



Why natural language is next step in the AI evolution

Nico Kreiling

Karlsruhe, 1.10.2019



Nico Kreiling

- › Data Scientist @ inovex
- › Twitter: nicokreiling
- › www.techtiefen.de

Language is important...

- › Most information is available in either written or spoken words
- › Language projects fundamental real world concepts
- › Speech is everywhere

...but really difficult



Vision is easy

Walking or even climbing can be learned fast



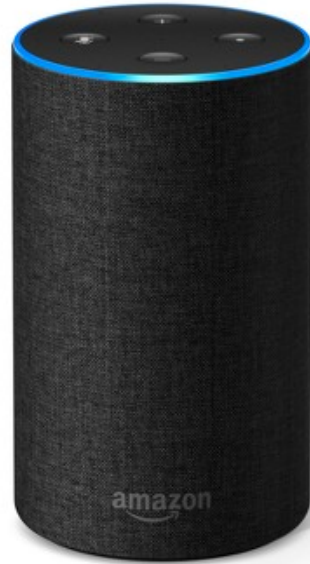
But language
is complicated



Why is language complicated ?

- › Language has many syntactic relationships
Stoiber probably thought, he was making a clear point
- › Language understanding often requires external knowledge
Pippi feeds Mr. Nilsson regularly.
- › Language heavily depends on the context
„The president has a great vision“
- › Irony and sarcasm are topics on their own

But it works...



...how does it?

How does modern NLP work?

Word-Vectors

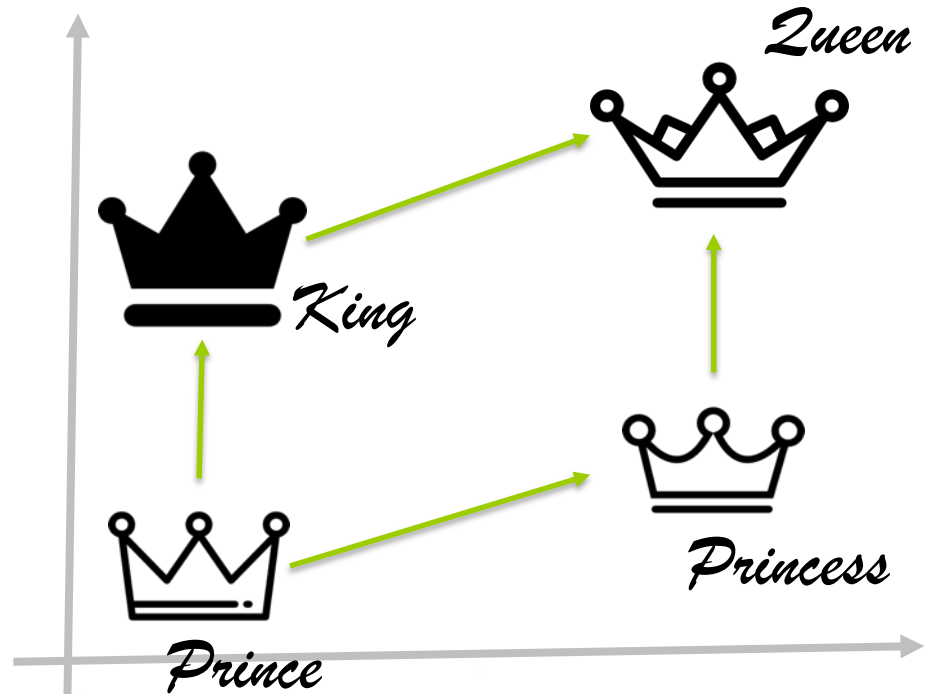
Language
is
more
than
words



0,11349057	0,84879879	0,93261575	...	0,83844097	0,79705175
0,60343267	0,26889299	0,51101431	...	0,79602685	0,24950905
0,18796149	0,83380881	0,65787064	...	0,29058833	0,06187716
0,34244444	0,50918356	0,03611215	...	0,6951714	0,87274548
0,07952056	0,35248701	0,86835106	...	0,06748431	0,80904259

How does modern NLP work?

Word-Vectors



How does modern NLP work?

