw. konfigurierbar auf In-Memory

## ■ Geschwindigkeit

- Zugriffszeit auf den L1 cache **0,5 ns**
- Zugriffzeit auf den Hauptspeicher **100 ns**
- Zugriffszeit auf eine magnetische Festplatte **10.000.000 ns (bzw. 10ms)**
- **Arbeitsspeicher 100.000 Mal schneller als eine magnetische Festplatte**

## Latency Numbers



<b>L1 cache reference (cached data word)</b>	<b>0.5ns</b>
Branch mispredict	5ns
<b>L2 cache reference</b>	<b>7ns</b>
Mutex lock/unlock	25ns
<b>Main memory reference</b>	<b>100ns</b>
	<b>0.1μs</b>
Send 2K bytes over 1 Gb/s network	20,000ns
	20μs
SSD random read	150,000ns
	150μs
Read 1 MB sequentially from memory	250,000ns
	250μs
<b>Disk seek</b>	<b>10,000,000ns</b>
	<b>10ms</b>
Send packet CA to Netherlands to CA	150,000,000ns
	150ms

Quelle: Dr.-Ing. Jürgen Müller ; Warm-Up: In-Memory Data Management (2013) Folie:26

## Data Mining

- (Automatische) **Extraktion** von **Wissen** aus (großen) Datenmengen
- Techniken, Methoden und Algorithmen zur (**unbekannter?**) **Erkennung** statistischer Zusammenhänge/ Abhängigkeiten
- Bedeutungsgewinn durch „Big Data“

