

## DER AUTOR



**Kai Spriestersbach** ist Online Strategy Consultant bei der eology GmbH sowie Inhaber von SEARCH ONE.

## DER AUTOR



**Tobias Aubele** ist Professor für E-Commerce, insbesondere Conversion-Optimierung und Usability an der Hochschule Würzburg-Schweinfurt. Darüber hinaus berät er KMU im Bereich Webanalytics & Website-Optimierung.

## DER AUTOR



**Mario Fischer** ist Professor für E-Commerce an der Hochschule Würzburg-Schweinfurt und Herausgeber der Website Boosting. Er beschäftigt sich u. a. seit 1996 intensiv mit dem Thema Suchmaschinen.

# GOOGLES NEU

Tobias Aubele, Kai Spriestersbach, Mario Fischer

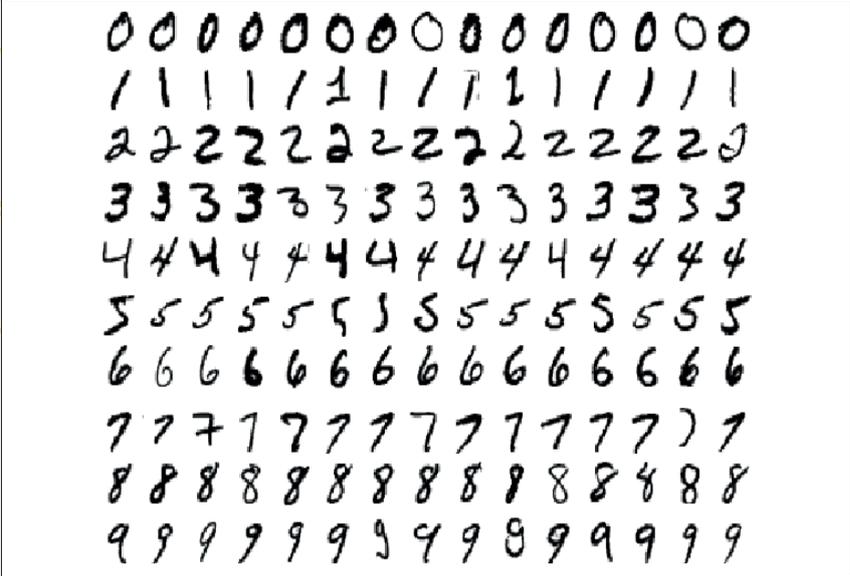
Am 26. Oktober letzten Jahres platzte eine kleine Bombe, als der Google-Mitarbeiter Greg Corrado in einem kurzen Interview mit Bloomberg erwähnte, man habe ein System mit künstlicher Intelligenz entwickelt, das intern den Namen „RankBrain“ trage und mit dem bereits seit Anfang 2015 verdeckte Tests gefahren werden. Er erklärte, man könne damit die ca. 15 % Suchanfragen, die zum ersten Mal als Suchbegriffe eingegeben werden, besser bearbeiten. Es sei, so Corrado, das dritt wichtigste Ranking-signal. Seither gibt es wenig offizielle Informationen dazu. RankBrain lässt aber nicht nur viele Fragen offen, sondern hat das Potenzial, die bisherigen Regeln der Suchmaschinenoptimierung künftig ordentlich durcheinanderzuwirbeln. Dies zu erkennen, gelingt allerdings nur, wenn man sich etwas tiefer mit dem Thema künstliche Intelligenz (AI, Artificial Intelligence), Machine Learning und Entitäten beschäftigt. Wenn man dann noch die am Horizont stehenden neuen Möglichkeiten der unvorstellbaren Rechenbeschleunigung durch Quantencomputer dazunimmt, könnte aus der kleinen Bombe ganz schnell eine richtig große werden.

Kai Spriestersbach, Tobias Aubele und Mario Fischer versuchen in dieser Ausgabe, für Sie etwas Licht in das verwirrend dunkle Grau zu bringen und auch einige Ecken besser und vor allem leicht verständlich auszuleuchten.

# RANK E AI-SUPERWAFFE?

# BRAIN

Die Organisation und Verarbeitung großer Datenmengen ist seit der Existenz der ersten Rechenmaschinen das wichtigste Aufgabenfeld von Computern. Insbesondere für Menschen aufwendige Rechenschritte können schnell und wiederholbar durchgeführt werden. Die ersten Computer wurden genau für diesen Zweck erdacht und konstruiert, doch die Fähigkeiten der maschinellen Datenverarbeitung waren stets begrenzt. Große Datenmengen und besonders rechenintensive Aufgaben konnten oft nicht bewältigt werden. Als schließlich die Prozessoren schneller, der Arbeitsspeicher größer und die Festplatten immer riesiger wurden, blieb



0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9

Abb. 1: Die ersten Ziffern aus dem MNIST-Datensatz von Y. LeCun; Quelle: <http://einfach.st/lecun>

## VERBREITETE ANWENDUNGSFÄLLE FÜR MASCHINELLES LERNEN

- » **Spracherkennung:** Egal ob Apples Siri, Googles Voice Search oder Microsofts Cortana – alle diese Dienste benutzen künstliche neuronale Netze.
- » **Bildererkennung:** Die massenhafte Anonymisierung von Nummernschildern, Gesichtern und die Erkennung von Hausnummern in allen Fotos von Google Streetview.
- » **Videoerkennung:** Jeff Dean brachte dem Google Brain mittels Videodaten von YouTube bei, Katzen zu erkennen. Der Algorithmus funktioniert so gut, dass er alle Katzen findet und nichts, was keine Katze ist.
- » **Selbstfahrende Autos:** Parallele Datenverarbeitung in Echtzeit der unterschiedlichen Sensoren der autonomen Fahrzeuge ist nur mithilfe neuronaler Netze möglich. Erst kürzlich stellte Nvidia einen eigenen Supercomputer für neuronale Netze auf der CES 2016 vor.
- » **Komponieren:** Neuronale Netze können einfach Gema-freie Musik in Echtzeit komponieren. Dienste wie JukeDeck ([www.jukeDeck.com](http://www.jukeDeck.com)) können sogar sekundengenau Musik für Videos oder Spiele erstellen, die einer bestimmten Stimmung und einem gewünschten Genre entspricht.
- » **Betrügererkennung:** Mit Machine Learning ist es einfach, Prognosemodelle aufzubauen, mit deren Hilfe potenzielle betrügerische Einzelhandelstransaktionen bzw. betrügerische oder unangemessene Artikelrezensionen erkannt werden können. Selbst über die Besonderheiten beim Eintippen falscher Adressdaten in Formulare können heute mit hoher Wahrscheinlichkeit Fakeanmeldungen erkannt werden.
- » **Personalisierung von Inhalten:** Mithilfe von Amazons Machine Learning (<http://einfach.st/amaz4>) kann eine Website eine besser personalisierte Kundenerfahrung bieten, indem prädiktive Analysemodelle verwendet werden, um Artikel vorzuschlagen oder den Website-Ablauf auf der Grundlage vorheriger Kundenaktionen zu optimieren.
- » **Dokumentenklassifizierung:** Die Verarbeitung unstrukturierter Texte ist eine besondere Stärke der KNN. So können beispielsweise Produktrezensionen als positiv, negativ oder neutral eingestuft werden oder bei Bedarf der Autor eines beliebigen Textes ermittelt werden.
- » **ChatBots und Inbox Autoresponder:** Für die Funktion des „Smart Reply“ in Googles Inbox-App (<https://www.google.com/inbox/>), welche erst kürzlich vorgestellt wurde, trainierte Google ein rückgekoppeltes neuronales Netz mit Millionen und Abermillionen von E-Mails und deren Antworten und kann nun zu jeder E-Mail passende Antwortmöglichkeiten vorschlagen. Schwierig bei der Entwicklung war laut Google der Fakt, dass die Entwickler aus Datenschutzgründen selbst keinen Zugriff auf die E-Mails haben durften und so quasi blind das System anlernen mussten.
- » **Autonome Finanzagenten:** Automatische Systeme lesen Unternehmensnachrichten und reagieren in Echtzeit auf gute Nachrichten mit Aktienkäufen oder auf schlechte Nachrichten mit Verkäufen. So konnte ein einzelner Bot 2,4 Millionen US-Dollar Gewinne erwirtschaften, indem er Sekunden nach der Veröffentlichung von Übernahmege-rüchten durch Intel Aktien der Firma Altera kaufte.

neuronalen Netze. Hier wurde dasselbe Verfahren wie im visuellen Kortex ausprobiert. Aus den gescannten Bildpunkten und dessen Grauwerten sollten die Zahlen von 0 bis 9 erkannt werden.

Wenn man sich die unterschiedlichen Schreibweisen anschaut, die Menschen verwenden (Abb. 1), eckige Zweien, runde Zweien etc., wird schnell klar, dass es sich dabei um ein sehr komplexes Problem handelt. Bei 28 x 28 Bildpunkten für eine gescannte Ziffer mit jeweils 256 Graustufen ergibt sich somit eine komplexe Datenreduktion von 200.704 bit auf nur noch zehn Werte.

Die Topologie eines Netzes, also die Zuordnung von Verbindungen zu Knoten, muss abhängig von seiner Aufgabe gut durchdacht sein. Nach der Konstruktion eines Netzes folgt die Trainingsphase, in der das Netz „lernt“. Theoretisch kann ein Netz durch folgende Methoden lernen:

- » Entwicklung neuer Verbindungen, Löschen bestehender Verbindungen
- » Ändern der Gewichtung
- » Anpassen der Schwellenwerte der Neuronen
- » Hinzufügen oder Löschen von Neuronen

Leider waren diese Versuche damals nicht besonders erfolgreich, denn bis 2007 hatten alle neuronalen Lernverfahren ein gemeinsames Problem: Bis zu einer bestimmten gelernten Teilmenge funktionierte alles wunderbar, doch ab einer gewissen Schwelle sorgten neue Lerndaten dann plötzlich und abrupt für eine drastische Verschlechterung der Ergebnisse.

