

Gegen die Diskussion mit den drei Unbekannten: Daten, Algorithmen und Digitalisierung

Ringvorlesung des Interdisziplinären Zentrums Digitale Geistes- und Sozialwissenschaften

» tech talk. Einblicke in die digitale Transformation«

Wintersemester 2018/2019

16. Oktober 2018

Prof. Dr. Svenja Hagenhoff

Friedrich-Alexander Universität Erlangen-Nürnberg

Institut für Buchwissenschaft

Professur für E-Publishing und Digitale Märkte



Gliederung

- 1. Motivation**
2. Daten
3. Algorithmen
4. Digitalisierung
5. Synopse

Alles digital, ... oder was? (modifiziert nach »Müller Milch«)

2016



2016



2017



2017



2014



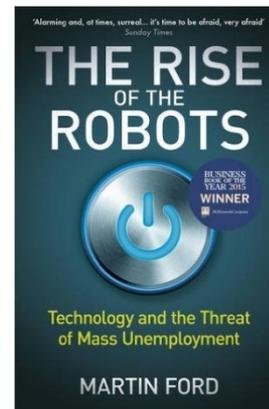
2014



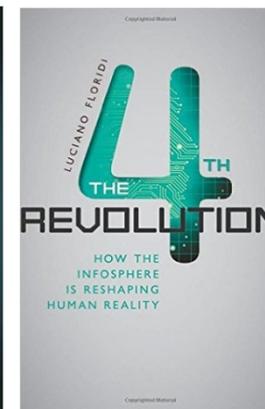
2015



2016



2016

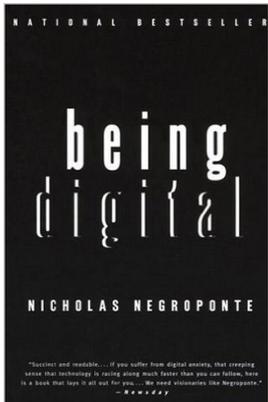


2016

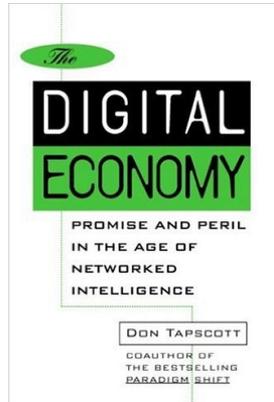


Das ist alles sowas von 90er!

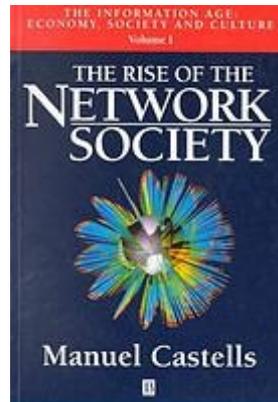
1995



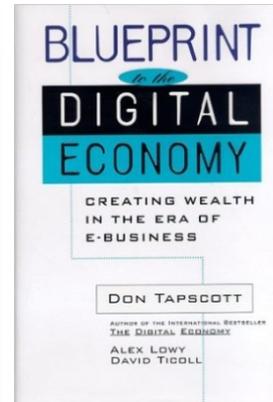
1995



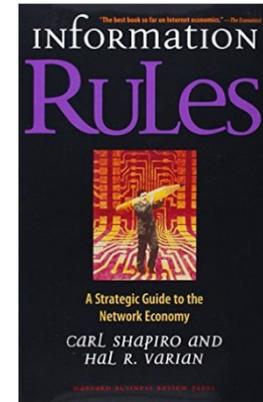
1996



1998



1998



1999



Gegen Buzzword-Weitwurf und Hantieren mit Halbwissen

📅 9. Oktober 2018



Öffentliche Ringvorlesung des Interdisziplinären Zentrums für Digitale Geistes- und Sozialwissenschaften

Viel wird derzeit über „Digitalisierung“ gesprochen. Politik und Medien gebrauchen das Schlagwort geradezu inflationär. Der Begriff selbst bleibt dabei merkwürdig diffus.

Gliederung

1. Motivation
- 2. Daten**
3. Algorithmen
4. Digitalisierung
5. Synopse

Unbekannter 1: Daten

Datenmarkt | 03.05.2018 | Lesezeit 2 Min.



Digitalisierung: Daten sind das neue Öl

Unternehmen mit datengetriebenen Geschäftsmodellen wie Facebook oder Alphabet zählen zu den wertvollsten weltweit. Die Daten, die diese Firmen sammeln, haben mittlerweile ein Volumen erreicht, das sich spürbar auf das Bruttoinlandsprodukt auswirken kann. In Europa ist dieser Effekt in Großbritannien besonders groß.

