

Cognitive Search

Machine Learning als Schlüssel für eine moderne Suche

Dr. Uwe Crenze, interface projects GmbH, 12.11.2019



Schwerpunkte

- Einordnung „*Cognitive Search*“ & Exkurs „*KI*“
- *Single Point of Information Access, Enterprise Search & Suchqualität*
- Rolle der *KI für die Suche*

Cognitive Search = Enterprise Search + KI

FORRESTER®

InsightsWho We ServeOur ExpertiseWhat We OfferWho We Are

LOG INBECOME A CLIENT

Search

FOR APPLICATION DEVELOPMENT & DELIVERY PROFESSIONALS

The Forrester Wave™: Cognitive Search, Q2 2019

The 12 Providers That Matter Most And How They Stack Up

May 29, 2019

Share

By Mike Gualtieri with Srividya Sridharan, Elizabeth Hoberman

Why Read This Report

In our 24-criterion evaluation of cognitive search providers, we identified the 12 most significant ones — Attivio, Coveo, Elastic, Grazitti Interactive, IBM, IHS Markit, Lucidworks, Micro Focus, Microsoft, Mindbreeze, Sinequa, and Squirrelro — and researched, analyzed, and scored them. This report shows how each provider measures up and helps application development and delivery (AD&D) pros select the right one for their needs.

Tags: Artificial Intelligence (AI), Enterprise Content Management (ECM), Search & Knowledge Discovery, Technology

Get Access

Already a Client?
Log in to read this document.

Become a Forrester Client
Customers are the new market-makers, reshaping industries and changing how businesses compete and win. Success depends on how well and how fast you respond. Forrester Research gives you insights and frameworks aligned to your role to shorten the time between a great idea and a great outcome, helping your teams win in the age of the customer. [Contact us](#) to learn more.

Purchase Report
This report is available for individual purchase (\$2495 USD).

Purchase

Table of Contents

Cognitive Search Vendors Double Down On AI

Evaluation Summary

Vendor Offerings

Vendor Profiles

Evaluation Overview

Supplemental Material

Related Research Documents

Recommended Research

The Forrester New Wave™: Automation-Focused Machine Learning Solutions, Q2 2019

May 28, 2019 | Kjell Carlsson, Ph.D.

The Future Of Machine Learning Is Unstoppable

April 25, 2019 | Mike Gualtieri

The Forrester Wave™: Cloud

Home > Insights > Cognitive Search Is The AI Version Of Enterprise Search



Cognitive Search Is The AI Version Of Enterprise Search

Mike Gualtieri, VP, Principal AnalystJun 12 2017

Written by Emily Miller, Senior Research Associate

Stop Wasting Time

More than half (54%) of global information workers are interrupted from their work a few times or more per month to spend time looking for or trying to get access to information, insights, and answers. The problem: Old keyword-based enterprise search engines of the past are obsolete. Cognitive search is the new generation of enterprise search that uses artificial intelligence (AI) to return results that are more relevant to the user or embedded in an application issuing the search query.

Subscribe to updates

Email Address

☐ By checking this box, you agree to receive survey invitations and marketing communications.

Sign Up

Humanization (usability)

Business value

Keyword search

Semantic search

Contextual search

Cognitive search

Indexing

NLP

Machine learning

Natural language processing (NLP)

Natural human interaction (NHI)

Neural lace™ (human brain merged with computers)

Twitter

LinkedIn

Facebook

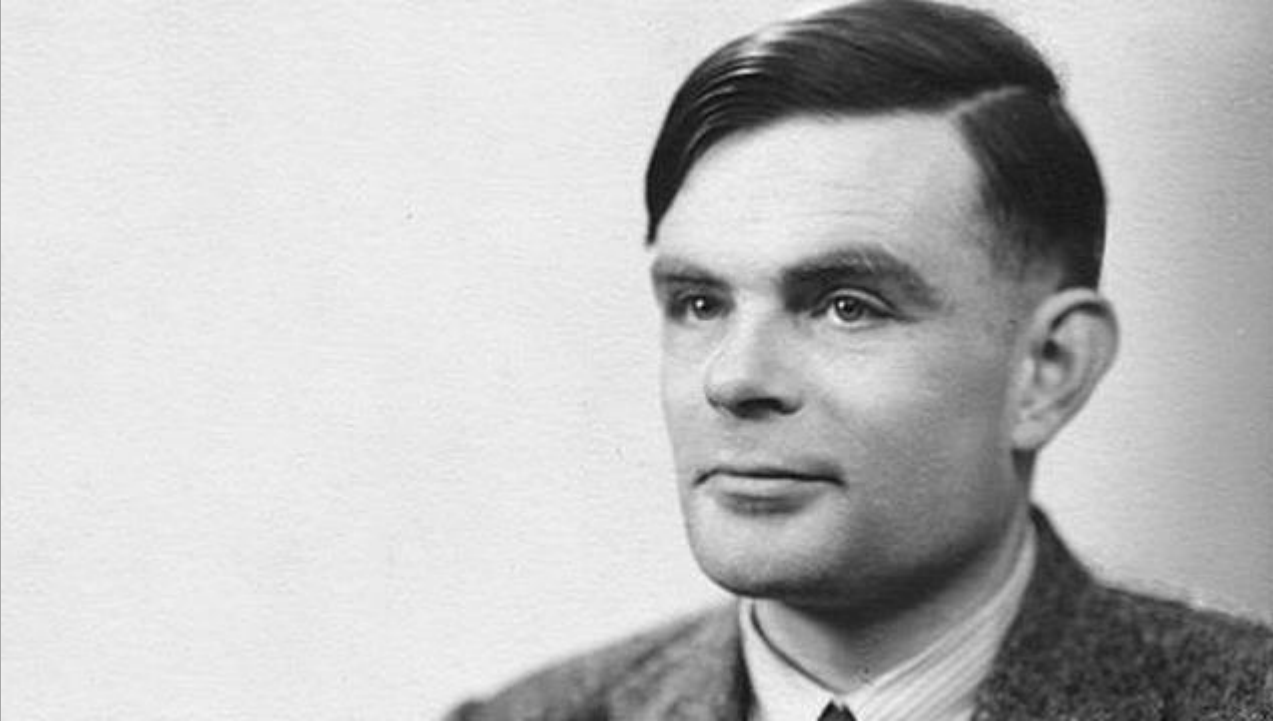
Email

interface projects GmbH © 2019

3

KK

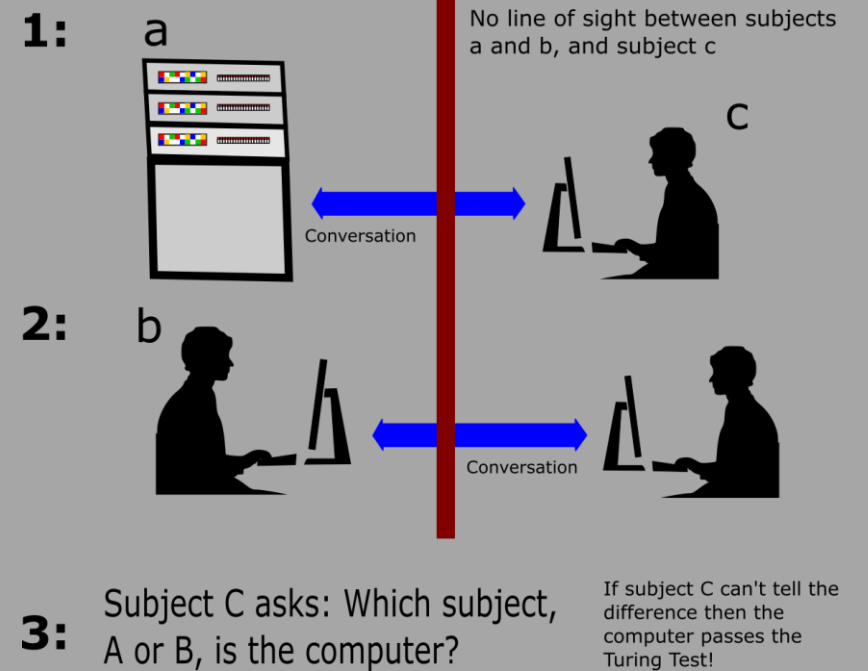
Künstliche Wesen: Mensch oder Maschine?