Das Hauptproblem damals: Die Rechenkapazität fehlte. Man kürzte bei der Berechnung der Schwellenwerte Nachkommastellen ab und wurde damit zu ungenau. Außerdem waren die Netze für die meisten Aufgaben nicht komplex genug gestaltbar.

ein Problem bestehen: die Verarbeitung unstrukturierter Daten.

Während Entwickler sich bei der Verarbeitung strukturierter Daten, beispielsweise aus einer relationalen Datenbank, auf das Vorhandensein gewisser Strukturen und Regelmäßigkeiten verlassen können, ist die Verarbeitung unstrukturierter Daten eine vollkommen andere Herausforderung. Hier kommt der bisherige Ansatz, detaillierte Regeln, Anweisungen und Ausnahmen im Vorfeld zu definieren und den Computer diese „nur noch“ ausführen zu lassen, an seine Grenzen. Ein vollkommen neuer Ansatz wurde notwendig, denn „Big Data“ und „Real-time-Verarbeitung“ sind aus modernen

Anwendungen nicht mehr wegzudenken. Der Ansatz, mit dem viele dieser Probleme gelöst werden konnten, ist das „Machine Learning“, also das maschinelle Lernen.

### Neuronale Netze in den 1980er-Jahren

In den 1980er-Jahren, als die ersten Computer auf die Schreibtische gelangten, hatte man schließlich genug Rechenleistung verfügbar und versuchte, Handschriften und Unterschriften, die vorher gescannt wurden, mithilfe der Rechner zu erkennen. Dabei wurde eine Vielzahl von Verfahren ausprobiert und etwa 1985 gab es den ersten großen Run auf die künstlichen



Abb. 2: Googles Antwort auf eine spezifische Frage

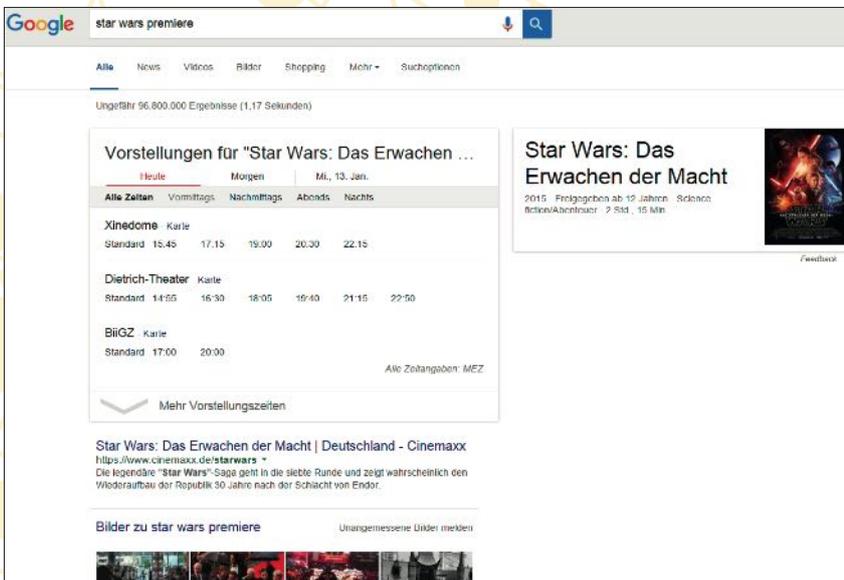


Abb. 3: Lokale und aktuelle Ereignisse einer Suchphrase



Abb. 4: Google Suggest mit Anzeige eigener Entitäten (hier Google.com)

**INFO**

„Machine Learning“ oder zu Deutsch maschinelles Lernen ist der Oberbegriff für die „künstliche“ Generierung von Wissen aus Erfahrung. Ein künstliches System lernt aus Beispielen und kann nach Beendigung der Lernphase verallgemeinern. Das Besondere daran ist, dass nicht einfach Beispiele auswendig gelernt werden, sondern das System quasi selbstständig Gesetzmäßigkeiten in den Lerndaten „erkennt“. Anschließend kann das System dann auch bislang noch unbekannte Daten beurteilen. Beim maschinellen Lernen bringt sich der Computer also selbst bei, etwas zu tun, anstatt von einem Menschen angeleitet zu werden oder stumpf dessen Anweisungen zu befolgen.

**Eine neuronale Revolution? Warum erst jetzt?**

Der große Durchbruch gelang 2007 dem britischen Professor Geoffrey Hinton an der Universität Toronto. Er nahm sich des alten Problems der neuronalen Netze an und löste es. Das von ihm eingesetzte Backpropagation-Lernverfahren mit mehreren Netzwerken (Back Propagation through time) gab es zwar bereits und es wurde zum Beginn der 1980-Jahre auch mathematisch gut verstanden, aber das Zusammensetzen dieser Funktionsprinzipien zu etwas Funktionierendem bekam erst Geoffrey Hinton so richtig in den Griff.

So konnte er auf einen Schlag gigantische Verbesserungen erzielen. Seine von ein paar Studenten zusammengebauten und angeleiteteten Netze waren auf Anhieb besser als alles, was die Support-Vektor-Maschinen nach 30-jähriger Optimierung zu leisten imstande waren. Sein Verfahren, das sog. „Deep Learning“ auf Basis künstlicher neuronaler Netze, ist so viel besser als alles andere und dabei auch noch so viel einfacher, dass es heute für die meisten Bereiche des maschinellen Lernens verwendet wird.

**State of the Art – Deep Learning**

Hinton setzte bei der Bilderkennung beispielsweise künstliche neuronale Netze (KNN) mit mehreren Hundert oder Tausend Ebenen ein, die nach oben hin aus immer weniger Neuronen bestehen.

Diese Netze können so von der komplexen Eingangsschicht nach oben zur einfachen Ausgangsschicht immer besser abstrahieren.

Die erste Schicht erkennt nur einzelne Bildpunkte und deren Farbe, die Ebene darüber verbindet Pixel (zum Beispiel zwei schwarze Bildpunkte zu einem Strich). Darüber folgen Schichten, die aus den Strichen Objekte kombinieren, und so durchläuft die Information Schicht für Schicht nach oben bis zur Ausgabe von Gesichtern oder Gegenständen.

Der Vorteil dieser Methode: Die Antwort erfolgt rasend schnell. Aber das Training dafür dauert relativ lange und man benötigt viele Trainingsdaten. Die bislang verwendeten Support-Vektor-Maschinen sind im Ergebnis fast gleich gut, aber man muss vorher die wichtigen und relevanten Teile der Daten herausuchen. Es muss manuell festgelegt werden, was von Bedeutung ist und was nicht. Das sogenannte Feature Engineering war letztlich auch die klassische Herangehensweise Googles an die Konstruktion von Algorithmen. Doch Deep Learning ermöglicht eine vollkommen neue Art des Trainings.

**Unsupervised Deep Learning ist eine echte Revolution**

Hier müssen nun nicht mehr Menschen das Problem vorher genau analysieren und zerlegen, sondern es reicht, einfach nur die Daten und das gewünschte Ergebnis in das Netzwerk zu geben und der Computer löst das Problem selbst. Neuronale Netzwerke erkennen und lernen jetzt von selbst, welche Features wichtig sind und welche nicht. Und das automatisch und ohne menschliche Eingriffe!

Wie gut diese Methode mittlerweile funktioniert, lässt sich auch an Stanfords „Natural Language Parser“ deutlich machen. Dieser wurde 25 Jahre lang von den besten Linguisten entwickelt, ständig verbessert und erweitert.

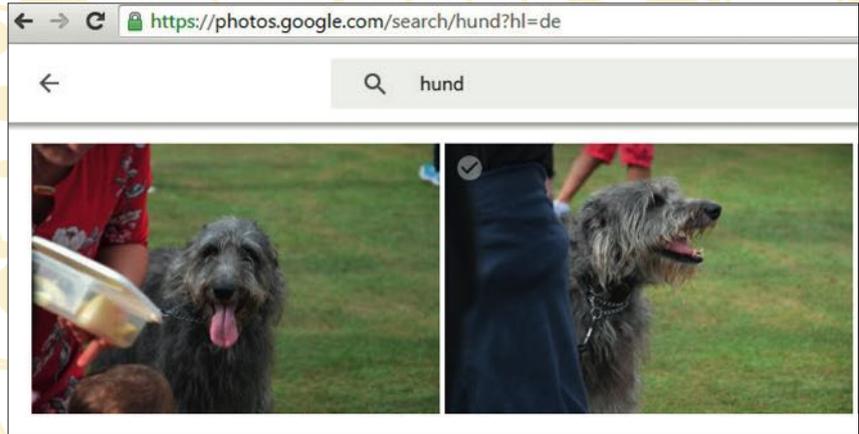


Abb. 5: Unbekannte Bilder können trotz fehlender Metainformationen dank Machine Learning inhaltlich korrekt erkannt werden

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Abb. 6: Beziehungen und Logik von Wörtern (Quelle: Google Research <http://einfach.st/googres6>)

In der neuen Version wurde der alte Code vollständig durch ein neuronales Netzwerk ersetzt. Ein selbstlernendes Netzwerk muss nichts über Grammatik oder Sprache wissen und funktioniert somit auch in jeder beliebigen Sprache. Man braucht nicht einmal mehr Spezialisten oder eigene lokale Teams. Es reicht aus, eine große Textmenge in das Netzwerk zu geben und es daraus lernen zu lassen. Menschliche Sprache folgt bestimmten Regeln und bildet daher bestimmte Strukturen aus. Diese müssen sich dann von ganz allein in den verschiedenen Ebenen des neuronalen Netzes abbilden.

**Google und Geoffrey Hinton**

Google wurde 2007 auf Geoffrey Hinton und seine Erkenntnisse aufmerksam und lud ihn zu einer Vorlesung ein, um seinen Durchbruch den Google-Mitarbeitern nahezubringen. Diese erkannten sofort das Potenzial, das diese Technologie für die eigenen Dienste

bereithielt, und so versuchte man mehrfach erfolglos, Hinton an Bord zu holen. Erst als man dem Professor schließlich Zugriff auf Googles Daten und nahezu unbegrenzte Rechenpower in Aussicht stellte, konnte dieser zumindest als Teilzeitmitarbeiter gewonnen werden.