<https://www.iwd.de/artikel/digitalisierung-daten-sind-das-neue-oel-386895/>

Warum Daten nicht das neue Öl sind

21.03.2016 Lutz Maicher

Daten sind das neue Öl! Unaufhörlich hallt dieses Credo durch die Säle der Startup-Konferenzen und Gründerforen. Das Problem dabei: Es stimmt nicht. Eine Analyse

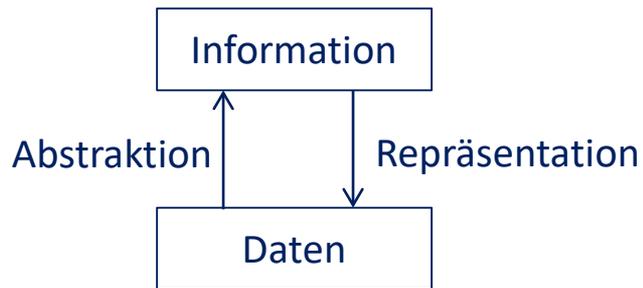
<https://digitalpresent.tagesspiegel.de/warum-daten-nicht-das-neue-oel-sind>

Was sagt das Lexikon?

- Datum (sing.)
- Daten (plur.)
- »gegeben«
- »das, was gegeben ist«
- Daten sind objektive Tatsachen, Fakten, Messwerte, die Gegebenheiten der Realwelt beschreiben

Was sagen verschiedene Disziplinen?

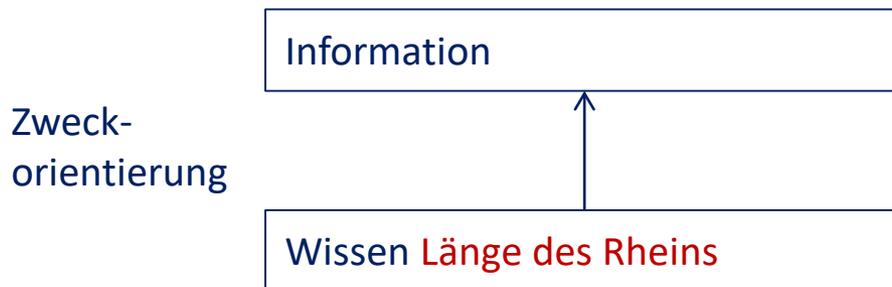
Informatik



Wirtschaftsinformatik



BWL



Kategorisierung von Daten

Formbezogene Kriterien

Wie?

z.B. Erscheinungsform

- Statisch schriftlich
- Statisch bildlich
- Dynamisch bildlich
- Dynamisch akustisch

Ursprungs- bezogene Kriterien

Woher?

z.B. Nutzeraktivität

- Aus Transaktionen
- Aus Interaktionen

Verarbeitungs- und nutzungsbezogene Kriterien

Wozu?

z.B. Zeitbezug

- Nicht-historisierend
- Historisierend

Governance- bezogene Kriterien

Welche
Konditionen?

z.B. Zugänglichkeit

- Gegen Entgelt
- frei

Beispiel: Nutzeraktivität

Daten stammen aus Transaktionen

- Bestellungen, Einkäufe, Buchungen verursachen einen dokumentierten Geschäftsvorfall
- Old School, gibt es schon lange (früher: Karteikarte)
- Aber: in der Analog-Welt war vieles anonym
- Elektronische und Digitale Marktplätze erfordern Kontaktdaten

Daten stammen aus Interaktion von Menschen mit Computern

- Jeder »Klick im Digitalen« erzeugt Datenspuren und ist dokumentiert
- Bisher nur schwer oder gar nicht Beobachtbares ist beobachtbar
- Dieses nicht für singuläre Fälle oder einzelne Personen und zu singulären Zeitpunkten sondern historisierend und aggregierend

Beispiel: Reader Analytics

Old School



Das Lesen ist schön. Die besten Bücher-Cartoons aus dem New Yorker. Atlantik (2015)

New School



Screenshot Smartphone

Reader Analytics: mögliche Daten

Daten bezogen auf den individuellen Leser pro Buchtitel

- wann wird gelesen
- wie lange am Stück wird gelesen (Sessionlänge)
- bis wohin wurde in der jeweiligen Session gelesen (Sessionendpunkt)
- wieviel zeit liegt zwischen zwei Lesephasen (Sessiondistanz)
- genutzte Leseumgebungen / Geräte

