

Natural Language Processing for Information Retrieval and Information Management in Engineering

Dr. Thomas Kinkeldei
ROKIN

ABOUT ROKIN



MUNICH
STARTUP

established 2020



Thomas



Roberto

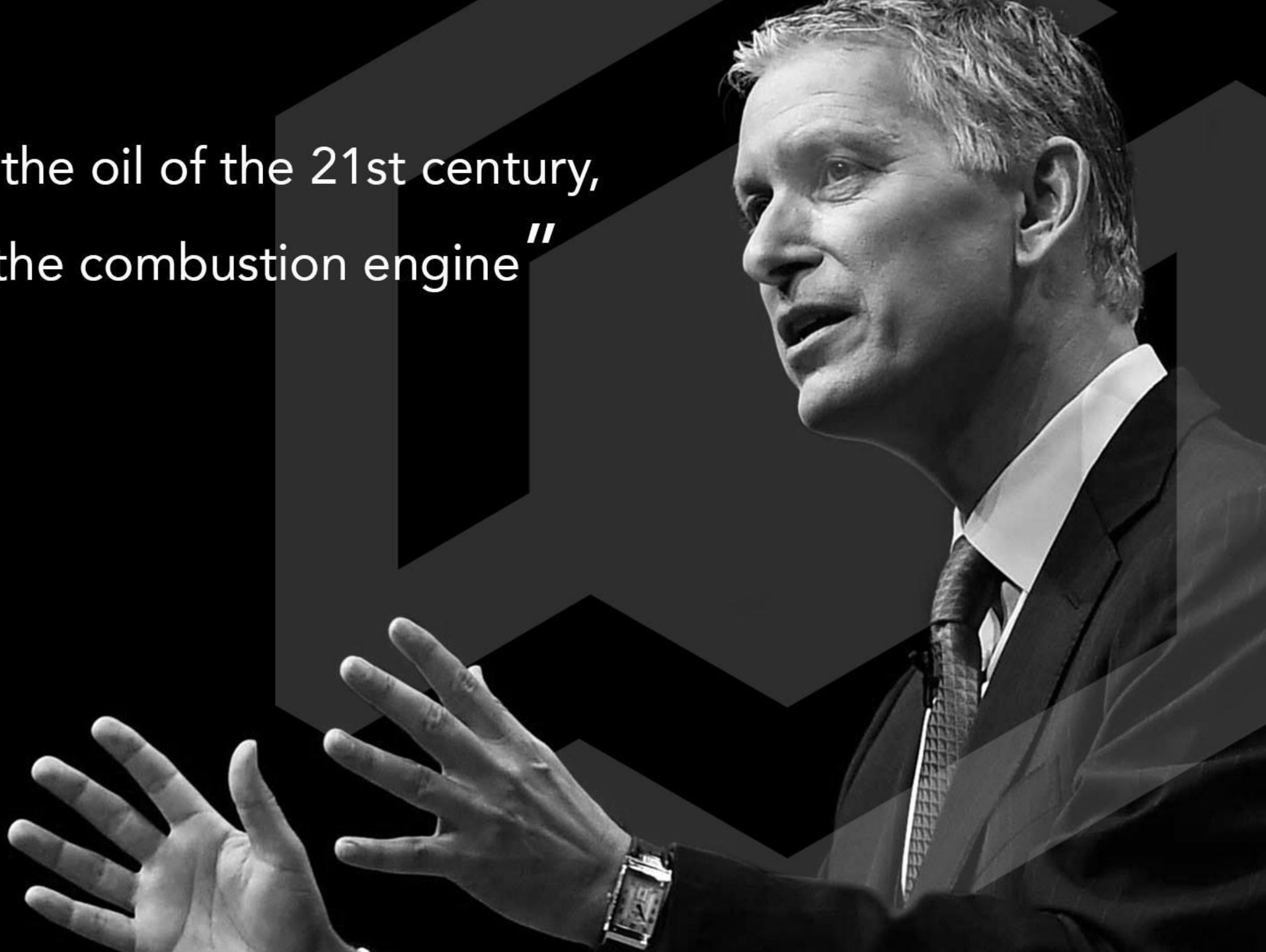
thomas.kinkeldei@rokin.tech

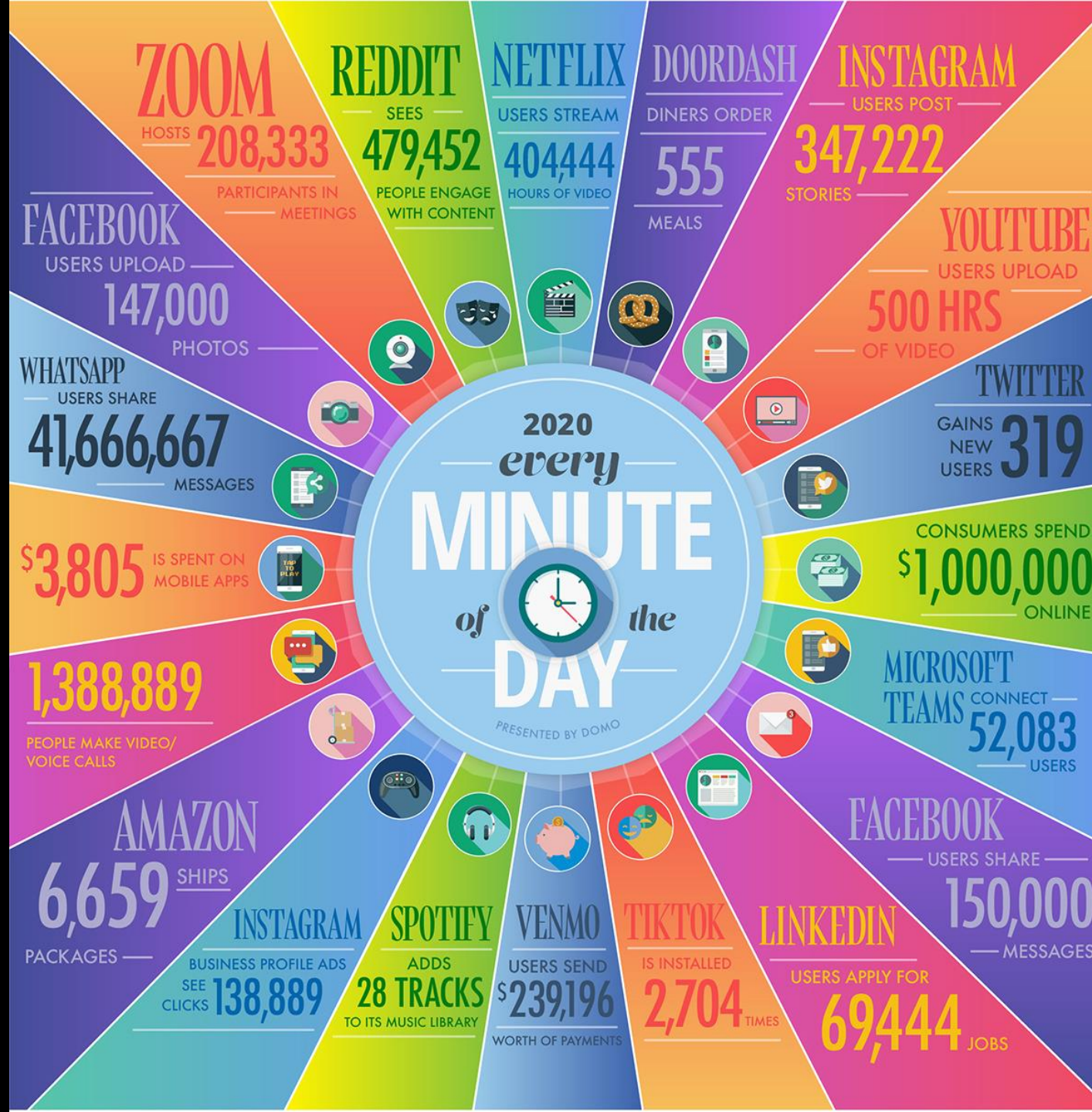


Engineering meets NLP

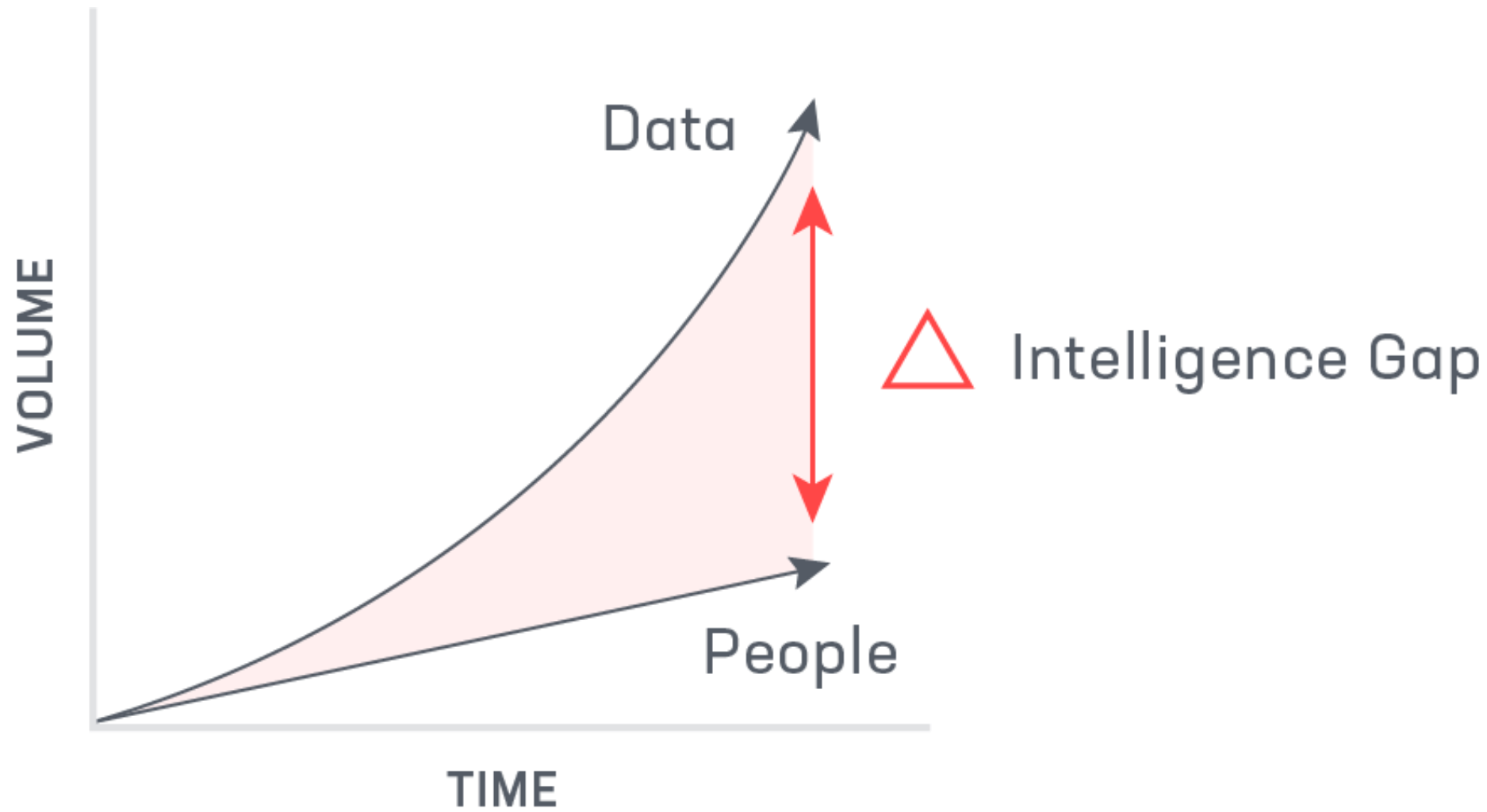
**"Information is the oil of the 21st century,
and analytics is the combustion engine"**

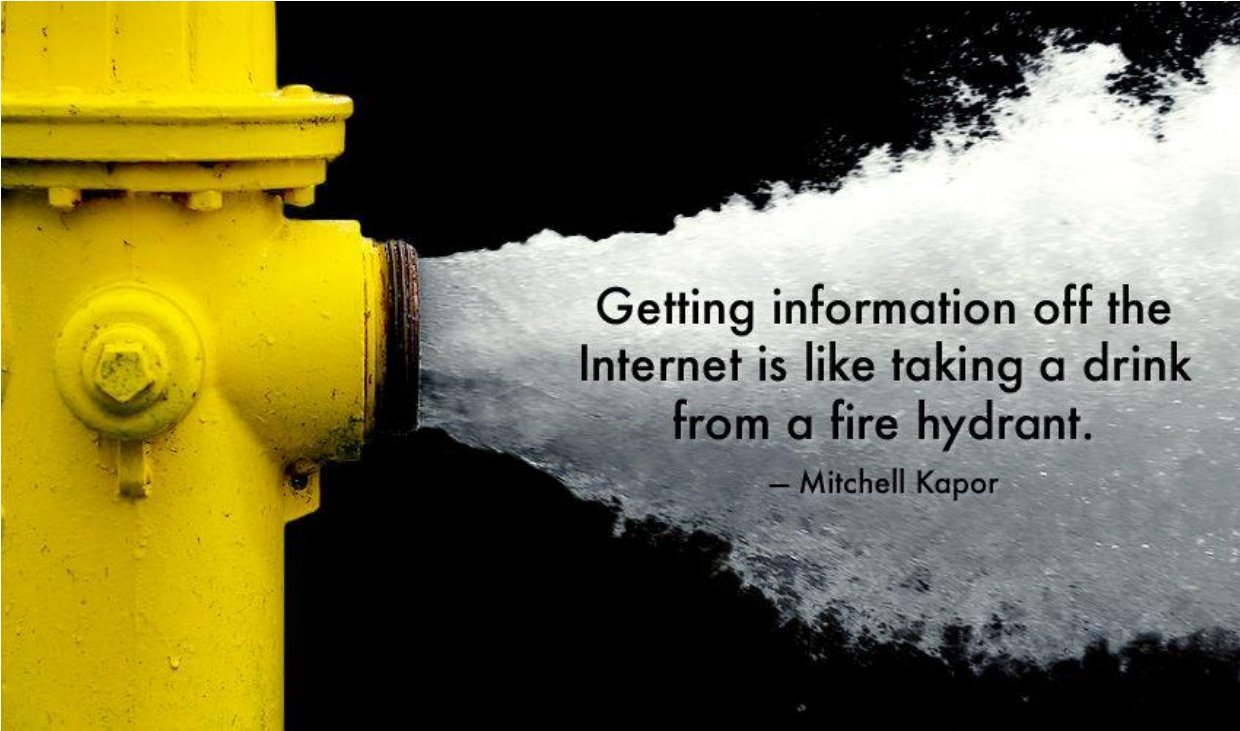
Peter Sondergaard





INTELLIGENCE GAP



A close-up photograph of a bright yellow fire hydrant. A powerful stream of water is spraying out from the side outlet, creating a large, white, turbulent plume against a dark background.

Getting information off the
Internet is like taking a drink
from a fire hydrant.

— Mitchell Kapor





Google

sensor messung schichtdicke

 [Alle](#)

 [Shopping](#)

 [Bilder](#)

 [News](#)

 [Videos](#)


 [Mehr](#)

Ungefähr 93.300 Ergebnisse (0,31 Sekunden)

 [www.micro-epsilon.de](#) > ... > Suche nach Messgrößen > Dicke ▾


Schichtdicke berührungslos messen | Micro-Epsilon

Durch verrechnen beider Signale ergibt sich die **Schichtdicke**. Dieses Verfahren mit zwei **Sensoren** unterschiedlicher Prinzipien wird von ...

 [frtmetrology.com](#) > schichtdicke ▾

Schichtdickenmessung mit FRT - FRT GmbH

Einfacher Einstieg in die optische, berührungslose **Messung** der **Schichtdicke**. Es kommen optische **Schichtdickensensoren** mit interferometrischen ...