**Google lernt tatsächlich, menschliche Aktionen zu verstehen**

Google schickt seinen Algorithmus zur Schule. Was vielleicht etwas flapsig klingt, wird in den letzten Monaten immer mehr zur Realität. Die Maschine versucht, eine Suchanfrage inhaltlich zu verstehen und weniger ein Keyword mit dem invertierten Index einer Datenbank abzugleichen. Ein Text wird quasi in ein mathematisches Geflecht von Vektoren umgewandelt und Beziehungen der Worte über Winkel und Entfernungen zueinander berechnet. Die Antwort auf die Frage „Bekanntester Film von George Lucas?“ ist demnach keine Auf-

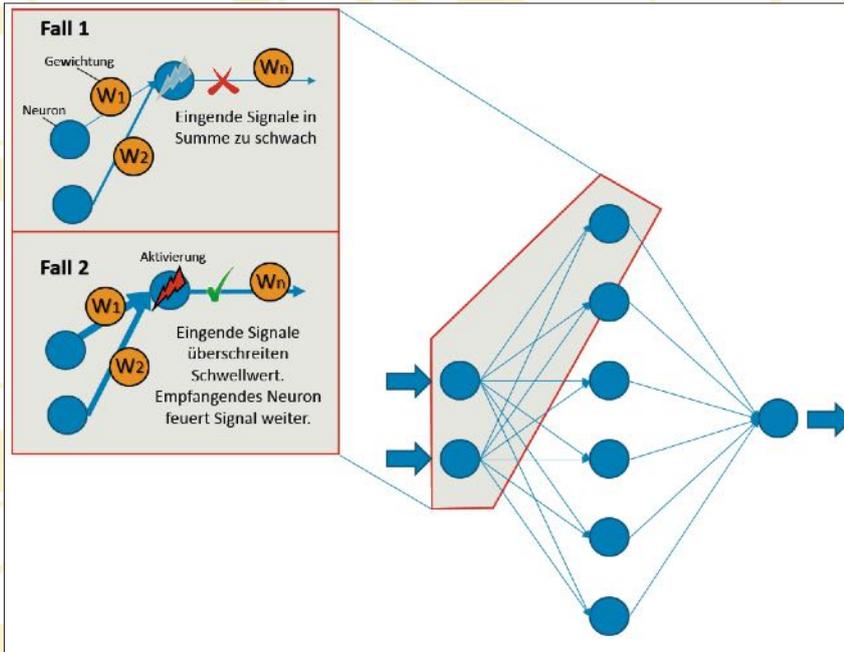


Abb. 7: Prinzipielle Funktionsweise eines neuronalen Netzes (stark vereinfacht)



Abb. 8: Was Hilfesuchende bei Google eingeben

listung persönlicher Informationen zu George Lucas oder genereller Dokumente zu generell bekannten Filmen, sondern eindeutig: „Star Wars.“ Das Zeitalter der Optimierung auf einzelne Keywords dürfte damit spätestens seit Hummingbird Geschichte sein.

Darüber hinaus wird auf Lokalität und Aktualität geachtet. Eine Suche nach der Star-Wars-Premiere liefert folgerichtig die Spielzeiten des nächstliegenden Kinos des Suchenden sowie brandaktuelle Kritiken, Bilder und Trailer zum aktuellen Film. Die Premiere bzw. Kritiken der letzten Episode bzw. Spielzeiten der Premieren in Berlin sind höchstwahrscheinlich für die breite Masse nicht relevant. Google denkt langsam, aber sicher mit – selbstverständlich auch zum Selbstzweck, dass die

Menschen wegen der ständig besseren Ergebnisse Google anderen Suchmaschinen vorziehen.

Bereits bei Google Suggest wird sofort versucht, den Besucher schnellstmöglich auf die richtige Zielseite zu bringen. Mehrdeutige Begriffe, was für Google aus technischer Sicht eigene Entitäten sind, werden explizit vorgeschlagen (Abb. 4).

### Warum setzt Google maschinelles Lernen bzw. RankBrain ein?

Google befasst sich bereits seit vielen Jahren mit Machine Learning und der Anwendung im Text- und Sprachverständnis (bspw. Word2vec als Open Source zum maschinellen Erlernen von Wortbedeutungen <http://einfach.st/>

gmach2). Es war abzusehen, dass ein weiteres System, welches auf maschinellem Lernen fußt – RankBrain – im Oktober 2015 nun auch offiziell bestätigt wurde. Panda basiert ebenfalls auf maschinellem Lernen und erlebt in verschiedenen Updates eine permanente Weiterentwicklung, vergleichbar mit einem Grundschüler, der geistig beständig in Richtung Gymnasium geht – unpräzises Wissen wird zu Detailwissen. Anwendungen wie Google Now bzw. die Suche in den eigenen Bildern bei Google Fotos zeigt aktuelle Ansätze dieser Lernverfahren.

Basierend auf umfassend vorliegenden Rohdaten (Bild, Text, Sprache, Videos etc.) müssen zur Effektivitäts- und Effizienzsteigerung Lernverfahren eingesetzt werden, welche den Algorithmus permanent weiterentwickeln – im Idealfall ohne die Notwendigkeit eines Inputs über maßgebliche Bewertungsverfahren. Jeff Dean von Google bestätigte bereits 2014, dass nicht für jede Aufgabe ein eigener Algorithmus eingesetzt werden kann bzw. der Entwicklungsaufwand minimiert werden soll. Ein „Reagieren“ durch einen von Menschen entwickelten Algorithmus ist logischerweise langsamer und aufwendiger als ein automatisiertes, autarkes Agieren auf Basis sich ändernder Muster. Spam kann dadurch schneller bekämpft werden bzw. relevante Seiten können besser identifiziert werden. Schnelle Entwicklungsergebnisse der Programmierer bis hin zu Realtime-Modifikation wären damit langfristig denkbar. Ein Tipp: Die Präsentation von Jeff Dean zu Deep Learning unter <http://einfach.st/deep3> ist wirklich sehenswert.

Beim Lernen geht es im Grunde immer um Beziehungen. Die Verbindung zwischen den Wörtern ist notwendige Voraussetzung für das Verständnis des Inhalts. Unbekanntes kann anschließend durch Analogien erschlossen werden. Mittels mathematischer Verfahren (Vektoren) können Wörter im n-dimen-

sionalen Raum abgebildet werden. Dieses theoretische Konstrukt erlaubt, mittels Beziehungen bzw. Logik Wissen und Analogien zu generieren und abzurufen. Basierend auf den Beziehungen in Abb. 6 ergibt sich bspw. Paris – France + Italy = Rome. Das System wird durch weitere Informationen permanent intelligenter und kann letztendlich bei Unbekanntem einen „educated guess“ abliefern. Apropos maschinelle Intelligenz: Das System lernt, was es als Input erhält, und so kann auch die Bratwurst typischerweise Germany zugeordnet werden.

### RankBrain – Googles Antwort für Neues und Vages

Schon immer wurden seitens Google große Anstrengungen unternommen, den bestehenden Content, mit all seinen Facetten, besser zu verstehen. Neben der Bilderkennung (inkl. Video) war dies insbesondere der Themenbereich Sprachen und Übersetzungen. Die Möglichkeiten der künstlichen Intelligenz wurden dabei sukzessive weiterentwickelt und finden nun auch im Bereich der Suche Widerhall. In einem Bloomberg-Video (<http://einfach.st/bloom3>) bestätigte ein Google-Mitarbeiter (Greg Corrado, Senior Research Scientist), dass „ein großer Anteil“ der Suchen von einem System, basierend auf künstlicher Intelligenz mit Namen RankBrain, interpretiert und damit beeinflusst werden. Die Maschine, bzw. dieser neue Teil im Google-Algorithmus, kann bei neuartigen oder unzureichenden Informationen in einer Suchphrase eine Vermutung über deren Bedeutung anstellen. Im Grunde kann das System auf ein riesiges Textarchiv (exakter: Entitäten mit mehreren Schichten) zurückgreifen, die in logischer Verbindung zueinander stehen. Sie agieren wie Menschen, die sich auf den Erfahrungsschatz bzw. gelernte Muster berufen.

Genauer gesagt erfolgen Aktivitäten der Neuronen im Gehirn, welche konti-

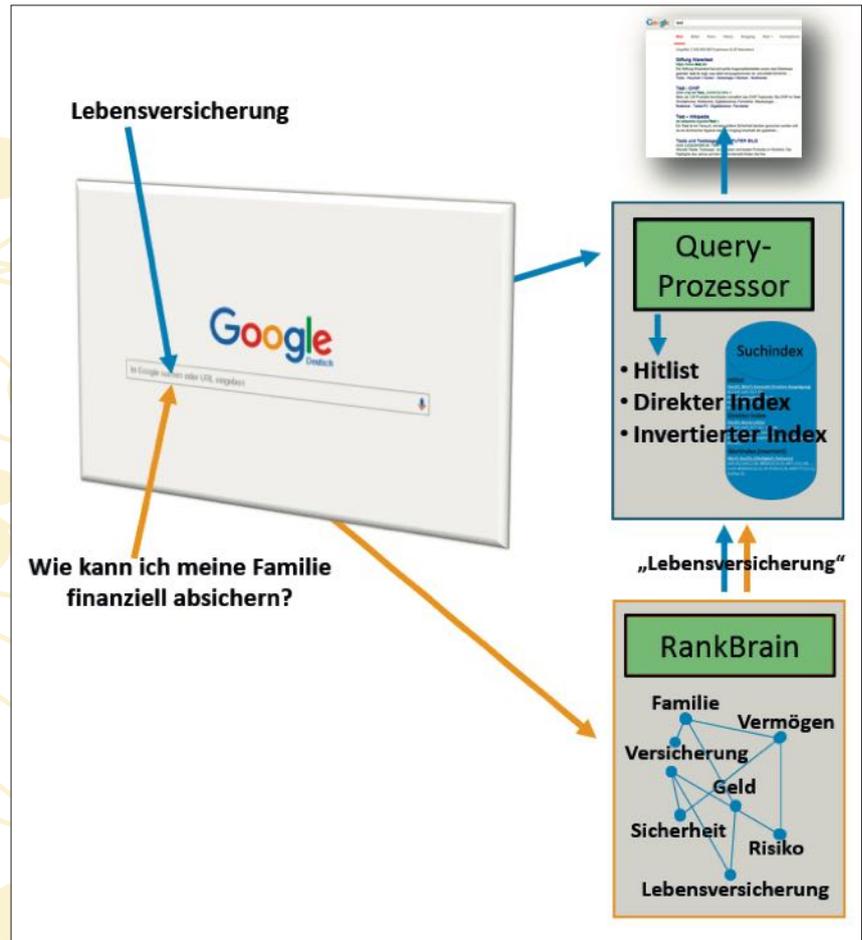


Abb. 9: RankBrain „übersetzt“ bestimmte Suchanfragen über Entitäten und deren Beziehungen untereinander und gibt sie dann erst an den Query-Prozessor weiter (stark vereinfachte Darstellung)

nuierlich neue Verbindungen zueinander erzeugen und dabei einzelne Bahnen verstärken (siehe im Vergleich Abb. 7).

