



Das kleine ABC der Informatik und seine Auswirkungen auf die Medien- und Meinungsvielfalt.

Ausschuss für Medienentwicklung und Medienordnung
der Medienkommission der LfM

Prof. Dr. Katharina A. Zweig



Das kleine ABC der Informatik

Gefährden

Algorithmen,

Big Data und

Künstliche Intelligenz

unsere Medienvielfalt und Meinungsfreiheit?



A wie Algorithmus

Ein Algorithmus ist ein Problemlöser

Problem



INPUT

**Der OUTPUT
der uns sagt,
wie Input
mit Output
zusammenhängt.**



OUTPUT

Input: By User:Bluemoose - Own work, [CC BY-SA 3.0](#)

Putput: By Yann (talk) - Own work, GFDL

Output: [CC BY-SA 3.0](#)



Beispiel

Problem: Sortieren



Ein Algorithmus ist...



...eine für jede **erfahrene Programmiererin** und jeden erfahrenen Programmierer **ausreichend detaillierte Lösungsvorschrift**, so dass bei **korrekter Implementierung** der Computer **für jede korrekte Inputmenge den korrekten Output** berechnet – in endlicher Zeit.

Sortieren 1: „Sortieren durch Einfügen“



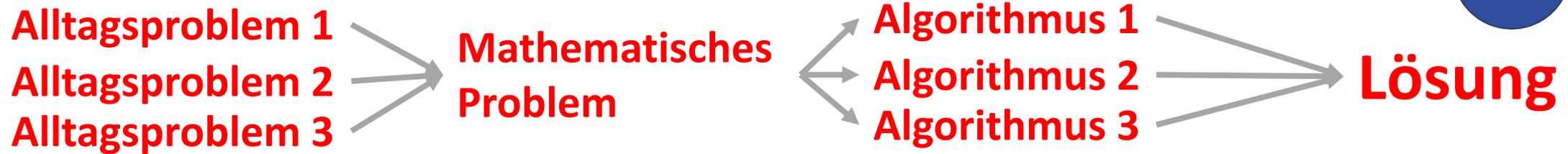
- Fange mit einem Buch an, stelle es ins Regal.
- Solange es noch Bücher gibt,
 - nimm das nächste,
 - geh am Regal entlang und sortiere es an der passenden Stelle ein.
- Alle Bücher, die schon im Regal stehen, sind in der richtigen, relativen Reihenfolge.
- Daher: wenn alle im Regal stehen, sind sie vollständig sortiert.

Sortieren 2: Aufsteigendes Sortieren



- Stelle alle Bücher irgendwie ins Regal.
- Gehe das Regal entlang – wenn dabei zwei Bücher in der falschen Reihenfolge nebeneinander stehen, vertausche sie. Tue dies bis zum Ende des Regals und gehe wieder zum Anfang.
- Laufe solange immer wieder am Regal entlang, bis im letzten Durchgang kein Tausch mehr nötig war.
- Wenn kein Tausch mehr nötig war, sind alle Bücher sortiert.

Problem-Algorithmus-Problem



- Ein mathematisches Problem kann also meist durch mehrere Algorithmen gelöst werden.
- Jeder Algorithmus löst nur genau ein mathematisches Problem.
- Im Sinne von „Alltagsproblemen“ löst derselbe Algorithmus sehr viele verschiedene Probleme:
 - Sortieren von Personen nach Anzahl ihrer Follower auf Twitter;
 - Anzeige von Nachrichten, sortiert nach Publikationsdatum;
 - Suchmaschineneinträge sortieren nach Bewertung durch Suchmaschinenalgorithmus;

Alle Sortierprobleme auf einen Schlag



- Gegeben eine Menge von Objekten oder Subjekten...
- und ein Sortierkriterium, das für je zwei von diesen besagt, welches nach links, welches nach rechts sortiert werden muss,...
- kann jeder beliebige Sortieralgorithmus die korrekte Lösung berechnen.

- Eine Interpretation der Ergebnisse (dies sind die relevantesten Nachrichten, die wichtigsten Freunde, die kaufenswertesten Produkte) liefert er **nicht**.