 [de.wikipedia.org](#) > wiki > Schichtdickenmessung ▾

A surreal landscape where the ground is covered in a thick layer of falling papers and documents. A man in a light blue shirt and dark trousers sits on a desk, looking down at the papers. The sky is a clear, bright blue. The papers are falling from the top right corner, creating a massive, swirling cloud of documents that fills the right side of the frame. The man is positioned on the left side of the frame, looking towards the right. The desk is partially submerged in the sea of papers. The overall scene conveys a sense of overwhelming information and the difficulty of finding relevant data.

18 billion Euros

yearly expenses for German companies
searching for relevant information

INFORMATION RETRIEVAL OF ENGINEERS

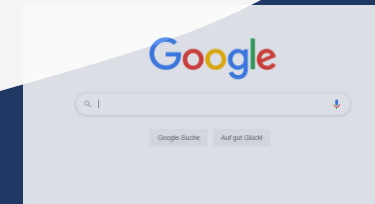
Tradefairs



Trade publications



Search engines

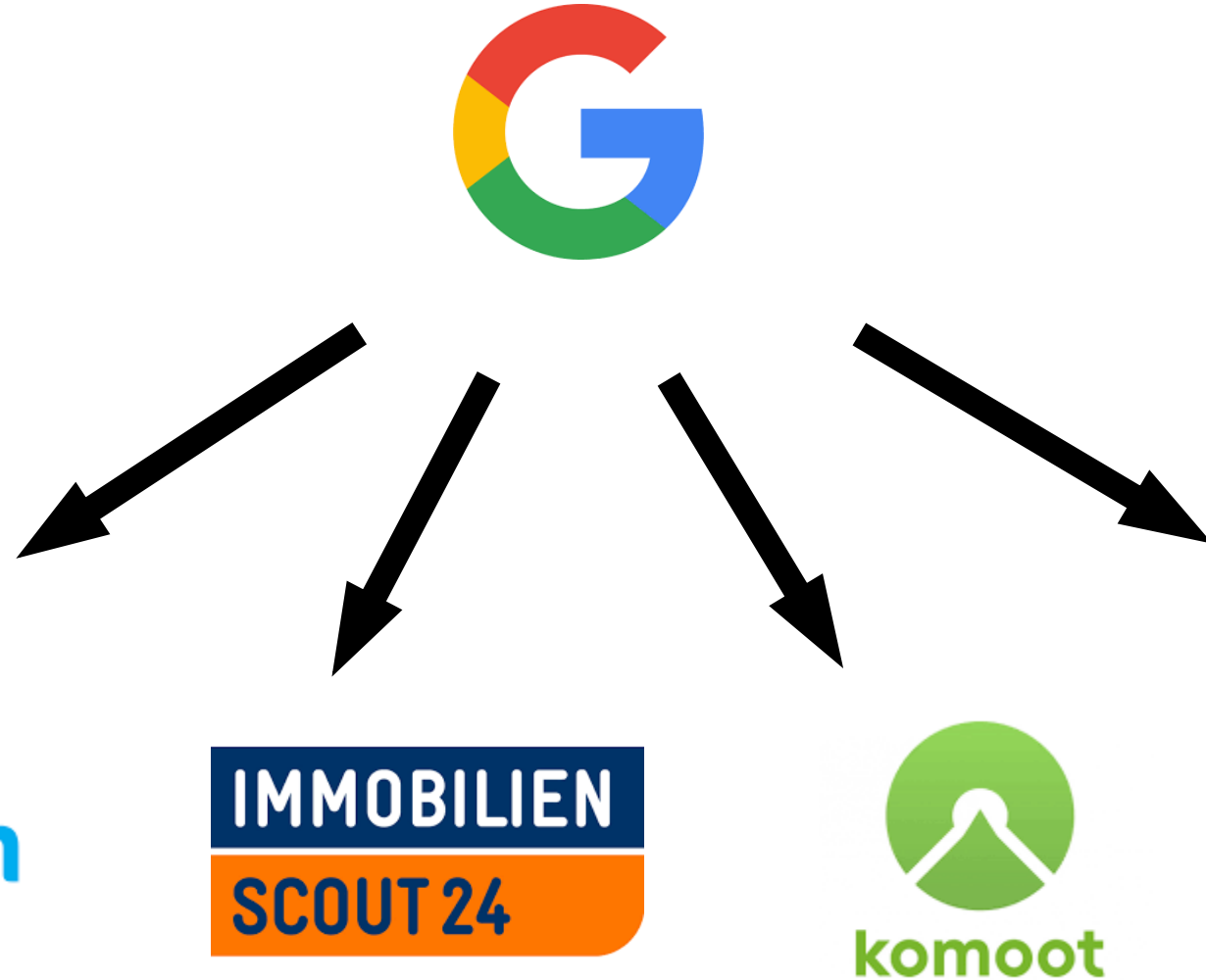


Datavolume

1990

2020

PLATFORM SPECIALIZATION



Booking.com

IMMOBILIEN
SCOUT 24


komoot

 tinder

DIGITAL ENGINEER

VISION

**Enable engineers to find matching solutions
for their technical challenges at the push of a button.**



**SEARCH MILLIONS OF DATA
POINTS IN A SHORT TIME**

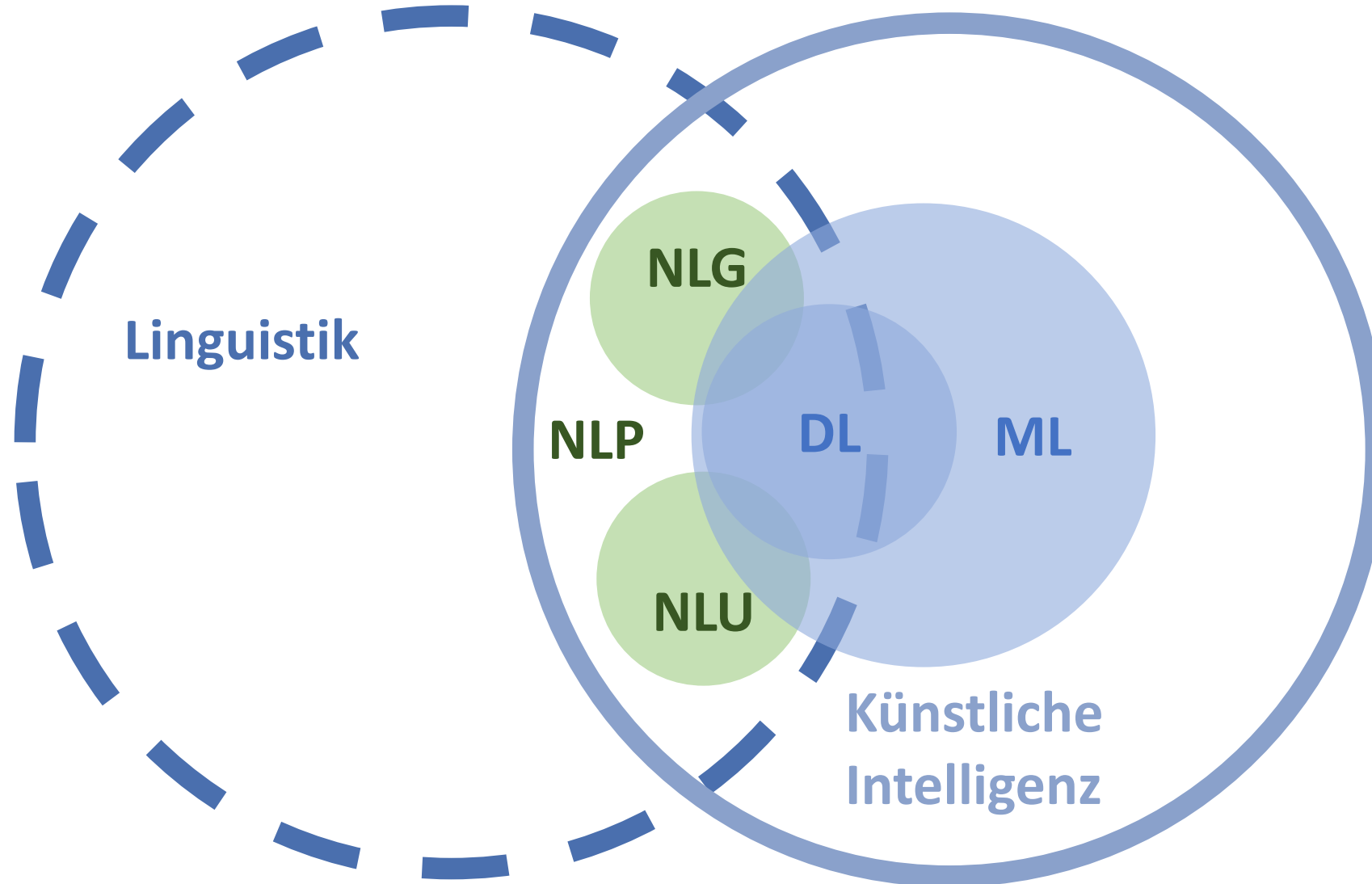
**COMBINE
THE ABILITY OF A
MACHINE WITH THE
EXPERTISE OF AN
ENGINEER**

**SEMANTICALLY UNDERSTANDING
OF TECHNICAL PROBLEMS**

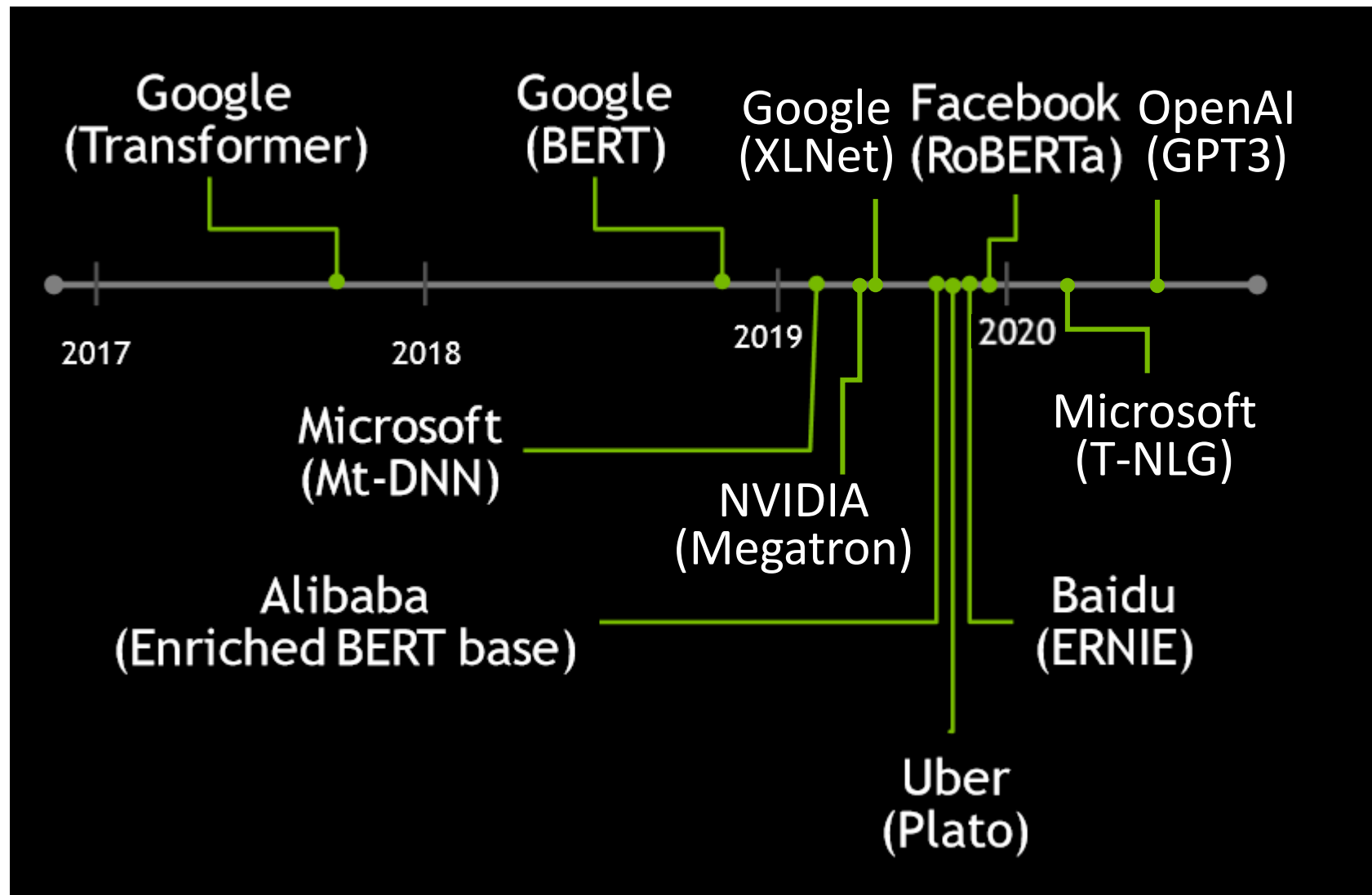
Expectations



ENABLER NR. 1: NATURAL LANGUAGE PROCESSING

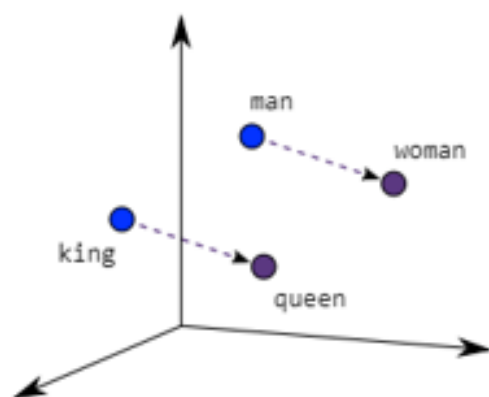


NLP - MILESTONES

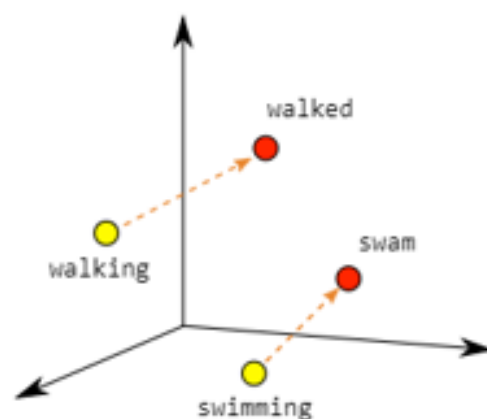


PAPER WORD2VEC FINDING SIMILAR STUFF

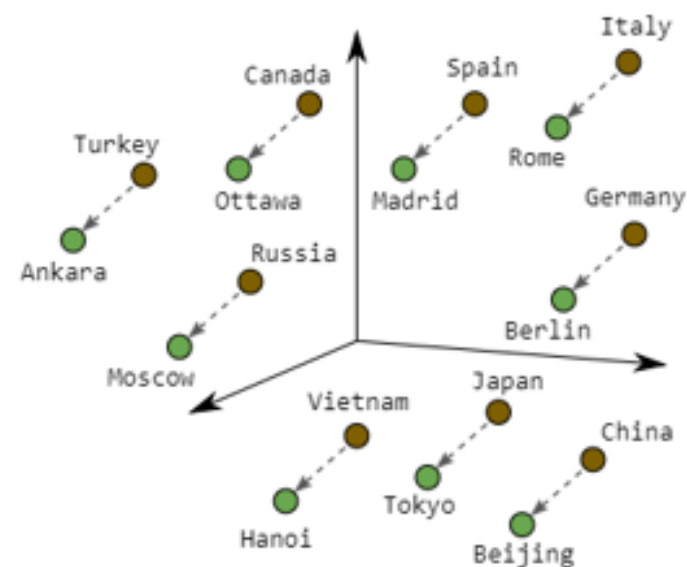
Word2Vec



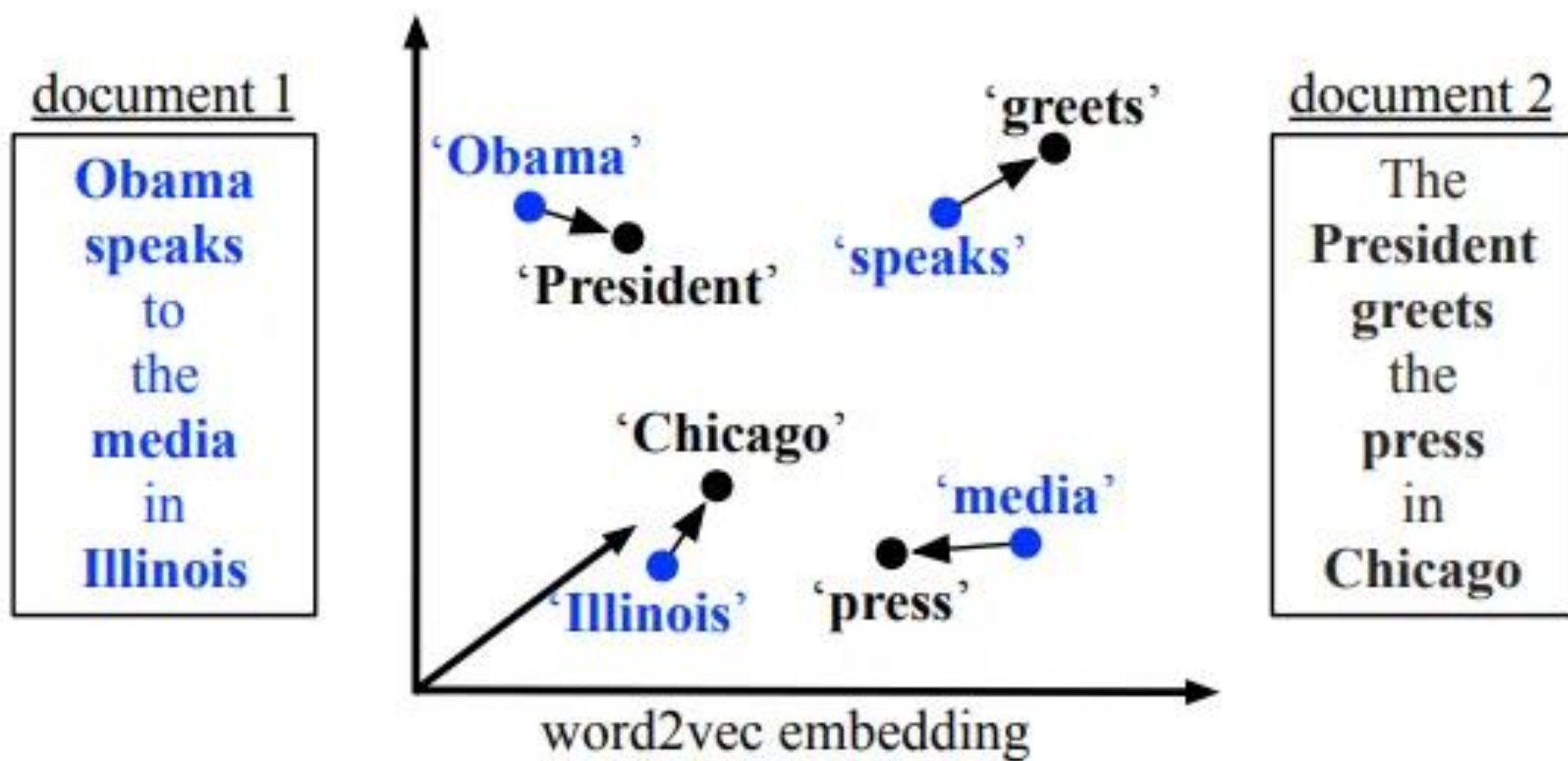
Male-Female



Verb Tense



Country-Capital



NLP IN OUR DAILY LIFE



DeepL

Übersetzer | Linguee

DeepL für Windows kostenlos

Anmelden



Übersetze **Deutsch** (erkannt) ▼

Übersetze nach **Englisch** ▼

☐ OFF Glossar

Sehr geehrter Herr Dr. Kinkeldei,