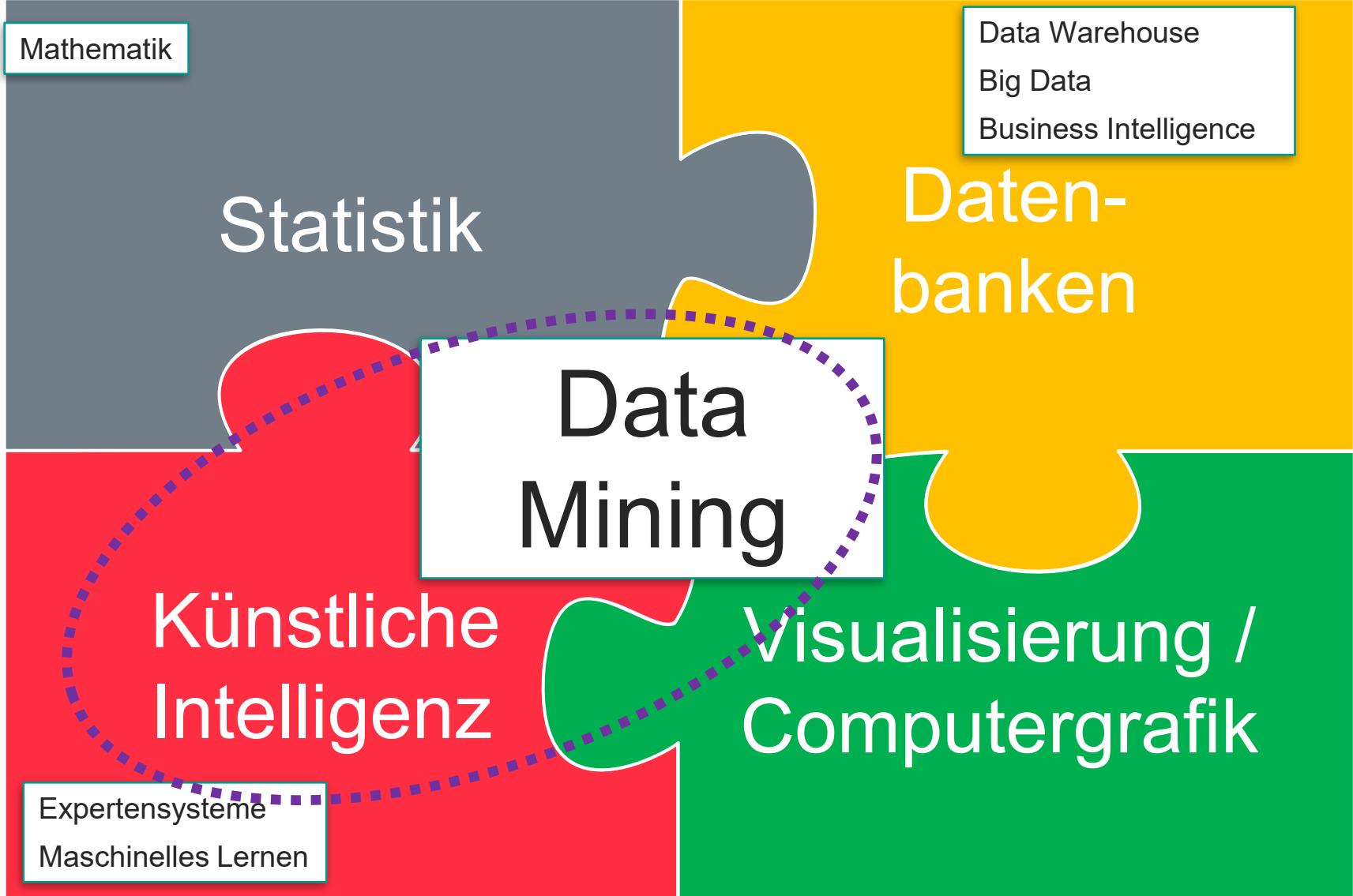
### Teilgebiete

- **Webmining**, insbesondere Nutzungsdaten
- **Text Mining**, Analyse unstrukturierter Daten
- **Data Mining** (im engeren Sinne), Analyse strukturierter Daten

### Welche Daten?

- Prozessdaten (IoT, I4.0)
- Umsatzdaten
- Genom-Daten
- Medizinische Daten
- Bilder, Videos
- Textdaten

## Zu- und Einordnung Data Mining in andere Fachgebiete



## Predictive Analytics

- Verfahren um **Vorhersagen** auf Basis von Daten zu machen
- Basis sind **historische Daten**, auf denen entsprechende (oft mathematische oder statistische) **Modelle** entwickelt werden
- Anwendung finden verschiedene Verfahren des Data Minings, überwachtes Machine Learning



Quelle: Wikipedia

## Beispiele

- Energiebedarf vorhersagen
- Kriminalität vorhersagen
- Krankheiten, Epidemien vorhersagen
- Maschinenausfälle
- Wettervorhersage

## Sentiment-Analyse

- Automatische Auswertung von Texten mit dem Ziel, eine geäußerte Haltung als positiv oder negativ zu erkennen
- Teilgebiet des Textminings
- Einsatz von NLP Technologien

### Beispiele

- Produktbewertungen
- Dienstleistungen (Hotelaufenthalte, Destinationen)
- Kundengespräche