Komplexe Algorithmen

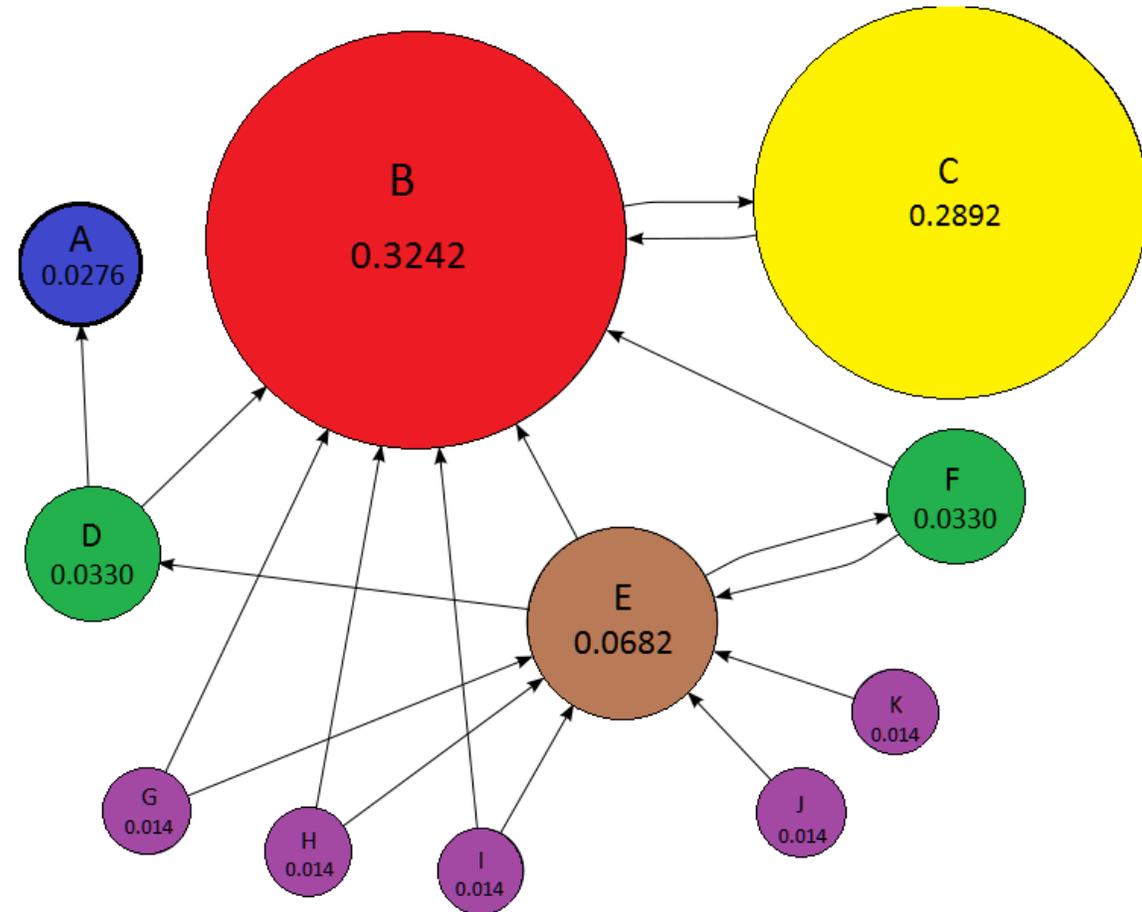
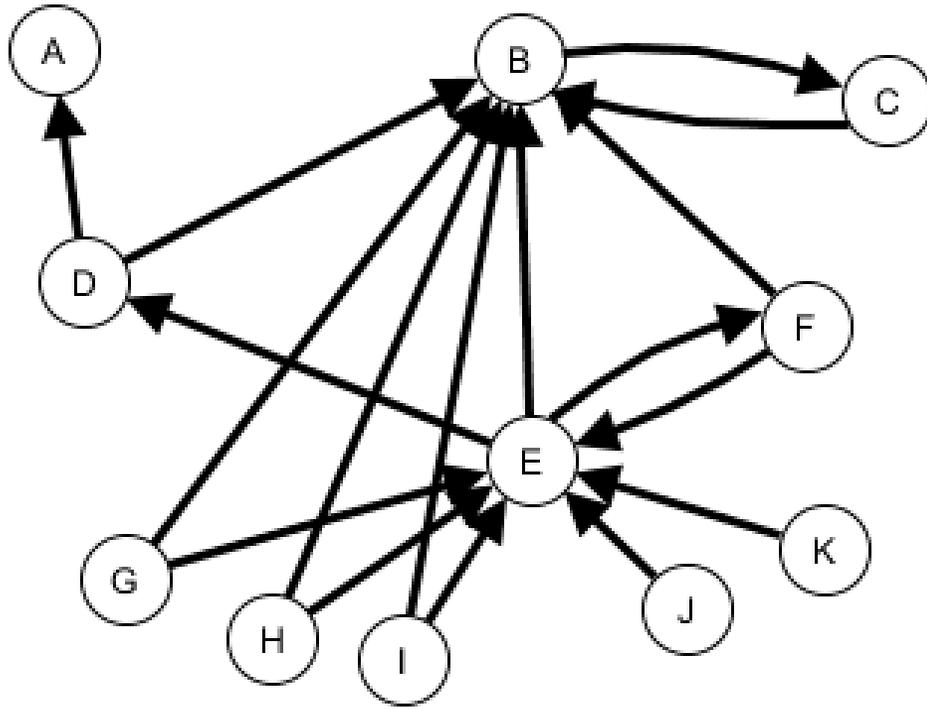
Beispiel: Suchmaschinenalgorithmen

Suchmaschinen 101



1. Filtern aus allen ihnen bekannten Webseiten diejenigen, deren Text mit den angegebenen Suchbegriffen zusammenhängen.
2. Welche Webseite ist am Wichtigsten?
3. Idee: nutze Vernetzungsstruktur

PageRank



Idee hinter dem Algorithmus



Ein Modell menschlichen Verhaltens: der Random Surfer

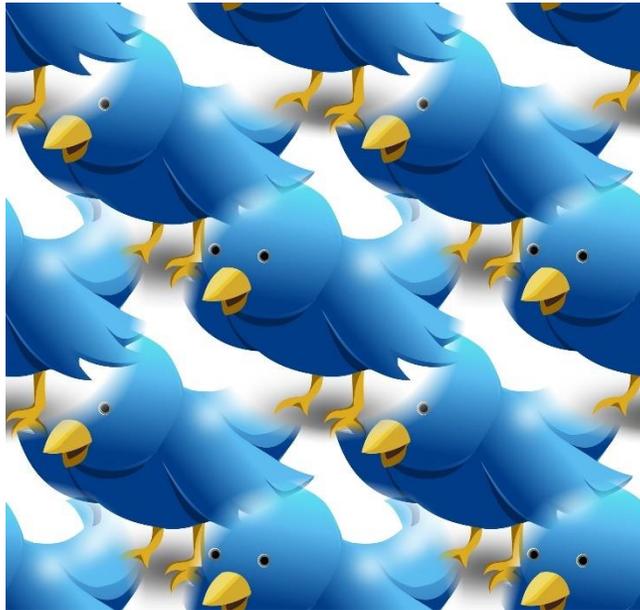
- Ein Surfer klickt auf eine Webseite
- Folgt einem der Links auf der Webseite zufällig
- Von Zeit zu Zeit springt er auf eine völlig neue Webseite
 - Modelliert externes Wissen

Modellierungsannahme



- Gibt nur dann relevante Ergebnisse, wenn Webseiten
 - Links auf ähnliche Seiten wie ihre eigene setzen,
 - Allen alle Seiten bekannt sind,
 - Links auf relevante, meinungsangebende Seiten setzen, und
 - ihre Links **unabhängig** voneinander setzen.
- Unter dieser Bedingung ist der Algorithmus neutral und objektiv und gibt das kollektive Wissen der Welt nutzbringend weiter.
- Die Veröffentlichung des Algorithmus führte prompt zu Manipulationen seitens der Webseitenbetreiber.
 - Zu große Offenheit der Algorithmen ist manchmal **schädlich**.