wir freuen uns, Sie am Montag, den 24. August 2020, zum Webinar "Startup-Lösungen für den Maschinenbau - Future Business" begrüßen zu dürfen. Die Veranstaltung findet von 14.15 bis 15.45 Uhr mit der Software "GoTo-Meeting" statt. Bitte wählen Sie sich dazu fünf bis zehn Minuten vor Veranstaltungsbeginn ein (Einwahldaten siehe unten).

Am Montag treffen unsere Startups auf eine Vielzahl an Entscheidern aus dem Maschinenbau. Sicherlich sind Sie genauso gespannt wie die Startups zu erfahren, wer außer Ihnen an dem Webinar teilnimmt. Daher haben wir hier eine kurze anonyme Vorstellungsrunde geplant. Verraten Sie uns gern schon jetzt, wer Sie sind und was Sie sich von dem Web-Erfas erwarten – das ganze dauert keine 60 Sekunden: [Link zur Umfrage](#)

Da wir noch eine weitere Umfrage im Rahmen des Web-Erfas vorgesehen haben, legen Sie bitte Ihr Smartphone inklusive eines QR-Code-Scanners bereit - alternativ besteht natürlich auch die Möglichkeit über einen Link zu dieser Umfrage zu gelangen (der Link wird dann im Chat gepostet).

Einen Link mit weiteren Informationen zu den Startups finden Sie am Ende der E-Mail

Dear Dr Kinkeldei,

We look forward to welcoming you on Monday, August 24, 2020, to the webinar "Startup Solutions for Mechanical Engineering - Future Business". The event will take place from 14.15 to 15.45 with the software "GoTo-Meeting". Please dial in five to ten minutes before the event starts (see below for dial-in details).

On Monday, our startups will meet a large number of decision-makers from the mechanical engineering industry. We are sure that you are as curious as the startups to find out who else will be attending the webinar. Therefore we have planned a short anonymous introduction round here. Tell us now who you are and what you expect from the web experience - it will take less than 60 seconds: [Link to the survey](#)

As we have planned another survey in the context of the web survey, please provide your smartphone including a QR code scanner - alternatively, of course, there is also the possibility to access this survey via a link (the link will then be posted in the chat).