**15 % der täglichen Suchphrasen hat Google vorher noch nie gesehen!**

Warum das Ganze? Google hat enorm viele Nutzerdaten und kann ein Suchergebnis anhand dieser sehr gut beurteilen. Ja, aber 15 % der täglichen Suchphrasen sind selbst für Google komplett neu (was schätzungsweise 450 Millionen Suchen pro Tag sind). Schaut man sich die Fragen an, die über die Web-analysesysteme noch angezeigt werden, ist das auch nicht weiter verwunderlich. Menschen tippen sehr lange Fragen ein, sind nicht immer der Rechtschreibung mächtig, nutzen Abkürzungen bzw. regionale Begriffe, schaffen immer wieder neue Wortschöpfungen etc. Google Suggest liefert heute schon eine Idee, was Menschen alles bei Google eingeben bzw. erwarten (Abb. 8).

### PageRank: Linksignale – RankBrain: Contentssignale?

Obwohl zu RankBrain seitens Google bisher nichts veröffentlicht wurde, lassen Patente (siehe z. B. <http://einfach.st/gpat9>), Expertenmeinungen und generelle Funktionsweisen von maschinellem Lernen im Information-Retrieval-Umfeld einige begründete Vermutungen zu.

Ähnlich wie George Clooney wird Google gezwungen, einen „educated guess“ über die eigentliche Nutzungsintention des Besuchers zu liefern. Die Vermutung über den Kern der Frage wird anhand von Analogien in der Sprache auf Wort- bzw. Phrasenebene berechnet (mittels der Vektoren über diverse Ebenen) und beeinflusst damit mittelbar das Suchergebnis massiv. Überwiegend bei Longtail-Suchbegriffen soll RankBrain zunächst großen Nutzen bringen. Damit werden Menschen, die mobil suchen und weniger Zeit für aktives Suchen haben,

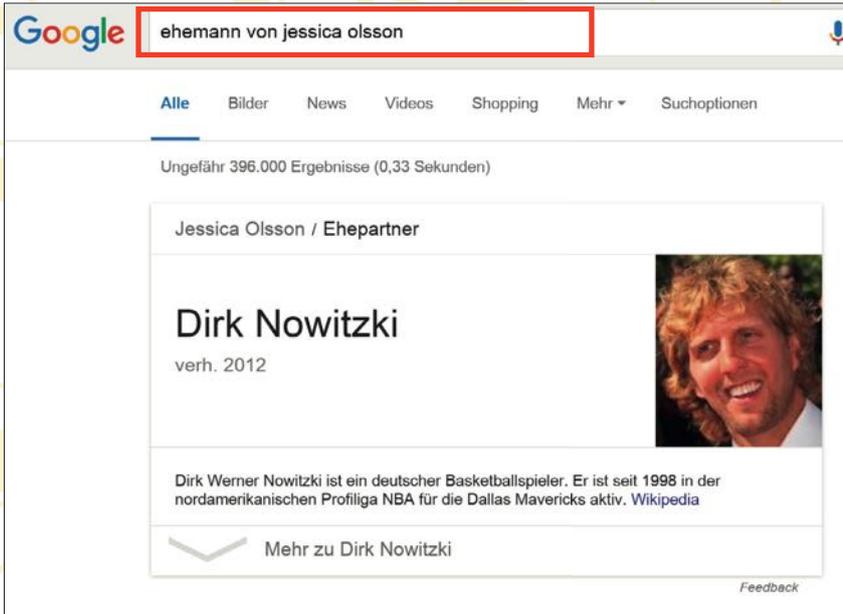


Abb. 10: Beispielhafte Darstellung einer inhaltlich verstandenen Frage



Abb. 11: Diese Antwort zeigt, dass Entitäten, nicht Worte erkannt bzw. verglichen werden

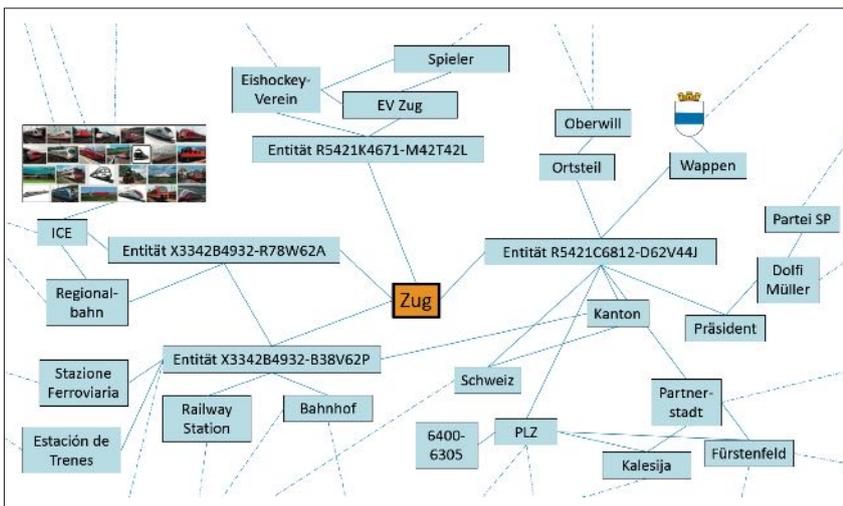


Abb. 12: Entitäten sind über Beziehungsgraphen miteinander verbunden (schematische Darstellung)

langfristig mit besseren Ergebnissen versorgt und dann auch in Zukunft weiterhin Google nutzen. Nachdem die mobilen Suchen 2015 die stationären Suchen übertrafen, ist das ein Muss für eine Suchmaschine: bessere Ergebnisse in einer mobilen Umgebung unter einem geänderten Suchverhalten bzw. Informationskonsumverhalten. Facebook setzt für den gefilterten Newsfeed ebenfalls künstliche Intelligenz ein, da ansonsten ein durchschnittlicher Facebook-Nutzer 1.500 Meldungen im Feed hätte, bei mehreren Hundert Freunden bis zu 10.000.

### RankBrain als „Vermittler“ neuartiger Fragen

Der Unterschied dieser Weiterentwicklung zum bestehenden Algorithmus ist immens. War das bisherige Suchergebnis abhängig vom Datenbankwissen bzw. der Erfahrung/den Erkenntnissen der Entwickler, lernt das Subsystem RankBrain ständig hinzu und erzeugt kontinuierlich neue Querverbindungen und Analogien innerhalb der Daten – gerade für unbekannte Suchanfragen ist dies unerlässlich. Diese Querverbindungen schaffen es somit, die Suchphrase besser zu deuten und damit die Datenbankabfragen zielgerichteter auszuführen. Ein Teil der Entwicklungintelligenz wird direkt auf die Maschinen verlagert. RankBrain kann als Bindeglied verstanden werden, welches die Sprache der Besucher mit der im invertierten Index der Suchmaschine in Einklang bringt. Bevor die eigentliche Suchphrase an die Datenbank weitergereicht wird, wird sie damit in eine „passende“ Sprache über-

#### INFO

RankBrain ist streng genommen im klassischen Verständnis eigentlich gar kein Rankingsignal, obwohl Greg Corrado es als solches bezeichnete. Es versucht, bestimmte eingetippte Suchphrasen über den Rückgriff auf vorhandene Entitäten zu verstehen, und stellt also aktuell eine Art Übersetzer dar, der die modifizierten Begriffe dann an den sog. Query-Prozessor übergibt.

führt – der Kern der Suchphrase wird verstanden und das Ergebnis ist damit zwangsläufig besser. Die Suchphrase nach dem „Ehemann von Jessica Olsson“ hätte früher wahrscheinlich Dokumente zu Ehemann bzw. Jessica Olsson zurückgegeben. Heutzutage versteht die Maschine die Zusammenhänge und kann die richtige Antwort liefern. Für Menschen ist intuitiv klar, wie die „süße, gelbe, gebogene Frucht, die am Anfang noch ganz grün ist“ heißt, die Maschine muss dies mittels Vektoren mathematisch herleiten bzw. letztendlich im Laufe der Zeit durch Wort-Sinn-Verknüpfungen erlernen. Das heißt, das System muss permanent mit historischen Suchphrasen und entsprechenden Ergebnissen gefüttert werden – kombiniert mit Synonymen und Homonymen. Sofern RankBrain als vorgelagertes System den perfekten Match zwischen übersetzter Suchphrase und gespeicherten Termen der Dokumente schafft, ist das erhaltene Ergebnis für den Besucher phänomenal. RankBrain liefert vereinfacht gesagt die Suchphrase „Banane“ anstelle der komplexen Suche und bekommt damit hoffentlich die relevanten Dokumente, Bilder und Videos zur eigentlichen Frage. Durch künstliche Intelligenz hat RankBrain langfristig die Chance, die Intention mit hoher statistischer Wahrscheinlichkeit abzuleiten und damit an die Suchmaschine die „richtige“ Frage zu stellen. Hier steht Google noch am Anfang und muss tatsächlich noch viel lernen – so wie auch Dirk Nowitzki nicht mit 2,13 Metern Körpergröße zur Welt kam.