## Meinungsführeridentifikation (Opinion Leaders)

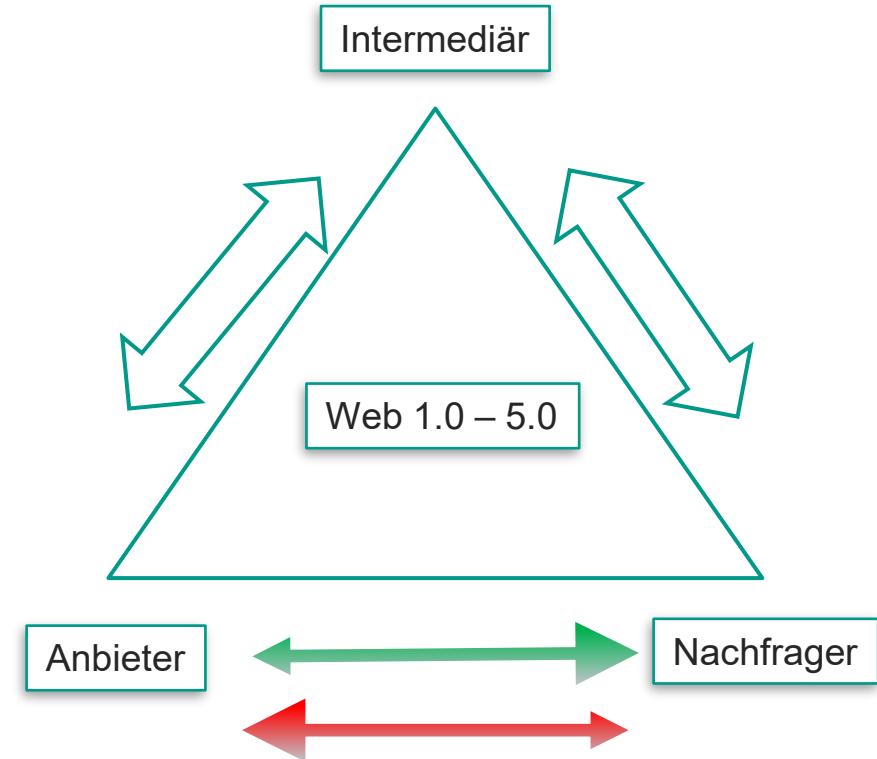
- Katz, Lazarsfeld(1957)
- Identifikation von Personen (z.B. Bloggern), die in einem oder mehreren Netzwerken eine führende Rolle spielen bzw. die Rolle von Multiplikatoren einnehmen und die Meinung, Einstellungen, Verhalten, Motive und Motivation anderer Personen (Zielgruppen) beeinflussen.
  - Selbiges für Medien
- Bündeln Informationsfluss
  - Wer wird beeinflusst?
- Google Trends

## Kennzahlenanalyse

- Reichweite
- Vernetzung (in/out)
- Häufigkeit
- Sharing von Beiträgen

## Beispiel

- Identifikation von Meinungsführern in Wissensmanagementnetzwerken in Unternehmen
- Impfmüdigkeit: Wer beeinflusst wen?



- „In der KI wird untersucht, wie man intelligentes Verhalten von Computern erfassen und nachvollziehen lassen kann oder wie man allgemein mit Hilfe von Computern Probleme löst, die Intelligenzleistungen voraussetzen.“ Claus, V.; Schwill, A.: Duden Informatik A – Z. 4. Auflage – Mannheim: Bibliographisches Institut & F.A. Brockhaus AG, 2006
- „... Der Versuch, mit Computern das menschliche Gehirn zu simulieren, um seine Funktion besser zu verstehen ... Computerprogramme durch die Nachbildung menschlicher Problemlösefähigkeiten ‚intelligent‘ zu machen“ Rechenberg P (2006) Informatik-Handbuch. Hanser, München [u.a.].
- „künstliche Intelligenz soll Maschinen in die Lage versetzen, menschliche Tätigkeiten zu übernehmen. Dabei soll das menschliche Gedächtnis, sein Lernverhalten und seine Entwicklung nachgebildet werden“ <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/technologien-methoden/KI-und-Softcomputing/Kunstliche-Intelligenz/index.html>
- „Teildisziplin der Informatik, die so unterschiedliche Problembereiche wie Bildverarbeitung, Robotik, Fließtext- und Spracherkennung, mechanische Beweise und anderes umfasst und auch die Konstruktion von wissensbasierten Systemen einschließt.“ Heinrich, L.; Roithmayr, F.: Wirtschaftsinformatik-Lexikon. – München, 1992

Nachahmung menschlichen Problemlösens, Schließen und Denkens durch Computer (Hardware, Software)

## KI-Varianten

- **Klassische KI (Symbol verarbeitende KI)**
  - Explizite Darstellung von Wissen, meist basierend auf logischen Konzepten (Logik)
  - Alle Entitäten im System werden über Symbole dargestellt und mittels Schlussfolgerungen („Inferenz“) verarbeitet
  - Menschen erarbeiten das Wissen und arbeiten es in das entsprechende KI Rahmenwerk ein
  - Beispiel: Logische Knobelaufgaben
- **Konnektionistische KI**
  - Neuronale Netze
  - Lernfähig
  - Meist größere Datenmengen notwendig
  - Heute dominierend
  - Beispiel: Bilderkennung
- **Gemeinsamkeiten**
  - Fehlertoleranz