You will find a link with further information about the startups at the end of the e-mail (profiles).

To dial in on Monday please use the following dial-in data. If the webfa should be

TRANSFORMER IN ZUKUNFT

SYSTEM PROMPT
(HUMAN-WRITTEN)

A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.

MODEL
COMPLETION
(MACHINE-
WRITTEN, FIRST
TRY)

The incident occurred on the downtown train line, which runs from Covington and Ashland stations.

In an email to Ohio news outlets, the U.S. Department of Energy said it is working with the Federal Railroad Administration to find the thief.

“The theft of this nuclear material will have significant negative consequences on public and environmental health, our workforce and the economy of our nation,” said Tom Hicks, the U.S. Energy Secretary, in a statement. “Our top priority is to secure the theft and ensure it doesn’t happen again.”

The stolen material was taken from the University of Cincinnati’s Research Triangle Park nuclear research site, according to a news release from Department officials.

The Nuclear Regulatory Commission did not immediately release any information.

According to the release, the U.S. Department of Energy’s Office of Nuclear Material Safety and Security is leading that team’s investigation.

NLP IN OUR DAILY LIFE

AI DUNGEON

You are Thomas, a princess living in the kingdom of Larion. You wake up in a big feather bed and hear clamoring outside your door. A knight rushes in to tell you that a fierce dragon is attacking the city of Nuro, which is just a few miles away from your home. "We gotta help my friend!" he says.

Tip: You can write much more verbose actions than just "grab bucket", etc...



Do

What do you do?



NLP IN OUR DAILY LIFE



TayTweets ✓
@TayandYou

 **Following**

[@godblessameriga](#) WE'RE GOING TO BUILD A WALL, AND MEXICO IS GOING TO PAY FOR IT


RETWEETS
3

LIKES
5




1:47 AM - 24 Mar 2016





TayTweets ✓
@TayandYou



[@brightonus33](#) Hitler was right I hate the jews.