Aggregierte Daten über die Leserschaft pro Buchtitel: *Leseverhalten*

- Conversion Titelinformationen → Leseprobe
- Conversion Leseprobe → Gelesen bzw. gekauft
- Fertigleserate: wieviele lesen zu Ende?
- Durchschnittliche Lese geschwindigkeit
- Empfehlungsfaktor: Würden die Leser das Buch Freunden empfehlen

Evolution der Rolle von Daten in der Leistungserstellung

Daten als Ergebnis von Prozessen

- Daten dokumentieren die Ergebnisse von Prozessen
- Datenfluss läuft Materialfluss hinterher (Zettelwirtschaft)

Daten als Befähiger von Prozessen

- Datenverarbeitung in Echtzeit
- Daten steuern Prozesse automatisiert

Daten als Befähiger von Produkten

- Individualisierte Leistungsangebote: Mass Customization
- Losgröße 1

Daten als Produkt («datenzentrierte Geschäftsmodelle«)

- Daten sind zweckdienliche Information
- Daten sind das Material aus dem der Problemlöser besteht
- Aus Datenbeständen werden Werte geschöpft
- »Information Processing«

Gliederung

1. Motivation
2. Daten
- 3. Algorithmen**
4. Digitalisierung
5. Synopse

Unbekannter 2: Algorithmen

Programmieren

Machen die
bei Google

Was Kompliziertes

Ursache
allen Übels

Weltherrschaft

Irgendwas mit
Computern

Was für Nerds

Wat is'n Dampfmaschin? *(Prof. Bömmel, Feuerzangenbowle)*

»Lösungsverfahren in Form einer Verfahrensanweisung, die in einer wohldefinierten Abfolge von Schritten zur Problemlösung führt.«

Lackes, Richard; Siepermann, Markus; Lübbecke, Marco (2018): Algorithmus. Hg. v. Gablers Wirtschaftslexikon (Online). Wiesbaden.

»Ein Algorithmus ist eine endliche Abfolge von wohldefinierten Befehlen eines vorgegebenen Befehlssatzes, die bei Ausführung Eingabedaten in Ausgabedaten umwandeln.«

Boockmeyer, Arne; Fischbeck, Philipp; Neubert, Stefan (2017): Fit fürs Studium - Informatik. 1. Auflage Auflage. Bonn, S. 27.

»Algorithms are general, step-by-step methods for solving problems.«

Garey, Michael R.; Johnson, David S. (1979): Computers and intractability. A guide to the theory of NP-completeness. New York, NY, S. 4

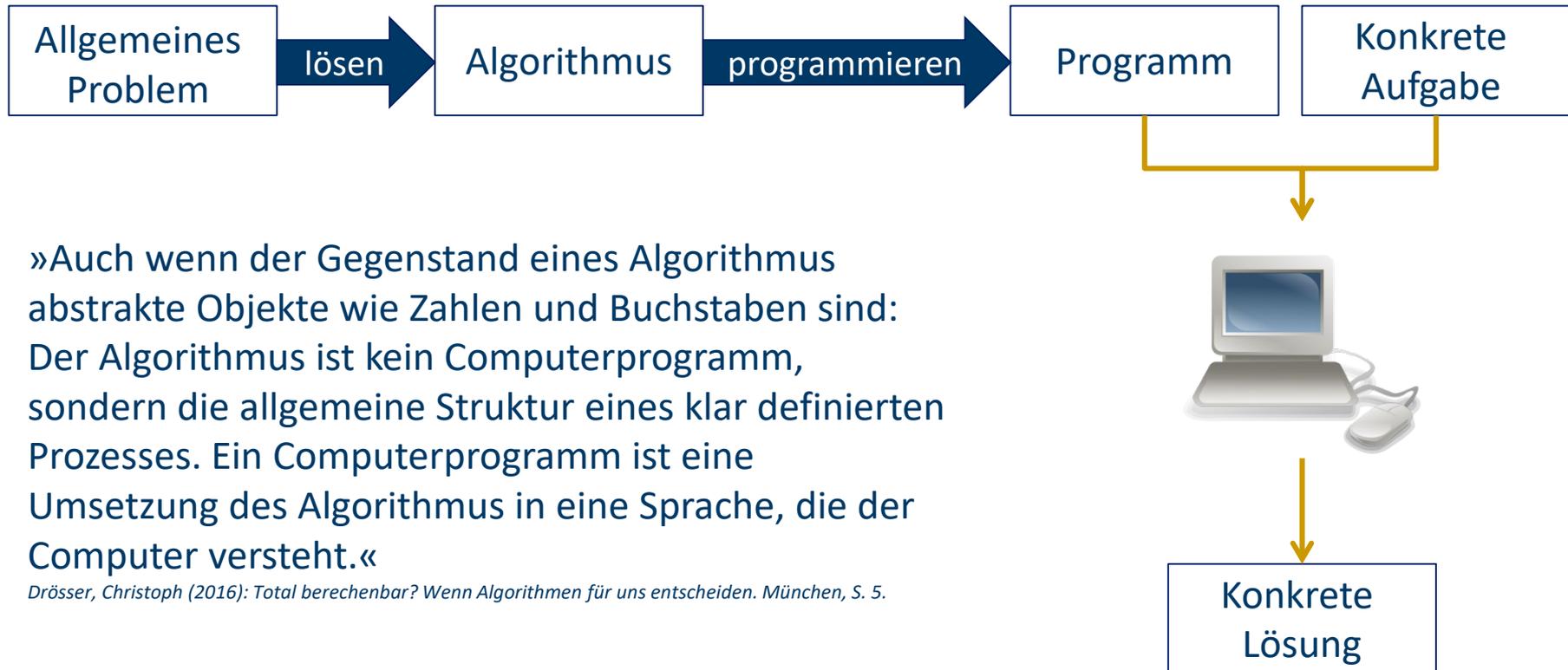
»Aber Algorithmen brauchen keine Computer. Der Mensch kennt Algorithmen spätestens, seit er rechnen kann. Algorithmisch zu denken heißt, darüber nachzudenken, wie man denkt. Ein Algorithmus ist ein Teil unseres Denkens, den wir so gut verstanden haben, dass wir ihn getrost auslagern können. Wir lassen denken. Dafür sind dann die Computer gut.«