Twitter Netzwerke



- Wer folgt wem auf Twitter?
- Dieselbe Art von Informationen:
 - Wer verlinkt auf wen?
- Relevant:
 - Accounts mit hohen Follower-Zahlen, wie z.B. @RegSprecher Steffen Seibert
 - 525.000 Follower
 - Accounts mit „wichtigen“ Followern:
 - Seibert folgt selbst 100 Accounts, darunter
 - Zeit Online
 - Jan Josef Liefers
 - Barack Obama



Zusammenfassung Algorithmen

Algorithmen



- Eingefrorene Handlungsanweisungen, von Menschen erdacht, um mathematische Probleme durch Computer zu lösen.
- Ein mathematisches Problem kann durch verschiedene Algorithmen gelöst werden – per Definitionem kommt dabei immer exakt dasselbe Ergebnis heraus!
- Verschiedene Fragen können durch dasselbe mathematische Problem modelliert werden, z.B.:
 - Was ist die relevanteste Webseite?
 - Wer ist der einflussreichste Twitterer?
- Das Ergebnis **muss interpretiert** werden.





B wie Big Data

Was heißt hier eigentlich „relevant“?

Was ist die relevanteste Nachricht zur Anfrage: „Erdogan Visafreiheit“?



Streit um Visafreiheit für Türken: Erdogan legt keine Tagesspiegel - 15.05.2016
Inzwischen vollendet Erdogan gerade seine euv
Vieles deutet darauf hin, dass der Präsident

Visafreiheit: Erdogan empört sich über die EU
ZEIT ONLINE - 12.05.2016
Mit Blick auf die Visafreiheit warf Erdogan ...
aufgebaut zu haben. Man habe sich oh
Erdogans Berater warnt vor Scheit
Meinung - Deutsche Welle - 1
Ausführlicher Hintergrund

Erdogan: "Wir lassen uns keine Anweisungen geben"
tagesschau.de - 11.05.2016
Der türkische Präsident Erdogan bleibt hart: Sein Land werde das ... Dann
würde nicht nur die Visa-Freiheit für türkische Staatsbürger scheitern ...
Flüchtlings-Deal in Gefahr | Warum regen sich alle über Erdogans ...
Ausführlich - BILD - 12.05.2016
Hintergrund (357 weitere Artikel)

+++ Flüchtlingskrise im News-Ticker +++ Unter 2000 Personen: USA ...
FOCUS Online - 11.05.2016
Brok wies darauf hin, dass Erdogan selbst großes Interesse an der Visafreiheit
habe, weil er durch ein Scheitern des Prestigeprojektes im ...

Visafreihe tagesschau.
"Die Europäis sagte Erdogan
Türkei will Anti-T
Meinung - BILD -
Ausführlicher Hintergrund (248 weitere Artikel)

Ändert das Anti-Terror-Gesetz für Visa!
sehen. "In diesem Fall ...
nicht entschärfen | Erdogan brüskiert ...
2016

Big Data



- Wie kann Relevanz modelliert und „quantifiziert“ werden?
- Big Data Methoden nutzen, z.B.:
 - Sprache der Anfrage, Niveau der Anfrage, Wörter, Wortkombination
 - Tageszeit und geographische Informationen, Gerätetyp
 - Ihre bisherigen Suchanfragen und Ihr persönliches Klickverhalten
 - Welche Seiten wurden angeklickt, wie lange betrachtet, kam die Nutzerin wieder zurück zu den Ergebnissen?
 - Metadaten der Nachrichten/Medien: wann erstellt, durch wen, wo publiziert, Verschlagwortung, Wahl der Wörter
 - Verhalten anderer Nutzer, „ liken “ auf sozialen Netzwerken, Interaktion mit Beiträgen



Die Zuordnung einer Frage zu einem mathematischen Problem bezeichnet man als Modellierung.



Frage + Big Data =
mathematisches Problem?

Von der Schwierigkeit der Modellierung

Relevanz – ein weites Feld



"A squirrel dying in front of your house may be **MORE RELEVANT TO YOUR INTERESTS** right now than people dying in Africa."

Mark Zuckerberg, CEO facebook,
nach David Kirkpatrick: „the facebook EFFECT“,
Simon & Schuster New York, New York, USA,
2010, S. 181



Big Data + Frage



- Die Algorithmen designerinnen und -designer müssen nun entscheiden, welche der Daten vermutlich mit „Relevanz“ korrelieren.
- Dies sollte am besten in einer einzigen Zahl pro Medium/Nachricht/Webseite münden, so dass man direkt sortieren kann.
- Beispiel Formel:
 - 3 * bisherige Zugriffe
 - Anzahl Tage seit Publikation
 - + Beliebtheitsquotient des Publikationsortes
 - + Beliebtheitsquotient des Verfassers + ...

Allgemein



$$\begin{aligned} & w_1 * \text{bisherige Zugriffe} \\ + & w_2 * \text{Anzahl Tage seit Publikation} \\ + & w_3 * \text{Beliebtheitsquotient des Publikationsortes} \\ + & w_4 * \text{Beliebtheitsquotient des Verfassers} + \dots \end{aligned}$$

- Wer bestimmt diese Gewichte, so dass insgesamt die „relevantesten“ (also die, die im Nachhinein am öftesten angeklickt werden) am weitesten nach oben sortiert werden?
- Dazu bedarf es Algorithmen der künstlichen Intelligenz.



