

LERNENDE
MASCHINEN
02.05.2017

INDUSTRIE
4.0
23.05.2017

SPRACH-
DIALOGUE
09.05.2017

KÜNSTLICHE
INTELLIGENZ

KI

**BIG
DATA
13.06.2017**

TEAM-
ROBOTIK
30.05.2017

AUTONOME
SYSTEME
16.05.2017

**ALTERS-
ASSISTENZ**

**SMART
SERVICE**

**SICHER-
HEIT**

**EMOTION &
VERHALTEN**

Vorlesungsreihe 2017: Künstliche Intelligenz für den Menschen: Digitalisierung mit Verstand

Mainz, 13. Juni 2017



BIG DATA: Maschinelles Lernen und Wissenextraktion aus großen Datenmengen

Prof. Dr. rer. nat. Dr. h.c. mult.

Wolfgang Wahlster



Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Saarbrücken/Kaiserslautern/Bremen/Berlin/Osnabrück

Tel.: (0681) 85775-5252

E-mail: wahlster@dfki.de

www.dfki.de/~wahlster



BIG DATA: Massendaten als Wirtschaftsgut Daten als das Öl/Gold der Zukunft



Big Data: Daten als Wirtschaftsgut



95% der 1,2 Zettabyte weltweiten digitalen Daten sind unstrukturiert - bei einem Datenwachstum pro Jahr von 62%.

Echtzeit-Analyse von Massendaten als Basis für Smart City Technologien



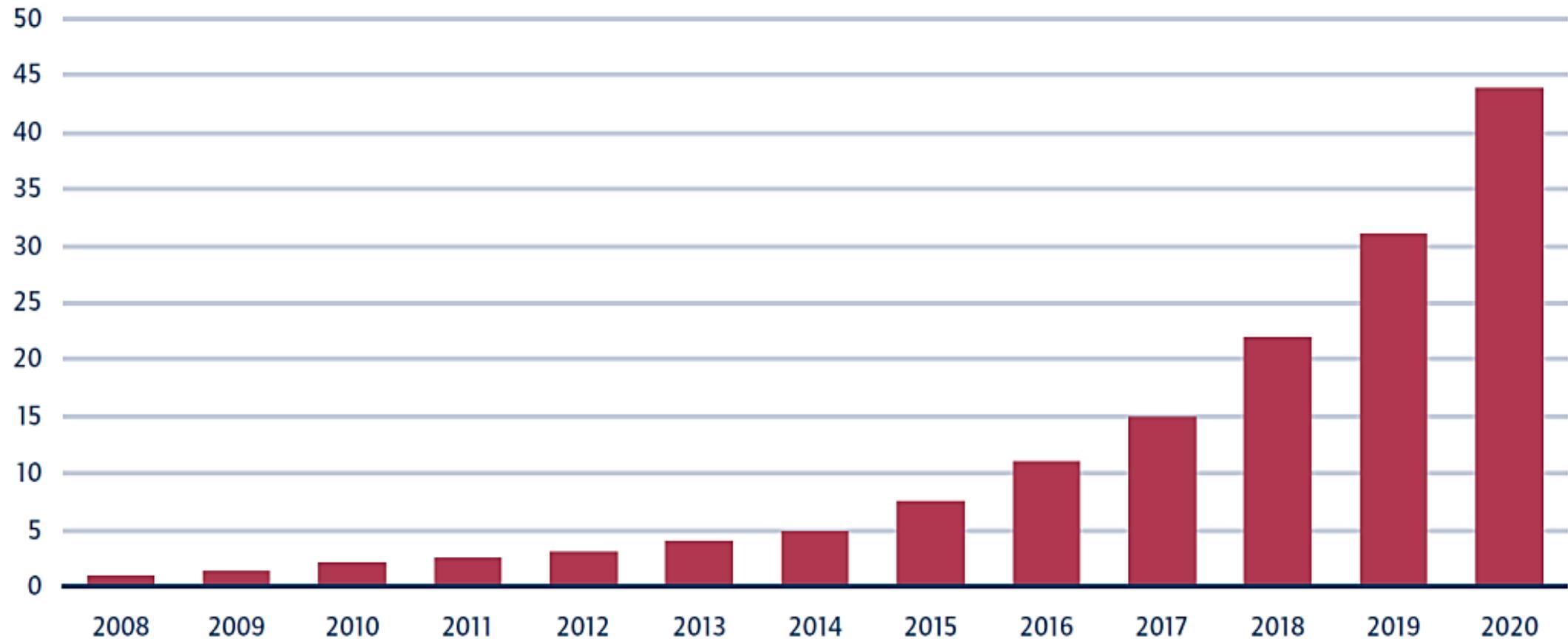
Datengetriebenes Management von Massenveranstaltungen



• bietet interessante **Zusatzfunktionen** an

• **Kommunikationsplattform**, um Besucher stets direkt erreichen zu können

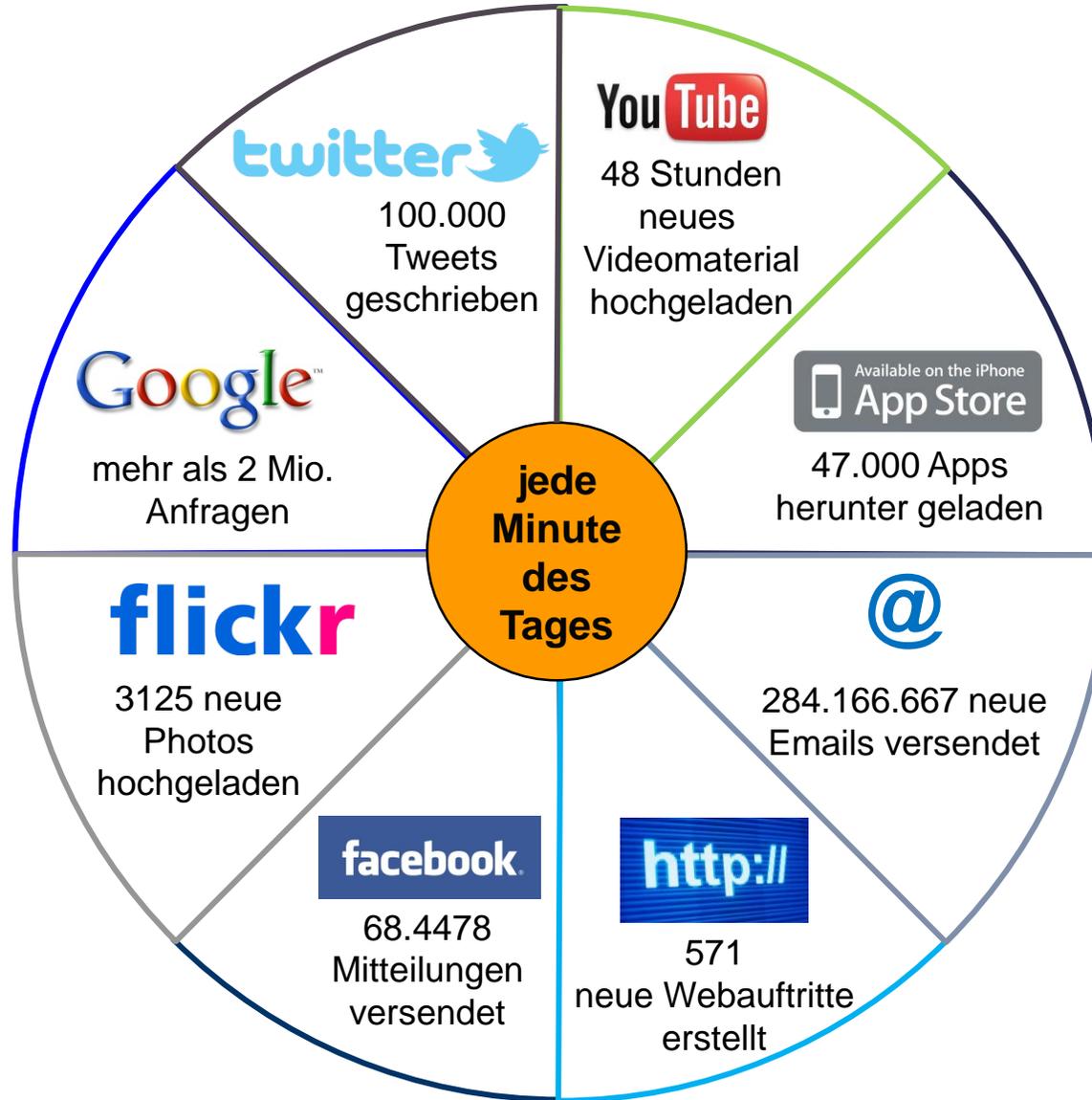
Entwicklung globaler Datenvolumina bis 2020 (in Zettabyte)



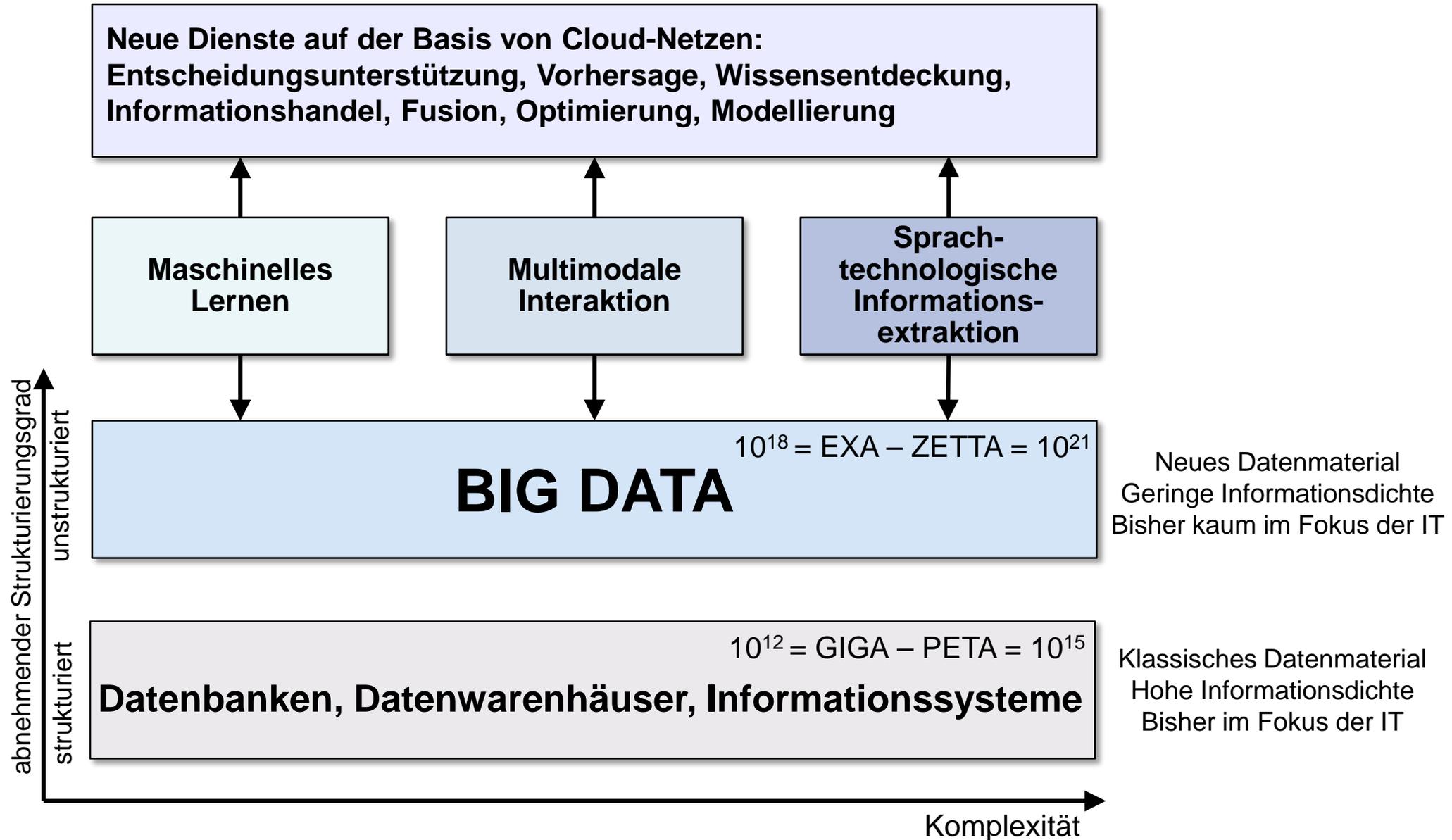
Quelle: AT Kearney 2013: 2

Weiterhin exponentielles Wachstum der weltweiten IKT-Nutzung generiert Zettabytes an unstrukturierten Daten: BIG DATA