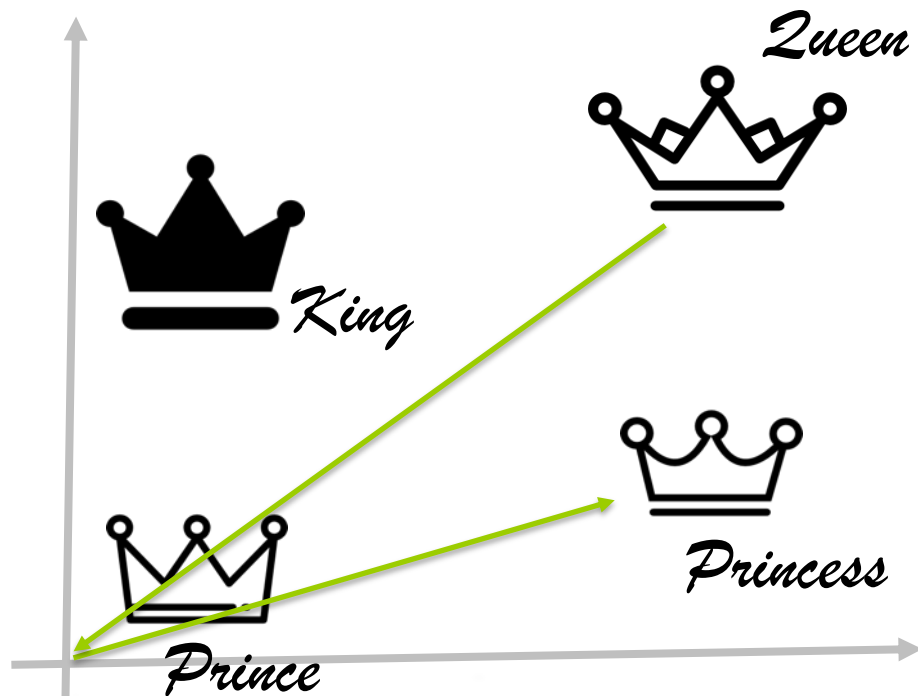
Word-Vectors

King

+ Princess

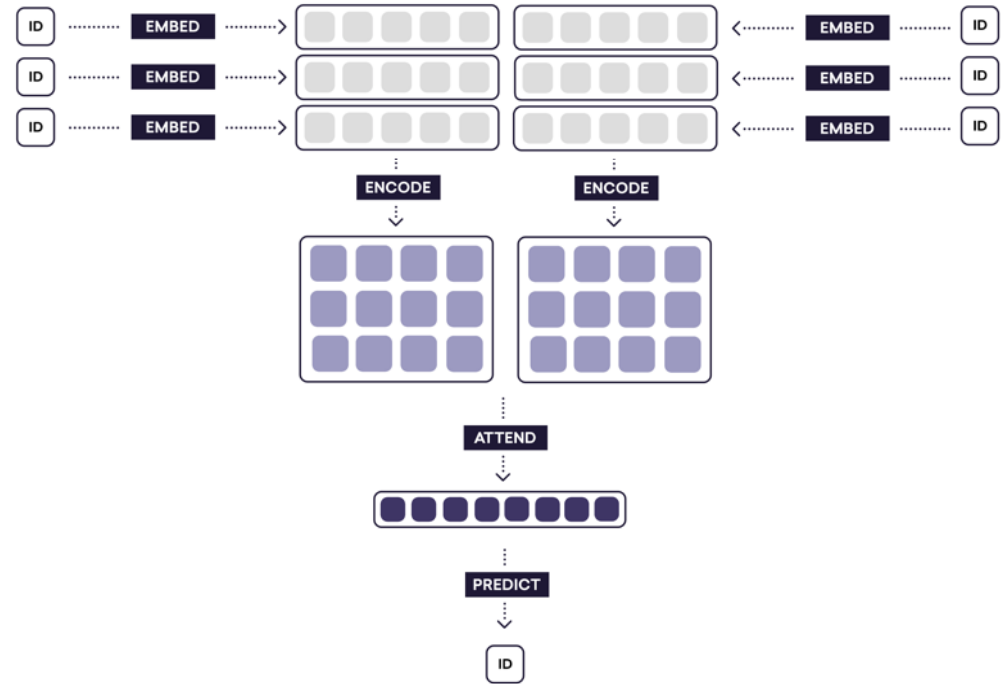
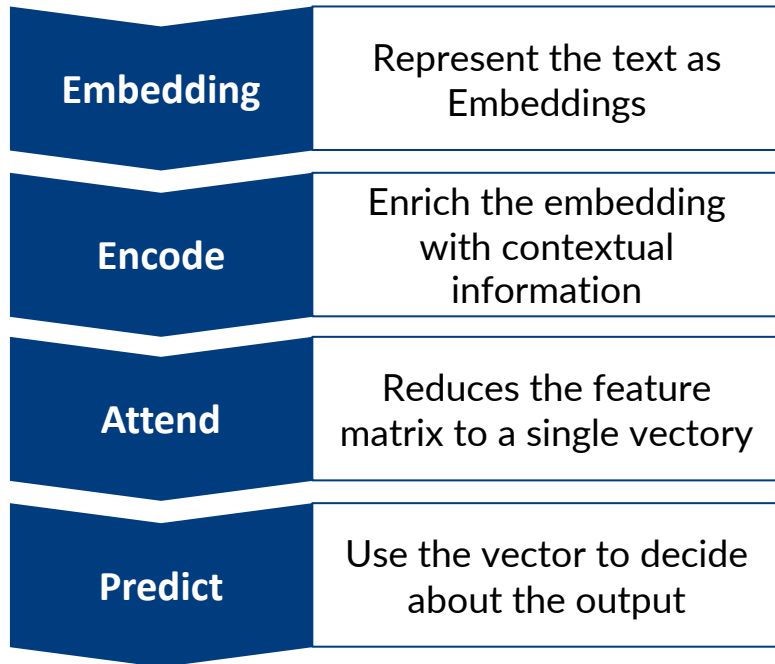
- Queen