Sucht man – ohne die Lebensumstände zu kennen – nach der Frau von Guido Westerwelle, bringt Google korrekt seinen Ehemann Michael Mronz als Ergebnis (Abb. 11). Dies zeigt noch mal deutlich den Unterschied: Es werden nicht textuelle Einträge auf Webseiten gesucht, welche die Worte „Frau von“ und „Westerwelle“ enthalten. Die dürfte es vernünftigerweise nicht geben. Da

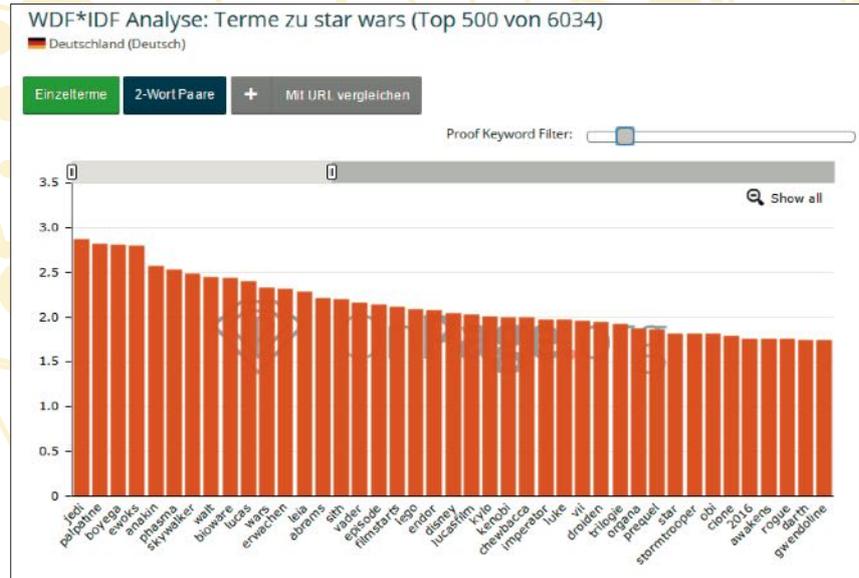


Abb. 13: Proof Keywords zur Verifizierung und Ausgangspunkt für weitere Aspekte (Quelle: OnPage.org)

Westerwelle als Persönlichkeit bekannt und sein Mann ein bekannter Sportmanager ist, haben beide mit Sicherheit bei Google „Entitäten“-Status, d. h., sie werden als eindeutige Objekte verwaltet. Über entsprechende Attribute einer „Lebenspartnerschaft“ sind beide als verbunden gekennzeichnet. Unter diesem Oberbegriff Lebenspartnerschaft können dann geschlechtsneutral sowohl Ehen als auch ein festes Zusammenleben subsumiert werden. Das bringt vielerlei Vorteile und erlaubt unter anderem auch 1:n- oder n:m-Beziehungen, was je nach Liberalität eines Landes und den Vorstellungen des Zusammenlebens einzelner Menschen alles erlaubt – ohne einschränkende Denkbarrieren.

Abb. 12 zeigt schematisch, welche Vorteile Entitäten und deren Verbindungen bieten. Das Wort „Zug“ kann mehrere Bedeutungen tragen. Zum einen kann es der Zug sein, in den man am Bahnhof einsteigt. Es könnte aber auch der Eishockeyverein EV Zug gemeint sein oder der Kanton Zug in der Schweiz. Durch die Zuordnung des Wortes Zug zu unterschiedlichen Entitäten lassen sich Gemeinsamkeiten, aber vor allem auch Unterschiede maschinell feststellen. Hinter einer eindeutigen Entität liegt also nicht der Zug generell, sondern quasi die jeweils unterschiedliche Bedeutung, der das Wort Zug „nur“

zugeordnet ist. Dabei können Worte sowohl als Attribut einer Entität zugeordnet sein, als auch selbst eine Entität sein.

Sieht man sich das Beziehungsgeflecht aus dem Eisenbahnumfeld (links unten in Abb. 12) an, wird deutlich, dass mit dieser Zuordnung die jeweilige Sprache in den Hintergrund tritt. So werden inhaltliche Erkennungen sogar sprachunabhängig ermöglicht oder bessere Übersetzungen, sofern die Datenbasis groß genug ist. Der „Kanton“ Zug hat auch einen Bahnhof, auf dem der „andere“ Zug eine wichtige Rolle spielt. Über Beziehungsgraphen, die ausweisen, wie weit Entitäten voneinander entfernt sind, lässt sich auch das maschinell auseinanderhalten – oder es lassen sich eben bei Bedarf gezielt einzelne Zusammenhänge herstellen. Wer sich schon mit sog. holistischer Contentoptimierung (Stichwort WDF\*IDF) beschäftigte, wird hier unschwer sofort Zusammenhänge erkennen können. Über die Verbindungen der Entitäten kann eine Maschine bei voller Datenbank natürlich sofort feststellen, welche Begriffsnähe das Thema einer Webseite zu wichtigen anderen Begriffen hat oder ob z. B. die textliche oder bildliche Erwähnung des Wortes „Zug“ in keinem besonderen Zusammenhang zu den wichtigsten Entitäten steht.

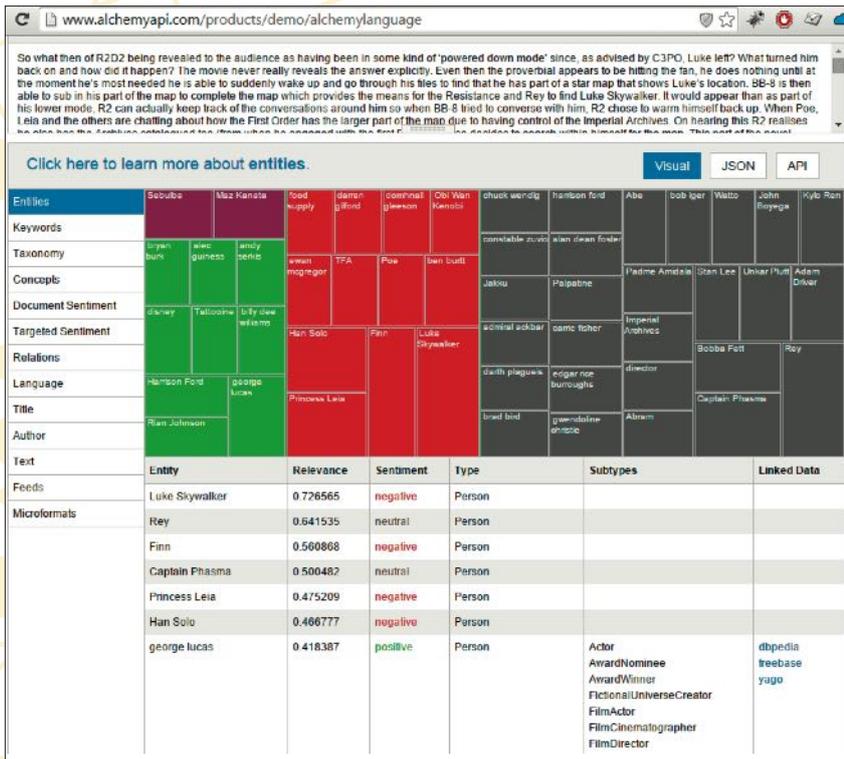


Abb. 14: Semantische Analyse einer Webseite mit Alchemy API

Ranking-„Fehler“ für eine Webseite, die sich mit dem Bahnhof im Kanton Zug beschäftigt, kann es nach dem Entitätenmodell nicht mehr geben. Die Seite taucht weder für „Kanton Zug“ noch für „Bahnhof“ in den Suchergebnissen auf. Wer nach dem Zug in Zug sucht, wird hingegen fündig, weil beide Entitäten (Ort Kanton Zug und über Bahnhof die Verbindung zu „Zug“) auf der Seite über die notwendigen Relationen identifizierbar sind.

### Quantencomputer machen künftig Notwendiges berechenbar

Natürlich ist einzusehen, dass ein kompletter Aufbau einer Entitäten-datenbank und vor allem die ständige Aktualisierung und Neuberechnung aller Beziehungen beim Auftreten neuer Entitäten derzeit noch nicht darstellbar sind. Dazu kämen noch das Umrechnen aller weit über 50.000 Suchanfragen pro Sekunde und der Abgleich mit allen infrage kommenden Entitätseinträgen und Beziehungen.

Die Lösung für dieses Problem liegt in der Nutzung von Quantencomputern.

Und so ist es kein Wunder, dass Google auch hier ganz vorn und in treibender Rolle mitspielt. Bereits vor einigen Jahren arbeitete man dort mit einem Expertenteam an der Entwicklung von Quantenalgorithmen. Die Programmiersprachen und das Paradigma dahinter (Zustände 0 oder 1) sind auf Quantencomputern nämlich gar nicht anwendbar. Letztere können theoretisch alle Zustände zwischen 0 und 1 gleichzeitig annehmen und daher auch schätzungsweise bei bestimmten Anwendungen bis zu 100 Mio. mal schneller rechnen. 2011 kaufte Google Deep Mind für 365 Mio. US\$; wie es aussieht, könnte sich das als wahres Schnäppchen erweisen. Deep Mind entwickelte u. a. Google Now mit. Aktuell arbeitet man dort an einer Art Kurzzeitgedächtnis, das dem des Menschen nachempfunden ist.

Seit letztem Jahr ist es nun so weit: Die Forscher, allen voran Hartmut Neven, Leiter des Quantum Artificial Intelligence Laboratory bei Google, verkündeten zusammen mit der NASA: Quantencomputer funktionieren! Ähnlich wie beim Gedankenspiel mit Schrödingers Katze, die gleichzeitig tot

und am Leben ist, können Quantencomputer auch mehrere sich überlagernde Zustände gleichzeitig berechnen. Damit können sie die reale Welt eben sehr viel besser abbilden als herkömmliche Computer, die nur mit zwei Zuständen arbeiten können. Solche „einfachen“ Zustände bilden die Natur nicht annähernd hinreichend ab, so Richard Feynman, einer der Väter der Quantenphysik. Neven bezeichnete die Quantenphysik sogar als Betriebssystem der Natur, was vielleicht für Laien etwas deutlicher macht, wie viel näher man sich mit dem neuen Computertyp an die Realität heranrechnen kann. Durch Quantencomputing kann man über neuronale Netze die Funktionsweise des menschlichen Gehirns besser abbilden. Dort feuern bei bestimmten Reizen ja mehrere variabel über Lernen verknüpfte Neuronen und ab einem bestimmten gelernten (!) Schwellwert wird das Signal eine Stufe weitergegeben. Dafür ist ein Quantencomputer geradezu geschaffen – alle diese Berechnungen bzw. eben auch Simulationen wirklich gleichzeitig ablaufen zu lassen. Herkömmliche Computer sind zwar schnell, können aber im Kern eben alles nur nacheinander rechnen.