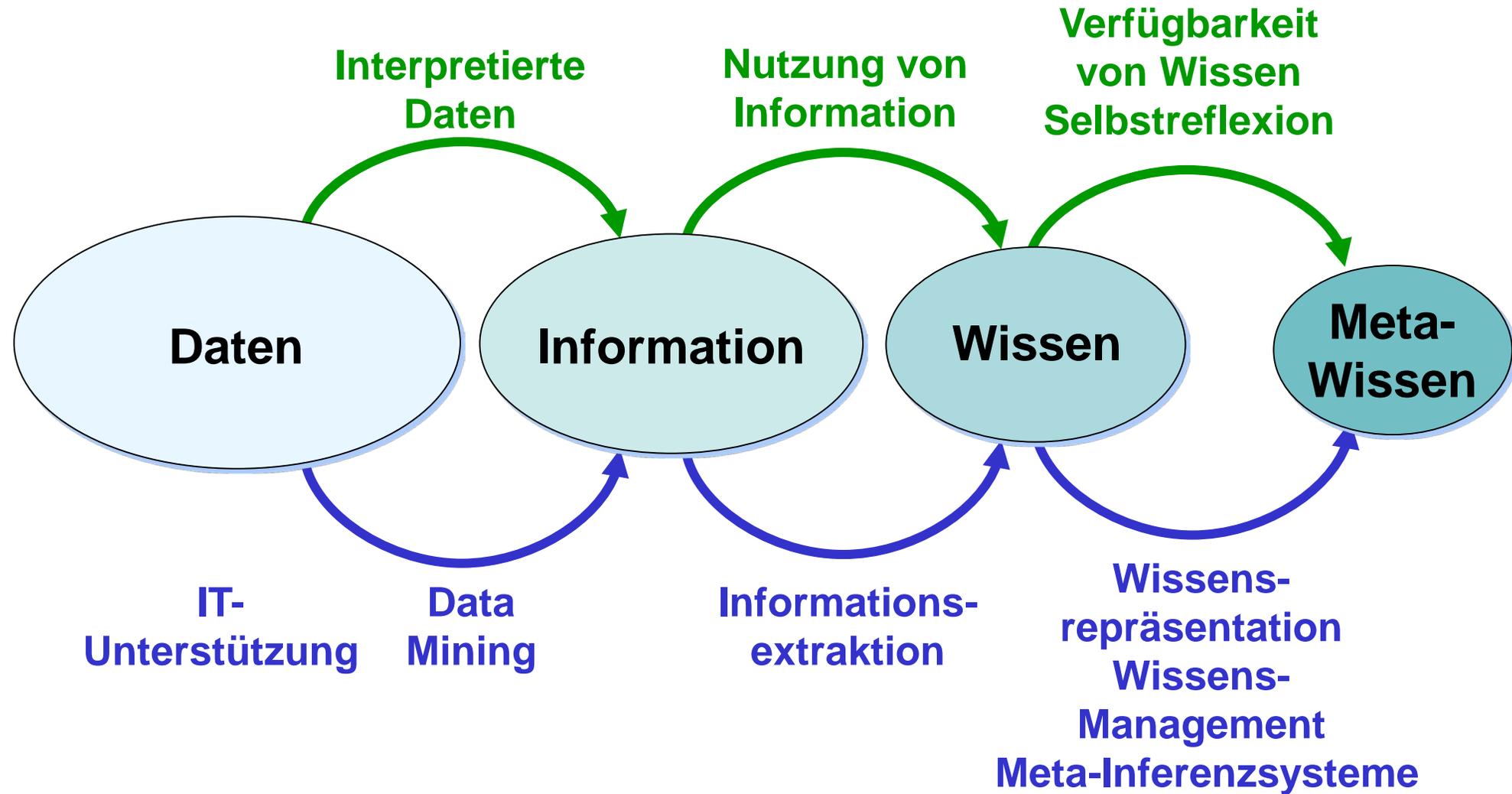
Bei Totalausfall der IKT würde die Weltwirtschaft binnen weniger Stunden zusammenbrechen



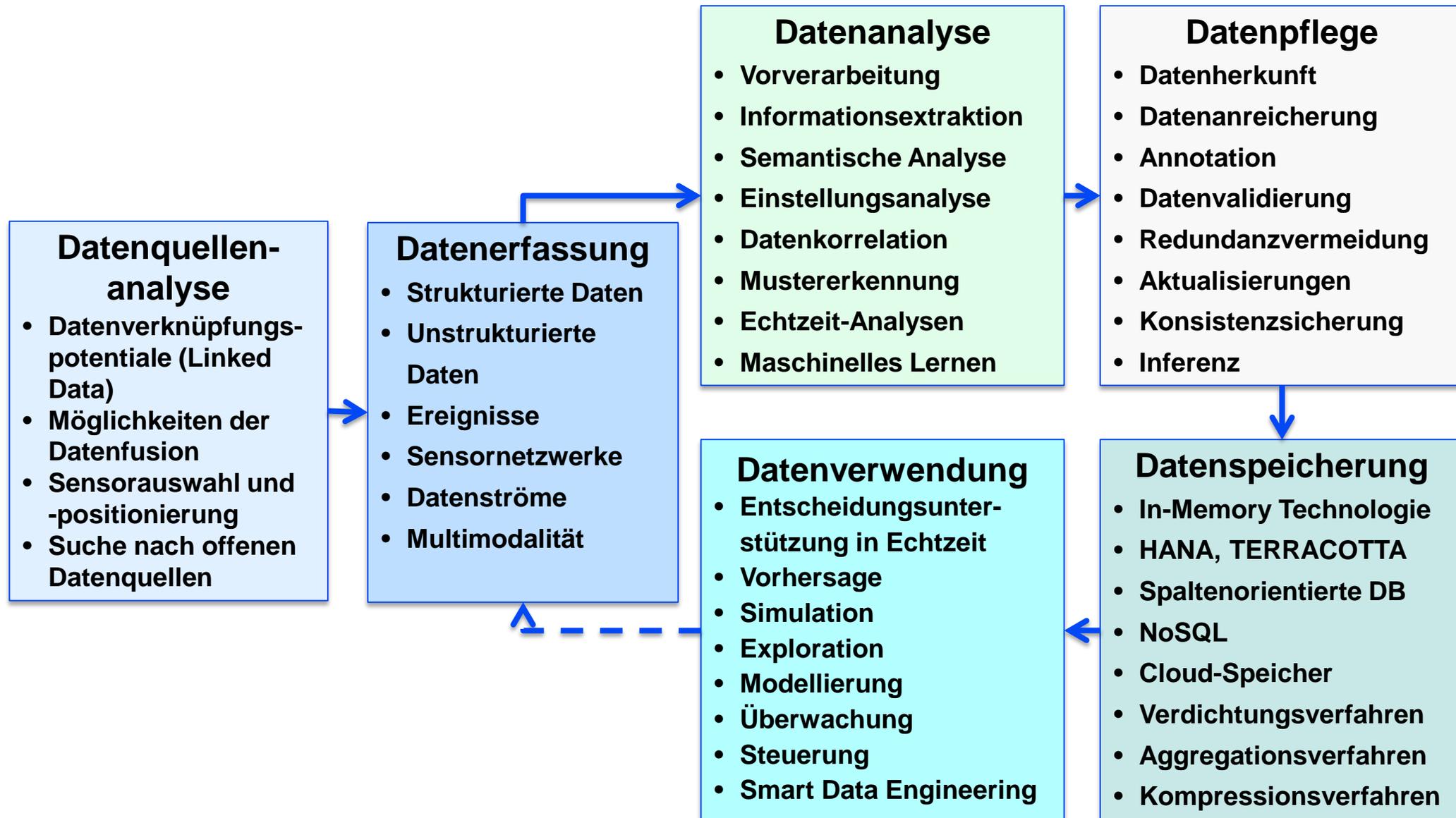
Unterscheidung zwischen klassischen Datenbanken und BIG DATA



Von Daten zu Metawissen



Der Verarbeitungszyklus für BIG DATA



Der Wissensresor von Google: Der größte Speicher von Wissenstripeln

→ Daten aus Webseiten

- Unstrukturierter Text
- Semistrukturierte Domänenontologien
- Strukturierte Tabelleninformation

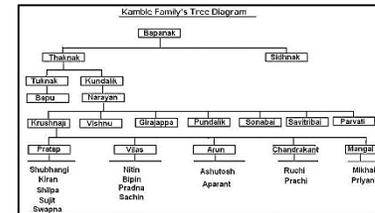
→ Daten aus sozialen Netzwerken



Beispiele Tripel:

((Saarbrücken, Hauptstadt_von, Saarland) 1)
((Koblenz, hält_Hitzerekord, Deutschland) .7)

<S, P, O> .99
<S, P, O> .96
<S, P, O> .76



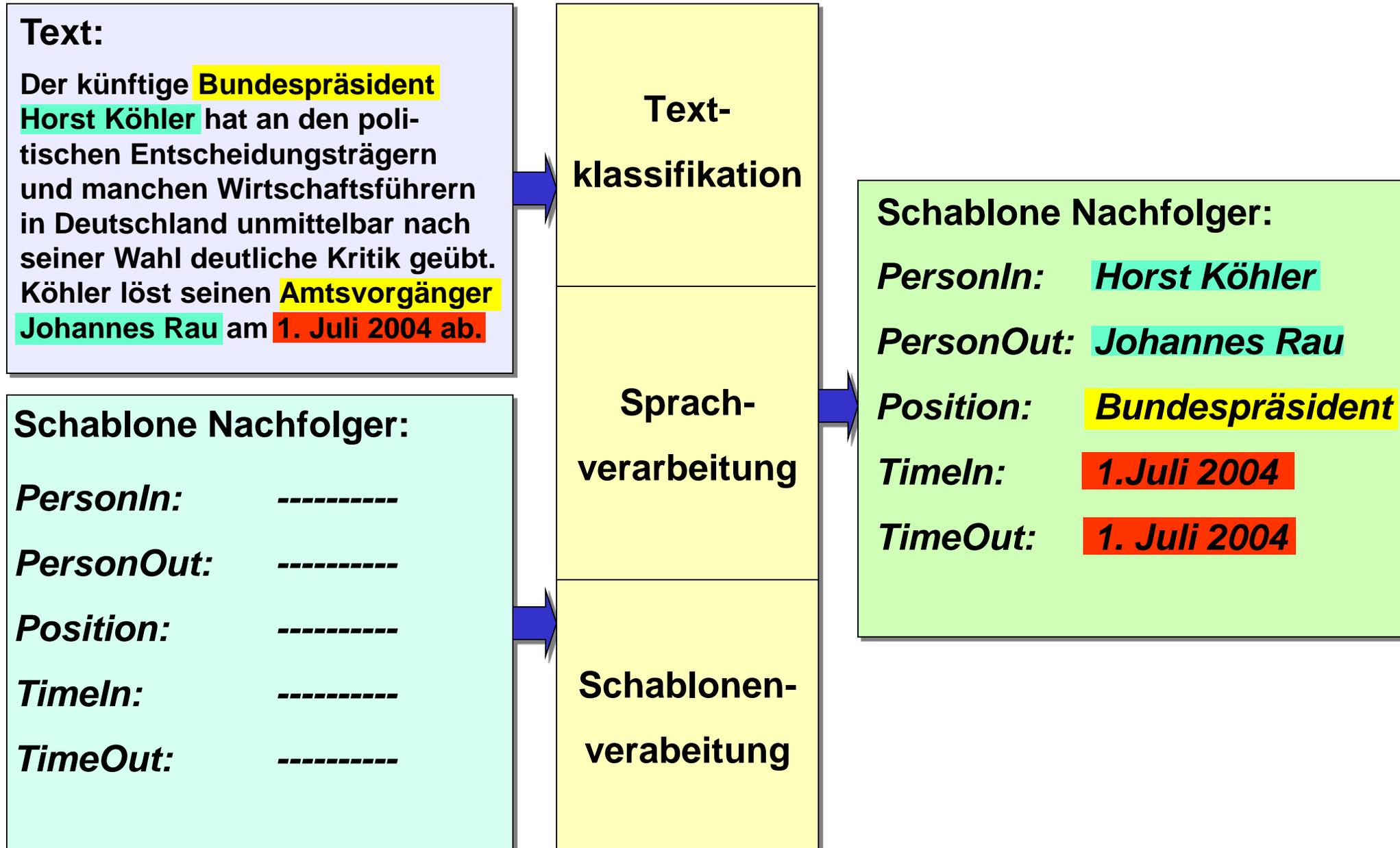
Year	Ranking	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	14th	15th	16th	17th	18th	19th	20th
2000	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2001	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2002	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2003	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2004	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2005	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2006	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2007	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2008	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2009	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2010	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2011	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2012	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2013	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2014	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2015	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2016	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2017	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2018	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2019	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2020	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2021	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2022	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21



Google
Knowledge
Vault

271 Mio. Wissenstripel
45 Mio. Entitäten
1100 Objektklassen
4469 Relationsklassen

Maschinelle Informationsextraktion aus Texten im Internet



Extraktion von Konsumentenkritik aus Blogs

Messages GATE document_0001C

Annotation Sets Annotations Co-reference Editor Text

Type	Set	Start	End	Features
negative	APPEARANCE	0	0	{}

73 Annotations (0 selected)

2005/06 VW Jetta
 Posted: Jun 7, 2004 3:41 PM, by user: Why?
 Wow, it looks like a Nissan Maxima. How boring.

Posted: Jun 7, 2004 3:54 PM, by user: AMG500SE
 No, more Corrola than Maxima... both boring.
 BTW, the poll is a bit biased isn't it? I mean since I dislike it I had to choose between Dislike and Previous Generation... "No" vote gets split 2-ways while "Yes" vote weighs twice of the "No's."

Posted: Jun 7, 2004 4:09 PM, by user: VDubMan2100
 It looks like they took a Honda and photoshopped a Golf V front end and VW badges onto it. Why buy this instead of the Golf?
 I don't dislike it, but it certainly doesn't inspire much in the way of excitement either, just like a Honda or Toyota.