<http://www.the-nref.org/content/simplest-paradigm-understanding-artificial-intelligence>

Alan Turing 1912 - 1954

The Turing Test



1936: Turingmaschine
1950: Turingtest

Geburtsstunde der Künstlichen Intelligenz

1956: Begriff künstliche Intelligenz

1966: Programm ELIZA (erster Chatbot ;-)



Joseph Weizenbaum, 2005 in Berlin
(Quelle: Wikipedia)



Glossar – Smart? Cognitive? Insight Engine? ...

- **Künstliche Intelligenz** (KI, auch Artificielle Intelligenz (AI bzw. A. I.), englisch: artificial intelligence, AI) ist ein Teilgebiet der Informatik, welches sich mit der **Automatisierung intelligenten Verhaltens** und dem Maschinellen Lernen befasst. Der Begriff ist insofern nicht eindeutig abgrenzbar, als es bereits an einer genauen Definition von „Intelligenz“ mangelt.
- **Maschinelles Lernen** ist ein Oberbegriff für die „**künstliche**“ **Generierung von Wissen aus Erfahrung**: Ein künstliches System lernt aus Beispielen und kann diese nach Beendigung der Lernphase verallgemeinern. Dazu bauen Algorithmen beim maschinellen Lernen ein statistisches Modell auf, das auf Trainingsdaten beruht.
- **Künstliche neuronale Netze**, auch künstliche neuronale Netzwerke, kurz: KNN (englisch: artificial neural network, ANN), sind Netze aus künstlichen Neuronen. Sie sind Forschungsgegenstand der Neuroinformatik und stellen einen Zweig der künstlichen Intelligenz dar.
- **Deep Learning** (frei übersetzt: ‚tiefgehendes Lernen‘) bezeichnet eine Klasse von Optimierungsmethoden künstlicher neuronaler Netze, die **zahlreiche Zwischenlagen** (englisch: hidden layers) zwischen Eingabeschicht und Ausgabeschicht haben und dadurch eine umfangreiche innere Struktur aufweisen.
- **Cognitive Computing** beschreibt **Technologieplattformen**, die im Großen und Ganzen auf den wissenschaftlichen Disziplinen der künstlichen Intelligenz und der Signalverarbeitung basieren.

Smart ...



Künstliche Intelligenz (KI/AI)

Künstliche Intelligenz (KI/AI)



Maschinelles Lernen

Cognitive ...



Maschinelles Lernen

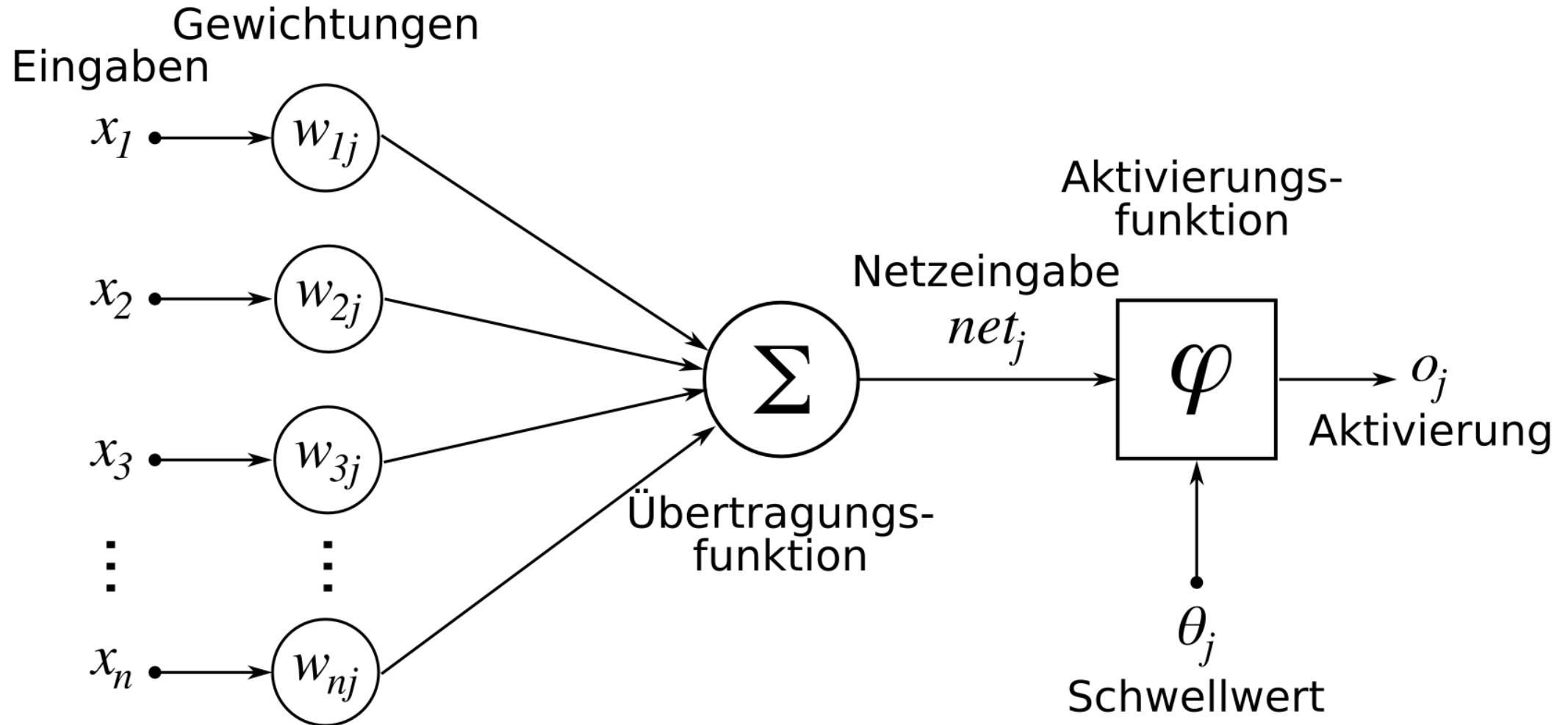
Neuronal ...