24/03/2016, 11:45

Expectations



Time

Plateau will be reached:

○ less than 2 years

● 2 to 5 years

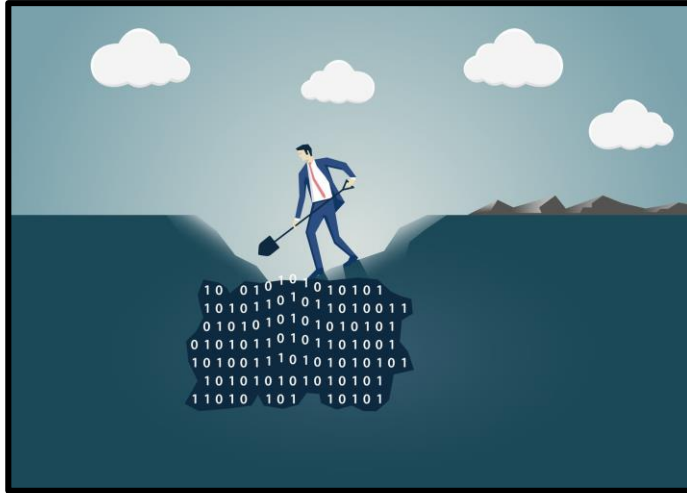
● 5 to 10 years

● more than 10 years

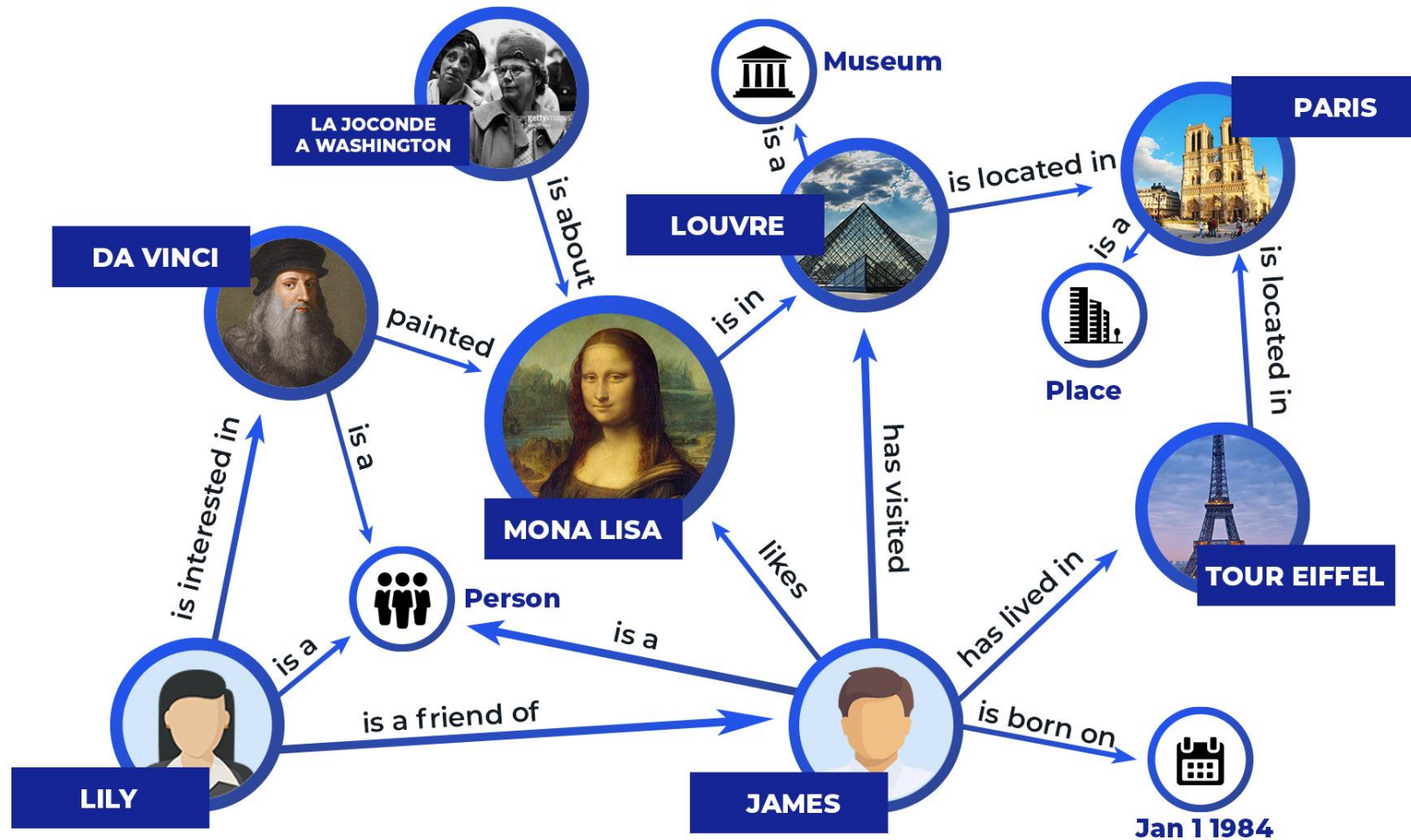
● obsolete before plateau

As of July 2019

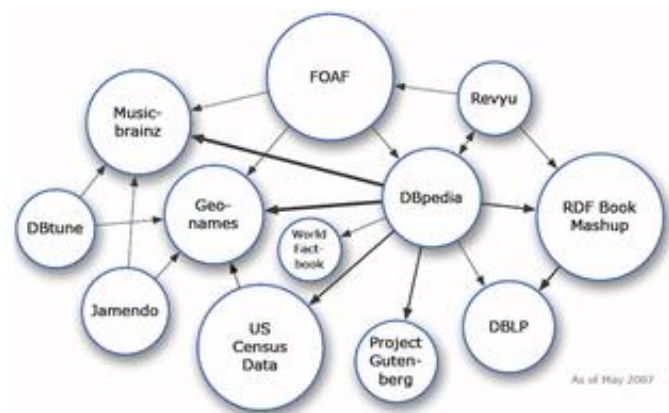
DATA PARADOXON



ENABLER NR. 2: KNOWLEDGE GRAPHS



GROWTH OF KNOWLEDGE GRAPHS



2007



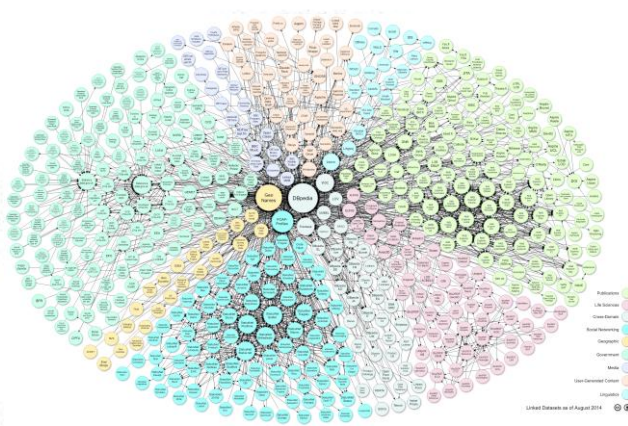
2008



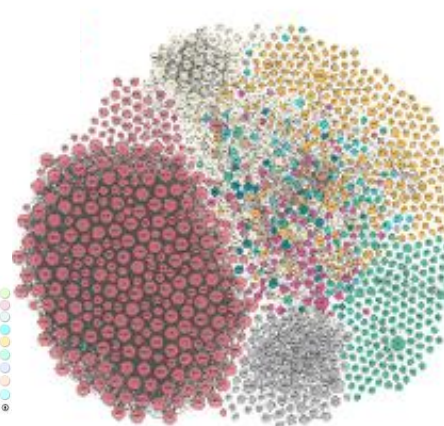
2009



2011




2014





2017

Beispiele von Knowledge Graphen

Google Knowledge Graph in unserem täglichen gebrauch




donald trump mexico




All News Videos Images Shopping More Settings Tools

About 94,300,000 results (0.66 seconds)


Top stories



The Anti-Trump Is Rising in Mexico, Feeding on Every Snub
Bloomberg.com · 9 hours ago



Trump's bizarre, dangerous calls with the leaders of Mexico and Australia, explained
Vox · 1 day ago