Posted: Jun 7, 2004 4:13 PM, by user: shaymart
 Looks a lot like a Toyota Corolla - boring

Posted: Jun 7, 2004 4:31 PM, by user: xesquire
 Ah... Always wondered what happened to the Catera. The new Golf/Jetta 5 are nothing exciting (and as an owner of a Golf 4, I think I can say that). They look too much like other cars, and not like VW's and that was part of the reason why I was a VW customer in the first place. VW's

- ▼ APPEARANCE
 - negative
 - neutral
 - positive
- ▼ OPINION
 - neutral
 - strongly_negative
 - strongly_positive
 - weakly_negative
 - weakly_positive
- ▶ Original markups
- ▶ REASON
- ▶ STYLE
- ▼ SUBJECT
 - car_component
 - car_make
 - car_type
- ▶ SUGGESTION

Corporate Real Estate (CRE) Management basierend auf Künstlicher Intelligenz

Herausforderungen

unstrukturierte /
unvollständige CRE
Daten

Sprachbarrieren
für die globale
Übersicht

IFRS- / FASB-
Rechnungslegung



Lösungen

strukturiert / vollständige
Daten
bereit zur Berichterstattung

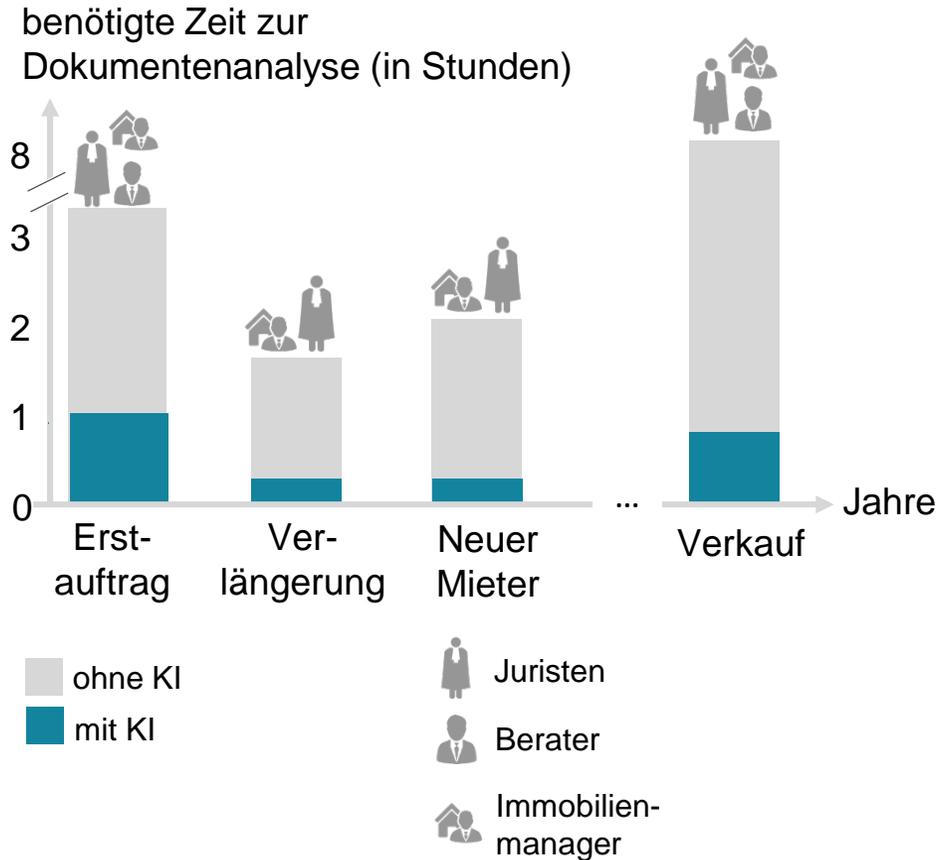
Dokumente in 20+ Sprachen
können in der gewählten
Sprache angesehen werden

Extraktion und Organisation
relevanter Infos für IFRS-/FASB-
Berichte

International Financial Reporting Standards (IFRS) /
Financial Accounting Standards Board (FASB)

Kosten- und Zeitersparnis beim CRE durch KI

Zeitgewinn entlang des Lebenszyklus für Immobilienverträge



Zeitgewinn bei den verschiedenen Prozessen



Überbrücken der Lücke zwischen unstrukturierten Dokumenten und strukturierten Systemen

Dokumentenmanagement

Ein Dokumenten-Management-System ist ein **zentrales digitales Archiv für Dokumente**.

Fokus auf Dateien

Unstrukturierte Dokumente

Weniger Inhaltsverwaltung

Versionierung

Maschinelles Lernen

Strukturierte Inhalte und Dokumente

Verknüpfung von Inhalten und Quellen

ERP/SAP-Systeme

Ein Enterprise-Resource-Planning System wird zur **Unterstützung der Ressourcenplanung des gesamten Unternehmens** eingesetzt.

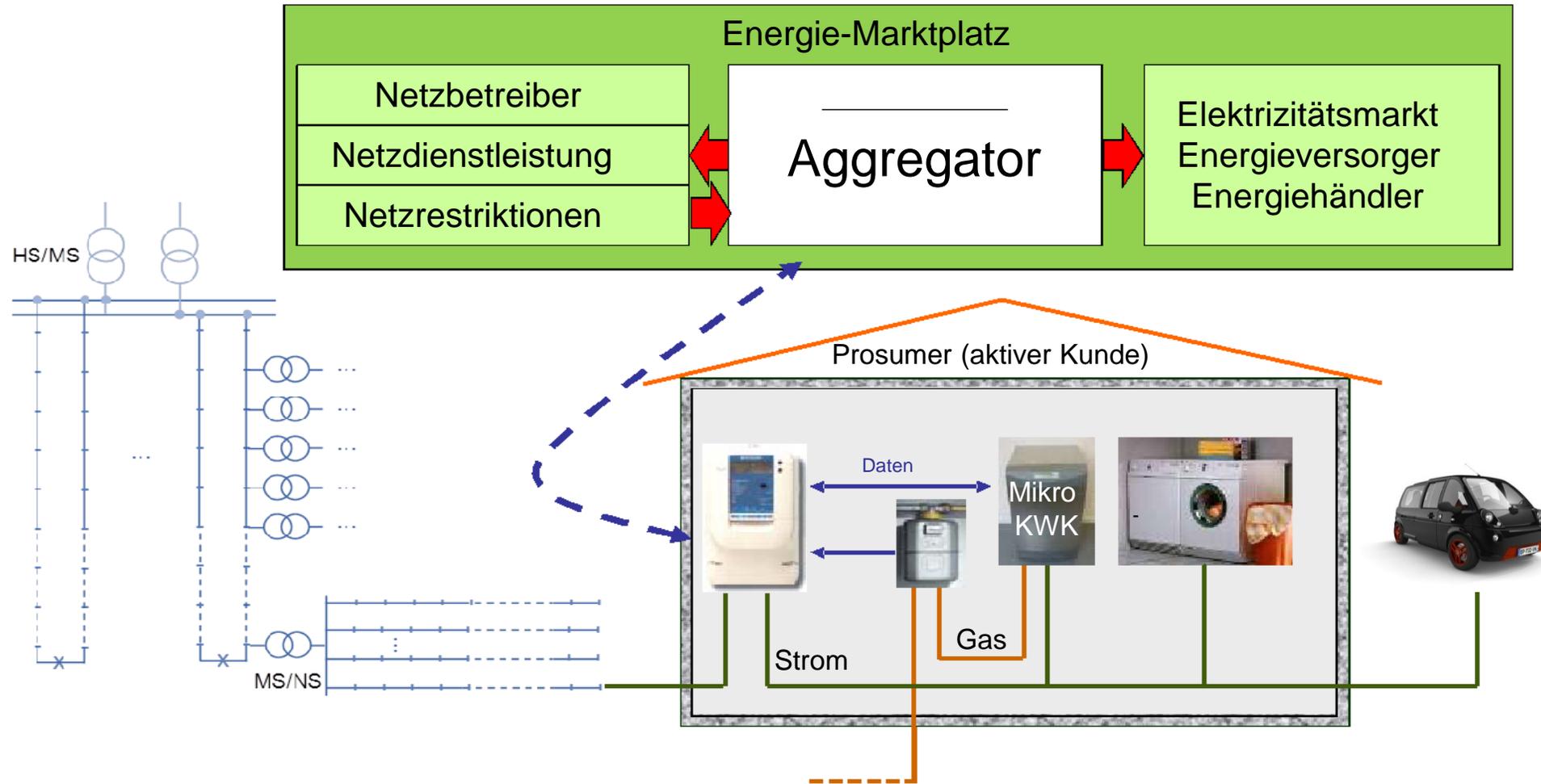
Fokus auf Inhalte

Strukturierte Inhalte

Quellen fehlen

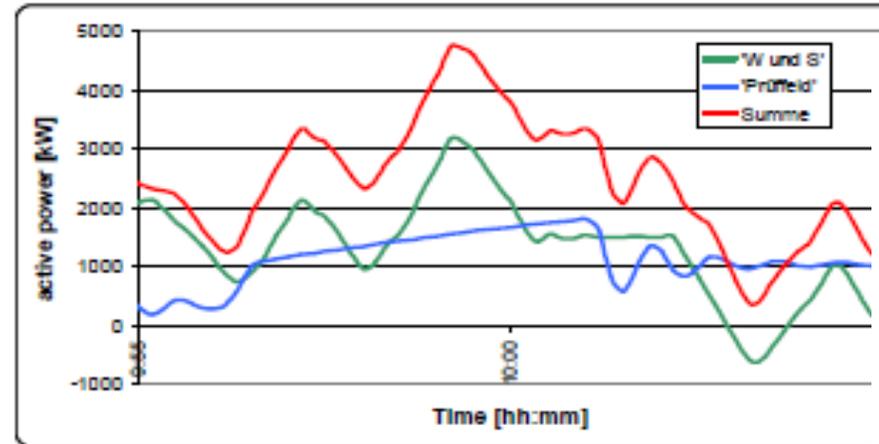
Accounting

BIG DATA aus Smart Grids: Aktivitätsprofile aus Verbrauchsdaten

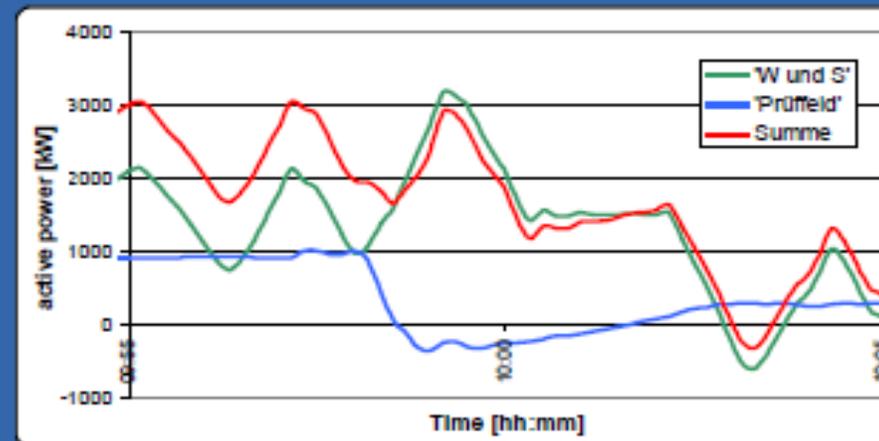


BIG DATA in Industrie 4.0: Lastoptimierte Steuerung von Produktionsprozessen

Ausgangssituation
Produktions-
prozesse ohne
Berücksichtigung
der Energie-
verbräuche in der
Planung



Ziel
Energetisch
koordinierte
Ausführung der
Prozesse



Lastmanagement

- Vermeiden von Lastabwürfen
- Kostensenkung durch reduzierte Leistungspreisanteile (z.B. max. 15-min-Mittelwert)

Energierückgewinnung und Netzzurückspeisung möglich

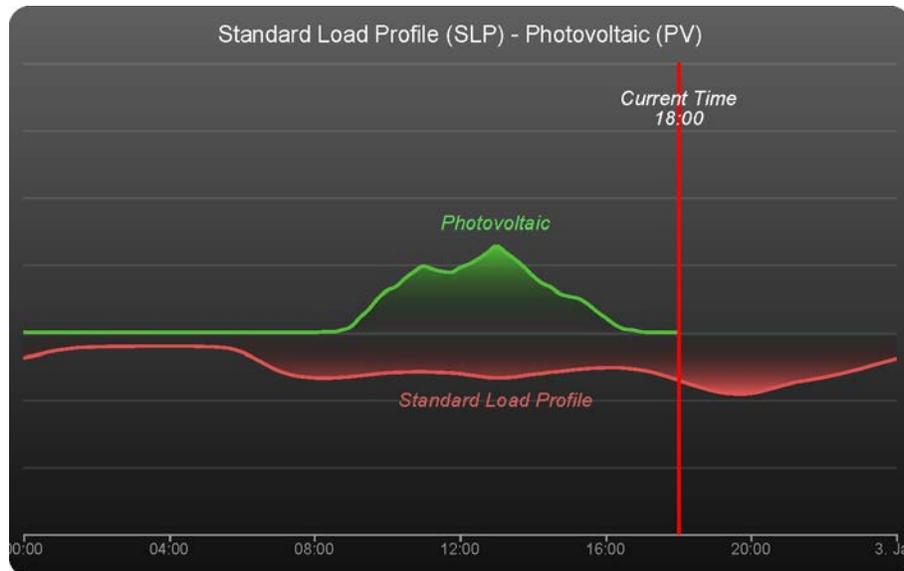
Smart Home: Das Internet der Dinge im Haus



Durch BIG DATA vom Standardlastprofil zum Individualprofil

Situation Heute

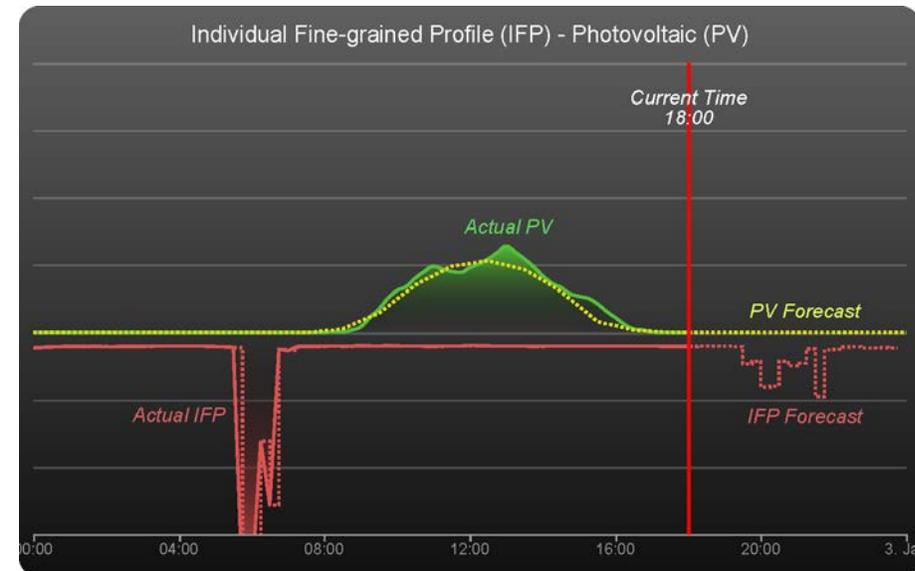
- Standardlastprofil
- Keine Vorhersage für Verbrauch / Erzeugung