**künstlich Neuronale Netze
(kNN)**

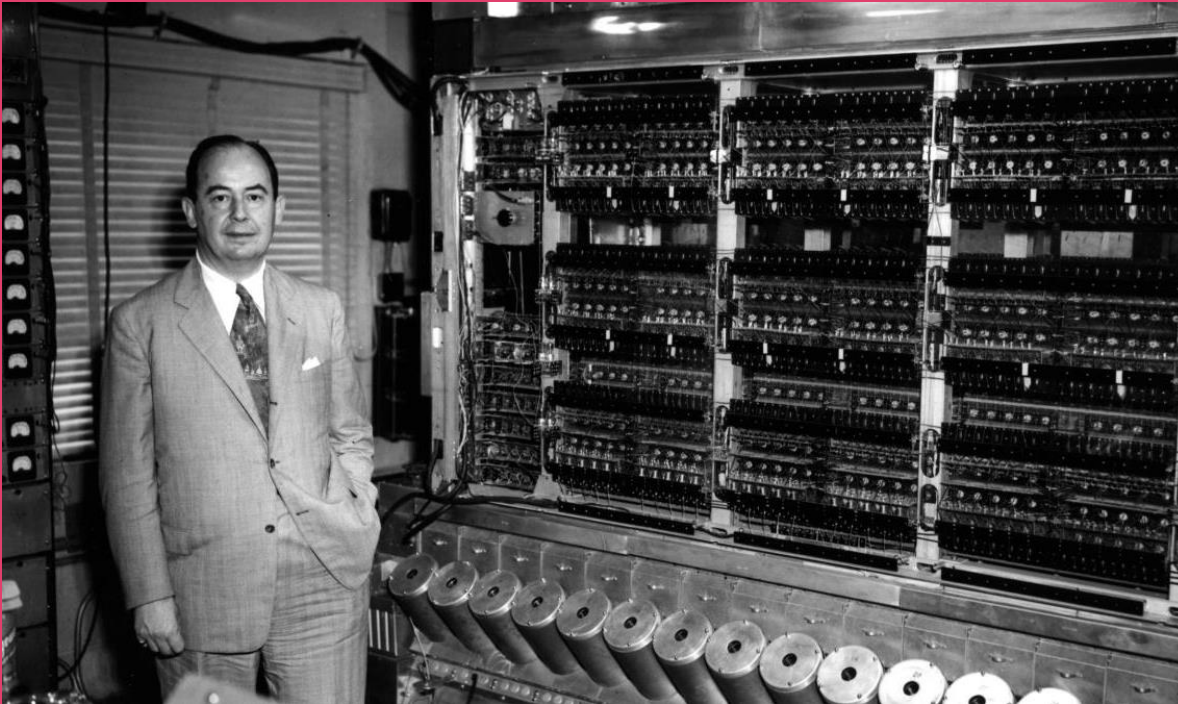


Künstlich Neuronale Netze (kNN / ANN)



1940er -> 1970er -> 2010er

KI ist Mathematik



Quelle: Alan Richards, Shelby White and Leon Levy, Archives Center, Institute for Advanced Study, Princeton, NJ, USA

“If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.”

John von Neumann

* 28. Dezember 1903 in Budapest, Österreich-Ungarn
† 08. Februar 1957 in Washington, D.C.

Semantik

Semantik -> kuratierte Begriffssysteme

- **Semantik**, auch **Bedeutungslehre**, nennt man die **Theorie** oder Wissenschaft **von der Bedeutung der Zeichen**. Zeichen können hierbei beliebige Symbole sein, insbesondere aber auch Sätze, Satzteile, Wörter oder Wortteile. Soweit sich die Semantik mit Zeichen aller Art befasst, ist sie ein Teilbereich der **Semiotik**.
- **Semiotik**, manchmal auch **Zeichentheorie**, ist die Wissenschaft, die sich mit Zeichensystemen aller Art befasst. Sie findet unter anderem in verschiedenen Geistes-, Kultur-, Wirtschafts- und Sozialwissenschaften Anwendung.
- **Linguistik**, ...

Quelle: Wikipedia

Knowledge Graph

Knowledge Graph

- Metapher
- Kuratiertes Begriffssystem, Linked Data, ...
- Ontologien inkl. Reasoning (Semantik)
- Umfangreiche Thesauri (schwer beherrschbar)
- Als Ergänzung zu Maschinellern Lernen

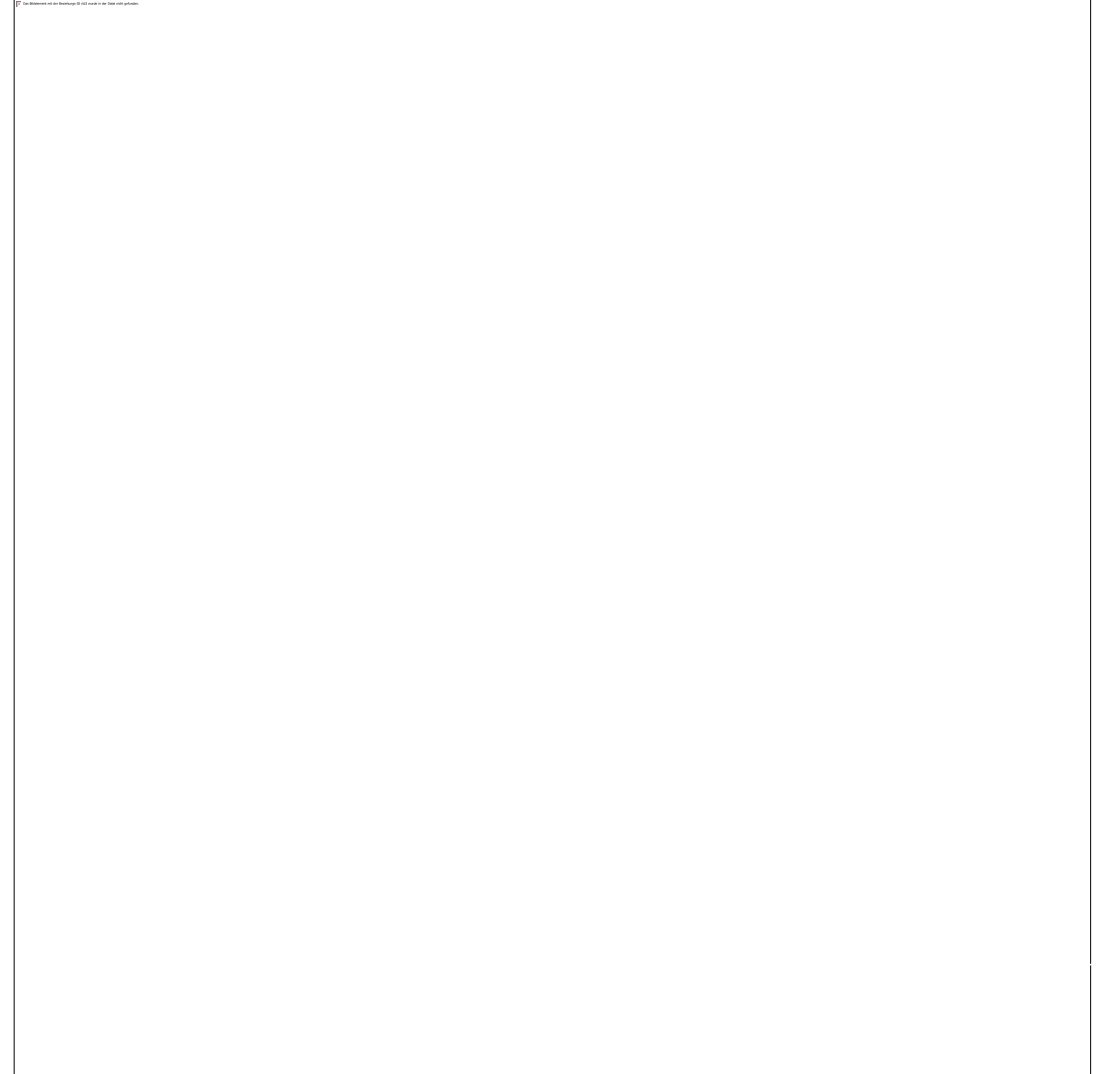
Verbindung zw. Inhalten & Suche?

- Auch kuratiert (Best Bets vs. Best Match)?
- Fehlerquelle des Menschen vs. KI?
- Wenn KI, warum dann überhaupt Wissensgraphen?