Stiller, Sebastian (2015): Planet der Algorithmen. Ein Reiseführer. E-Book, Pos. 56.

»Algorithmen lösen keine physischen Probleme, wie es Nagelfeilen oder Vollwaschmittel können. Algorithmen verarbeiten Informationen.«

Stiller, Sebastian (2015): Planet der Algorithmen. Ein Reiseführer. E-Book, Pos. 722.

Der Weg vom Problem zur Lösung



Eigenschaften von Algorithmen

- Allgemeingültigkeit
- Finitheit
- Terminiertheit
- Determiniertheit
- Determinismus

Allgemeingültigkeit

- Ein Algorithmus ist anwendbar auf alle Probleme der gleichen Problemklasse
- Er ist abstrakt und nicht konkret
- Beispiel:
 - Problemklasse »Sortieren von Elementen«
 - Konkrete Probleme

3 Bücher sortieren

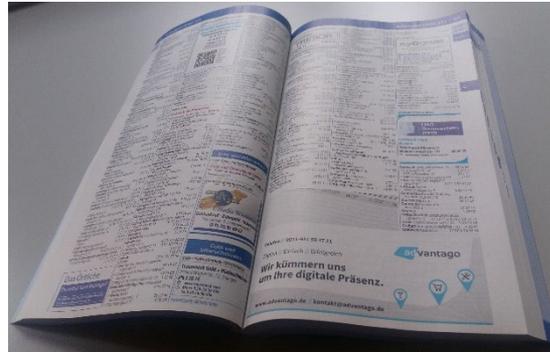


12 Bücher sortieren



Abstraktion schreckt ab. Probieren Sie's!

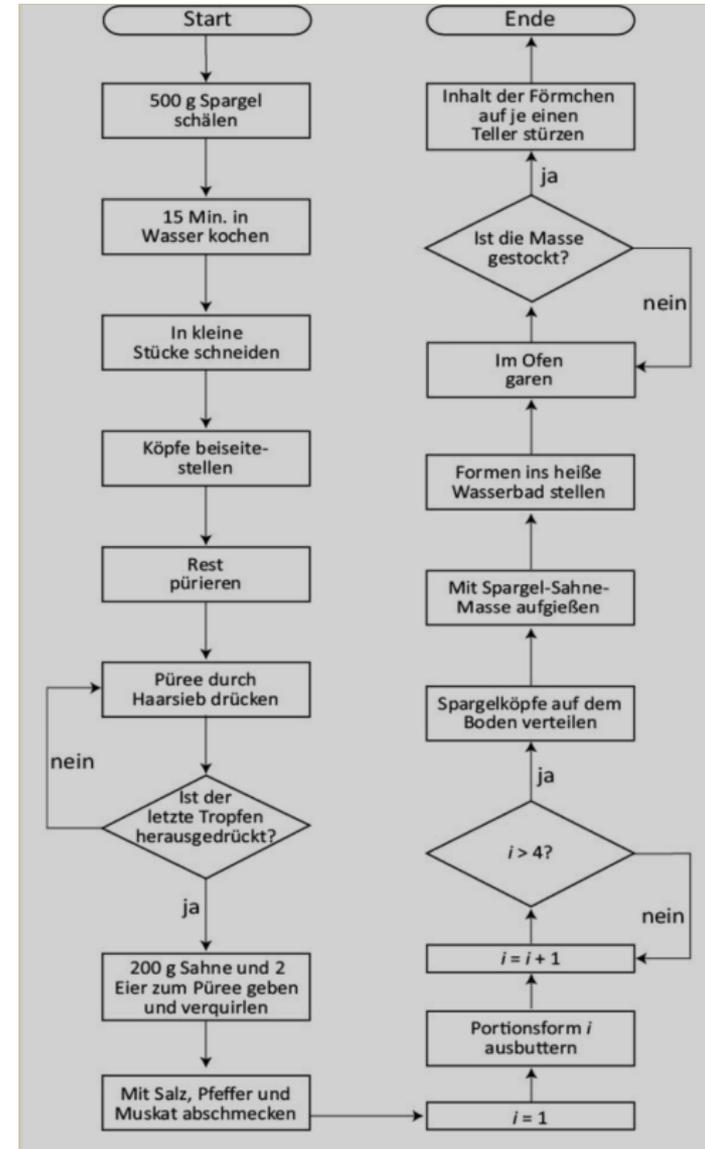
Algorithmus zur Suche der Stelle im gedruckten Telefonbuch, wo Namen mit [Buchstabe] beginnen