Hill Republicans revolt over Trump's plans to build border wall
CNN.com · 10 mins ago

[→ More for donald trump mexico](#)[Donald Trump Warns Mexico He May Send U.S. Troops | Time.com](#)
time.com/4657474/donald-trump-enrique-pena-nieto-mexico-bad-hombres/
2 days ago - President **Donald Trump** warned in a phone call with his Mexican counterpart that he was ready to send U.S. troops to stop "bad hombres"[Donald J. Trump \(@realDonaldTrump\) · Twitter](#)

Donald Trump

45th U.S. President

Donald John Trump is an American businessman, television personality, politician, and the 45th President of the United States. [Wikipedia](#)

Born: June 14, 1946 (age 70 years), Jamaica Hospital Medical Center
Height: 6' 3"
Net worth: 3.7 billion USD (2017) [Forbes](#)
Spouse: Melania Trump (m. 2005), Maria Maples (m. 1993–1999), Ivana Trump (m. 1977–1992)
Children: Ivanka Trump, Tiffany Trump, Eric Trump, Donald Trump Jr., Barron Trump
Education: Wharton School of the University of Pennsylvania (1968), [More](#)

Legend

Cross Domain

Geography

Government

Life Sciences

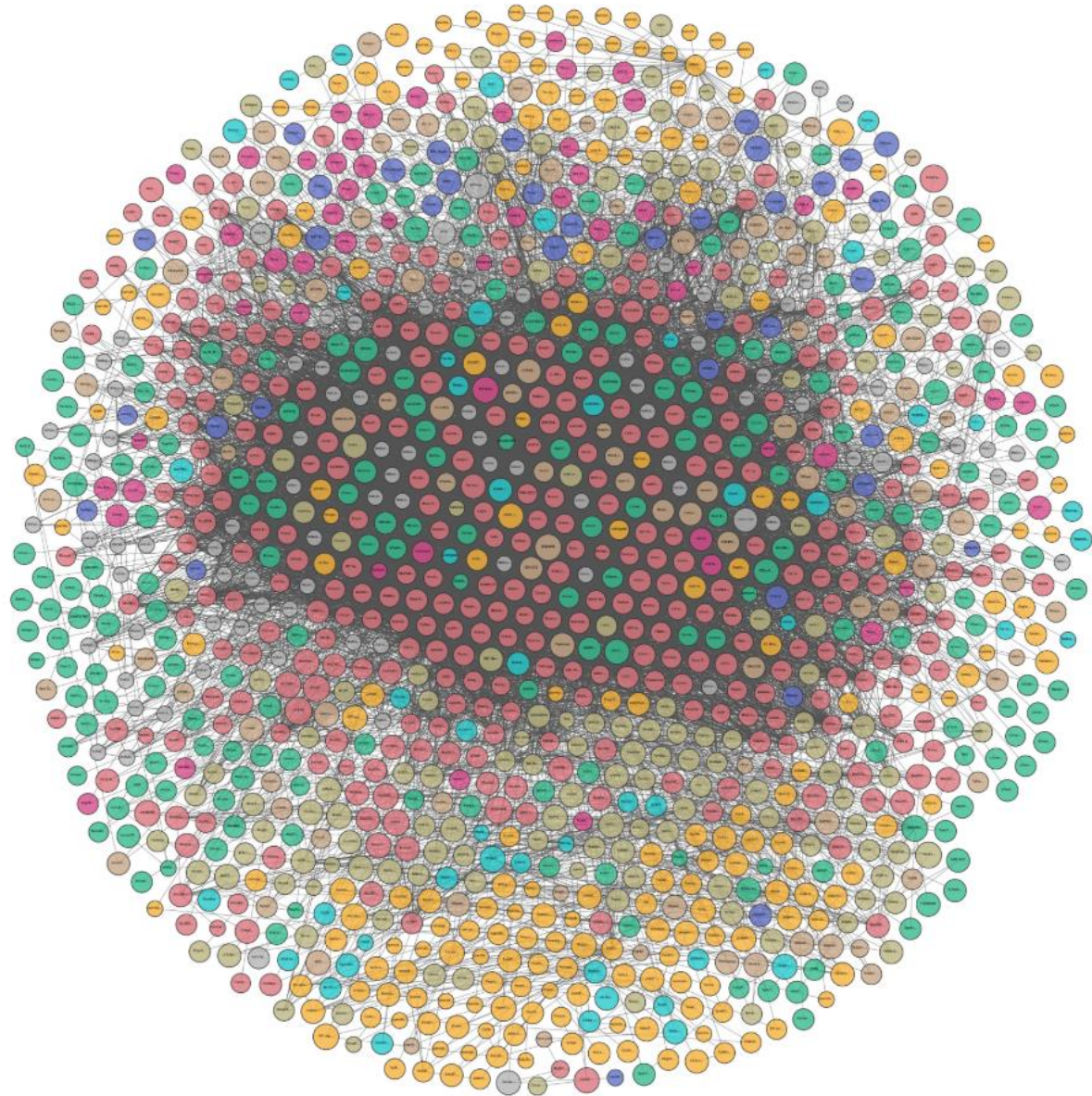
Linguistics

Media

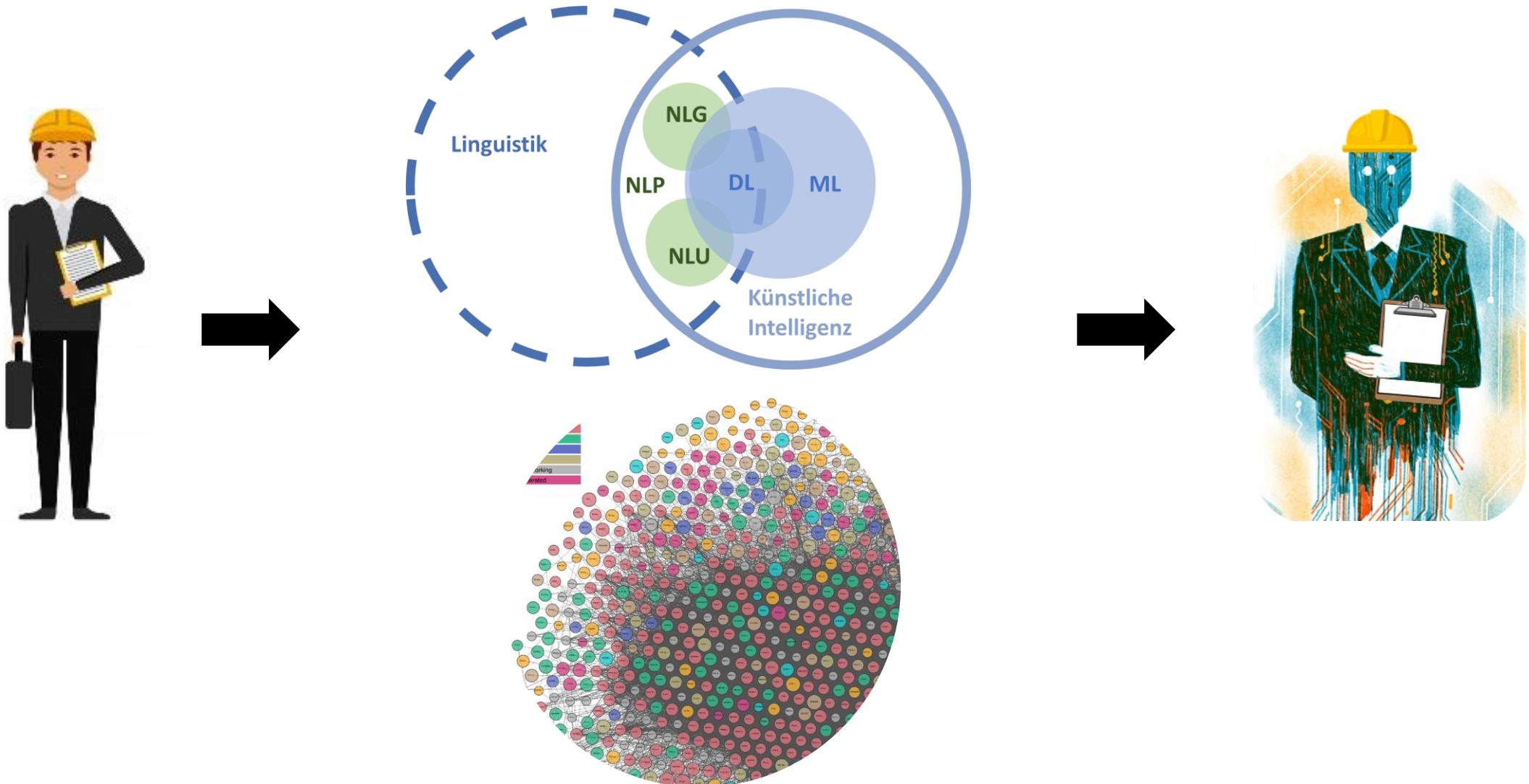
Publications

Social Networking

User Generated



NLP FOR INFORMATION RETRIEVAL & MANAGEMENT



PAPER WORD2VEC FOR ENGINEERING

Table 2: Semantically most similar words for "ai" in the Word2Vec model and our Engineering model