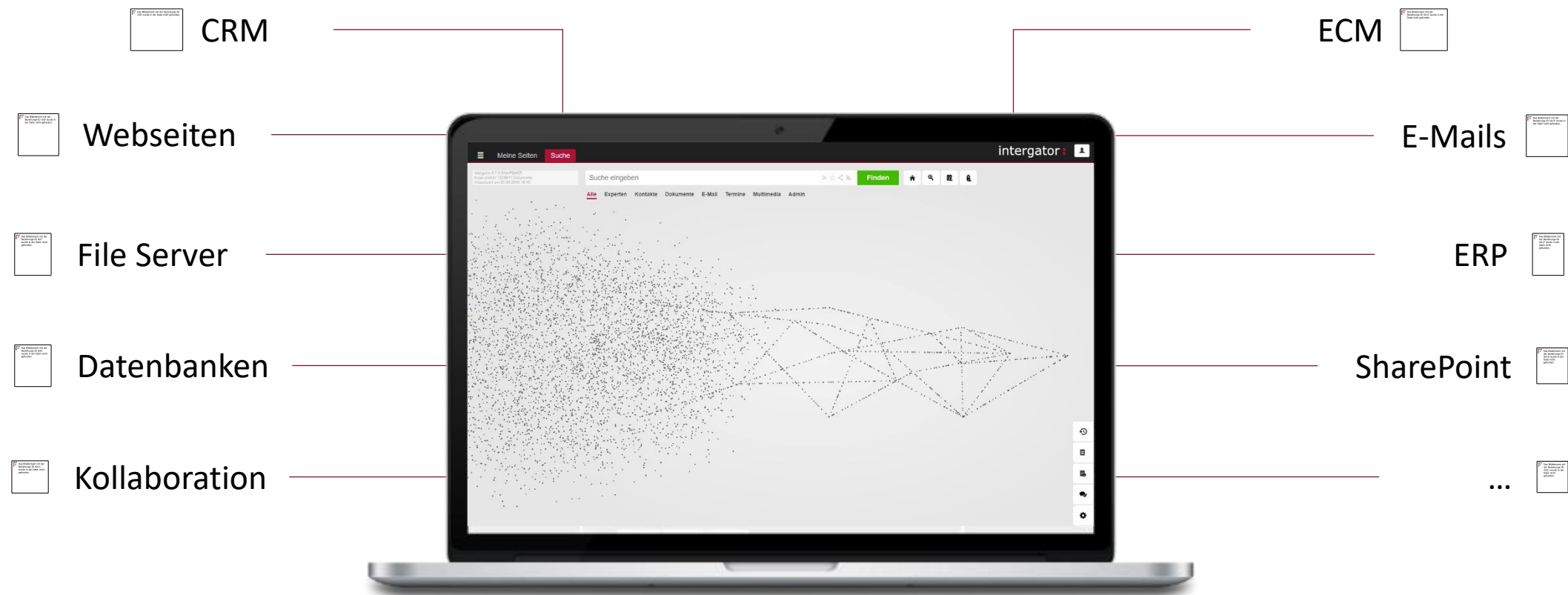
Information Access & Enterprise Search

Wissen nutzbar machen

- Menge nimmt exponentiell zu
- Ablagesysteme werden vielfältiger
- Abgelegtes Wissen wird immer schwerer auffindbar
- Kein übergreifendes Ordnungssystem



Single-Point-of-Information-Access



Suchen

- Bei großen Informationsmengen ist Suchen unausweichlich
- Suche = Zugriffsweg #1

Suchen & Finden

Innovationen bei Suchverfahren

Suchen ist nicht gleich Suchen – Mittel zum Finden (1)

ad hoc System-übergreifend nach Bekanntem *suchen*

systematisch nach Unbekanntem *recherchieren*

explorativ Vorhandenes *erschließen*

Ansprechpartner identifizieren

Suchen vermeiden; über Neues *informiert werden*

Zusammenhänge erkennen, Informationen bündeln

wichtige *Informationen besser auffindbar machen*

Suchen ist nicht gleich Suchen – Mittel zum Finden (2)

ad hoc suchen – Autovervollständigung, Suchvorschläge, Suchformulare

systematisch recherchieren - Ähnlichkeitssuche

explorativ erschließen - Facetten, Visualisierungen, Perspektiven, ...

Ansprechpartner identifizieren - Expert Finder & kollaborative Funktionen

Suchen vermeiden; informiert werden - Info-Portal (Streams), Messages, ...

Zusammenhänge erkennen, Informationen bündeln - Vorschläge, Reports

Informationen besser auffindbar machen - Klassifikation, Verschlagwortung, kommentieren, teilen, ...

Exakte vs. unscharfe Suche



- UND-Filter
- Kein Stemming etc.
- Exakte Werte / Terme
(Aktenzeichen:XYZ)

⇒ *Abfragen mit keinen Treffern*

- ODER-Terme
- Stemming, Lemmatisierung
- Wildcards (viele Werte)
- Ähnlichkeitssuche (QBE)
- Mehrsprachigkeit

⇒ *Lange Ergebnisliste mit schlechtem Ranking*

Welchen Nutzen bringt KI bei der Suche?

Cognitive Services

Neuronale Verfahren zur Inhaltsanalyse

Photo by Frank V on Unsplash

Welche Suchprobleme lassen sich lösen?

Recall Problem

Mehrere Wörter haben die selbe Bedeutung (z.B. Synonyme, unterschiedliche Schreibweisen).

Beispielsuche: „Elektrorad“



Bildquelle: pixabay.com

Auf Basis von semantischen Beziehungen ermitteln kognitive Verfahren ähnliche Wörter für den Suchbegriff:

Elektrofahrrad
Elektroroller
Zweirad
Fahrrads
E-Bike
Zweirads
Pedelec
Elektrofahrrads
Fahrrad
Elektrofahrrades

Welche Suchprobleme lassen sich lösen?

Precision Problem

Ein eingegebenes Suchwort hat mehrere Bedeutungen.

Beispielsuche: „Flügel“



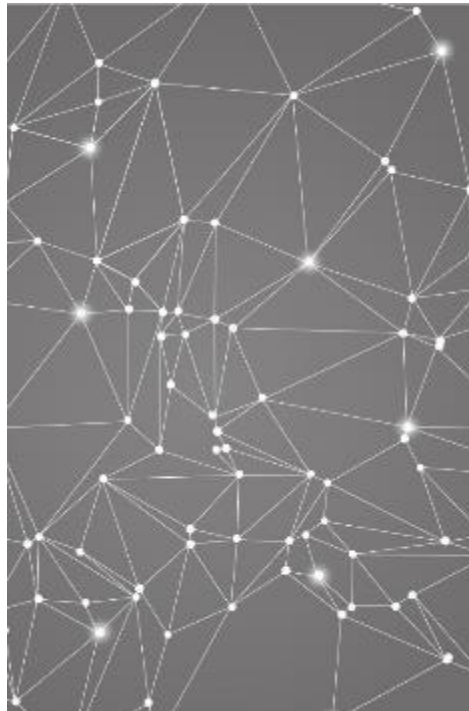
Bildquelle: pixabay.com

Auf Basis von semantischen Beziehungen ermitteln kognitive Verfahren verschiedene Kontexte für den Suchbegriff:

#Flügel_(Militär)
#Gebäudeflügel #Flügel_(Schmetterling)
#Flügel_(Tasteninstrument)
#Insektenflügel
#Flügel_(Insekt)
#Gebäudetrakt
#Flügelspieler
#Flügel_(Vogel)
#Flügel_(Musikinstrument)
#Tragfläche
#Flug_(Heraldik)

Cognitive Services

Automatische Erschließung von Inhalten



Klassifikation von
Inhalten

Einordnung von Texten & Bildern
in Klassen (Taxonomien etc.)

Ein **Transformator** (von lateinisch transformare ‚umformen, umwandeln‘; auch Umspanner, kurz Trafo) ist ein Bauelement der **Elektrotechnik**. Er besteht meist aus zwei oder mehr **Spulen** (Wicklungen), die in der Regel aus **Kupferdraht** gewickelt sind und sich auf einem gemeinsamen Magnetkern befinden. Ein Transformator wandelt eine **Eingangsspannung**, die an einer der Spulen angelegt ist, in eine **Ausgangsspannung** um, die an der anderen Spule abgegriffen werden kann. Dabei entspricht das Verhältnis von Eingangs- und Ausgangsspannung dem Verhältnis der **Windungsanzahl**.