→ Zum Handeln ungeeignet

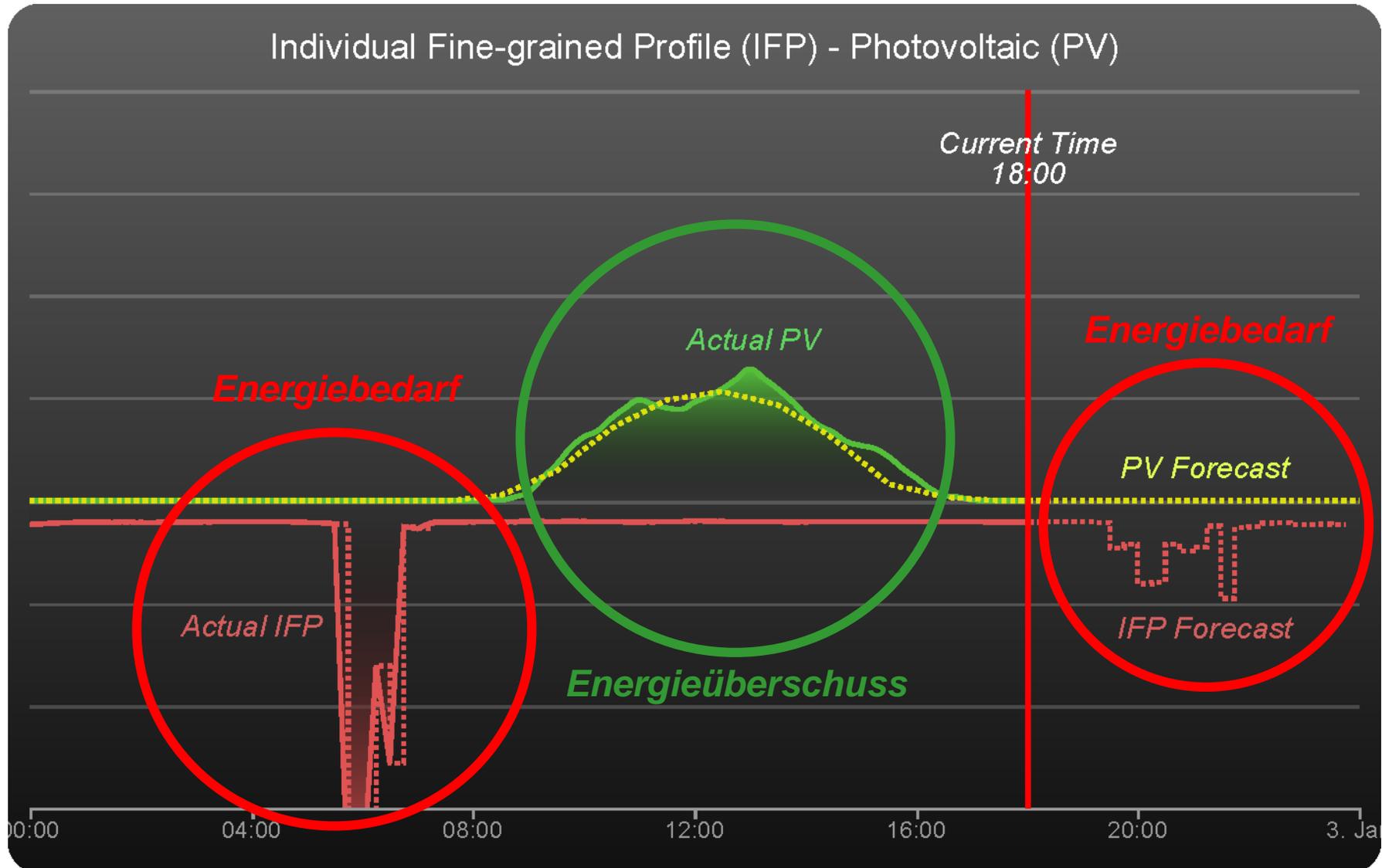
Situation mit PEC

- Individuelles Feingranulares Profil
- Tagesaktuelle Vorhersage
- Anpassung der Vorhersage über den Tag

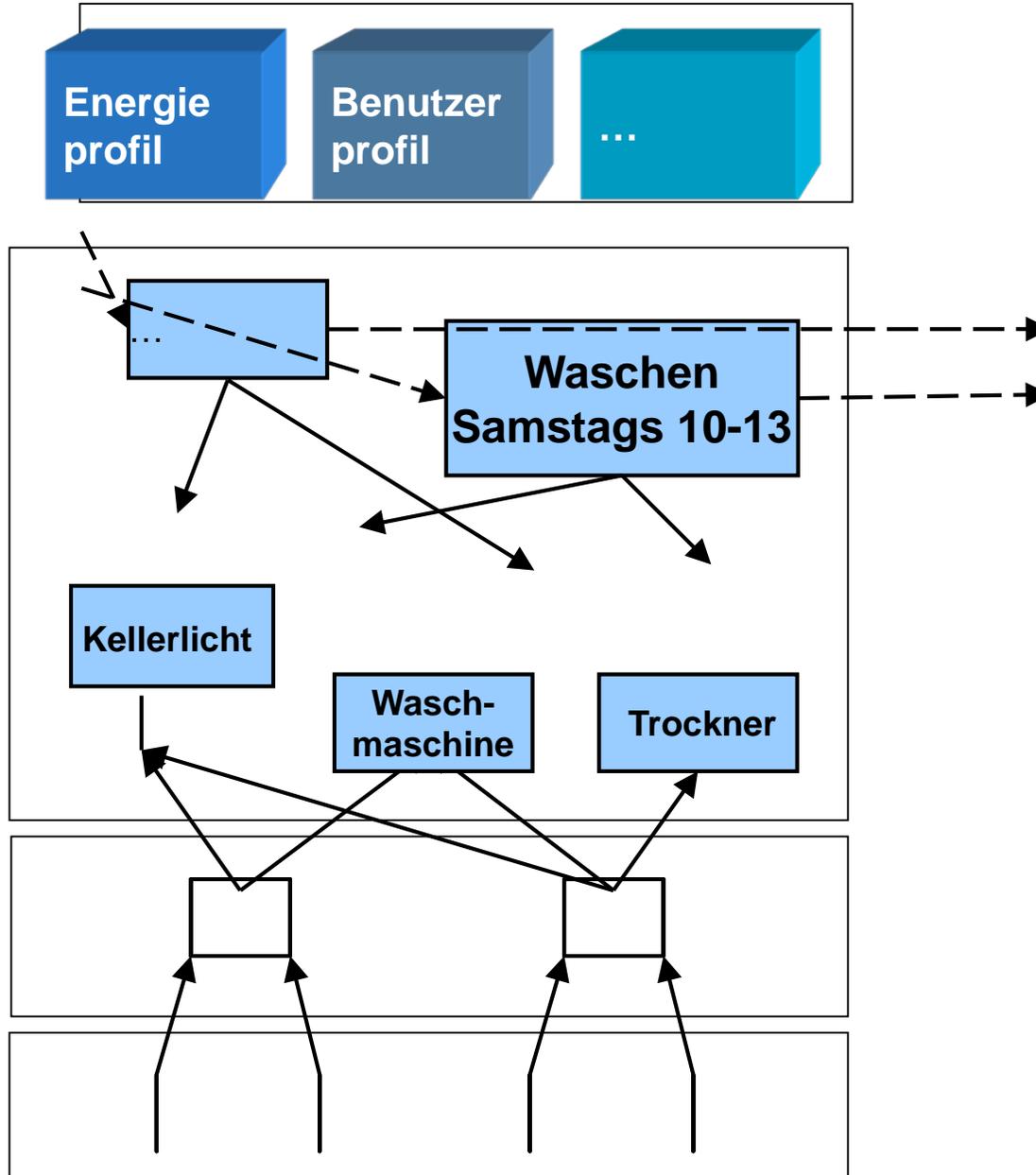


→ Zum Handeln geeignet

Individuelles Feingranulares Profil



Automatisierte Benutzer- und Energieprofile



Digitale Modellierung

Maschinelles Lernen
aus unsicheren
Beobachtungen

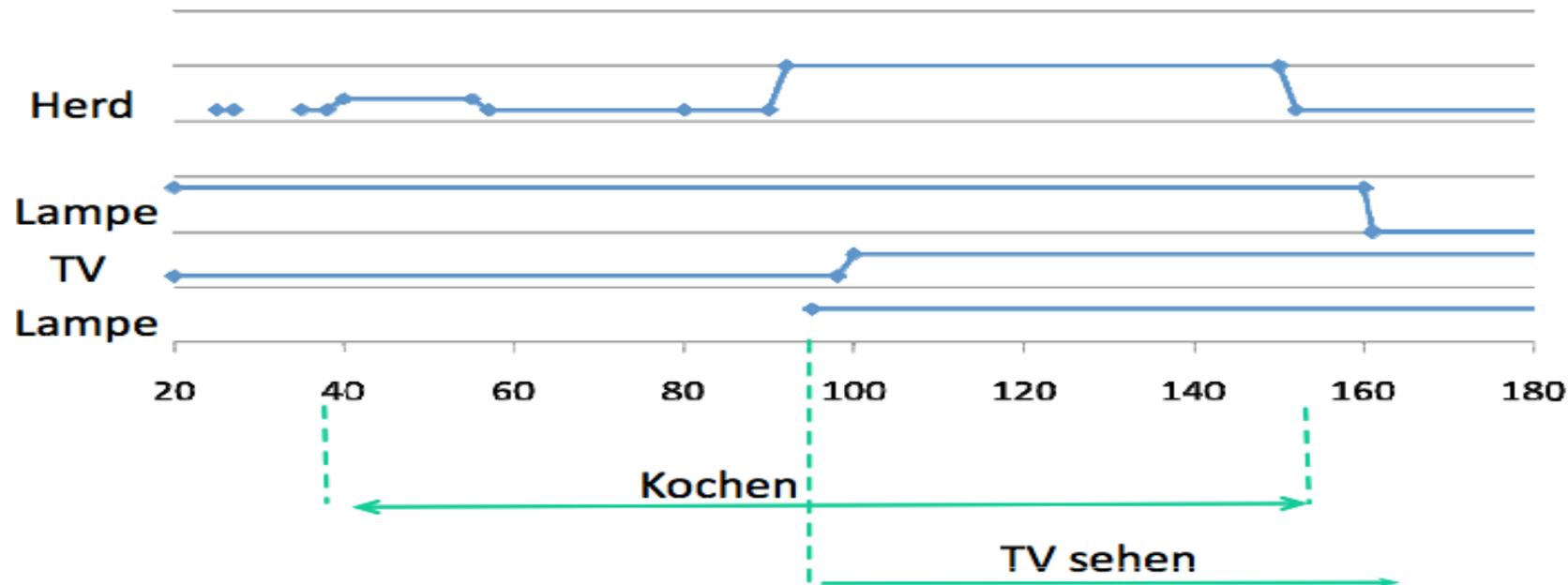
Dynamisches
Bayessches Netz
für algorithmische Verarbeitung
Von Unsicherheit

Vorverarbeitung
Statistische Klassifikation

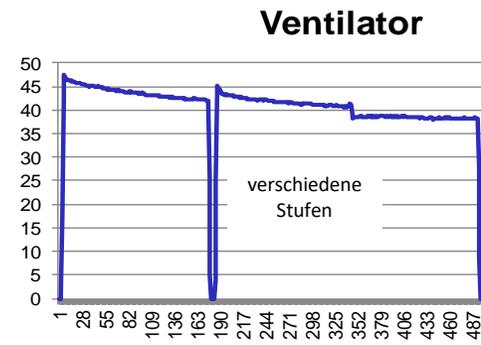
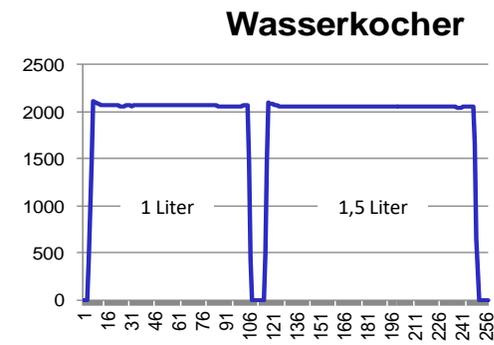
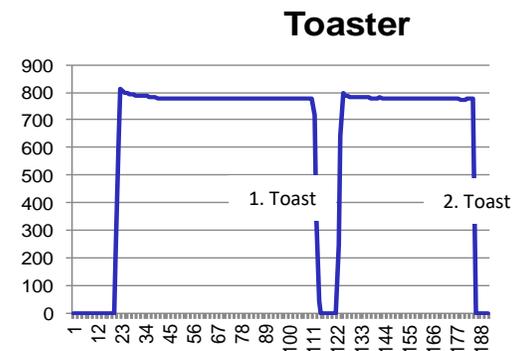
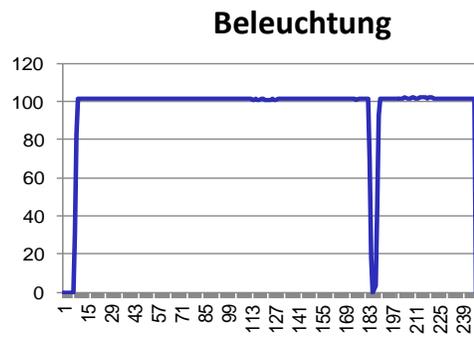
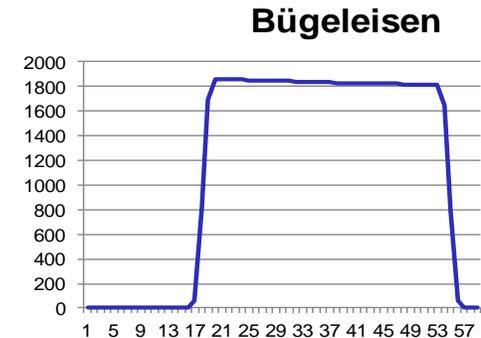
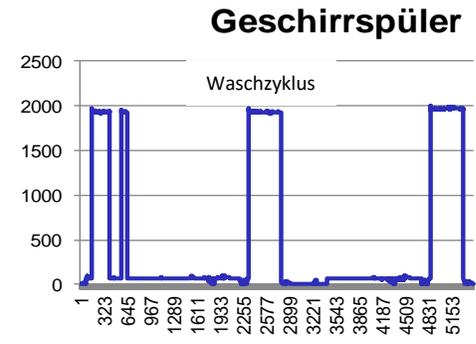
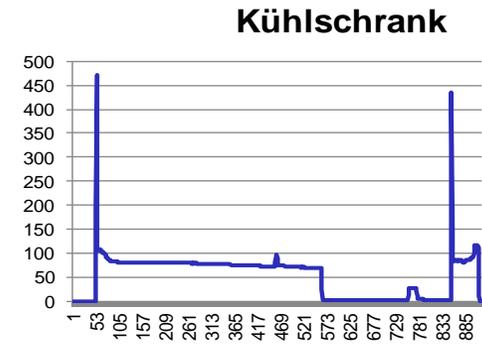
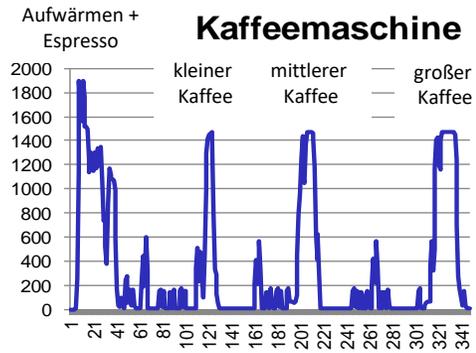
Sensordaten
über Digitalstrom