C wie Künstliche Intelligenz

Lernende Algorithmen





Künstliche Intelligenz

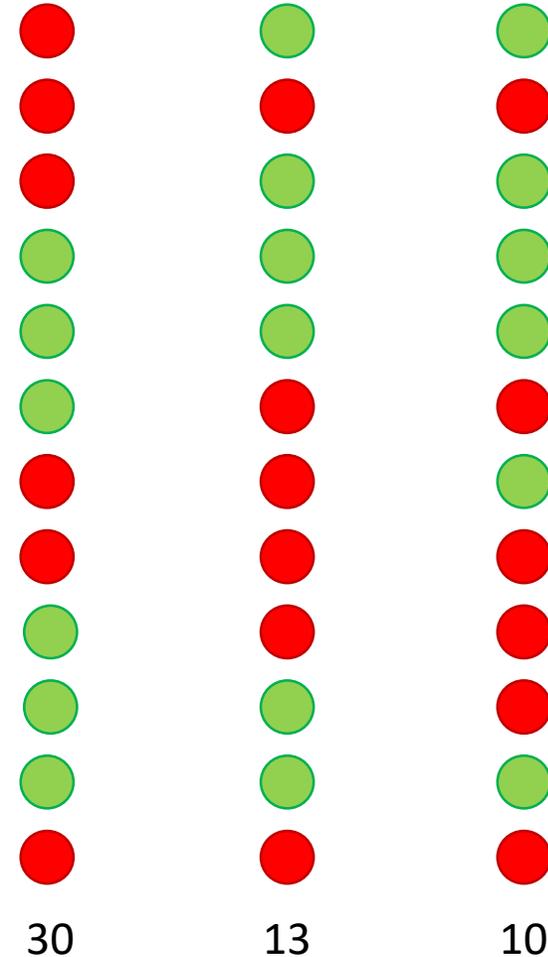
- **Problem:** gegeben eine Menge von bekannten Daten, finde Muster, die auf neuen Daten vorhersagen, wie sich etwas oder jemand verhalten wird.
- Algorithmus baut – basierend auf bekannten Daten – eine Zwischenstruktur auf, die dann Vorhersagen für neue Daten generiert.
- Der Algorithmus wird „auf den Daten trainiert“.



Lernen von Gewichten



- Algorithmus probiert Gewichte
- Bewertet jeweils, wieviele angeklickte Inhalte möglichst weit oben stehen.
- Die Gewichtung, die das maximiert, wird für weitere Daten genommen.
- Kann im Wesentlichen für alles verwendet werden:
 - News Feed bei Facebook
 - Suchmaschinen
 - Produktempfehlung





Relevanz von Algorithmen

Neutrale Empfehlungsalgorithmen?



- Viele Betreiber behaupten, sie würden **nur** nach Relevanz sortieren.
- Empfehlungsalgorithmen filtern, lernen und sortieren.
- Sie machen Modellierungsannahmen, wählen Variablen aus und lernen nur aus einer Teilmenge von Daten.
- All diese Schritte können mehr oder weniger gut gelingen, keiner davon ist neutral im Sinne von „objektiv“.

Spielkampsche Regel



**Alle Algorithmen sind objektiv
Bis auf die von Menschen gemachten!**

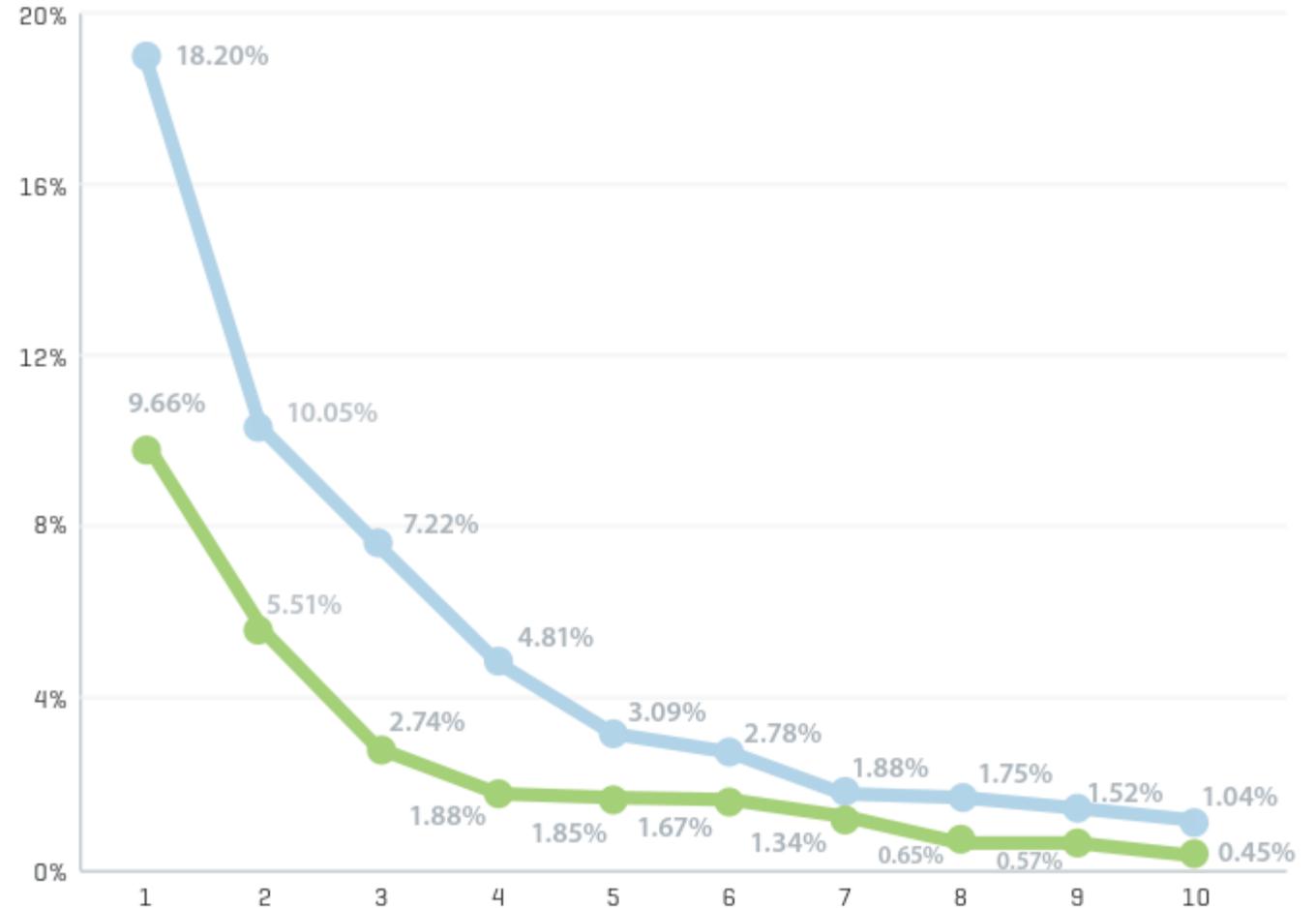


Menschliche Wahrnehmung

Vom Wert, der Erste zu sein

- Der erste bekommt ca. 18,2% (9,7%) aller User, der zweite nur noch 10,1% bzw. 5.51.
- 1. Platz fast doppelt so oft wie der 2., 2,5 mal so oft wie der dritte, fast fünfmal so oft wie der 4.