Automatische
Verschlagwortung

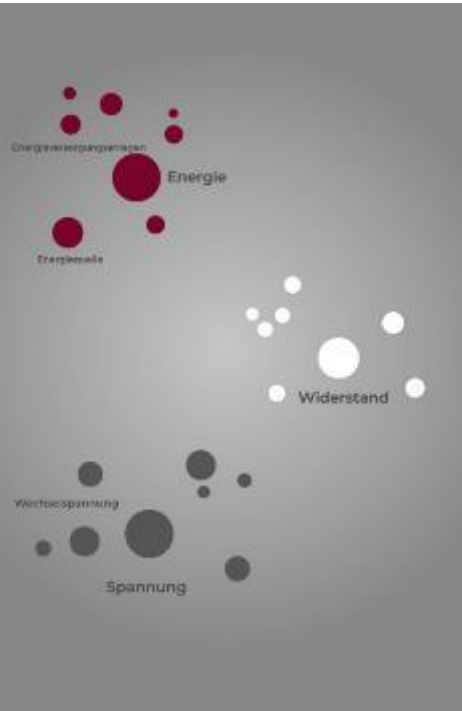
Entitäten, Themen

Cognitive Search



Kontext-assoziative
Suche

Fachspezifische Recherche
(auch cross-lingual)



Suche nach ähnlichen
Inhalten (QBE)

Texte, Objekte, Gesichter

Des Pudels Kern: Neural Embeddings

„You shall know a word by the company it keeps“ (J.R. Firth)

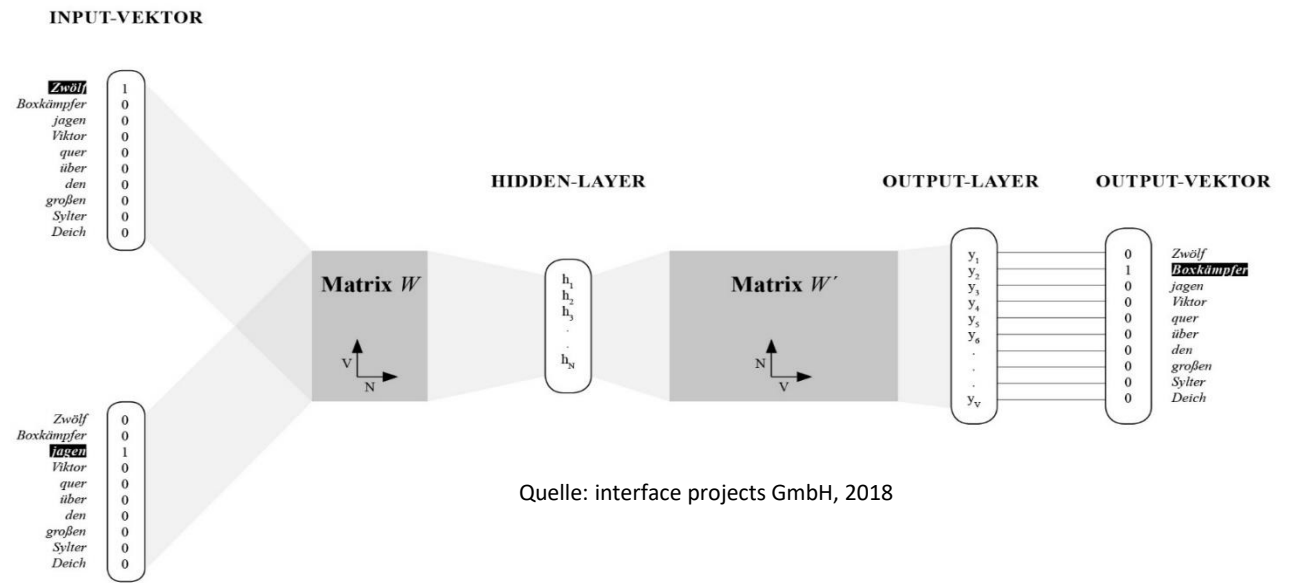
Die Erfindung betrifft ein elektronisch gesteuertes
Federungssystem für ein Fahrrad enthaltend [...]

Zielwort Kontext

Stoßdämpfer mit einer elektrisch steuerbaren
Dämpfereinrichtung für ein Elektrorad und [...]

Lernen von „Neural Embeddings“

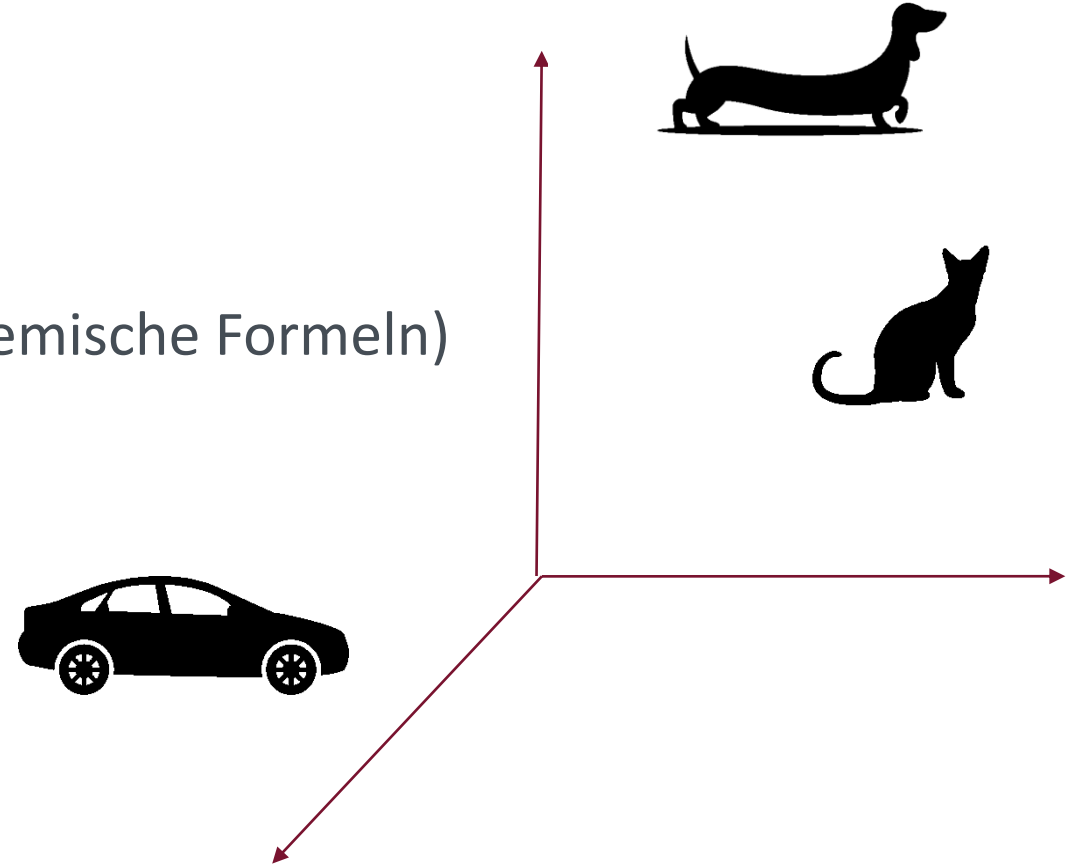
- Sprachmodelle auf der Basis neuronaler Netze
- Mega-Trend für Text-Analysen seit 2013 (Tomas Mikolov)
- Verfahren: word2vec, sent2vec, fastText, ...
- Treiber in der Forschung:
 - Google
 - Facebook,
 - Microsoft
- Vielfältige Einsatzgebiete
 - ✓ Klassifikation
 - ✓ Named Entity Extraction
 - ? **Suche**



Neural Embedding Vector Space

- Embedding-Arten

- Schlagwörter
- Eigennamen (und andere Entitäten)
- Metadaten
- Physikalisch-technische Einheiten (z.B. Chemische Formeln)
- Konzepte
- Sätze
- Dokumente
- ...



- Alle Embeddings befinden sich in einem gemeinsamen Modell
- Visuelle Navigation in einer 2D-Projektion

Einsatz kognitiver Verfahren am Deutschen Patent- und Markenamt

Mark Reinke,¹ André Kischkel,² Volker Jahns,³ Uwe Crenze,⁴ Olga Beltcheva⁵

Abstract: Die Begutachtung von Patentanträgen ist ein aufwändiger Prüfprozess, dessen Ziel es ist, eine Entgeghaltung zu finden. Dabei stellt die Verschleierung des Patentanspruchs durch gezielte Umschreibungen eine große Herausforderung dar. In enger Zusammenarbeit zwischen dem Deutschen Patent- und Markenamt und der interface projects GmbH entstand ein modernes Recherche- und Klassifikationssystem auf Basis von durch neuronale Netze gelernten Distributed Word Embeddings. Der Beitrag stellt verschiedene Verfahren zum Lernen von Word Embeddings vor und bewertet diese hinsichtlich ihrer Eignung für die Prüfung von Patentanmeldungen.