Anwendung auf Microgrids

- Modellierung eines Haushalts
 - Erfassung aller Stromverbraucher (evtl. eigene Internetadresse)
 - Erfassung aller Energieproduzenten



Erkennung verschiedener Verbraucher im Haushalt anhand von Lastkurven (Wirkleistung): x-Achse Watt, y-Achse Zeitverlauf

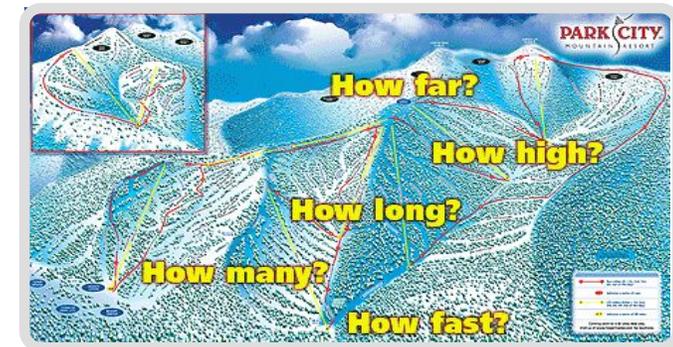


Anwendungsbeispiel: Unterhaltung und Sport

- Bewegungsmuster ermöglichen die Analyse der Performanz bzw. Korrektheit von Abläufen
 - Sportler können Fehler erkennen und ihr Verhalten verbessern
 - Laufcomputer zeichnen biometrische Daten auf, die mit Trainingsdaten abgeglichen werden können
 - SlopeTracker™: Aufzeichnung von Abfahrtsläufen
- Derartige Informationen besitzen nicht nur einen persönlichen Wert → Kommunikation beispielsweise über Blogging und Social Networking
 - Austausch aktueller Routenbedingungen beim Skiwandern
 - *Avesani et al. (2004): A trust-enhanced recommender system application: Moleskiing*



Polar Laufuhr RS 100



© <http://www.slopetracker.com/>

Trend: Tracking von Gegenständen

- Beispiel Logistik
 - DHL Sendungsverfolgung: Wo ist das Paket?
- Beispiel: Authentizität von Produkten
 - EPCGlobal ePedigree: Ein „Stammbaum“ soll die Fälschung von Medikamenten erschweren
- Beispiel: Kundeninformation
 - Iglo: Woher kommt der Spinat?



Quelle: <http://iglo.de/dasigloversprechen/gemueseundmehr/woherkommtihrspinat/index.asp>

*“America is becoming a nation of track-a-holics. We want to go online and track the whereabouts of everything we order — or do. It's sometimes because **we need to know**, but often it's simply because **we want to know**.”* Bruce Horovitz, USA Today

Produktgedächtnisse liefern als digitale Zwillinge für jede Instanz eines Produktes Rückverfolgungsinformation

Der Internet-Code kennt jedes Erntefeld



Mit dem Lasermarker kann individuelle Produktinformation abgerufen werden

Ihr Spinat wurde von Bauer Bernd Müller am am 4. Juli auf seinem Feld Deichblick geerntet.



Domino's Pizza Tracker : Zeigt sogar den Namen des Mitarbeiters an, der die Pizza gebacken hat

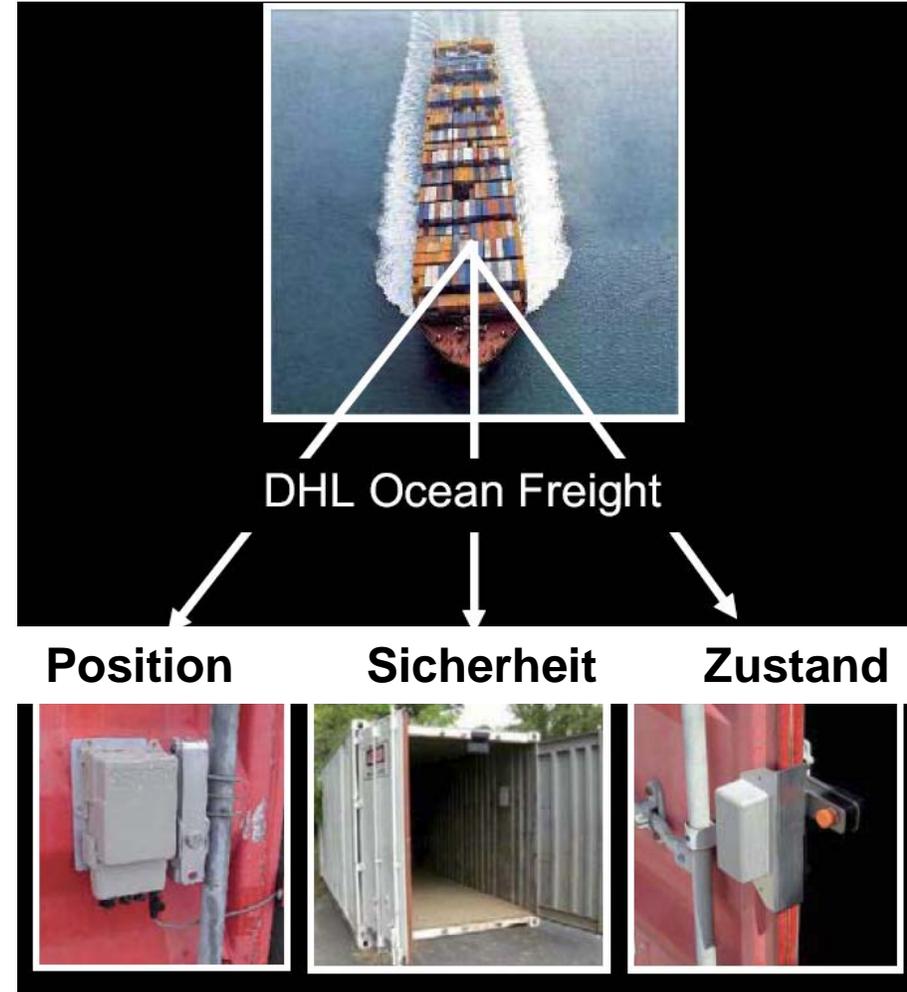
Von 75% aller Kunden genutzt

The screenshot shows the Domino's Pizza Tracker web interface. At the top left is the Domino's logo. The navigation bar includes 'ORDER', 'MENU', 'COUPONS', 'LOCATIONS', and 'SEE LOCAL OFFERINGS'. The current time is 'SUN 9:05 AM' and the language is 'ESPAÑOL'. The main heading is 'PIZZA TRACKER'. A text block explains that the tracker shows the status of the order from preparation to delivery. Below this is a progress bar with five stages: 1. ORDER PLACED, 2. PREP, 3. BAKE, 4. BOX, and 5. DELIVERY. The progress bar is currently empty. Below the progress bar is a form to 'ENTER YOUR PHONE NUMBER*' with fields for area code, number, and extension. There is a checkbox for 'I AGREE TO ONLY TRACK MY OWN ORDERS AND TO THE TERMS & CONDITIONS' and a 'TRACK YOUR ORDER' button. A 'REMEMBER ME' checkbox is also present. A vertical red banner on the right side of the interface reads 'DOMINO'S // PIZZA TRACKER'.