Ein Baby muss niemand lehren, wie man „hört“ oder „sieht“ und Gegenstände erkennt. Das passiert ganz natürlich, indem sich im Gehirn Neuronen miteinander verbinden und durch Aktion – Reaktion diese Verbindungen verstärken oder wieder auflösen. Exakt dies passiert auch in den neuen neuronalen Netzen. Der Computer „lernt“ ohne vorherige Regeln. So schaffte Google es mit der DeepMind-Software z. B., dass Computer über solche neuronalen Netze von selbst lernten, wie die alten Atari-Spiele wie Breakout oder Space Invaders funktionieren. Die künstliche Intelligenz drehte einfach so lange (virtuell bzw. im Hauptspeicher) an den jeweils verfügbaren Reglern herum und maß die Reaktion des Spiels, bis sie alle Spiele fehler-

frei bis zum Ende durchspielen konnte. Wer diese Spiele kennt, der weiß, dass alle komplett anders aufgebaut sind und unterschiedlich funktionieren. Der Computer lernte bzw. erkundete selbstständig, wie die Regeln aufgebaut sind – ohne jede Vorgabe! Damit hätte man „a general-learning algorithm that should be applicable to many other tasks“, meinte Koray Kavukcuoglu von Google. Gibt es in unseren Köpfen verborgene Regeln, nach denen wir entscheiden, ob eine Webseite nützlich ist oder nicht? Wenn ja, wird man diesen in jedem Fall mit solchen Methoden deutlich näherkommen können.

Der erste Quantenchip selbst ist übrigens kaum daumennagelgroß und wurde von D-Wave entwickelt, die u. a. von Jeff Bezos (Amazon) und einer Tarnfirma der CIA finanziert werden. Und, ach ja – man stehe dabei noch ganz am Anfang, betonen die Forscher. Wer sich den Stand der Computer von vor 50 Jahren vergegenwärtigt und was seither passiert ist, kann vielleicht abschätzen, welche Explosion an Geschwindigkeit und Komplexitätsberechnung da auf uns zukommt. Und schon vor der Serienreife des ersten Quantencomputers kaufte sich Google die nötigen Spezialisten ein und arbeitet schon präventiv in Eigenregie an der Entwicklung noch besserer Quantencomputer.

### Fokussierung auf Themen und messbare Kompetenz

Marketers sollten sich also darauf einstellen, dass sich die Spielregeln in vielleicht gar nicht allzu ferner Zukunft ändern, denn das Zusammenspiel von Machine Learning mit neuronalen Netzen und Quantencomputern bringt wohl einen heute noch unvorstellbaren „Intelligenz“-Schub für Google & Co.

Websitebesucher suchen in der Regel so lange, bis sie ihr Informationsbedürfnis gestillt haben. Typischerweise tauchen während der Recherche bzw. des Konsumierens von Content weitere

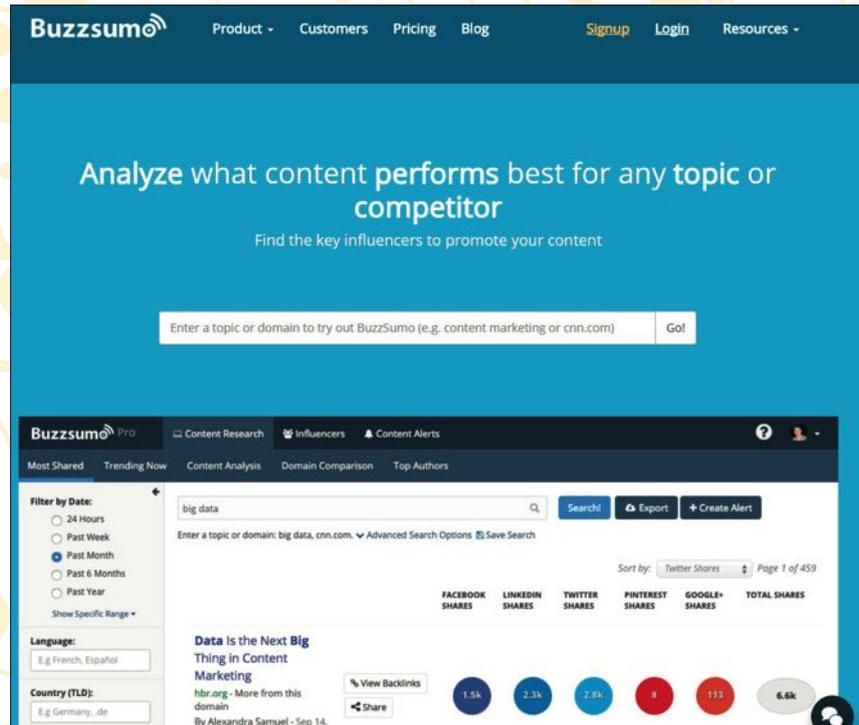


Abb. 15: Analyse sozialer Interaktionen

Fragen auf, die es ebenfalls zu beantworten gilt. Im Idealfall ist dabei kein weiteres „Googeln“ notwendig bzw. es sollte vermieden werden, dass durch fehlende Informationen ein Mitbewerber in das Relevant-Set des Besuchers kommt. Die abschließende Befriedigung einer Suchintention sollte durch den Content einer Webseite, am besten der eigenen, möglich sein. Wird nach „Star Wars“ gesucht, sollte in einem umfassenden Bericht auf jeden Fall „Jedi“, „Anakin“ oder „Darth Vader“ vorkommen. Das würden wir als Menschen erwarten und letztendlich auch die Maschine. Die eigene Webseite sollte daher immer zu einem Thema eine Autorität sein, so wie wir auch auf Aussagen von Ärzten als Autorität vertrauen. Welche Fragen werden Kunden im Themengebiet wohl als Nächstes stellen? Der Content bzw. die entsprechende interne Verlinkung sollte damit bereits passende Fragen und Antworten vorsehen, die der Suchende in der aktuellen Suchphase noch gar nicht stellt. Ein guter Arzt kennt die nächsten Fragen seiner Patienten, wenn sie eine Diagnose erhalten – so sollte auch Content im Web aufgebaut sein.

Brands haben durch RankBrain eine weitere große Chance: Sobald das System gelernt hat, dass bei bestimmten Fragetypen eine entsprechende Marke bzw. Website die perfekte Antwort war, könnte dies auch bei unbekanntem Fragen ggf. als Lösungsvorschlag in Betracht gezogen werden. Panda und Penguin bestrafen Webmaster für unzureichenden Content bzw. Linkspam. RankBrain versorgt im Gegensatz Webmaster von gutem Content mit hoch relevanten Suchanfragen bzw. Besuchern. In Summe wird dadurch ein höheres qualitatives Niveau der Suchergebnisse zu erwarten sein: Die Suchphrase wird inhaltlich verstanden, auf dem spambereinigten qualitativ hochwertigen Content wird eine entsprechend zugeschnittene Suchphrase platziert, bevor letztlich für das individuelle Ranking relevante Filter (lokal, aktuell, mobil etc.) angewandt werden. Backlinks sollte sich der Content verdienen, sie sollten nicht manuell von Linkbuildern gesetzt werden: Qualität erntet Link statt Link übertüncht Qualitätsmanko.

Daher könnte bspw. verteilter Content thematisch aggregiert werden bzw. bestehender Content auf Erweiterung

bzw. Themenrelevanz überprüft werden (W-Fragen, Proof Keywords, verwandte Suchanfragen). Vor dem Hintergrund der Nutzersignale wäre es ideal, wenn die eigene Webseite durch die Besucher später nochmals besucht würde: „Unter allen Seiten ist diese die beste.“ Denkbar ist auch eine tiefere algorithmische Anpassung der nachfolgenden Suchergebnisse, basierend auf den besuchten Dokumenten, da sich bspw. Dokumente gegenseitig ergänzen. Microdata bzw. Schema.org ist damit immer noch hilfreich und sollte auf jeden Fall auf der To-do-Liste eines SEOs stehen.

Tipp: Alchemy API analysiert eine URL und liefert potenzielle Entitäten innerhalb des Dokuments. Die Links zu dbpedia eröffnen weiterführenden Content zum Themenbereich bzw. zur Entität und sind damit ein idealer Ausgangspunkt für die Erweiterung bzw. Fokussierung des Themas.

### Nutzersignale als Elementarfaktor für maschinelles Lernen

Sofern ein Kind auf eine heiße Herdplatte fasst, wird es die Hand schnellstmöglich zurückziehen und im Idealfall daraus lernen. Aus einem Input bzw. Impuls ist eine Reaktion (gut/schlecht) erkennbar und demnach auch für die Zukunft prognostizierbar. Überlegt das Kind, nochmals auf die Platte zu fassen, wird die Reaktion klar sein. Diese „Schmerzen“ durchläuft sinnbildlich auch eine Maschine: Content wird durch Algorithmus bewertet und dem Nutzer ausgespielt. Wenn ein Suchergebnis für die entsprechende Position in der SERP eine höhere als die erwartete CTR hat und der Besucher anschließend eine entsprechende Verweildauer aufzeigt, wird er ein positives Nutzererlebnis haben. Eine Belohnung in Form konstanten Rankings bzw. einer Steigerung ist wahrscheinlich.