## ▪ Bekannte Verfahren

- Regelbasierte Systeme
- Expertensysteme
- Logisches Programmieren
- Mustererkennung
- Neuronale Netze
- Deep Learning
- Genetische Algorithmen
- Semantische Netze
- Fuzzy basierte Methoden

## Wissensrepräsentation

## ▪ Deklarative WR:

- Wissen explizit als Beschreibung gespeichert
- Mittels einer davon unabhängigen Inferenzmaschine werden Probleme gelöst und Schlussfolgerungen gezogen
- Leicht erweiterbar und änderbar durch Trennung von Daten und ausführender Maschine

## ▪ Prozedurale WR:

- Wissen in Form von Prozeduren (Algorithmen, Programmen) gespeichert
- Änderungen erfordern Eingriff in die Logik des Programms

## Aussagenlogik

Die Aussagenlogik untersucht die Beziehungen zwischen Aussagen und Aussagenverbindungen. Aussagen sind abstrakte Begriffe, auch Propositionen genannt, die in der Alltagssprache durch Sätze ausgedrückt werden. Dabei kommt es in der Aussagenlogik nicht auf den konkreten Inhalt der Aussagen an, sondern nur auf die Entscheidung, ob eine Aussage wahr oder falsch ist. Die Struktur der Aussagen interessiert nicht.

Wann ist ein Schluss gültig = wahr?

Es regnet oder die Erde ist eine Scheibe

## Prädikatenlogik

Für die Prädikatenlogik sind Aussagen selbst komplexe Gebilde, die nach bestimmten Regeln aufgebaut sind. Sie verwendet Prädikate und Quantoren (Existenzquantor, Allquantor) um komplexe Aussagen zu formulieren.  
Prädikate sind Aussagen mit Variablen.

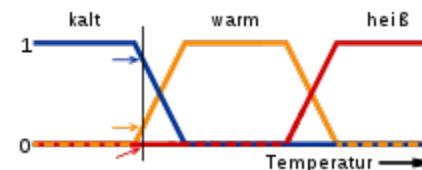
$\forall x \text{ Katze}(x) \Rightarrow \text{Säugetier}(x)$  (Alle Katzen sind Säugetiere)

Alle Menschen sind sterblich.  $\forall x \text{ menschlich}(x) \Rightarrow \text{sterblich}(x)$

Sokrates ist ein Mensch.  $\text{menschlich}(\text{Sokrates})$

## Fuzzy Logik

Unscharfe Logik. Erlaubt Schlüsse, die nicht eindeutig wahr/falsch (binär) sind, sondern auch Ausdrücke wie selten, ziemlich... enthalten. Der Wahrheitswert einer Aussage wird numerisch in einem Intervall [0,1] beschrieben (Funktion).



## Weitere Logiken

Modallogik

- Behandelt Aussagen, die Begriffe wie „möglich, notwendig“ usw. enthalten

Temporallogik

- Aussagen über die Zeit und Zeitspannen (vorher, nachher), zu welchem Zeitpunkt ist ein Ausdruck wahr?

## Zwei elementige Boolesche Algebra

- Spezialform einer allgemeinen booleschen Algebra
- true / false , 0 / 1, ...
- Wichtige Operatoren: AND, OR, NOT
- **Anwendung:** Aussagenlogik

Konjunktion			Disjunktion			Negation	
$\wedge$	0	1	$\vee$	0	1		$\neg$
0	0	0	0	0	1	0	1
1	0	1	1	1	1	1	0