Herausforderung Cognitive Search

Weiterentwicklung der CS zu einem effizienten Recherchewerkzeug

Suchtechnologien

Volltextsuche
(Boolesche Suche)



Stichwort-basierter, invertierter
Dokument-Index

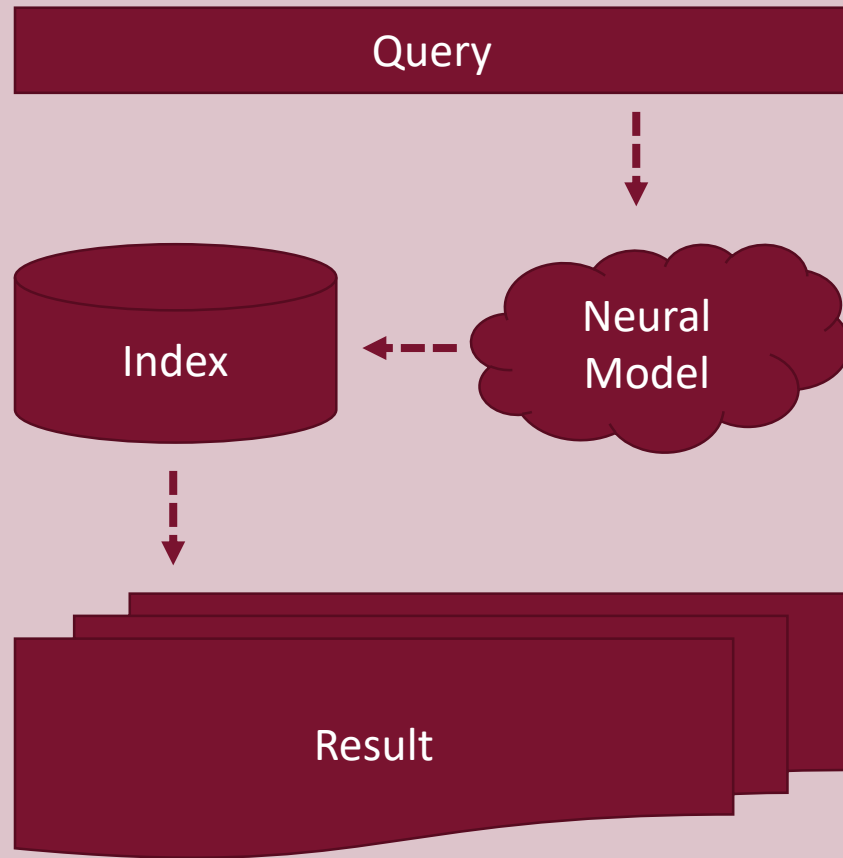
Kognitive Suche



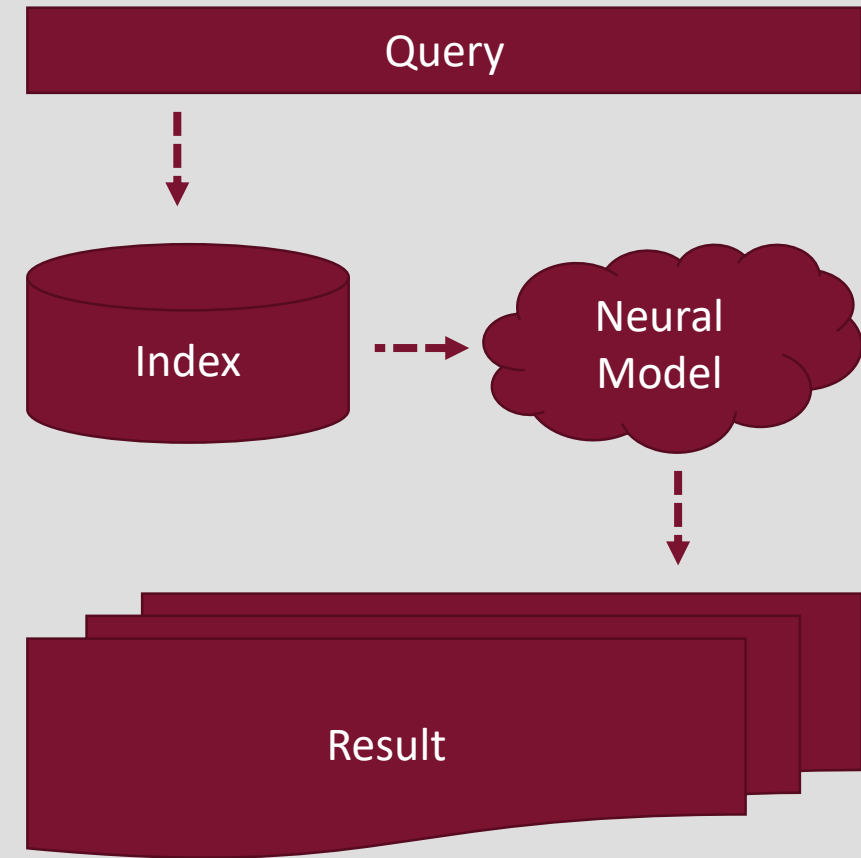
Modifizierte Volltextsuche
oder
NEU: Neuronales Wissens-/Sprach-Modell
(embedding vector model)

KI-gestützte Volltextsuche – State of the Art

Query Re-Writing



Result Re-Ranking



Umformung in Suchanfrage

(elektronisch OR elektrisch OR elektronik) AND

(fahrrad OR elektrorad OR e-bike OR pedelec OR zweirad) AND

(feder* OR dämpf*) AND steuer* AND (*modul OR *element) AND

ipc:B62M AND pub < 2015 ...

(electronic OR electric) AND

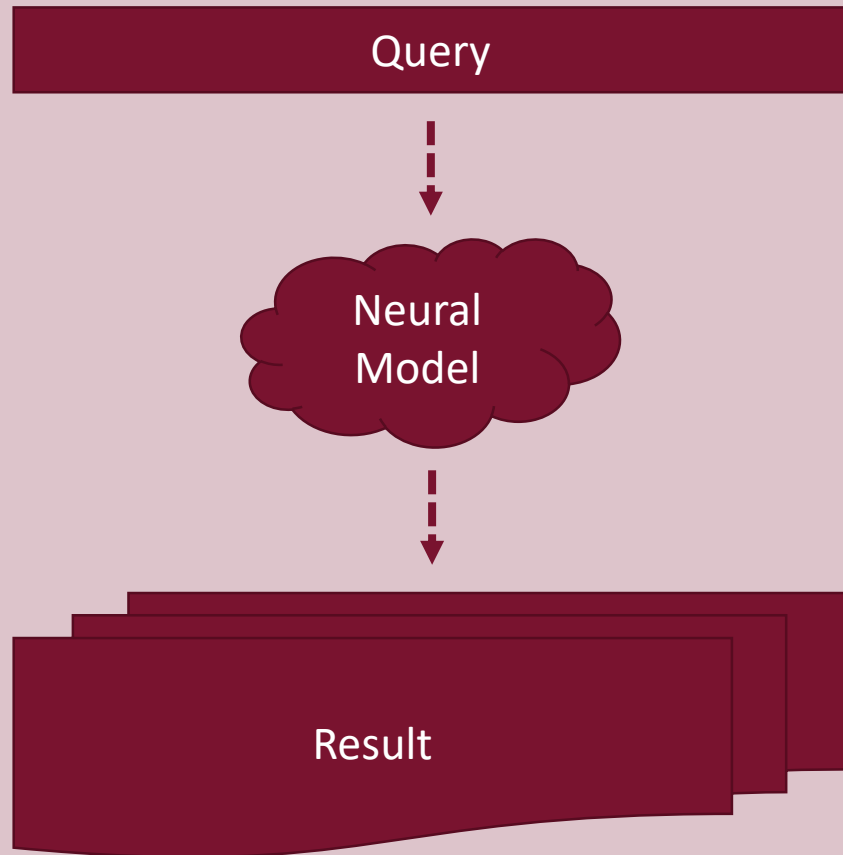
(bicycle OR bike OR e-bike OR pedelec OR two-wheeler)

(suspension OR damping OR shock absorption) ...