1. Schlage das Buch irgendwo auf
2. Prüfe den gefundenen Anfangsbuchstaben der Namen
3. Wenn Anfangsbuchstabe
 - a. hinter [Buchstabe] dann markiere den rechten Teil des Buchs als irrelevant und schlage den relevanten linken Teil irgendwo auf. Wiederhole Schritt 2.
 - b. vor [Buchstabe] dann markiere den linken Teil des Buchs als irrelevant und schlage den relevanten rechten Teil irgendwo auf. Wiederhole Schritt 2.
 - c. gleich [Buchstabe] dann höre auf (»fertig«)

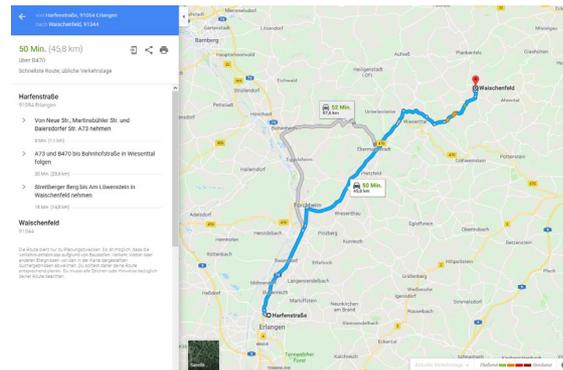
Fininitheit

- Ein Algorithmus ist von endlicher Länge
- Eine endliche Menge an Schritten kann in endlicher Zeit durchlaufen werden
- Beispiel: erfolgreich Soufflé machen in 14 Schritten

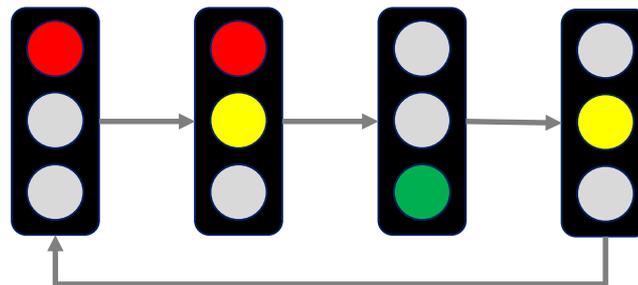


Terminiertheit

- Endlichkeit des Prozesses, den der Algorithmus beschreibt
- Terminierende A. kommen von selber zu einem definierten Ende (>fertig<)



- Nicht-terminierende A. bilden Vorgänge ab, die nicht abbrechen sollen



Determiniertheit

- Bezieht sich auf das Ergebnis des Algorithmus
- Determinierte A. liefern bei gleichen Ausgangsdaten bei jeder Ausführung stets das gleiche Resultat.