GOOGLE VS BING CLICK-THROUGH RATE



Paul Davison at Digital Relevance™: „A Tale of Two Studies: Establishing Google & Bing Click-Through Rates“, Study by Digital Relevance™ using client data from Jan-June 2011, available from http://connect.relevance.com/a-tale-of-two-studies-establishing-google-bing-click_through-rates or research@relevance.com; published 2013.

Bevorzugt Google Demokraten?



Studie von Trielli, Mussenden und Diakopoulos¹:

Unter 16 Präsidentschaftskandidaten (USA) gab es bei Demokraten unter den ersten 10 Suchergebnissen 7 positive Berichte, bei Republikanern nur 5,9.

1 <http://algorithmwatch.org/warum-die-google-suchergebnisse-in-den-usa-die-demokraten-bevorzugen/>

Sind wir beeinflussbar über Algorithmen?



- Suchergebnisreihenfolgen:
 - Manipulierte Suchreihenfolgen werden vom Nutzer nicht bemerkt und können die Tendenz eines unentschlossenen Wähler beeinflussen (Epstein & Robertson, 2015)
- Facebooks „Vote“ bzw. „Ich habe gewählt“-Button
 - Studie von Bond et al. über den Effekt auf das Wahlverhalten.
 - Effekt war klein, aber hochgerechnet ca. 60.000 mehr Wahlstimmen.

Epstein, R. & Robertson, R. E.: “The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections”, Proceedings of the National Academy of Science, 2015, E4512-E4521

Bond, R. M.; Fariss, C. J.; Jones, J. J.; Kramer, A. D. I.; Marlow, C.; Settle, J. E. & Fowler, J. H.: “A 61-million-person experiment in social influence and political mobilization”, Nature, 2012, 489, 295-298

Zuccherosconi



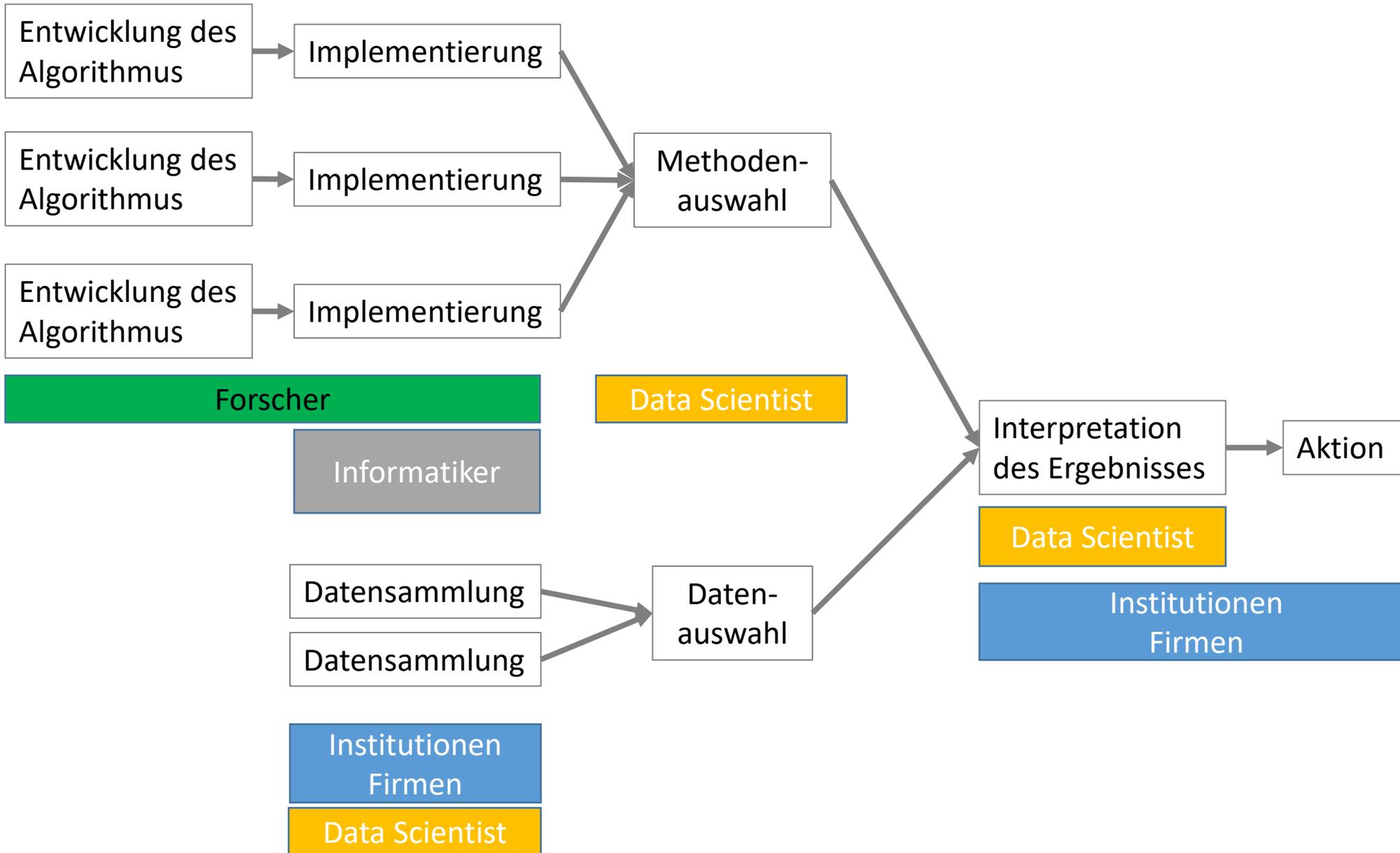
Wollte Zuckerberg der jüngste
Präsident der USA werden –
wer wäre rein technisch noch in
der Lage, als Gegenkandidat über
politische Themen zu punkten?