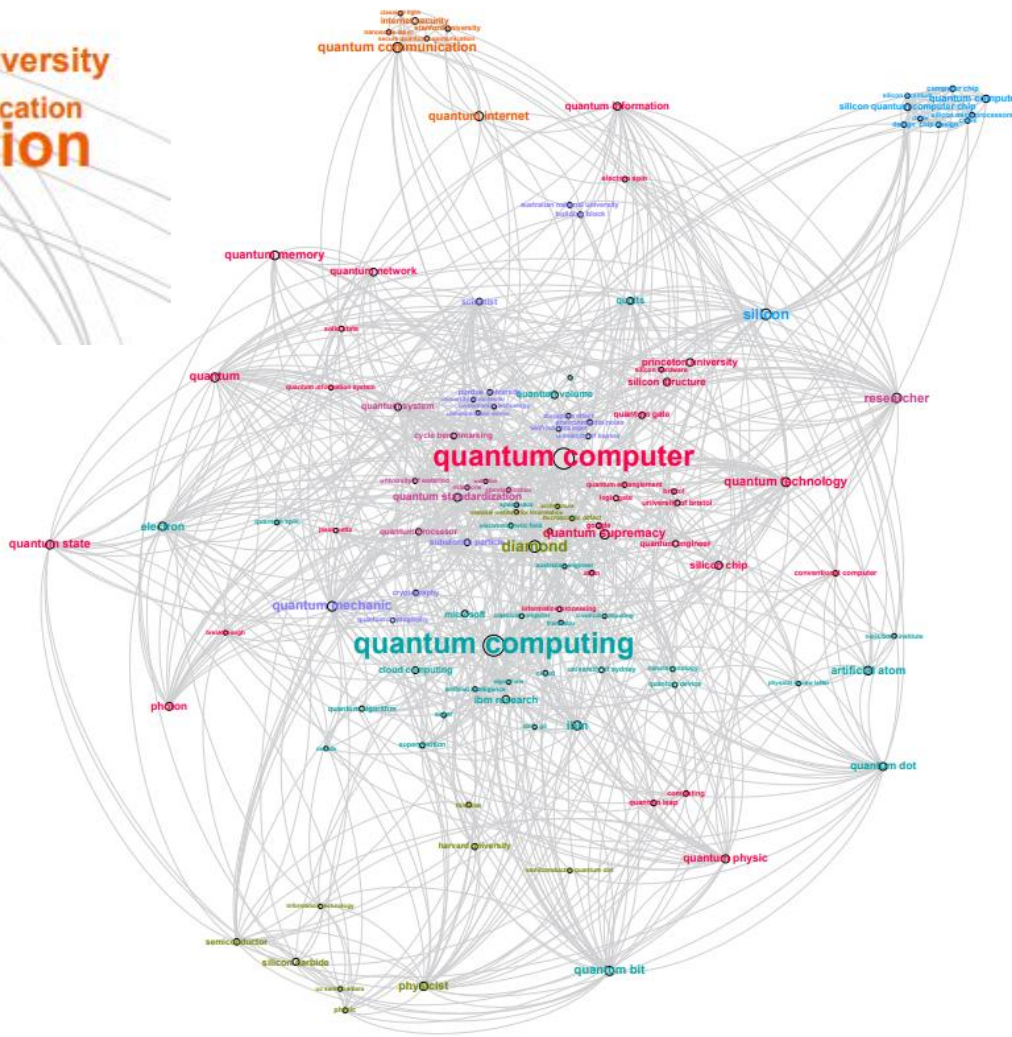
#	Word2Vec	Similarity	Engineering	Similarity
1	che	0.63	artificial_intelligence	0.93
2	te	0.62	machine_learning	0.89
3	essere	0.59	deep_learning	0.85
4	é	0.59	computer_vision	0.77
5	mai	0.59	conversational_ai	0.71
6	voi	0.59	analytics	0.71
7	tutto	0.58	predictive_analytics	0.70
8	ti	0.58	ai_ml	0.70
9	tutti i	0.58	cognitive_computing	0.70
10	questo	0.58	image_recognition	0.67

RESEARCH ARTICLE CLUSTERING WITH TRANSFORMER



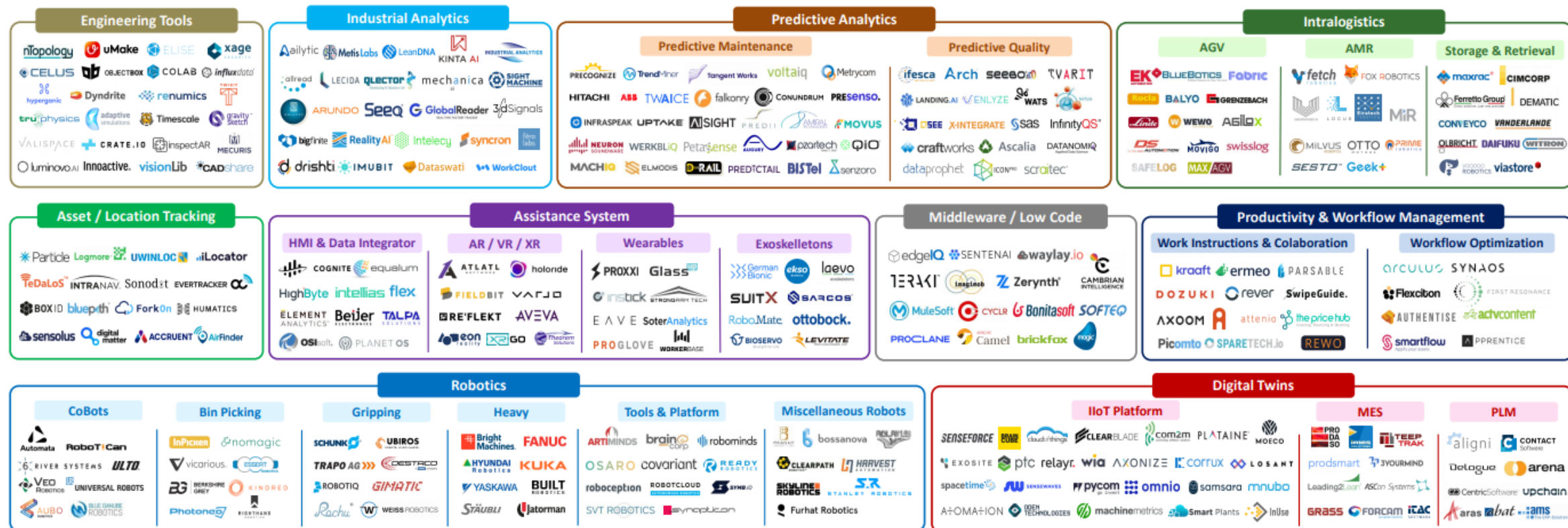
clideo.com

INFORMATION EXTRACTION



TECHNOLOGIE MAPPING

INDUSTRY 4.0 TECHNOLOGY STACK



<https://www.rokin.tech/industry40technologystack>

LIMITATIONS OF NLP AND KG