= Prince



How does modern NLP work?

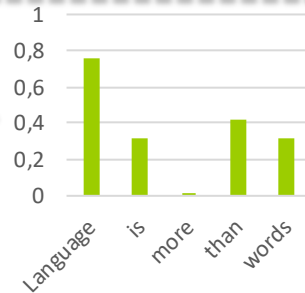
Encoder / Decoder Model



How does modern NLP work?

Transformer and Self-Supervised Learning

Language is **more** than words. **Grammar** is also important.



- › Self Attention
- › Mask some words
- › Randomly add a second sentence

How does modern NLP work?

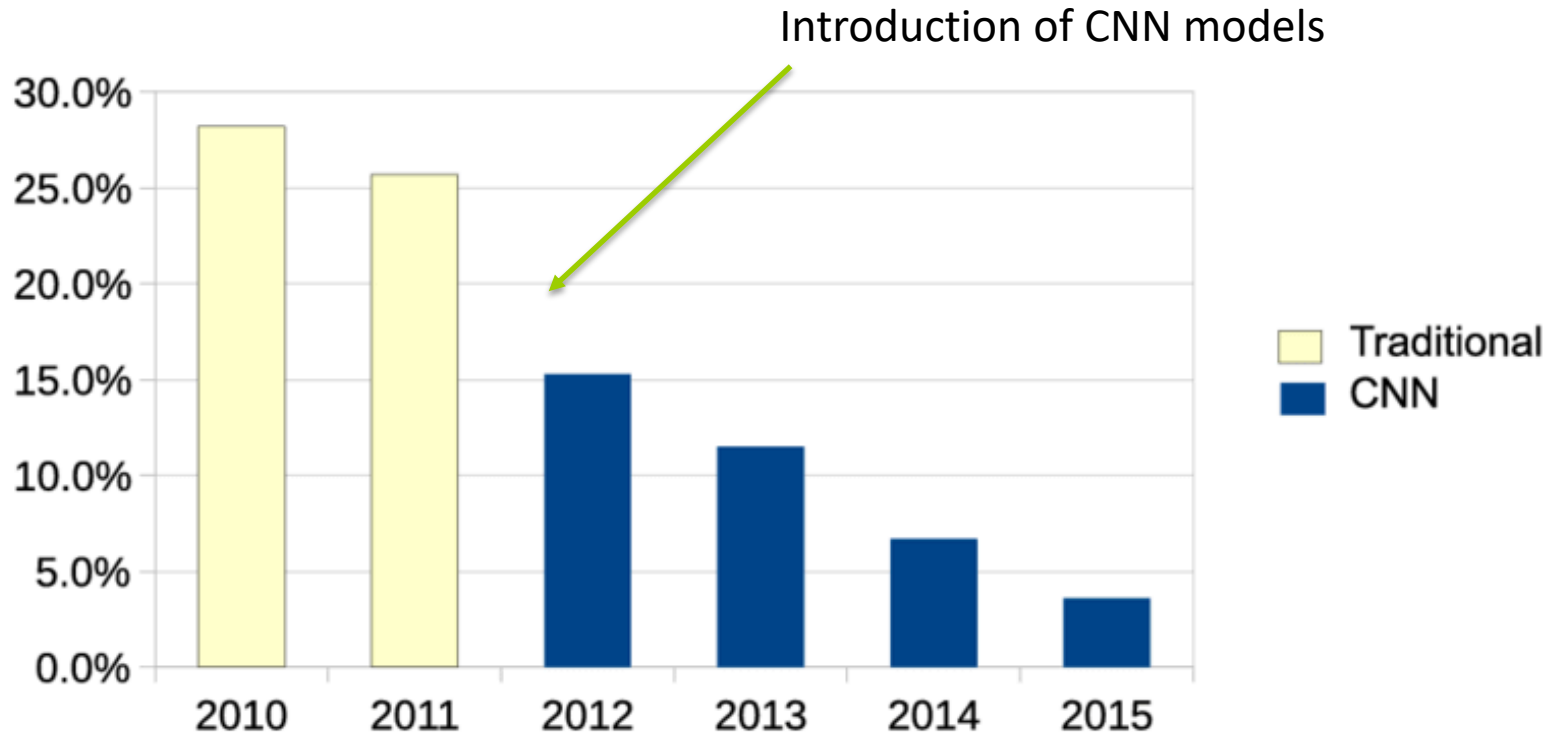
Big Language Models + Transfer learning

- › Bert Large
 - › 24-layer, 340M parameters
- › Bert Base (104 languages)
 - › 12-layer, 110M parameters
 - › 64 Tesla V100 GPUs for 79.2 hours
 - › Cost of training \$2074–\$12,571