Anderenfalls wird der Besucher sich bildhaft gesprochen am „Content die

Finger verbrannt haben“ und schnellstmöglich ein anderes Suchergebnis ansteuern (Bounce/Return to SERP). Die Maschine kann permanent in den gesammelten Daten bzw. dem ausgespielten Content nach Mustern suchen, die auf ein voraussichtlich gutes bzw. schlechtes Ergebnis hinweisen (das Ergebnis aus Nutzersicht liegt vor). Die manuellen Einschätzungen durch Quality Rater helfen, guten von schlechtem Content zu unterscheiden (siehe aktuelle Rating Guidelines als PDF unter <http://einfach.st/goinside3>). Weitere Tests mit dem bestehenden bzw. ähnlichen Content (Near Duplicate Content) sind dann letztendlich nicht notwendig, um das voraussichtliche Nutzerverhalten abzuleiten. Der Content muss demnach seine Rankings kontinuierlich behaupten, indem er die Nutzererwartungen trifft bzw. im Idealfall übertrifft. Inhaltliche Erweiterungen bzw. Aktualisierungen des Contents sind damit ein notwendiger Bestandteil einer Rankingsicherung (siehe bspw. Artikel zu Content-Audit in WSB #35).

Tipp: Obwohl Social Media lt. Google nicht zum Ranking herangezogen werden, lohnt ein Blick in Tools wie Buzzsumo oder Talkwalker, um aktuelle Themen bzw. meistgeteilte Themen der Mitbewerber zu analysieren. Was treibt Nutzer um?

Ein Blick über den Tellerrand hinaus zeigt jedoch, welche Möglichkeiten sich langfristig abzeichnen könnten – nicht 2016, aber durchaus in naher Zukunft.

Die Rankingalgorithmen werden dynamischer und beliebig komplex

Die Google Engineers können zum Teil schon aktuell nicht mehr eindeutig voraussagen, wie welches der vielen Signale sich in Kombination mit Filtern und anderen Signalen tatsächlich auf das Ranking einer Webseite auswirkt. Die Komplexität ist dabei so hoch, dass ein einziger Mensch sie in Summe nicht mehr gedanklich nachvollziehen kann. Beim Deep Learning wird sich dies

# STATEMENT

## HANNS KRONENBERG

HEAD OF PRODUCT MANAGEMENT SISTRIX



**Hanns studierte in Münster Betriebswirtschaft mit den Schwerpunkten Marketing und Statistik. Er beschäftigt sich seit 15 Jahren mit dem Thema Suchmaschinen. Vor seinem Wechsel zu SISTRIX arbeitete Hanns als selbstständiger SEO-Unternehmensberater für mehrere Kunden in den SISTRIX-Top-100.**

Mit RankBrain steht Google ein weiteres Werkzeug zur Verfügung, um passende Inhalte zu den Suchanfragen zu liefern. Dazu werden Wörter und Phrasen in Vektoren umgewandelt, die sich mathematisch berechnen lassen. Wie man hört, funktioniert das sehr gut. Laut Google erfolgt das mit einer Art künstlicher Intelligenz. Auch wenn man über den Begriff „künstliche Intelligenz“ auf philosophischer Ebene sicherlich streiten kann, ist das Ganze technisch sehr faszinierend. Für vorausschauend denkende SEOs ändert das aber nichts. Langfristig ist man erfolgreich, wenn man die Ziele von Google zu seinen eigenen Zielen macht. Googles Ziel ist es, die besten Angebote ganz oben in den organischen Suchergebnissen anzuzeigen. Daher sollte man dafür sorgen, dass man das beste Angebot in seinem Segment bereitstellt. Jede Veränderung, die den Google-Algorithmus verbessert, spielt einem dann in die Karten. Natürlich müssen die Inhalte weiterhin barrierefrei, gut verständlich und richtig gewichtet für Google aufbereitet werden. Wer nach dieser Strategie handelt und nicht gegen die Google-Richtlinien verstößt, wird auch weiterhin bei Google erfolgreich sein. Die Orientierung an den Bedürfnissen der Nutzer ist dauerhaft der einzig wirksame Wettbewerbsvorteil. RankBrain ist ein weiterer Schritt von Google, dass diese Anstrengungen auch belohnt werden.

## STATEMENT

**SEBASTIAN ERLHOFER**

GESCHÄFTSFÜHRER DER MINDSHAPE GMBH



**Sebastian ist begeisterter Online-Marketer mit Fokus auf Suchmaschinen-Optimierung. Er ist als Buchautor, Speaker und Agenturinhaber bekannt. Vor allem die Kombination aus strategischer Optimierung, gezielter Onpage-Optimierung und speziellen technischen Umsetzungen, am liebsten mit TYPO3, ist sein Steckenpferd.**

Zur Berechnung der Rankings nutzt Google nicht immer die gleichen Rankingfaktoren in der gleichen Gewichtung. Je nach Suchanfrage versucht die Suchmaschine, die Suchintention zu bestimmen, und möchte dabei dann möglichst passende und hilfreiche Ergebnisse liefern. Trotz des Hummingbird-Updates ist Google allerdings nicht wirklich gut darin, komplexe oder sogar natürlichsprachige Anfragen passend zu „beantworten“.

RankBrain ist der logische Schritt, um die viele Jahre vorherrschenden statistischen Bewertungsverfahren im Suchalgorithmus mit mehr Dynamik zu füllen. Durch maschinelles Lernen, vielleicht sogar irgendwann künstliche Intelligenz, können Suchanfragen sinnvoller klassifiziert und damit besser bewertet werden. Noch lernt RankBrain die Klassifizierung offline, und wenn diese nach menschlichen Bewertungen zufriedenstellend läuft, darf RankBrain sich auch online austoben. Google hält seinen maschinellen Lernhund also noch an einer sehr kurzen Leine.

Doch was bedeutet dieser technologische Fortschritt? Für die Suchmaschinen-Optimierung bedeutet das eine nochmals erhöhte Komplexität. Die Einstufung durch RankBrain ist sicherlich nur ein weiterer Schritt, den gesamten Algorithmus immer „schlauer“ und „autonom“ in seinen Entscheidungen zu machen. Das bedeutet, dass weder Google noch sonst jemand mehr erklären kann, wie bestimmte Faktoren zusammenwirken und welche wichtig und weniger wichtig sind. Für die Suchmaschinen-Optimierung steht die Veröffentlichung von RankBrain meiner Meinung daher für den langsamen Wandel – weg von statischen technischen Verfahren hin zu nutzerorientierter Suchmaschinen-Optimierung. Als SEOs bleibt uns am Ende nichts anderes übrig, als Inhalte und Strukturen möglichst optimal für Suchende auszurichten. Denn das wird immer das Ziel jeglicher künstlicher Suchintelligenz bei Suchmaschinen sein. Ich freu mich drauf!

sicher noch verstärken. Die Maschine sucht ständig nach neuen Mustern, die einen Hinweis auf gut oder schlecht bringen können, und gebiert praktisch ständig neue Signale, die bisher noch niemand auf dem Schirm hatte.

Basierend auf der Idee, dass es den (einen) Algorithmus nicht mehr gibt, sondern ein permanent selbstlernendes System, werden die Innovationszyklen bei Google wahrscheinlich noch schneller und spezifischer vorangehen. So können z. B. spezielle Spam-Muster in einem Sprachraum erkannt werden, obwohl diese von der „Hausprache“ Englisch abweicht. Über den Abgleich von Entitäten wird dies weitgehend mit tatsächlich verwendeten Begriffen möglich (siehe Abb. 12). Ein klassischer Rollout von USA/UK (Englisch) in die jeweiligen Sprachen wäre damit künftig obsolet. Es könnten umgekehrt sogar Signale aus anderen Sprachen eine Modifikation im Kernmarkt bewirken. Änderungen würden damit schneller global aktiv, oder es könnten auch ganz spezifische länder- bzw. ortsspezifische Muster zum Tragen kommen.

### **Spam- und Blackhat-SEO wird automatisch bekämpft**

Der Vorteil von Machine Learning ist zusammenfassend darin zu sehen, dass permanent neue Verknüpfungen in den Daten entstehen. Dabei kommt insbesondere Mustern bzw. schwachen Signalen eine große Bedeutung zu, da diese erkannt und weiterentwickelt werden können. Genauso wie die Maschine lernt, eine Sprache zu verstehen, können auch Spam- und Blackhat-Techniken noch besser durch die Maschine über die Muster erkannt werden. Analysiert bspw. eine Maschine die Texte einer Website, so ist schon anhand des Sprachstils und der verwendeten Wörter ableitbar, ob es sich um Gastartikel handelt oder nicht. Künftig könnte es sogar möglich sein, einen Autor, der genügend publiziert hat, an seinem ganz persönlichen Sprachstil zu identifizieren. In kleinem Umfang geht dies übrigens schon heute. Sollte ein solcher Artikel damit identisch gewertet werden, wie der Rest der Seite? Hierzu wird es künftig sicherlich jede Menge Spekulationen geben.

### **Was sind die Implikationen für strategisches SEO?**

Lange wird es wohl nicht mehr dauern, bis ganz generell aus Sucheingaben wie „Wie groß ist Dirk Nowitzki?“ eine interne Umwandlung in <Entity:Nowitzki\_Dirk>.<Property:Height> stattfindet und statt eines Textvergleichs das Ergebnis aus Entitäten und ihren Beziehungen gezogen wird. Das bedingt, dass Webseiten ebenfalls in Entitäten „zerlegt“ werden. Oder ist das bei Google vielleicht schon in vollem Gange?

Aus Webtexten wie „Nowitzki ist 2,13 m groß“, „Dirk Nowitzki ist Basketballspieler“, „Jessica Olsson ist verheiratet mit Dirk Nowitzki“ und „Basketball = Sport“ wird die echte Erkennung der zugrunde liegenden Informationen über die Größe, den Lebenspartner, die berufliche Tätigkeit und in welche Kategorie diese fällt.