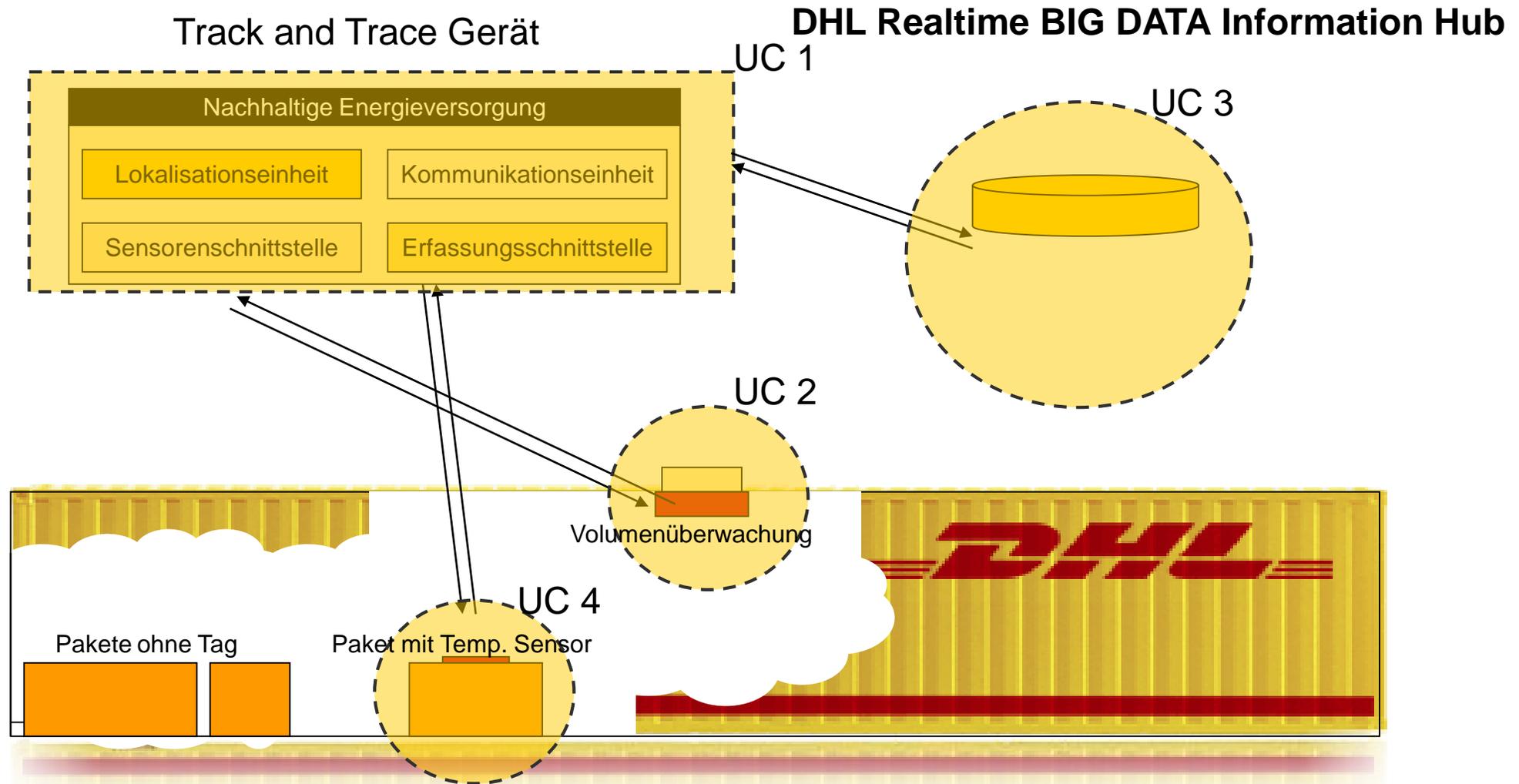
BIG DATA in der Paketverfolgung



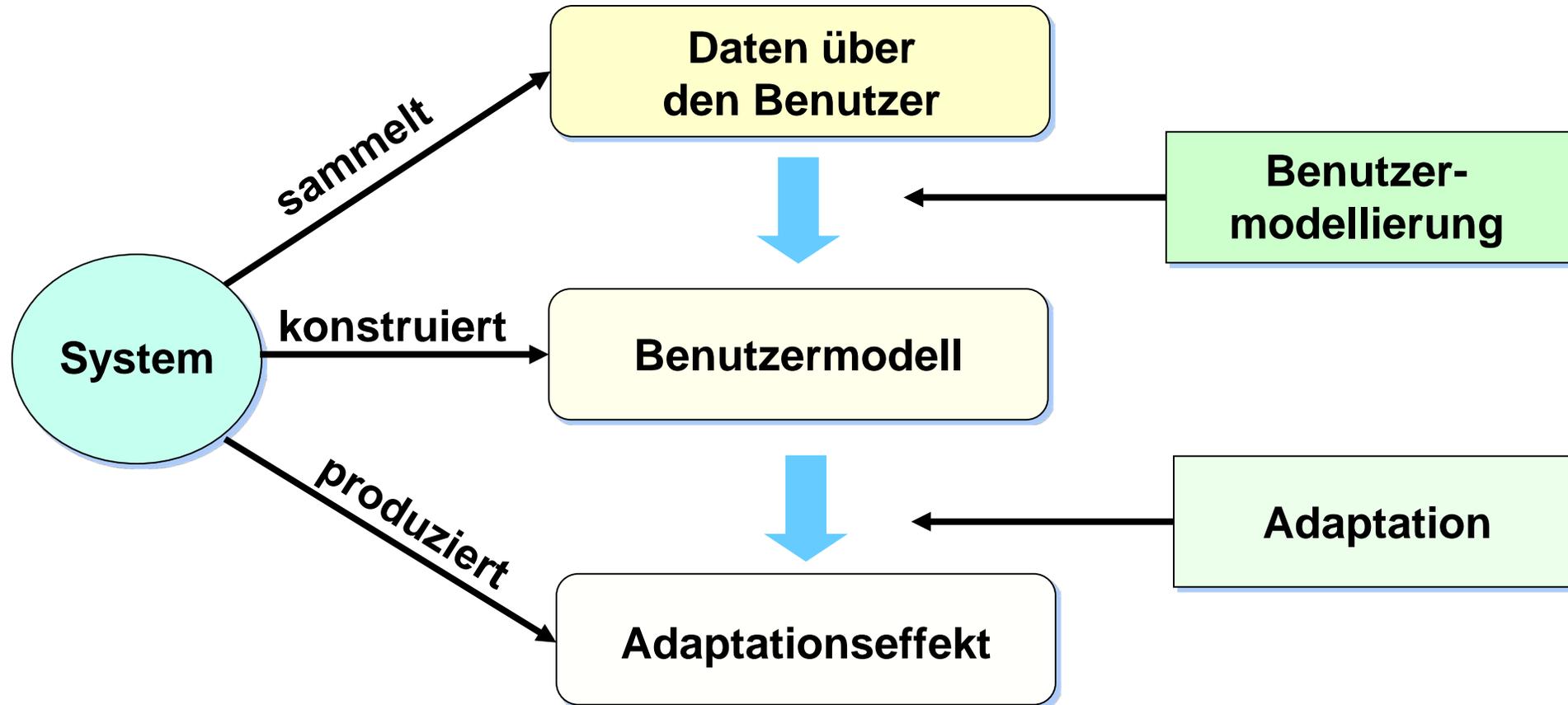
Temperatursensor
In DHL-Verpackung



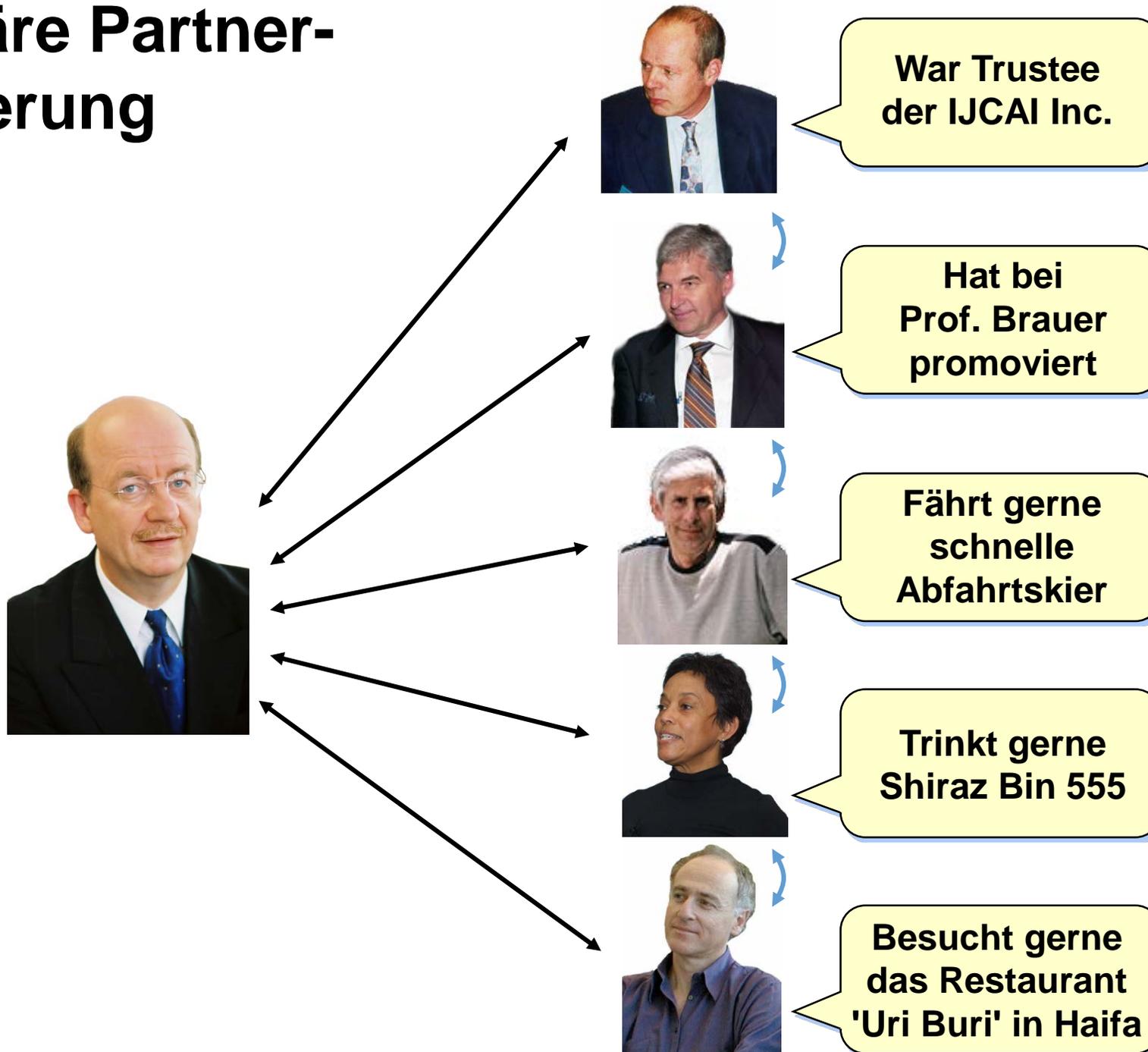
BIG DATA in der Logistik: Semantische Produktgedächtnisse



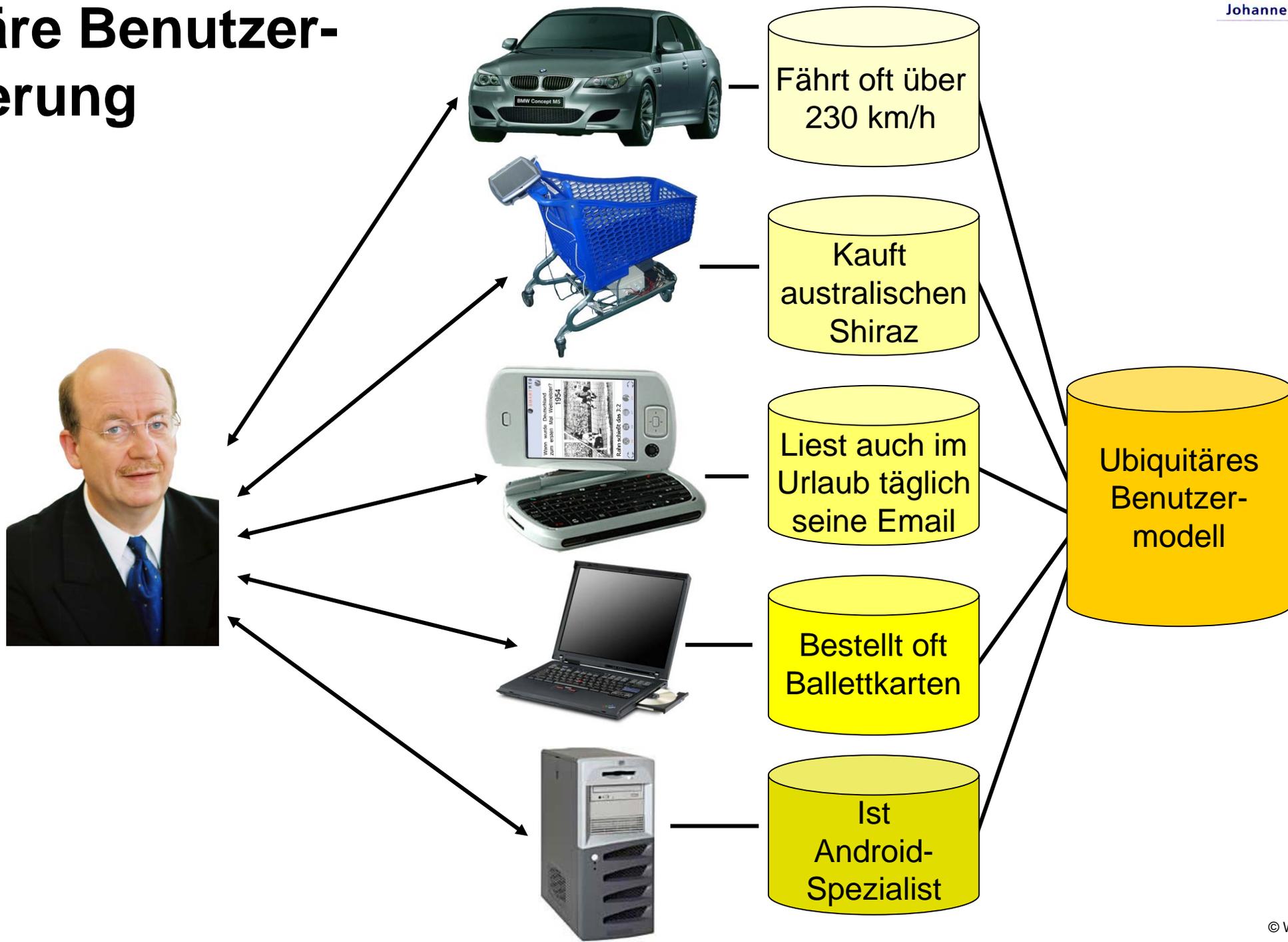
Benutzermodellierung und -adaption



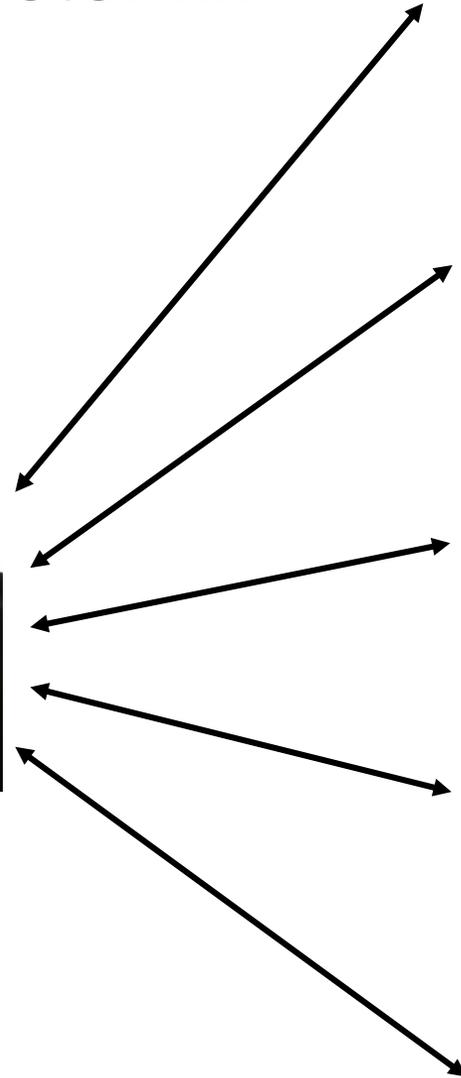
Ubiquitäre Partnermodellierung



Ubiquitäre Benutzermodellierung



Ubiquitäre Partnermodellierung durch Service-Roboter im Jahr 2050 ???



Schon dreimal
mußte ich dieses Jahr die
Heizung reparieren.



Ich muß bei Wanderungen mit
ihm immer die komplette
Picknick-Ausrüstung
schleppen.



Montags wird immer das
Wohnzimmer gesaugt.



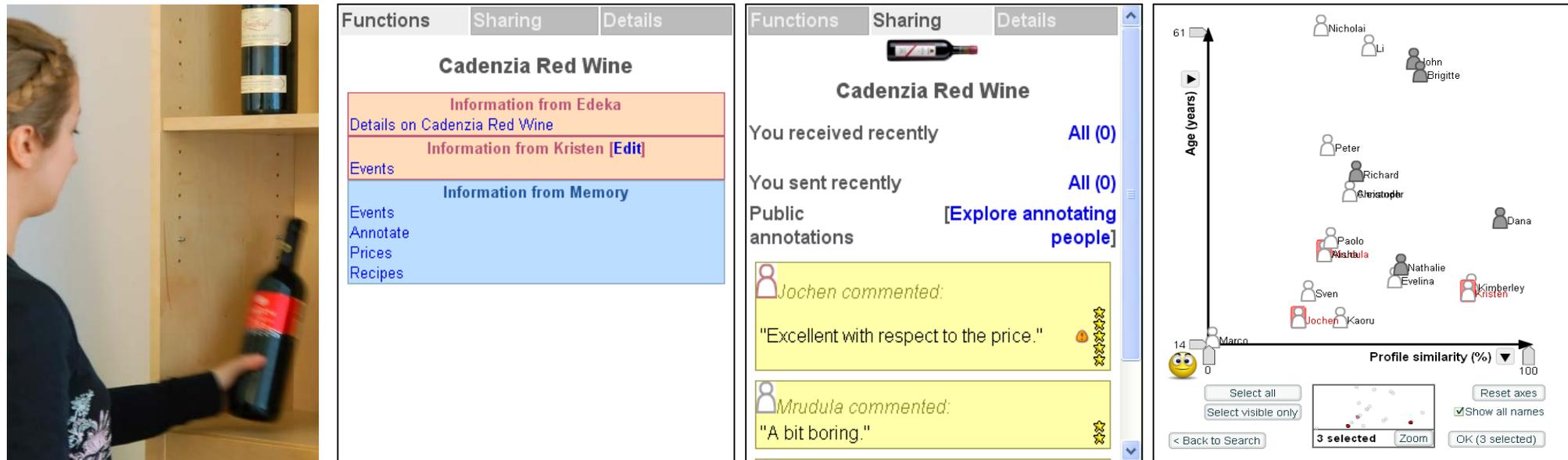
Ich wollte ihm letzte Woche
mal zeigen wie man richtig
Hip-Hop tanzt.



Beim Kindergeburtstag
muß ich immer beim
Topfschlagen mitmachen.

Interaktionsbeispiel

- Digitales Produktgedächtnis: Meinung des Herstellers
- Digitales Benutzertagebuch: Eigene Erfahrungen
- Community: Meinungen anderer Personen



The composite image illustrates the user interaction process. On the left, a woman selects a bottle of Cadenzia Red Wine from a shelf. The middle-left panel shows the product's digital memory page, categorized by source: 'Information from Edeka', 'Information from Kristen', and 'Information from Memory'. The middle-right panel displays user reviews and annotations for the same product, including comments from Jochen and Mrudula. The rightmost panel is a scatter plot showing the relationship between 'Age (years)' and 'Profile similarity (%)' for various users, with names like Jochen, Mrudula, and others plotted.