But we can build upon this model!



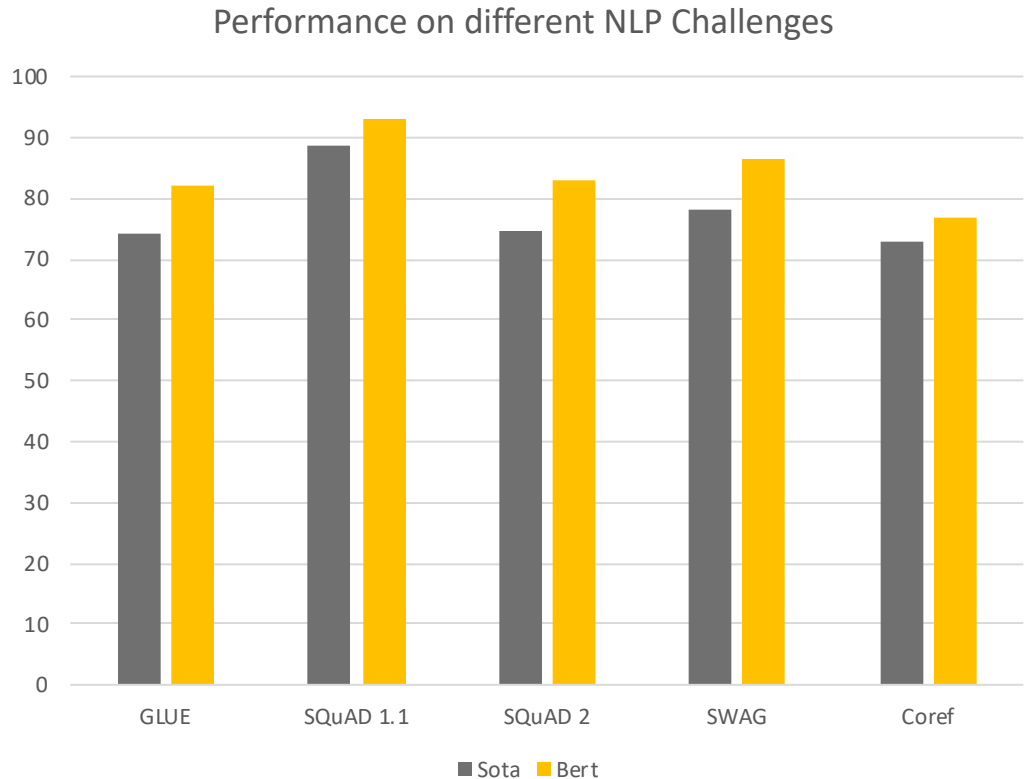
Solving Vision



top5 error rate of the annual winner of the ImageNet image classification challenge

Solving Speech

- › **GLUE:** Language Understanding
- › **SQuAD:** Question Answering
- › **SWAG:** Common Sense Reasoning
- › **Coref:** Co-Reference Resolution





Time for applications

Spotify



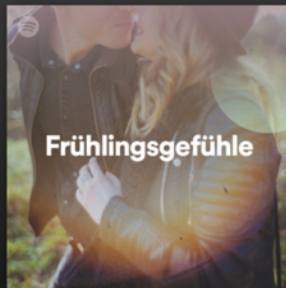
Songs as words



Word Vectors



Auto-Generated
Playlists



PLAYLIST

Frühlingsgefühle

Ganz vorsichtig streckt der Frühling seine Fühler aus. Mit dem passenden Soundtrack kommen die passende Stimmung.