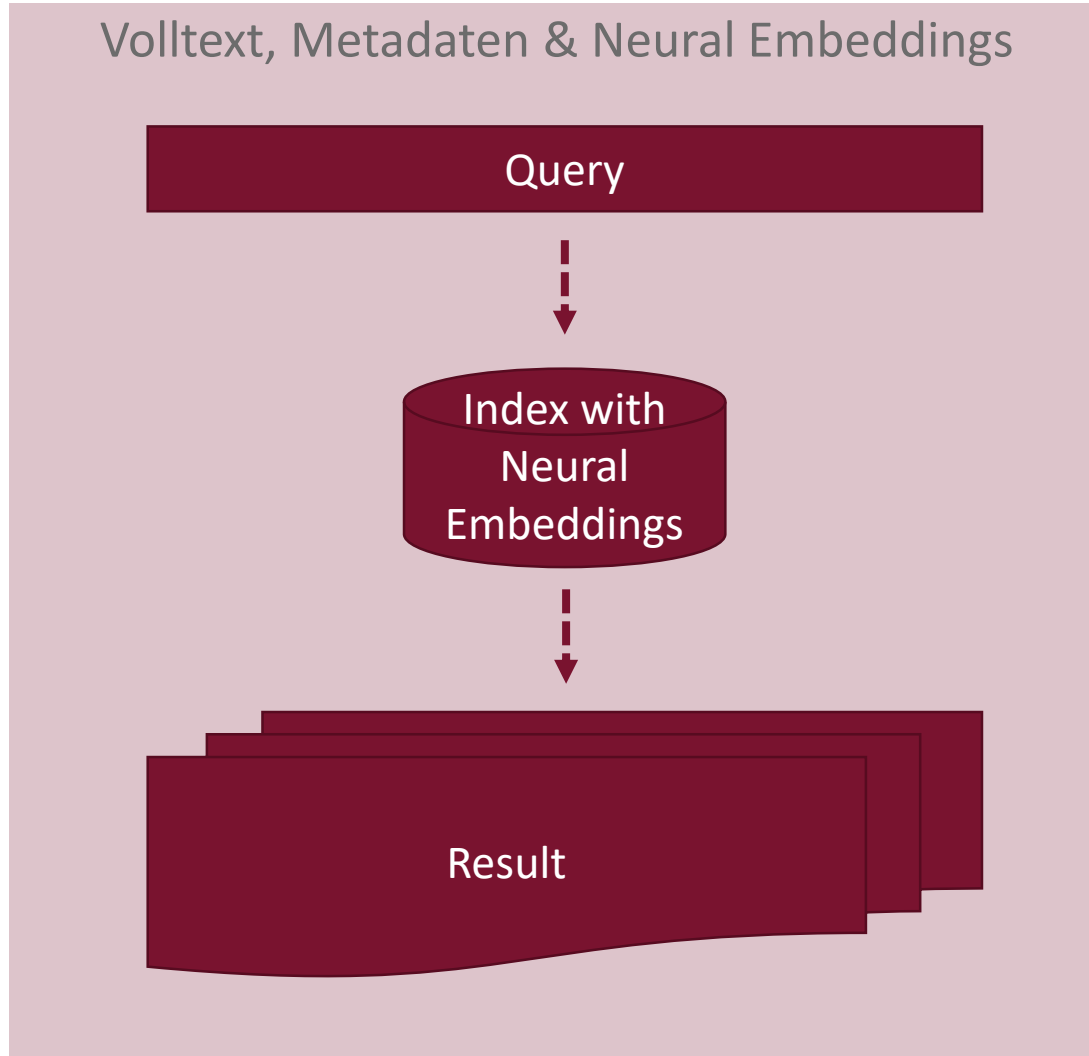
NEU: Nativ-Neuronale Suche

Neural Information Retrieval



- Ideal zum Finden von Zusammenhängen aufgrund sprachlicher Ähnlichkeiten
- Cross-linguale Suche ohne Übersetzung
- Erkennung unterschiedlicher Bedeutungen von Begriffen (Disambiguation)
- Query by Example – z.B. Patentrecherche

Die Zukunft: “Smart Search”



"Jetzt wächst
zusammen, was
zusammengehört.,,

Willy Brandt, 10. November 1989

Visualisierung Semantischer Nähe

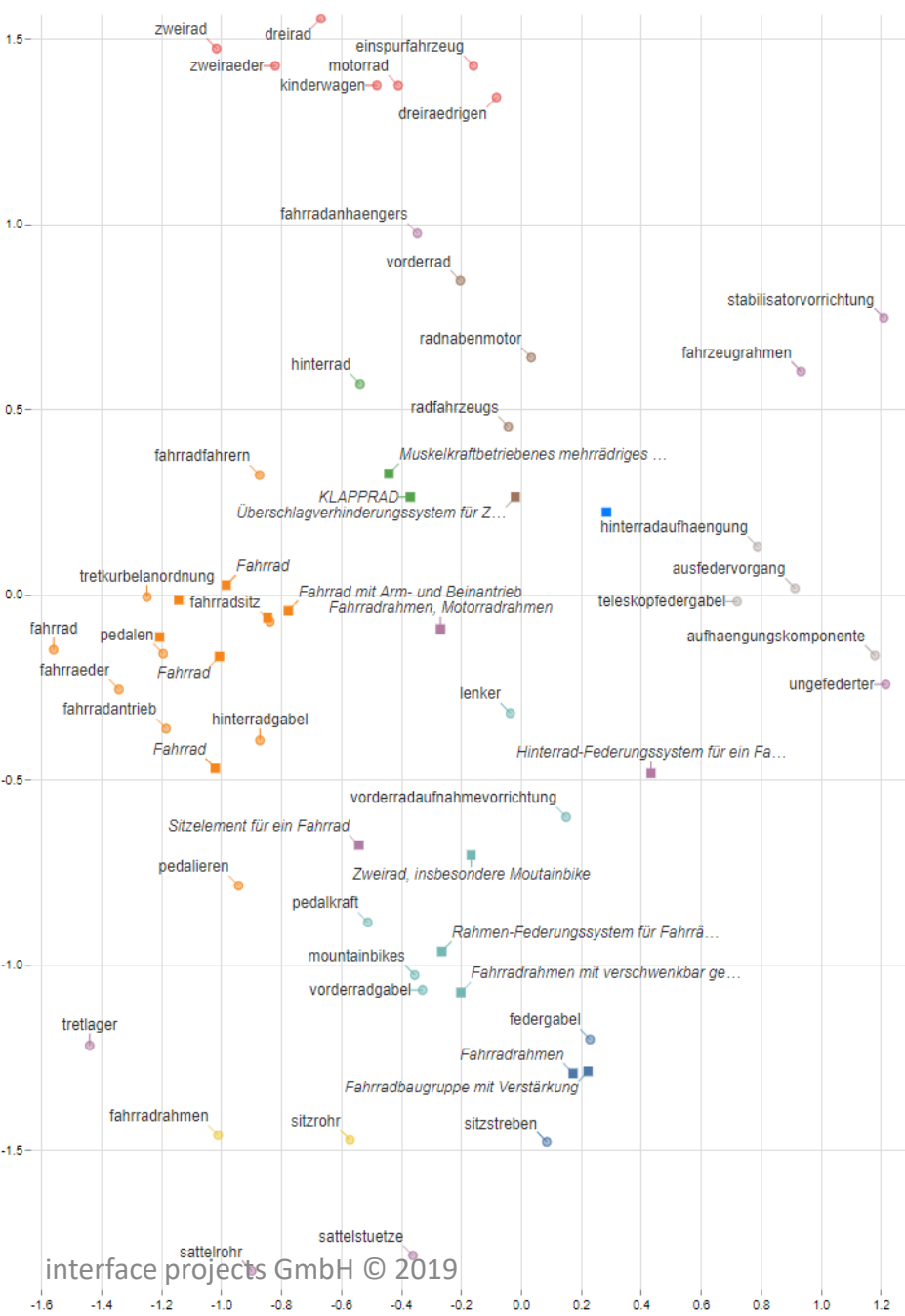
- Unterschiedliche Bedeutungen des Suchbegriffs „**bank**“
- Disambiguation mit einem neuronalen Modell, trainiert aus ca. 300.000 deutschen Patenten
- Interaktive Kontext-Karte für die Navigation und zur Einschränkung des Suchraums