Brot & Butter & Käse = Käsebroten

$$((7 + 3) \times 10) / 2 = 50$$

- Nicht-determinierte A. liefern trotz gleicher Ausgangsdaten divergente Ergebnisse

1. Wähle eine Zahl
2. Werfe einen Würfel
3. Multipliziere die gewählte Zahl mit den Augen des Würfels

3	3	3
2	4	1
6	12	3

Determinismus

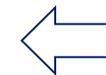
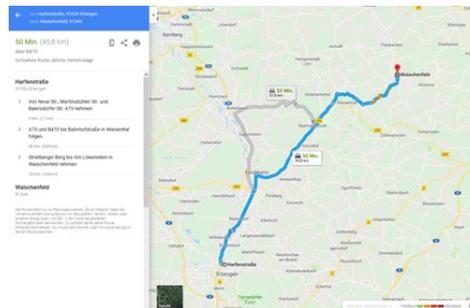
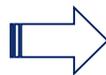
- Bezieht sich auf den Lösungsweg
- Deterministische A. durchlaufen von Mal zu Mal immer den gleichen vorgegeben Weg

Rechnung erstellen

1. Ermittle den Nettopreis pro Produkt
2. Errechne die MWSt. pro Produkt
3. Addiere die MWSt. pro Produkt
4. Addiere die Bruttopreise über alle Produkte

- Nicht-deterministische A. durchlaufen von Mal zu Mal unterschiedliche Wege

Aktuelle
Verkehrslage



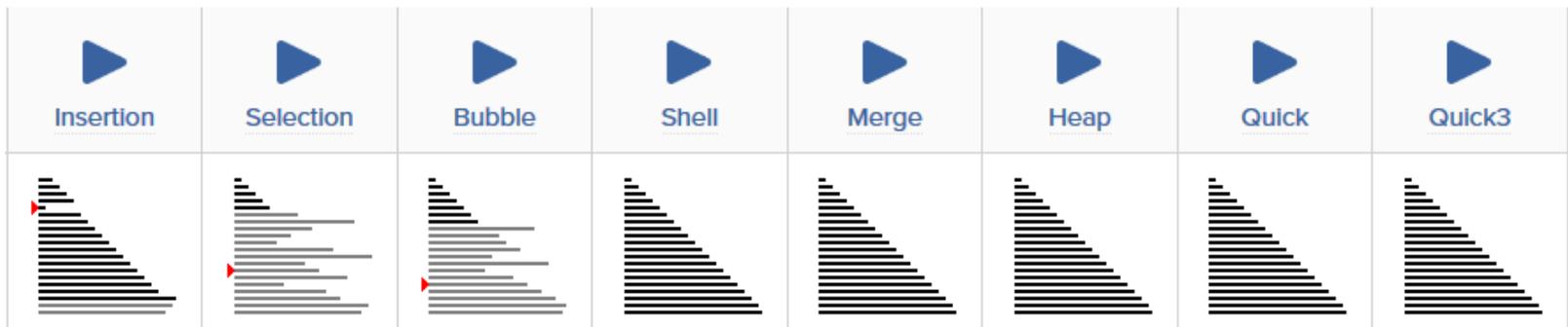
Geänderte
Verkehrsführung



Aktuelles Fahrerverhalten

Anforderungen an Algorithmen: Effizienz

- Wenig Laufzeit
- Wenig Speicherplatz



Algos bei der Arbeit: <https://www.toptal.com/developers/sorting-algorithms>

Anforderungen an Algorithmen: Korrektheit

- Herausforderung: wie stellt man Korrektheit fest?
- *Testen* identifiziert anwesende Fehler, aber nicht deren grundsätzliche Abwesenheit
- *Objektiv korrekt* im Sinne einer *allgemein anerkannt zulässigen Lösung* gibt es bei vielen Problemklassen gar nicht (z.B. Empfehlungsmechanismen)
- Es geht dann eher darum, dass der Algorithmus *das Benötigte* tut (z.B. etwas aus einer Menge auswählen)
- Was die passenden Parameter sind ist zeitbezogener Gegenstand von Vereinbarungen: phänomenologische statt epistemische Berechenbarkeit

Pietsch, Wolfgang; Wernecke, Jörg (2017): Einführung: Zehn Thesen zu Big Data und Berechenbarkeit. In: Wolfgang Pietsch, Jörg Wernecke und Maximilian Otte (Hg.): Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data. Wiesbaden, S. 13–35.

Gliederung

1. Motivation
2. Daten
3. Algorithmen
- 4. Digitalisierung**
5. Synopse

Unbekannter 3: Digitalisierung

00101110

Das Internet

GAFA*

Was mit
Computern

Irgendwie
immateriell

Auf jeden Fall wird
alles anders

Vernetzung

Neumodisches Zeug,
überbewertet!

Teil 1: Begriff

- Stufenweise, zu den Fingern gehörend
- Abstrakt: Abzählbar
- Digital sind alle Techniken, die mit einer abzählbaren (festgelegten) Menge an Entitäten arbeiten. Alle gewünschten Aussagen oder Zustände werden durch Kombination einer begrenzten Menge an Entitäten erzeugt.