Der Kunde wählt ein Produkt

Umgebung befragen, Freundeskreis befragen, Tagebuch durchsuchen

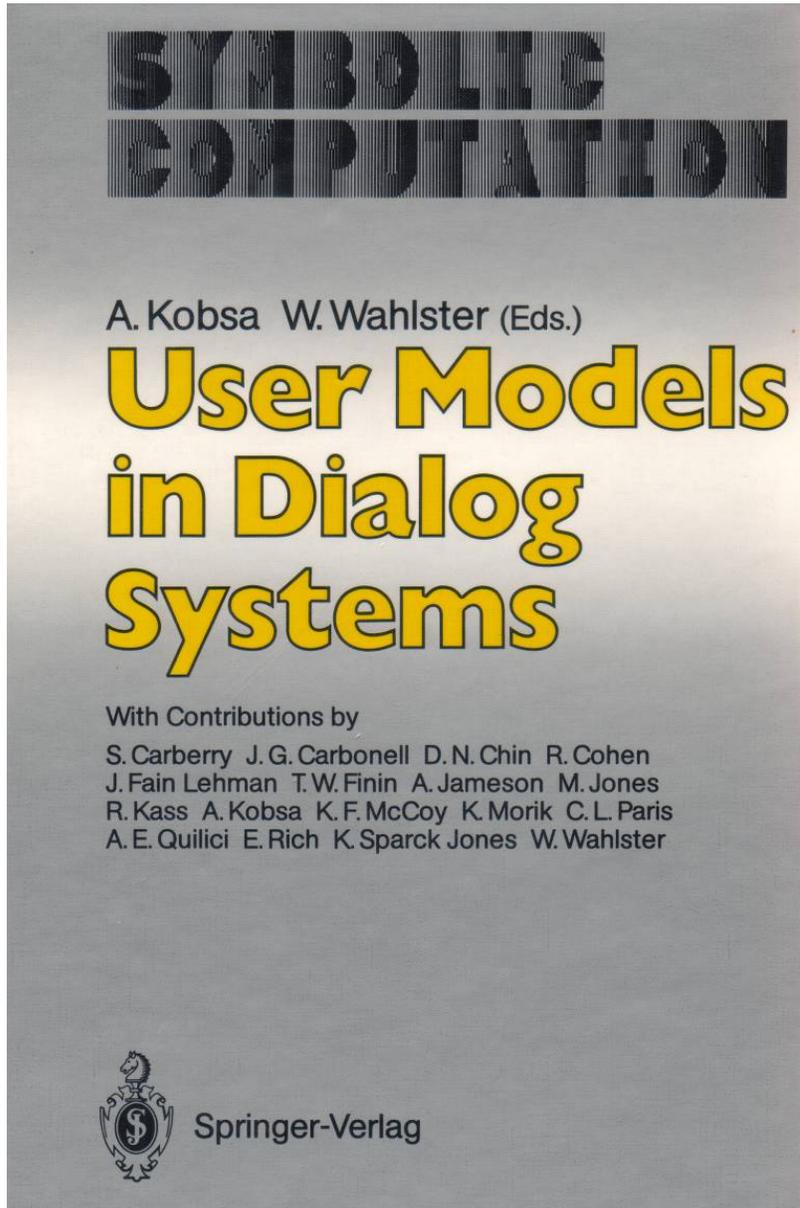
Bewertungen, Kommentare, Hinweise

Einordnung der Kommentatoren in der Community



Demo

Vor 20 Jahren 1986: Erste internationale Konferenz zur Benutzermodellierung in Deutschland



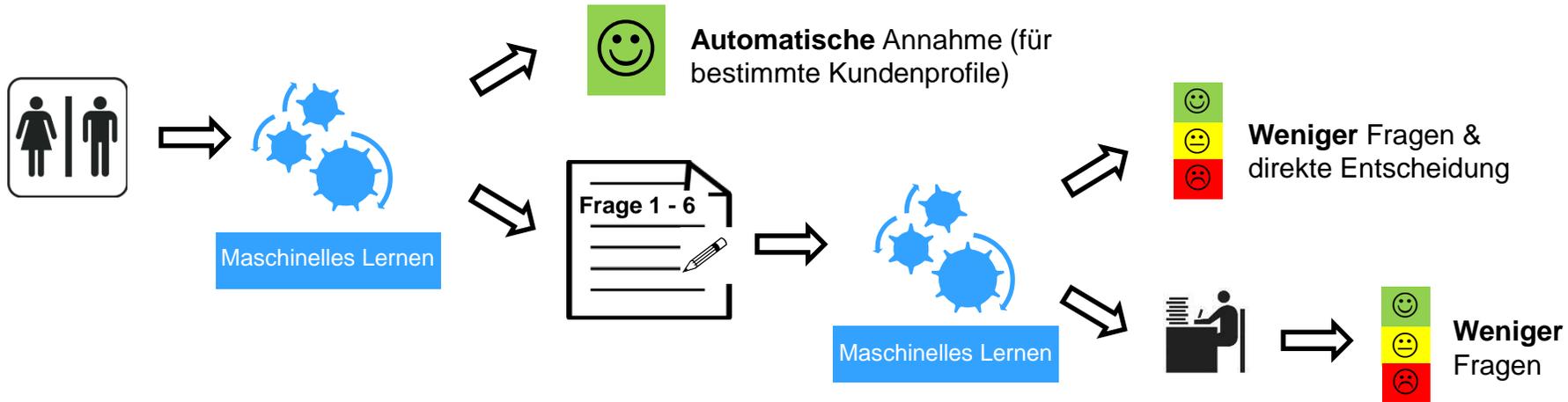
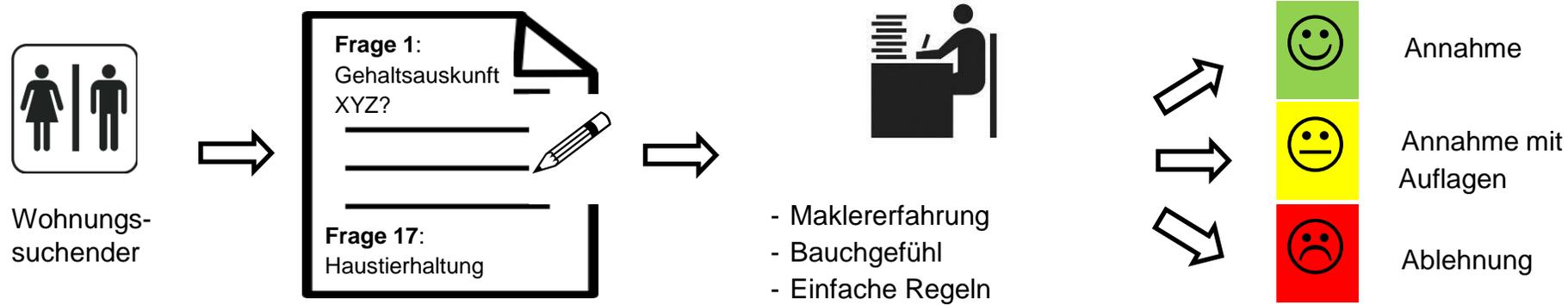
Im Kloster **Maria Laach**
organisiert von W. Wahlster
und A. Kobsa (SFB 314)

Heute Zeitschrift:
**User Modeling and
User-Adapted Interaction**

bewertet als eines der 5%
besten Informatik-Journale

und Konferenzserie:
User Modeling

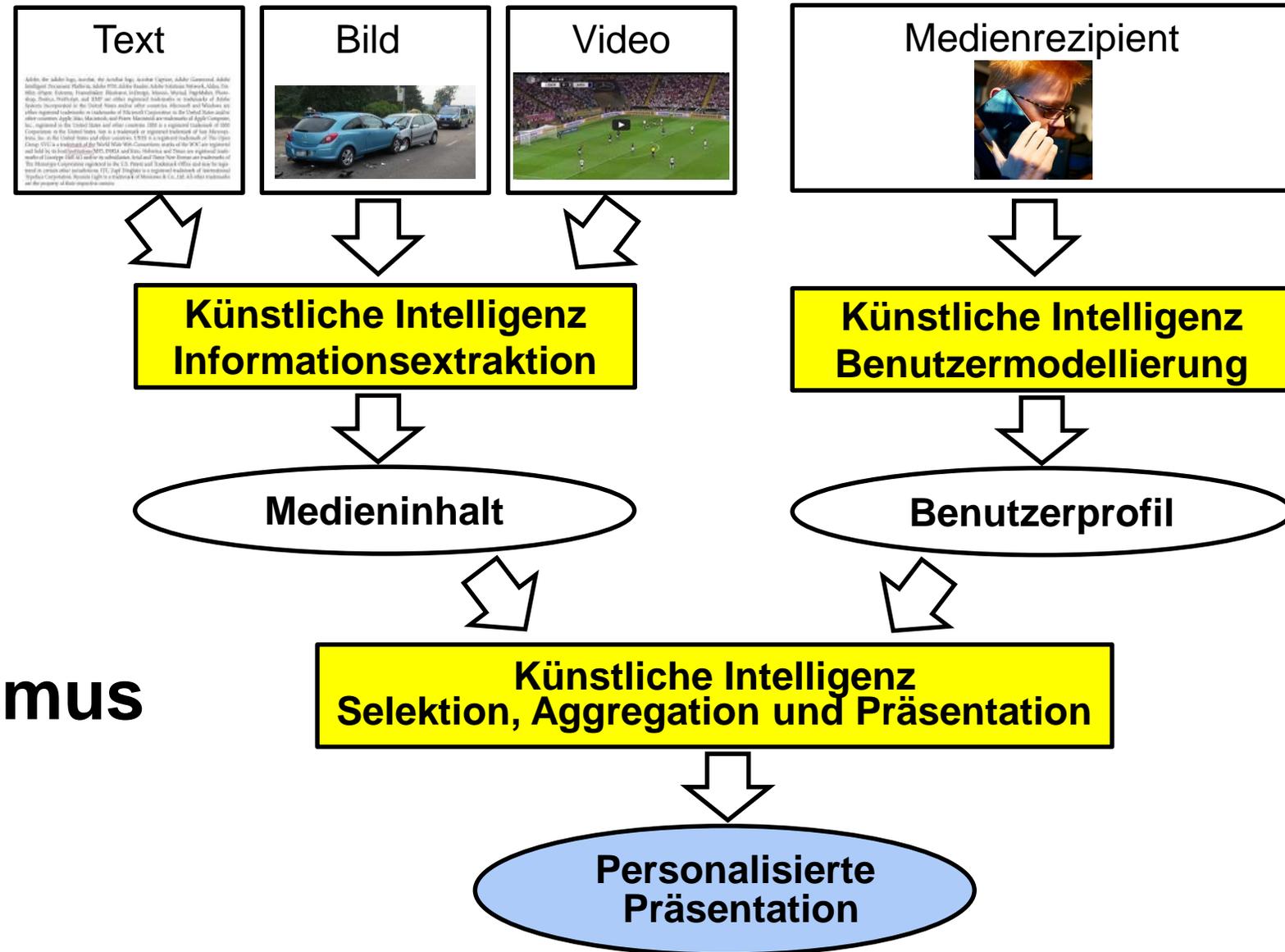
Intelligente Immobilienvertragsprüfung basierend auf KI



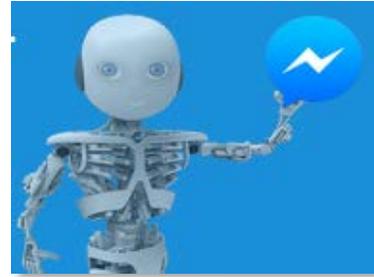
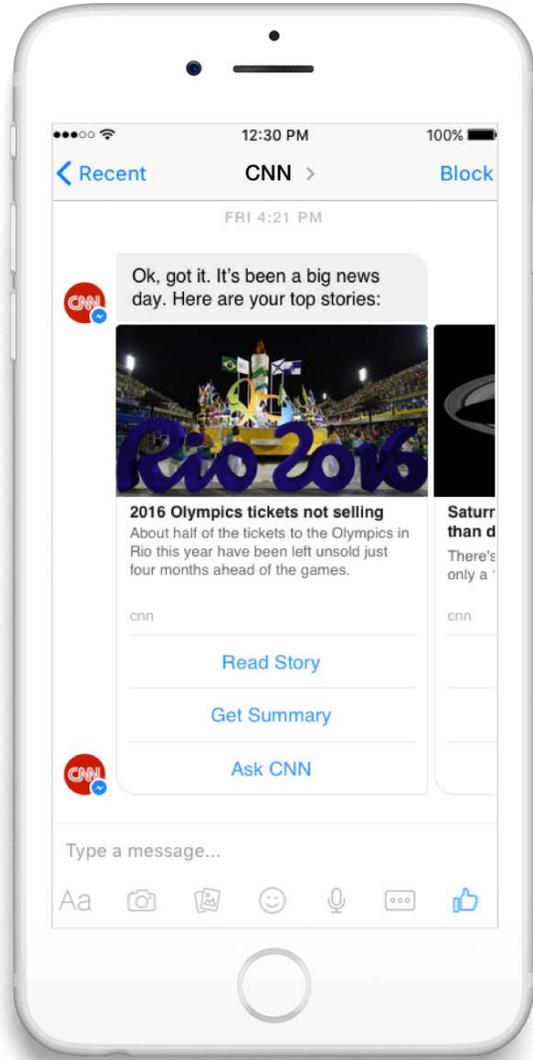
Künstliche Intelligenz ermöglicht die automatische Personalisierung von mobilen Medienpräsentationen

Medienbot

Robot Journalismus



Künstliche Intelligenz als Wegbereiter für Medien-Bots und Robo-Journalismus?



Facebooks Messenger Chat Bots
Beispielanwendung von CNN

Noch in der frühen Phase, aber werden durch Deep Learning immer besser.