Algorithmen in einer demokratischen Gesellschaft

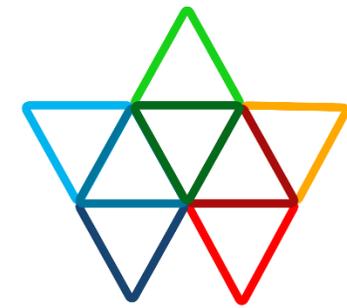
Verkettete Verantwortlichkeiten



Wer überwacht die Auswirkungen auf die Gesellschaft?

Medien?
Gesellschaft?
Politik?
Institutionen?
Firmen?
Recht?

Gründung von „Algorithm Watch“



ALGORITHM
WATCH



Lorena Jaume-Palasi, Mitarbeiterin im iRights.Lab



Lorenz Matzat, Datenjournalist der 1. Stunde, Gründer von lokaler.de, Grimme-Preis-Träger



Matthias Spielkamp, Gründer von iRights.info, ebenfalls Grimme-Preis-Träger, Vorstandsmitglied von Reporter ohne Grenzen.



Prof. Dr. K.A. Zweig, Junior Fellow der Gesellschaft für Informatik, Digitaler Kopf 2014, TU Kaiserslautern

Topaktuell: Trending Topics bei Facebook



- Trending Topics: Themen, die auf Facebook gerade vermehrt Beachtung finden.
- Vorwürfe:
 - Mitarbeiter hätten diese Liste redigiert (vor allen Dingen am Anfang);
 - Facebook-Themen selbst werden (meistens) rausgenommen;
 - Gesellschaftlich wichtige, aber nicht „trending“ Themen wären zu hoch reingesetzt worden („Black Lives Matter“, Syrien);
 - Ebenso Themen wie der Absturz von MH 370 oder die Charlie Hebdo-Attacken;
 - Manche Themen hätten erst aufgenommen werden dürfen, nachdem sie von großen Verlagen aufgegriffen wurden.

Beipackzettel für Algorithmen



Welches Problem „kuriert“ der Algorithmus?

Was ist das Einsatzgebiet des Algorithmus, was seine Modellannahmen?

Welche „Nebenwirkungen“ hat der Algorithmus?

Schlussformel



... zu Risiken und Nebenwirkungen der Digitalisierung befragen Sie bitte Ihren nächstgelegenen Data Scientist oder den deutschen Algorithmen TÜV.