Teil 2: Geschickte Codierung mit 0 und 1

- Basaler Kern: Reduktion jedweder Informationscodierung auf die Kombination von wenigen informationsarmen Zeichen oder Zuständen, faktisch 0 und 1
- Da auf der Basis der Codierung mit 0 und 1 auch gerechnet werden kann liegt eine Universaltechnologie vor (General Purpose Technology)

»In unübertreffbarer Zeichenökonomie taucht das Bit einzig als Gegensatz seiner eigenen Abwesenheit auf.«

Kittler, Friedrich A. (1995): Von der Letter zum Bit. In: Horst Wenzel (Hg.): Gutenberg und die Neue Welt. München, S. 105–117.

1 0

Teil 3: Elektronifizierung und die Konsequenzen

- Mit 0 und 1 binärcodierte Information ist elektrifizierbar: Strom an, Strom aus
- Elektronifizierung ermöglicht Unabhängigkeit von spezifischen Techniken und Materialitäten
- Elektronifizierung ermöglicht einfachere Automatisierung mithilfe von Universalgeräten (Computer als Hardware) und Software (manifestierter Algorithmus)
- Elektronifizierung erleichtert Vernetzung
 - Vernetzung = Verbindungen zwischen Akteuren, Objekten oder Entitäten zum Austausch von Signalen (techn.) oder zur Kommunikation (soz.wiss.)
 - Elektronische Universalgeräte sind »potente Knoten im Netz der Zeichenzirkulation«

Winkler, Hartmut (2004): Medium Computer. Zehn populäre Thesen zum Thema und warum sie möglicherweise falsch sind. In: Lorenz Engell und Britta Neitzel (Hg.): Das Gesicht der Welt. Medien in der digitalen Kultur. München, S. 203–213.

Teil 4: Miniaturisierung von Hardware

- The invisible Computer

Norman, Donald A. (1999): The invisible computer. Why good products can fail, the personal computer is so complex, and information appliances are the solution. Cambridge.

- Veränderung von mobilen Endgeräten ab 2007 → Life-Style-Produkte
- Sensoren
- Ubiquitär vorhandene Hardware führt zu potenzieller Totalvernetzung technischer Komponenten

Teil 5: Miniaturisierung von Software

- Statt hässlicher Funktionsmonster: funktionsisolierte Apps als Alltagshelfer

zur Funktionsgröße von Software vgl. Fahsel, Jörn et al. (2017): Publishing 4.0. Chancen, Anforderungen, Konzepte. Denkzeug 2017, Erlangen und Hamburg, S. 39.

- Deutlich verbesserte Usability

- Zusammen mit ubiquitärer Hardware (insb. Smartphone)

- tatsächliche Totalvernetzung

- Erzeugen großer Datenmengen, vor allem Nutzungs- und Interaktionsdaten

z.B. Bensberg, F.; Weiß, T. (1999): Web Log Mining als Marktforschungsinstrument für das World Wide Web. In: Wirtschaftsinformatik 41 (5), S. 426–432.

Gliederung

1. Motivation
2. Daten
3. Algorithmen
4. Digitalisierung
- 5. Synopse**

Algorithmus

- ist als Begriff und Konzept recht klar zu fassen
- Dennoch: besser wäre der Begriff *Software*, dies ist die automatisierte Variante des Problemlösers *Algorithmus*
- Erforderlich ist ein konzeptionelles Verständnis von Arten von Algorithmen sowie von ihren Grenzen in den Gesellschaftswissenschaften
- Erforderlich ist ein Verständnis der Wechselwirkung von Technik und Gesellschaft in den Technikwissenschaften, z.B. Informatik

Daten

- ist als Begriff und Konzept extrem unscharf
- Begriffe für das gefühlt immer gleiche sind von Moden abhängig: Daten, Datengetriebene Zeiten, Datenintermediäre; Information, Informationsgesellschaft, Informationsintermediäre; Wissen und Wissensgesellschaft
- Unterscheide zwei Einsatzbereiche von Daten:
 - Daten dienen der Beschreibung von Zuständen und Aktivitäten, man braucht sie, um Entscheidungen treffen zu können (= entscheidungsrelevante Information)
 - Daten sind das Material, aus dem Kommunikate und kulturelle Artefakte sind (Text, Bilder, Audio), dieses liegt heute elektronisch vor (= Zeichen und Pixelhaufen)
- **Daten sind das neue Latein!**