Vorteil: Parallele Verfolgung aller Nachrichtenströme in Echtzeit
Erkennung von wahrscheinlichen Falschmeldungen
Personalisierte Auswahl, Aggregation und Präsentation

Nachteil: Quellenlage, Nachrichtenauswahl und Informationsgenese oft intransparent. Kritische Distanz und Unabhängigkeit nicht garantiert. Keine Implementation des Pressekodex

Künstliche Intelligenz für das semantische TV: Inhaltserkennung in laufenden TV-Sendungen und parallele Zusatzinformationsangebote



Automatische Extraktion von Gesichtern, Namen, Orten, Produkten und semantische Suche im Web mit Sprach-, Bild- und Zeichenerkennung

aus Dissertation: Dr. M. Deru, Swoozy:
Intelligente und benutzerzentrierte Interaktionen
für das semantische Fernsehen, DFKI 2016 © w. Wahlster

Semantisches Fernsehen mit multimodaler Interaktion



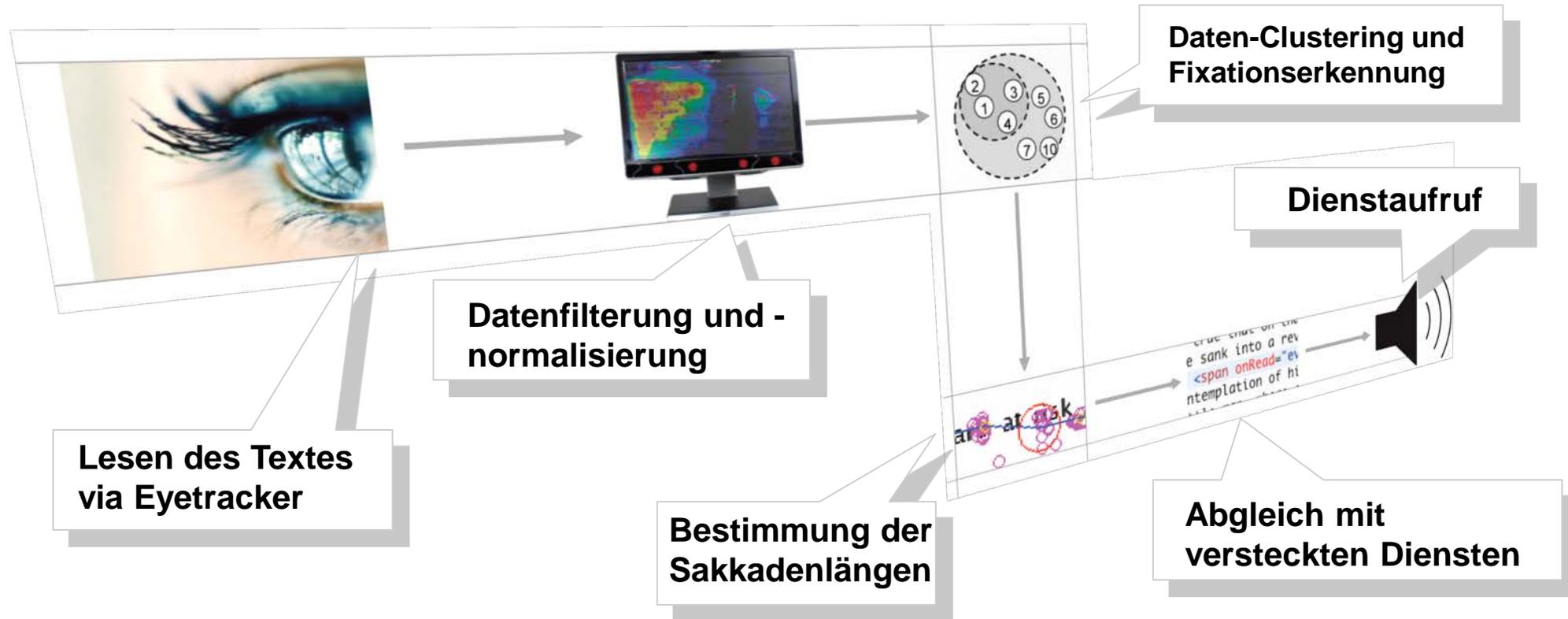
Faktensuche

TEXT 2.0: Stellen Sie sich vor, ein Text "weiß", dass er gerade von Ihnen lesen wird



- **Text 2.0** ist eine neue Interaktionsform mit Texten, die Internet-Dienste durch Blickbewegungsanalysen auslöst
- Es basiert auf der Idee, dass der Computer weiß, welche Zeile, welcher Satz oder welches Wort gerade gelesen wird.
- Während des Lesens werden im Text 2.0 versteckte Dienste ausgelöst.

Text 2.0 stellt ein effizientes Verfahren zur Verfügung, um Aufmerksamkeitsdaten beim Lesen nutzbar zu machen.



Text 2.0 des DFKI: Blickbewegungsanalyse per Webkamera während des Lesens

Chapter 1

Once when I was six years old I saw a magnificent picture in a book, called True Stories from Nature, about the primeval forest. It was a picture of a boa constrictor in the act of swallowing an animal. Here is a copy of the drawing.

In the book it said: "Boa constrictors swallow their prey whole, without chewing it. After that they are not able to move, and they sleep through the six months that they need for digestion."

I pondered deeply, then, over the adventures of the jungle. And after some work with a colored pencil I succeeded in making my first drawing. My Drawing Number One. It looked something like this.

I showed my masterpiece to the grown-ups, and asked them whether the drawing frightened them. But they answered: "Frighten? Why should any one be frightened by a hat?"



Von Google-Gründer Larry Page persönlich mit Begeisterung getestet

Neue Medien werden stets auch zur Radikalisierung bis hin zum „Mitmach-Terrorismus“ mißbraucht



Flugblatt aus der Reformationszeit mit dem Titel „Ego Sum Papa.“ gegen Papst Alexander VI. etwa um 1500.

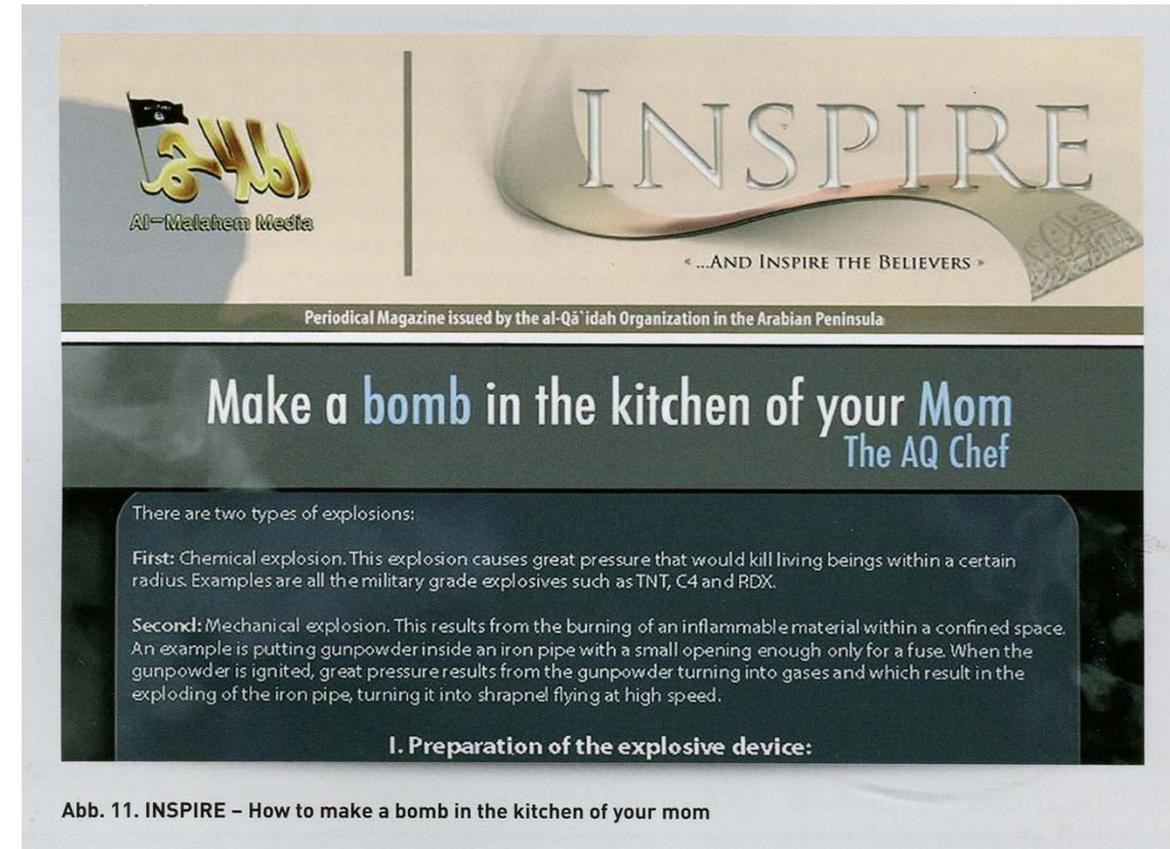


Abb. 11. INSPIRE – How to make a bomb in the kitchen of your mom

Dschihadistisches Online Magazine „Inspire“



(Image produced by ISIS supporters)

Twitter-Dienst des IS als Propagandamittel im Sekundentakt

Rollenanalyse in sozialen Netzwerken



- Wer hat den meisten Einfluß?
- Wer sind die Initiatoren?
- Wer verhält sich unsozial?
- Wer ist am aktivsten?
- Wie stabil sind diese Rollen?

Der Datenhunger staatlicher Behörden



NSA Datenzentrum in Utah:
Mehrere Terabyte für jeden
Menschen auf der Welt,, 65
Megawatt Stromverbrauch

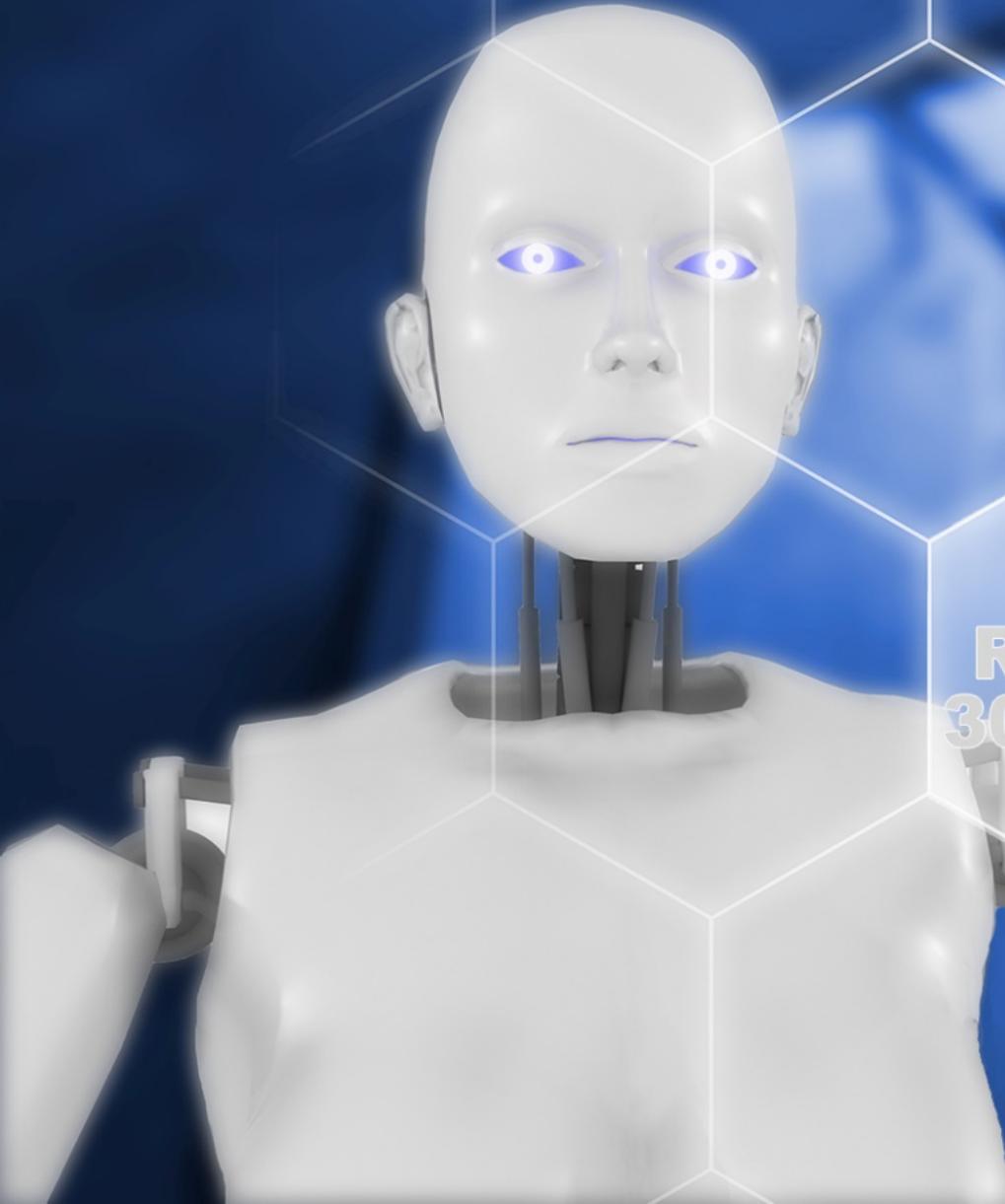
Angreifer können
verschlüsselte Daten jetzt
speichern und später
entschlüsseln

Neues facebook-Datenzentrum in Alaska: Senkung des Stromverbrauchs durch extrem kühle Umgebungen



Sorgen um Privatheit im Internet

- **Digital Natives** befürchten, von anderen Nutzern beobachtet zu werden.
- **Digital Immigrants** befürchten dass Betreiber persönliche Informationen sammeln, verarbeiten und weitergeben.
- **Digital Outsiders** fühlen sich den Gefahren im Internet hilflos ausgesetzt.



LERNENDE
MASCHINEN
02.05.2017

INDUSTRIE
4.0
23.05.2017

SPRACH-
DIALOGE
09.05.2017

KÜNSTLICHE
INTELLIGENZ

KI

BIG
DATA
13.06.2017

TEAM-
ROBOTIK
30.05.2017

AUTONOME
SYSTEME
16.05.2017

**ALTERS-
ASSISTENZ**

**SMART
SERVICE**

**SICHER-
HEIT
20.06.2017**

**EMOTION &
VERHALTEN**

Vielen Dank für Ihre Aufmerksamkeit