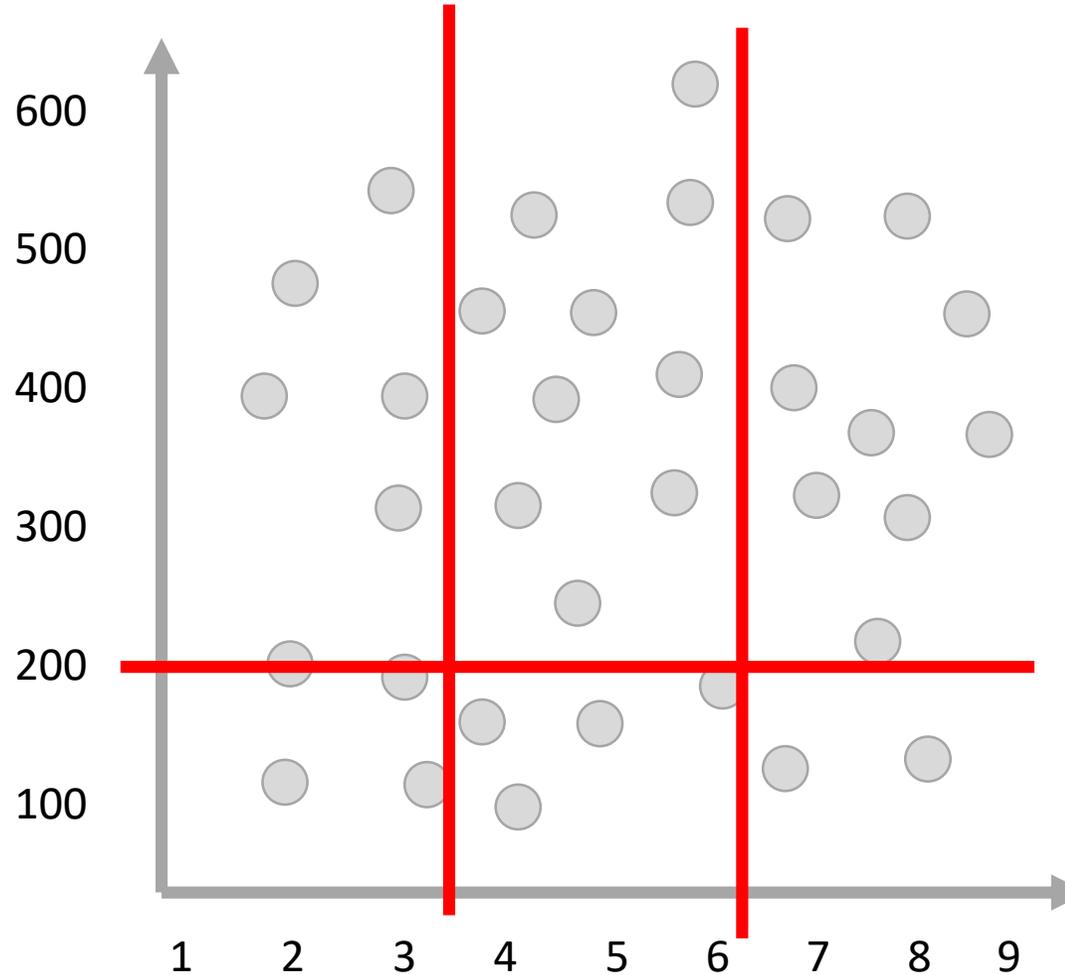


Entscheidungsbäume

Beispiel mit 2 Variablen



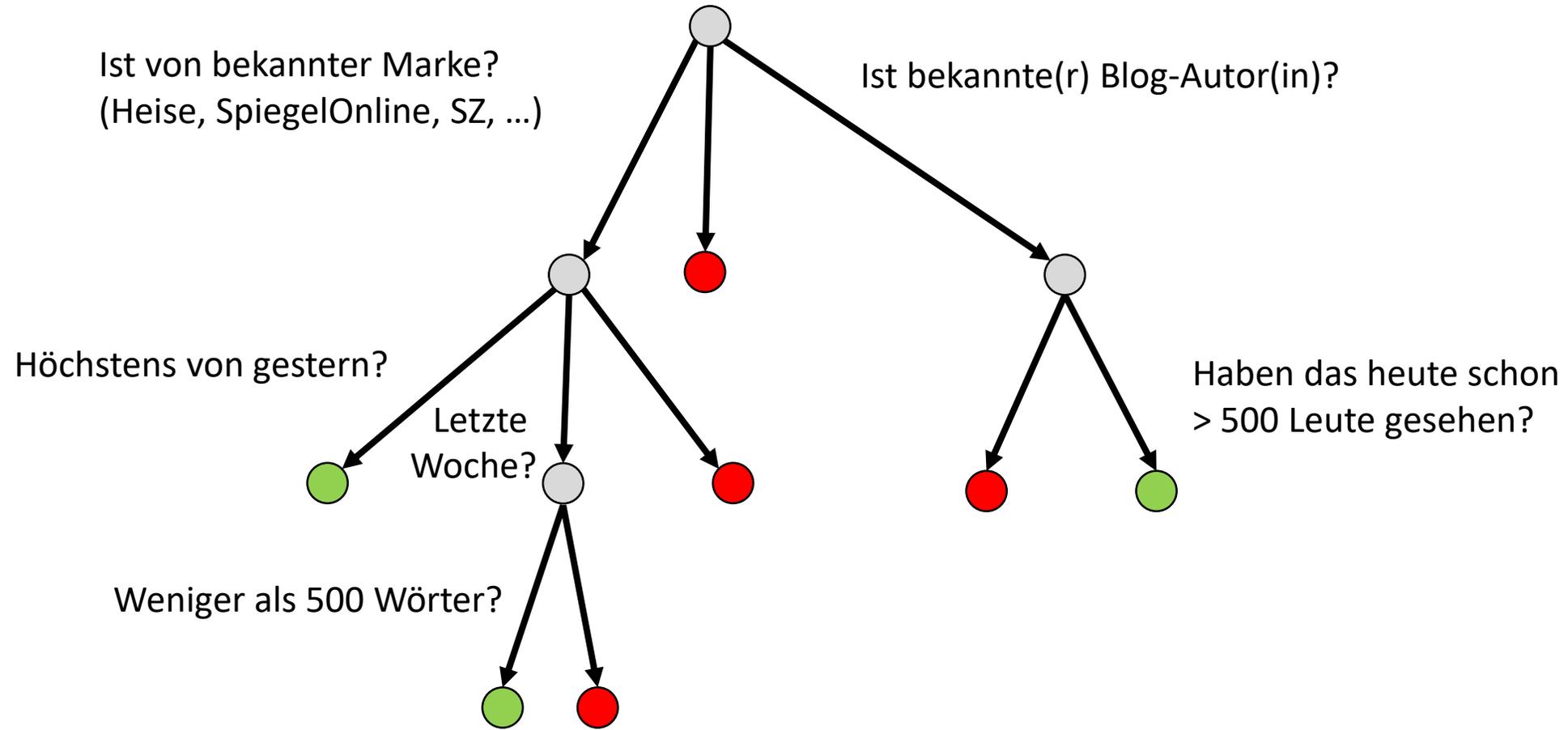
Länge des Textes [Worten]