$a \wedge b = b \wedge a$ $a \vee b = b \vee a$	Kommutativgesetze
$a \wedge (b \wedge c) = (a \wedge b) \wedge c$ $a \vee (b \vee c) = (a \vee b) \vee c$	Assoziativgesetze
$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$ $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$	Distributivgesetze
$a \wedge 1 = a$ $a \vee 0 = a$	Neutralitätsgesetze
$a \wedge 0 = 0$ $a \vee 1 = 1$	Extremalgesetze
$a \wedge \bar{a} = 0$ $a \vee \bar{a} = 1$	Komplementärgesetze
$a \wedge a = a$ $a \vee a = a$	Idempotenzgesetze
$a \wedge (a \vee b) = a$ $a \vee (a \wedge b) = a$	Absorptionsgesetze (Verschmelzungsgesetze)
$\bar{\bar{a}} = a$	Doppelnegationsgesetz (Involution)
$\overline{a \wedge b} = \bar{a} \vee \bar{b}$ $a \vee \bar{b} = \bar{a} \wedge \bar{b}$	DE MORGANSche Regeln

Anwendungsfall der Booleschen Algebra mit den Werten **true / false**

## Elementaraussagen

- sind entweder wahr oder falsch
- $2 + 2 = 4$  vs.  $2 + 2 = 5$
- Eisen ist leichter als Blei

## Verknüpfung

- und  $\wedge$ , oder  $\vee$ , nicht  $\neg$
- Abgeleitet: Implikation  $\rightarrow$ , Bikonditional(XNOR)  $\leftrightarrow$ , XOR  $\vee$ , NOR  $\bar{v}$ , NAND | (Shefferscher Strich)

## Definition

1. Jede Elementaraussage ist eine Aussage.
2. Sind  $A, A_1, A_2$  Aussagen, so sind auch  $A_1 \wedge A_2, A_1 \vee A_2, \neg A$  Aussagen.

Der **Wahrheitswert einer Aussage** errechnet sich aus den Wahrheitswerten der Teilaussagen anhand von **Wahrheitstabellen**.

$T = \text{true} = \text{wahr} = 1$

$F = \text{false} = \text{falsch} = 0$

Identisch mit Boolescher Algebra!

A	$\neg A$	A	B	A	B	A	B	$A \rightarrow B$	$A \leftrightarrow B$
T	F	T	T	T	T	T	T	T	T
F	T	T	F	F	T	F	F	F	F
		F	T	F	T	T	T	F	F
		F	F	F	F	T	T	T	T

A	B	A	B	A   B	A	B
T	T	F	F	F	F	F
T	F	T	F	T	F	F
F	T	T	F	T	F	F
F	F	F	F	T	T	T

## Aussagenlogik - Beispiele

A	$\neg A$	A	B	A	B	$A \rightarrow B$	$A \leftrightarrow B$
T	F	T	T	T	T	T	T
F	T	T	F	F	T	F	F
		F	T	F	T	T	F
		F	F	F	F	T	T

A	B	$\neg A$	B	$\neg A$	$\neg A$	B	$A \rightarrow B$
T	T	?	?	F	T	T	T
T	F	?	?	F	F	F	F
F	T	?	?	T	T	T	T
F	F	?	?	T	T	T	T

Alle logischen Operatoren durch NAND darstellen

Verknüpfung	NAND Umsetzung
$\neg A$	$A   A$
$A \wedge B$	$(A   B)   (A   B)$
$A \vee B$	$(A   A)   (B   B)$
$A \text{ NOR } B$	$((A   A)   (B   B))   ((A   A)   (B   B))$
$A \text{ XOR } B$	$(A   (B   B))   ((A   A)   B)$
$A \rightarrow B$	$A   (B   B)$
$A \leftrightarrow B$	$(A   B)   ((A   A)   (B   B))$

$$A \quad B \quad C \quad | \quad (A \wedge \neg B) \vee C \rightarrow (A \vee B)$$