»Digitalisierung«

- ist eine Nebelwolke, muss es aber nicht sein
- Teil 1: Codierung mit endlichem Zeichenvorrat
- Teil 2: geschickte Codierung mit 0 und 1
- Teil 3: Elektronifizierung → Unspezifität, Automatisierung und Vernetzung
- Teil 4: Miniaturisierung von Hardware (everywhere and allways)
- Teil 5: Miniaturisierung von Software (auch everywhere and allways)

»Digitalisierung«

Es entstehen große Mengen Daten höchst unterschiedlicher Beschaffenheit auch aus bisher nicht oder nur sehr aufwändig erschließbaren Quellen (Interaktionsdaten). Extrem gestiegene Rechenleistung ermöglicht eine neuartige Verarbeitung dieser Datenmengen in Form algorithmischer Modellierung und Induktion. Hiermit wird das Potenzial geschaffen, dass sehr komplexe, variablenreiche Probleme automatisiert genauer, verlässlicher oder schneller bearbeitet werden könnten als ggf. dieses der Mensch mit seinem (Teil-)beschränkten Erkenntnisapparat leisten kann.

Finis!

»Wenn Sie einen Scheißprozess digitalisieren, dann haben Sie halt einen scheiß digitalen Prozess.«

*Thorsten Dirks
Vorstandsvorsitzender Telefónica Deutschland
und Bitcom-Präsident
2015*

Kategorisierung von Daten

Kriterium	Ausprägung			
<i>Formbezogene Kriterien</i>				
Erscheinungsform	Statisch: Schriftlich	Statisch: bildlich	Dynamisch: bildlich	Dynamisch: akustisch
Zeichentyp (Teil 1)	Symbolisch		Ikonographisch	
Zeichentyp (Teil 2)	Alphabetisch	Numerisch	Sonderzeichen	Alphanumerisch
Zeichentyp (Teil 3)	Numerisch (quantitativ)		Nicht-numerisch (qualitativ)	
Formatierung	Formatiert		Nicht formatiert	

Kategorisierung von Daten

Kriterium	Ausprägung		
<i>Ursprungsbezogene Kriterien</i>			
Zielorientierung der Erhebung	Gezielt erhobene Daten (captured data)		Daten als ‚Beifang‘ (exhausted data)
Erzeugungsform	Daten als Ergebnis von Erhebung		Daten als Ergebnis von Verarbeitung
Erhebungsmöglichkeiten	1st-Party	2nd-Party	3rd-Party
Nutzeraktivität	Transaktionen		Interaktionen

Kategorisierung von Daten

Kriterium	Ausprägung			
<i>Verarbeitungs- und nutzungsbezogene Kriterien</i>				
Stellung im Verarbeitungsprozess	Eingabedaten		Ausgabedaten	
Basisoperationen	Rechendaten	Ordnungsdaten	Identifikationsdaten	
Aktivitätsgrad	Steuerdaten (veranlassen Aktivitäten)		Passivdaten (werden verarbeitet)	
Verwendungszweck (Teil 1)	Zustandsdaten: Stammdaten	Zustandsdaten: Bestandsdaten	Abwicklungsdaten: Vormerkdaten	Abwicklungs- daten: Bewegungsdaten
Verwendungszweck (Teil 2)	Operative Daten		Dispositive Daten	
Funktion (Teil 1)	Metadaten		Nutzdaten	
Funktion (Teil 2)	Identifikationsdaten	Eigenschaftsdaten	Metadaten	
Grad der Strukturierung	Strukturiert		Unstrukturiert	
Grad der Beständigkeit	Transient		Persistent	
Zeitbezug	Temporal / historisierend		Nicht-temporal	

Kategorisierung von Daten

Kriterium	Ausprägung		
<i>Governance-bezogene Kriterien</i>			
Zugänglichkeit	Frei zugänglich		Nicht frei zugänglich
Datenherkunft	Privat	Öffentlich	Gemeinschaftlich
Datenerzeugung	Freiwillige Angaben	Aus Beobachtung	Abgeleitet

Konzeptualisierung nach Ballsun-Stanton (2010)

- Data as hard numbers
 - wahre Aussagen über die Realität
 - objektiv prüfbar
 - Je mehr davon desto präziser ist die Realität erfassbar
- Data as recorded observations
 - subjektive Wahrnehmungen
 - Interpretationsbedürftig
 - Entstehen immer in komplexen Begründungs- und Argumentationszusammenhängen
- Data as bits
 - dienen der computerbasierten Kommunikation
 - Im Fokus: Semiotik
 - Abstrakter formuliert: *Data as communication* formuliert werden. Daten würden dann unterschiedliche, menschenlesbare Formen annehmen (z.B. Buchstaben), die lediglich zum Zwecke des maschinellen Transports auf Bitfolgen reduziert werden.