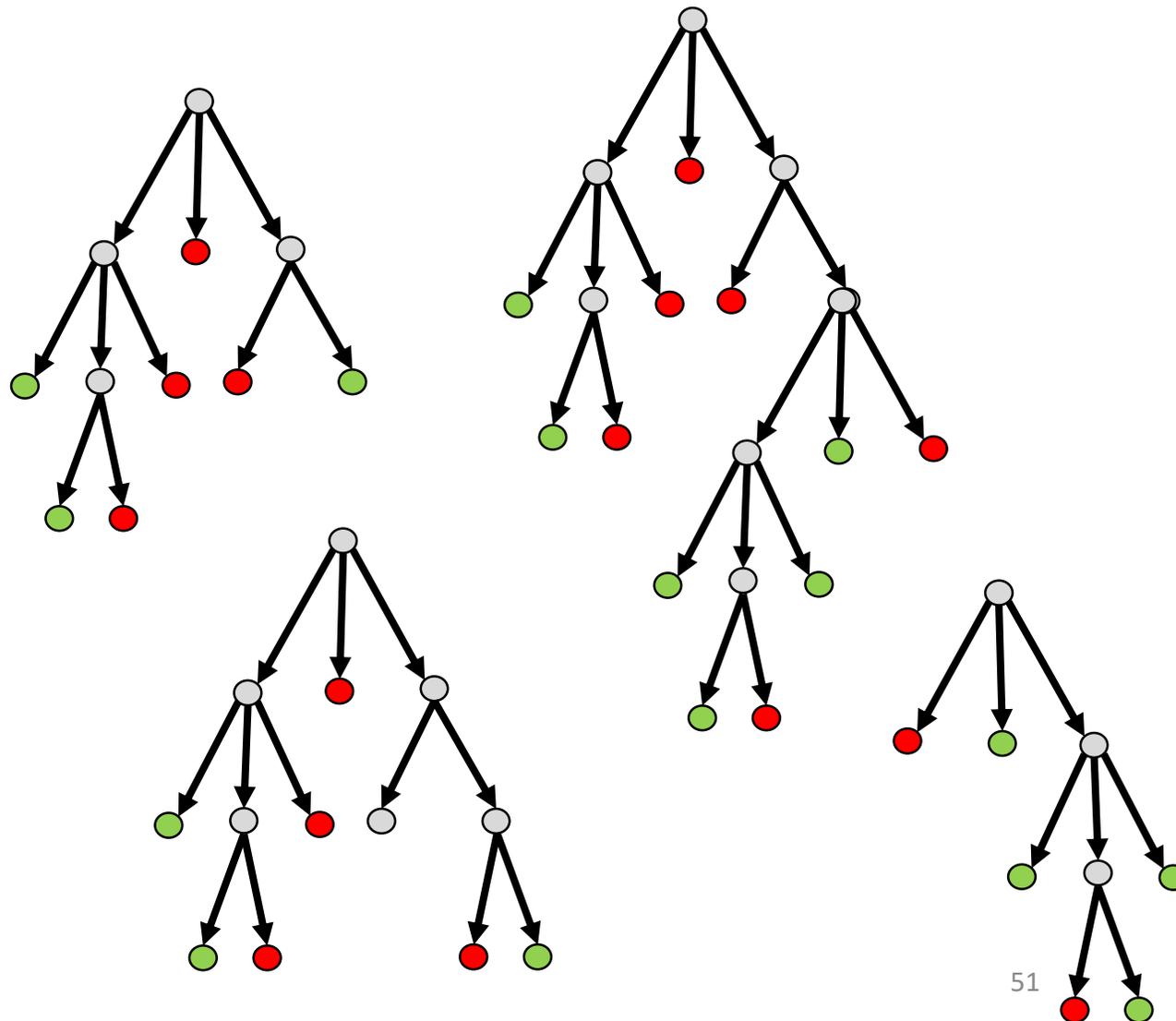
Aktualität

[Stunden seit Veröffentlichung]

Entscheidungsbaum



Entscheidungswald

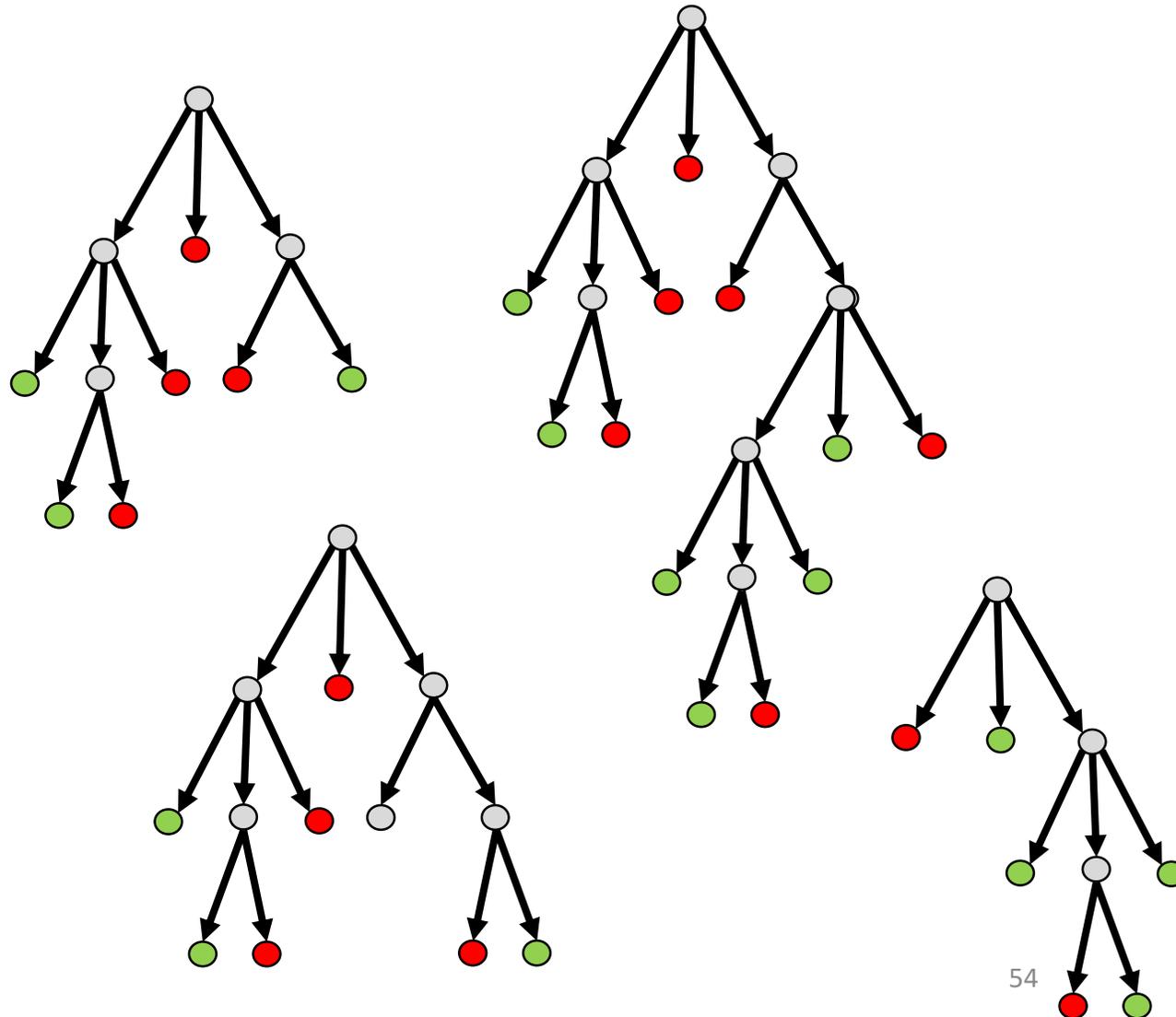


- Aus einem Datensatz werde Hunderte dieser Bäume gelernt:
 - Unterschiedliche Variablenauswahl
 - Andere Reihenfolgen der Variablen
- Bei neuen Daten (Webseiten) „laufen“ diese durch alle Bäume, jeder Baum sagt „ja“ oder „nein“
- Es kommt eine Vorhersage heraus: 330 von 1000 Bäumen sagen: „Dieser Link wird geklickt werden“



Personalisierung

Ihr persönlicher Entscheidungswald



Die maschinellen Möglichkeiten sind so groß, dass heute jede Nutzerin und jeder Nutzer seinen eigenen Entscheidungswald bekommen könnte.

Dies nennt man „personalisierte Algorithmen“