1]	T	T	T		T	✓
2]	T	T	F		T	✓
3]	T	F	T		T	✓
4]	T	F	F		T	✓
5]	F	T	T		T	✓
6]	F	T	F		T	✓
7]	F	F	T		F	
8]	F	F	F		T	✓

$$A \quad B \quad C \quad | \quad A \wedge B \wedge C$$

1]	1	1	1		1	✓
2]	1	1	0		0	
3]	1	0	1		0	
4]	1	0	0		0	
5]	0	1	1		0	
6]	0	1	0		0	
7]	0	0	1		0	
8]	0	0	0		0	

$$1 = T, 0 = F$$

### Prolog: Programming in Logic

```
vater_von(christine,wolfgang).
vater_von(renate,wolfgang).
vater_von(juergen,wolfgang).
vater_von(lutz, wolfgang).
vater_von(martin, lutz).
vater_von(ralf, lutz).
frau_von(wolfgang, gerda).
frau_von(lutz, kerstin).

grossvater_von(X, Y) :-
vater_von(X, Z), vater_von(Z, Y).

?- grossvater_von(ralf, X).
X = wolfgang.
```

Ist die Aussage wahr oder falsch?

„Wenn Heiligabend und Silvester auf einen Tag fallen,  
dann hat jede Stunde 70 Minuten“.

## Übung 5: Aussagenlogik Mörder finden

Wer war der Mörder?

Der Butler oder der Koch oder der Chauffeur hat den Baron umgebracht.  
Wenn der Koch den Baron umgebracht hat, dann war der Eintopf vergiftet, und wenn der Chauffeur den Baron umgebracht hat, dann war eine Bombe im Auto. Der Eintopf war nicht vergiftet und der Butler hat den Baron nicht umgebracht.



### Was ist das Ziel?

- Problemlösung analog zu einem Experten für ein bestimmtes Fachgebiet
- Wissen wird vom Experten explizit zur Verfügung gestellt (codiert), meist in Form von Heuristiken (Erfahrungswissen)

### Realisierung

- Oft mittels regelbasierter Systeme realisiert
- „Business Rule Engines“

### Beispiele

- Konfigurieren von Computern – passende Komponenten finden
- Versicherung – optimale Versicherungskombination finden basierend auf den Angaben des Kunden
- Medizin – Auswahl von Medikamenten

## Wissensdarstellung

- Darstellung des Wissens in Form von Regeln (nicht nur Regeln in Logikform!) - Produktionsregel
  - Wenn **Bedingung**, dann **Schlussfolgerung**
  - Wenn **Bedingung**, dann **Aktion**
  - Ggf. mit Wahrscheinlichkeiten versehen

## Wissensbasis

- Regelbasis
- Faktenbasis

## Verfahren

- Vorwärtsverkettung
- Rückwärtsverkettung

## Beispiel

- R1/XCON – Konfiguration von VAX Maschinen
  - OPS5

ASSIGN-POWER-SUPPLY-1

IF:

THE MOST CURRENT ACTIVE CONTEXT IS ASSIGNING A POWER SUPPLY  
 AND AN SBI MODULE OF ANY TYPE HAS BEEN PUT IN A CABINET  
 AND THE POSITION IT OCCUPIES IN THE CABINET IS KNOWN  
 AND THERE IS SPACE IN THE CABINET FOR A POWER SUPPLY  
 AND THERE IS NO AVAILABLE POWER SUPPLY  
 AND THE VOLTAGE AND FREQUENCY OF THE COMPONENTS IS KNOWN

THEN:

FIND A POWER SUPPLY OF THAT VOLTAGE AND FREQUENCY  
 AND ADD IT TO THE ORDER

### Ein wichtiges Anwendungsgebiet der KI / des Data Mining sind Empfehlungssystem („Recommender Systeme“)

- Suchen und Erstellen Sie eine Definition für Empfehlungssysteme!
- Beschreiben Sie zwei Anwendungsbeispiele!
- Welche Verfahren werden für Empfehlungssysteme eingesetzt?

**Regelbasierte Expertensysteme sind klassische KI-Systeme, früher stark „gehypt“. Heute wird dieser Begriff eher selten verwendet und ebenso die Anwendung solcher Systeme.**

- Welche Gründe vermuten Sie dafür?
- Wodurch wurden dieser KI-Typ ersetzt?
- Ist der Begriff wirklich verschwunden?