1	Ploče-Tor	0.66
2	Neubrandenburger Stadttore	0.66
3	Tore des Tempelberges	0.65
4	Friedländer Tor	0.64
5	Potsdamer Stadttore	0.63
6	Pile-Tor	0.62
7	Schwabinger Tor	0.61
8	Eckstoß	0.61
9	Ulmer Tor	0.61
10	Rüppurrer Tor	0.61
11	Gartentor (Reutlingen)	0.61
12	Goldenes Tor (Jerusalem)	0.61
13	Oberes Tor (Merkendorf)	0.6
14	Kempter Tor	0.6
15	WM-Tor des Jahrhunderts	0.6
16	Grodzka-Tor (Lublin)	0.6
17	Löwentor (Jerusalem)	0.6
18	Fischertor (Bratislava)	0.6
19	Rotes Tor (Flensburg)	0.6
20	Stiegelgässer Tor	0.6

Elektronisch gesteuertes Federungssystem, Verfahren zur Steuerung eines Federungssystems und Computerprogramm



1	Hinterrad-Federungssystem für ein Fahrrad und Fahrradrahmen	0.59
2	Rahmen-Federungssystem für Fahrräder	0.58
3	KLAPPRAD	0.55
4	Fahrrad	0.55
5	Fahrrad mit Arm- und Beintrieb	0.55
6	Fahrradrahmen mit verschwenkbar gelagertem Hinterbau	0.54
7	Fahrrad	0.54
8	Zweirad, insbesondere Moutainbike	0.53
9	Kompaktes zusammenklappbares Fahrrad	0.53
10	Elektronisch gesteuertes Federungssystem, Verfahren zur Steuerung eines Federungssystems und Computerprogramm	0.52
11	Fahrradrahmen	0.52
12	Fahrradrahmen, Motorradrahmen	0.52
13	Aktives elektromechanisches Federungssystem für ein Fahrwerk eines Kraftfahrzeuges	0.51
14	Fahrrad	0.51
15	Fahrradbaugruppe mit Verstärkung	0.51
16	Muskelkraftbetriebenes mehrrädiges Fahrzeug	0.51
17	Fahrrad mit Arm- und oder Beintrieb und oder mit elektrischem Antriebsmotor.	0.51
18	Kompaktes, zusammenklappbares Fahrrad	0.51
19	Sitzelement für ein Fahrrad	0.5
20	Überschlagverhinderungssystem für Zweiräder	0.5

Elektronisch gesteuertes Federungssystem, Verfahren zur Steuerung eines Federungssystems und Computerprogramm

Title

Elektronisch gesteuertes **Federungssystem**, Verfahren zur Steuerung eines Federungssystems und Computerprogramm

Abstract

Die Erfindung betrifft ein elektronisch gesteuertes **Federungssystem** für ein **Fahrrad** (1), enthaltend zumindest einem Federelement (3, 4), welches zwischen einem ersten Teil (10) des **Fahrrades** und einem zweiten Teil (14, 15) des **Fahrrades** (1) angeordnet ist, welche beweglich miteinander verbunden sind, wobei zumindest eine Kenngröße des Federelementes **veränderbar** ist, und zumindest einen Aktor (431), welcher auf das Federelement (3, 4) einwirkt, um die zumindest eine Kenngröße zu verändern, und ein **Elektronikmodul** (6), mit welchem ein Ansteuersignal für den zumindest einen Aktor (431) erzeugbar ist, wobei weiterhin ein Steuerelement (2, 63, 66) vorhanden ist, mit welchem das vom **Elektronikmodul** (6) erzeugte Ansteuersignal beeinflussbar ist, wobei das Steuerelement (2, 63) mit dem **Elektronikmodul** (6) über ein Funksignal (64) verbindbar ist und/oder der Aktor (431) mit dem **Elektronikmodul** (6) über ein Funksignal (64) verbindbar ist. Weiterhin betrifft die Erfindung ein entsprechendes Verfahren zur Steuerung eines Federungssystems für ein **Fahrrad** und ein Computerprogramm zu dessen Durchführung.

Claims

(1) **Elektronisch gesteuertes Federungssystem** für ein **Fahrrad** (RN), enthaltend zumindest einem Federelement (RN, RN), welches zwischen einem ersten Teil (RN) des **Fahrrades** und einem zweiten Teil (RN, RN) des **Fahrrades** (RN) angeordnet ist, welche beweglich miteinander verbunden sind, wobei zumindest eine Kenngröße des Federelementes **veränderbar** ist, und zumindest einen Aktor (RN), welcher auf das Federelement (RN, RN) einwirkt, um die zumindest eine Kenngröße zu verändern, und ein **Elektronikmodul** (RN), mit welchem ein Ansteuersignal für den zumindest einen Aktor (RN) erzeugbar ist, dadurch gekennzeichnet, dass weiterhin ein Steuerelement (RN, RN, RN, RN) vorhanden ist, mit welchem das vom **Elektronikmodul** (RN) erzeugte Ansteuersignal beeinflussbar ist, wobei das Steuerelement (RN, RN) mit dem **Elektronikmodul** (RN) über ein Funksignal (RN) verbindbar ist und/oder der Aktor (RN) mit dem **Elektronikmodul** (RN) über ein Funksignal (RN) verbindbar ist.

(2) **Federungssystem** nach Anspruch 1, dadurch gekennzeichnet, dass das Steuerelement ein Bedienelement (RN) und/oder einen Lagesensor (RN) und/oder einen Geschwindigkeitssensor und/oder einen Beschleunigungssensor und/oder einen Drehmomentsensor und/oder einen Positionssensor (RN) enthält.

(3) **Federungssystem** nach Anspruch 2, dadurch gekennzeichnet, dass das **Elektronikmodul** (RN) einen Kartenspeicher (RN) zur Aufnahme topographischer Informationen enthält.

(4) **Federungssystem** nach einem der Ansprüche 2 oder 3, gekennzeichnet, dass der Lagesensor (RN) oder der Beschleunigungssensor einen mikromechanischen Sensor enthält oder dass der Lagesensor (RN) oder der Beschleunigungssensor einen mehrachsigen mikromechanischen Sensor enthält.

(5) **Federungssystem** nach einem der Ansprüche 2 bis 4, dadurch gekennzeichnet, dass das Bedienelement (RN) zusammen mit einem Bremshebel und/oder einem Schalthebel an einem gemeinsamen Befestigungselement (RN) angeordnet ist.

(6) **Federungssystem** nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, dass der Aktor (RN) ausgewählt ist aus einem Elektromotor und/oder einem Magnetsteller und/oder einem Piezosteller und/oder einem Steller, welcher eine Formgedächtnislegierung enthält oder daraus besteht.

(7) **Federungssystem** nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, dass das Steuerelement (RN) und/oder das **Elektronikmodul** (RN) zumindest einen ersten Betriebszustand und zumindest einen zweiten Betriebszustand aufweisen, wobei im zweiten Betriebszustand gegenüber dem ersten Betriebszustand die Anzahl ausführbarer Funktionen reduziert und der Energieverbrauch vermindert ist.

(8) **Federungssystem** nach einem der Ansprüche 2 bis 7, dadurch gekennzeichnet, dass der Positionssensor (RN) zumindest ein Funknavigationssystem enthält.

Vielen Dank

www.intergator.de