Dies bedeutet auch, dass für immer mehr sog. Informational Searches, also Suchanfragen, die eine Information als Antwort erwarten, Google die Suchenden immer seltener auf Websites schicken muss, um diese Fragen zu beantworten. In Ansätzen ist da ja heute schon so. Die Frage nach dem Wetter morgen in Frankfurt wird von modernen Smartphone-Betriebssystemen ja schon heute direkt beantwortet – ohne den Umweg über einen Ergebnisklick und das „Selbstauf-der-Seite-Raussuchen“. Am Ende könnten also langfristig nur noch Navigational (Suche nach Websites oder Unternehmen/Marken) und Transactional Searches („ich will aktiv etwas tun“) als Trafficlieferanten für Websitebetreiber übrig bleiben. Und selbst das kann in Zukunft mehr und mehr obsolet werden. Dann genügt die geäußerte Aufforderung „Buch mir den günstigsten Flug am 15. Juni nach Hamburg, mein Termin beim Ottoversand geht von 10 bis 14 Uhr“ und eine Maschine – Google? – sucht nach der besten Problemlösung und wickelt alle nötigen Transaktionen ab – für eine Gebühr gegenüber den beteiligten Transportunternehmen. Wird dies dann vielleicht sogar das neue AdWords? Sind transaktionale Websites vielleicht sogar ein Auslaufmodell, weil die Menschen gar nicht mehr selbst suchen (möchten) und nur noch Maschinen miteinander über Clearingstellen – Google? – kommunizieren und Transaktionen abschließen? Bis dahin ist sicher noch ein weiter Weg – nachdenklich macht dies aber schon.

Für Websitebetreiber heute gilt: Wer austauschbar ist und ohne echtes bzw. eigenes Know-how Informationen und keinen Mehrwert gegenüber allgemein zugänglichen Aussagen anbietet, wird wohl zu den großen Verlierern zählen. Umgekehrt werden sicherlich diejenigen als von Google vermittelte Anlaufstelle im Web dominieren, die es schaffen, sich wirklich auch maschinell erkennbar holistisch zu bestimmten Themen darzustellen. Ach ja – die Zeiten, in denen man sich per Linkeinkauf Google gegenüber hübscher darstellt, als man wirklich ist, dürften dann wohl auch absehbar einem Ende entgegengehen – auch wenn Links heute noch mit zu den wichtigsten Rankingfaktoren zählen. „Beliebtheit“ durch künstliche Links zu simulieren, geht vergleichsweise einfach und schnell. Überzeugenden Content zu produzieren bei Weitem hingegen nicht. Wer sich damit schon heute wirklich ernsthaft auseinandersetzt, ist nicht nur für die Zukunft bestens gerüstet, sondern kann bereits heute durch verbesserte Rankings davon profitieren.

RankBrain erzielt nach Angaben von Google Ergebnisse, die die Erwartungen übertreffen. Verdeckte Vergleichstests hätten laut Greg Corrado gezeigt, dass sich die Ergebnisse ohne RankBrain spürbar verschlechtern. Das wird die Engineers ganz sicher anspornen, weiter nach Mustern und Möglichkeiten für noch mehr Machine Learning zu suchen. Es bedarf somit irgendwann keiner „Updates“ mehr, die Erkennung wird damit schneller, tiefgreifender und in Echtzeit möglich sein – die Nutzer werden dafür sicher dankbar sein.

RankBrain arbeitet wohl derzeit noch „unter Aufsicht“ und wird noch manuell von Menschen trainiert. Über Trainingsdaten verfügt Google ja weiß Gott genug. Kommen später wirklich Quantencomputer zum Einsatz, werden die Möglichkeiten schlicht explodieren.¶

# STATEMENT

## MARCUS TOBER

GESCHÄFTSFÜHRER SEARCHMETRICS GMBH



**Marcus Tober studierte Informatik in Berlin und gründete gleichzeitig 2005 die Searchmetrics GmbH. Ziel war es immer, eine datengetriebene Software dort einzusetzen, wo Marketer Zeit verlieren oder nur mit hohem Aufwand an Insights kommen. Searchmetrics ist mittlerweile ein internationales Softwareunternehmen mit über 220 Mitarbeitern.**

RankBrain ist eine Machine Learning Technology, implementiert in den Google-Algorithmus, um auf Basis fortwährenden Lernens relevantere Ergebnisse für den Nutzer zu bieten. Und dass Machine Learning eng mit Google verwoben ist und eigentlich schon sehr lange von Google benutzt wird, sollte jedem klar sein. Die größte Herausforderung für Google ist, aus der Masse der vorhandenen Webseiten diejenigen herauszufiltern, die für den Query eines Nutzers zu einer bestimmten Zeit, an einem bestimmten Ort, mit einem bestimmten Device wesentlich sind. RankBrain wird dafür sorgen, dass Usersignale und das Verhalten, das Google misst, zurück in den Algorithmus fließen. Google selbst behauptet sogar, dass RankBrain mittlerweile der drittstärkste Ranking-Faktor sei. Ich denke, es ist so, dass RankBrain alle Signale und Faktoren nimmt und zu gewissen Queries gewichtet. Die Gewichtung erfolgt anhand von Usersignalen. User verhalten sich anders je nach Query, der eine bestimmte Intention beinhaltet. Das wertet Google aus, um wiederum alle Signale querybasiert zu gewichten und bessere Ergebnisse, quasi maßgeschneidert, zu liefern. Und da RankBrain sich ständig weiterentwickelt, werden die Ergebnisse je Query auch immer besser und individueller, ohne dass Google generelle Algorithmusänderungen vornehmen muss.

## STATEMENT STATEMENT

**MARCUS TANDLER**

CO-GRÜNDER &amp; CEO ONPAGE.ORG GMBH



**Marcus Tandler, SEO-Interessierten auch als „Mediadonis“ wohlbekannt, liebt SEO! Er ist einer der Gründer und Geschäftsführer der beliebten und preisgekrönten Webseiten-Analyse-Software OnPage.org. Seine ersten Gehversuche im SEO-Bereich erfolgten bereits vor Googles Deutschlandstart mit dem damaligen Suchmaschinenprimus Altavista.**

Google macht sich schon länger die Vorzüge autonomer selbstlernender Algorithmen zunutze, vor allem bei der Erledigung zeit- und personalressourcenintensiver Aufgaben. Dennoch ist zumeist noch immer menschliche Aufsicht notwendig, um die Ergebnisse kontinuierlich zu verbessern. Das Bekanntwerden von Rankbrain hat eine Menge Hoffnungen und Wünsche innerhalb der SEO-Gemeinschaft geweckt, von intelligenten Spam-Cops, die schlechte Webseiten vollautomatisch aus dem Index kegeln, und von Linkbewertung in Echtzeit, mit unmittelbarer Sanktion. Bei Rankbrain wiederum geht es erst mal nur um Reformulierung von Suchanfragen, und zwar solchen, die Google noch nie zuvor gehört hat. Das explodierende mobile Nutzerverhalten und der Boom der virtuellen Assistenten hat Suchmaschinennutzung fundamental verändert. Die Leute haben mittlerweile verstanden, dass man Google nun auch richtige Fragen stellen kann. Bei solchen Suchanfragen gilt es, möglichst genau zu erraten, was der Suchende hier zu suchen versucht. Kluge Maschinen müssen also mindestens so klug wie dumme Menschen sein. Für SEOs gibt es wenig Kontaktpunkte mit Rankbrain. Bei SEO geht es ja zumeist eher um die Optimierung für das genaue Gegenteil, nämlich möglichst häufig gesuchte Suchworte und -phrasen. Ein handwerklich gut arbeitender SEO profitiert allerdings von solch einer schlaun Konsolidierung seitens Google – auf nutzerseitig möglichst hochwertige Suchergebnisse. Am Ende zählt der Webseitenbesucher an sich. Es gilt nicht nur, diesen durch eine möglichst hohe Position in den Suchergebnissen auf die eigene Seite zu locken, sondern dessen Suchbedürfnis bestmöglich zu befriedigen. Dieses Berufsmantre der SEOs wird auch mit künstlicher Intelligenz noch sehr lange Bestand haben.

**JONAS WEBER**

SEO CONSULTANT (JONASWEBER.COM)



**Diplom-Kaufmann Jonas Weber ist SEO & Online-Marketing-Berater. Er arbeitete u. a. bei Google im Search Quality Team in Dublin und ist aktiver „Google Developers Expert“ für die Bereiche SEO & Online-Marketing.**

Das grundsätzliche Ziel von Google ist es, so viel wie möglich zu automatisieren. Dies ist bei weltweit 2,1 Billionen Suchanfragen pro Jahr (2014, statista.de) auch gar nicht anders möglich. Manuelle Eingriffe in den Algorithmus sind nur als Notlösung gedacht, sobald dieser versagt. Somit ist es verständlich, dass Google mit dem Einsatz künstlicher Intelligenz den Algorithmus weiterentwickeln möchte. Nur so lässt sich langfristig qualitativ skalieren. Das Spannende am Rankingfaktor RankBrain ist, dass er vor allem bei neuen, zuvor noch nie gestellten Suchanfragen zum Einsatz kommt. Gerade bei den vielen Longtail-Suchanfragen ist es schwierig, die Qualität der Suchresultate zu steuern, es lohnt sich schon allein aus Kosten-Nutzen-Gründen nicht. In diesen Longtail-Nischen, die in der Summe sehr groß sind, hat Google das größte Webspamproblem. Des Weiteren erzielte RankBrain bei den ersten Tests qualitativ bessere Ergebnisse als eine manuelle Zuordnung der Programmierer. Höchstwahrscheinlich wird das Benutzerverhalten für RankBrain eine wichtige Rolle spielen. Der Rankingfaktor wird aus vergangenem Verhalten lernen und dies bei neuen Suchanfragen berücksichtigen, sowohl im positiven als auch im negativen Sinne. Eine weitere Möglichkeit, sich von Backlinks als wichtigem Rankingfaktor zu distanzieren. Dieser technische Fortschritt hat einen weiteren Riesenvorteil für Google: Zuletzt geriet das Unternehmen extrem stark in die Kritik, vor allem aus datenschutzrechtlichen Gründen und wegen der existierenden Monopolstellung. Eine selbst lernende mathematische Formel kann nicht für „böse“ Entscheidungen von Googles Management verantwortlich gemacht werden. Hoffentlich hilft es Google auch insofern, dass weniger manuelle Abstrafungen erfolgen.