Created by Spotify • 67 songs, 3 hr 58 min

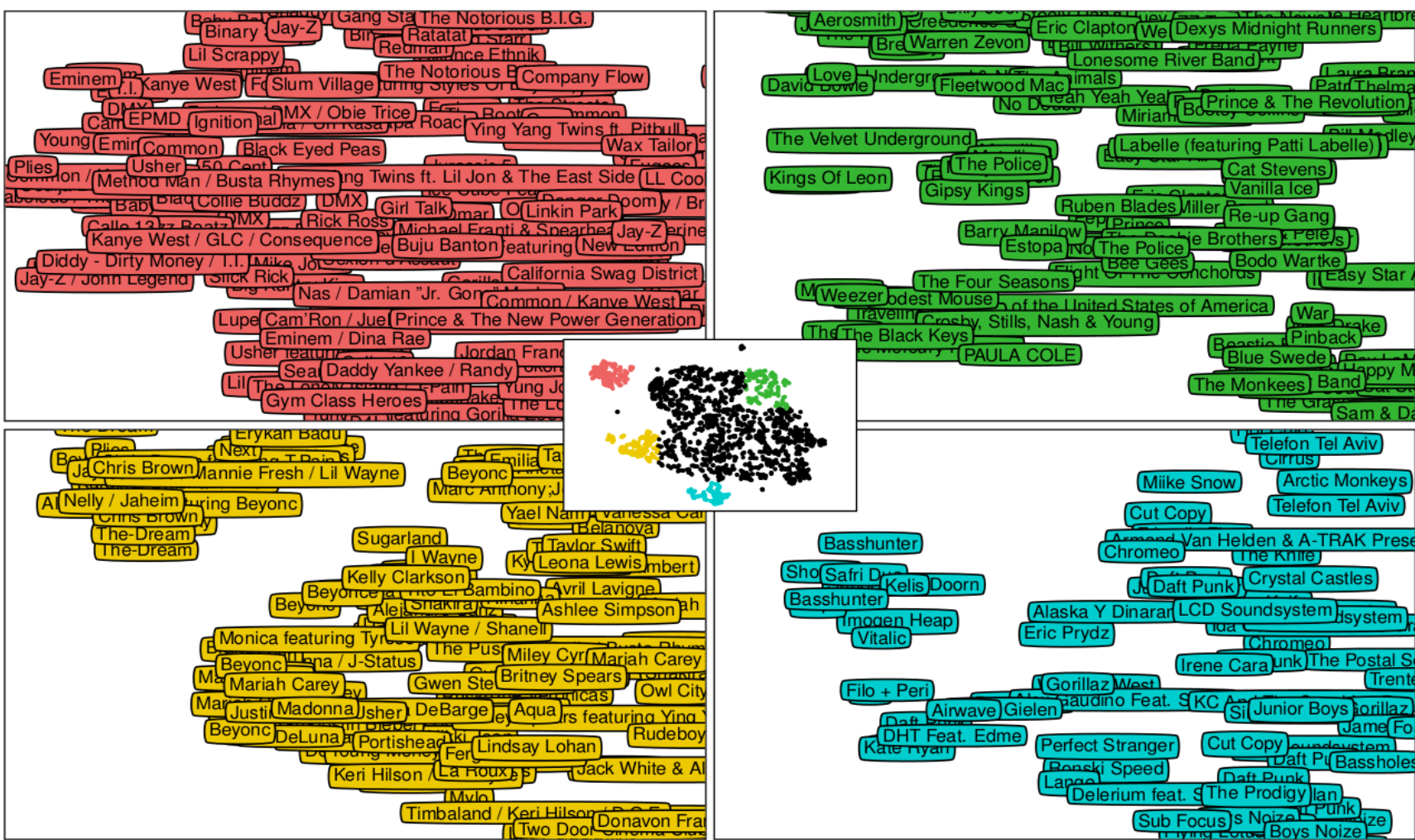
PLAY

FOLLOWING



Q Filter

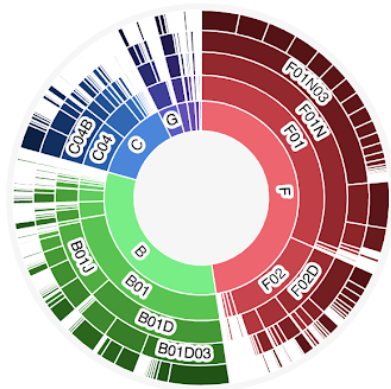
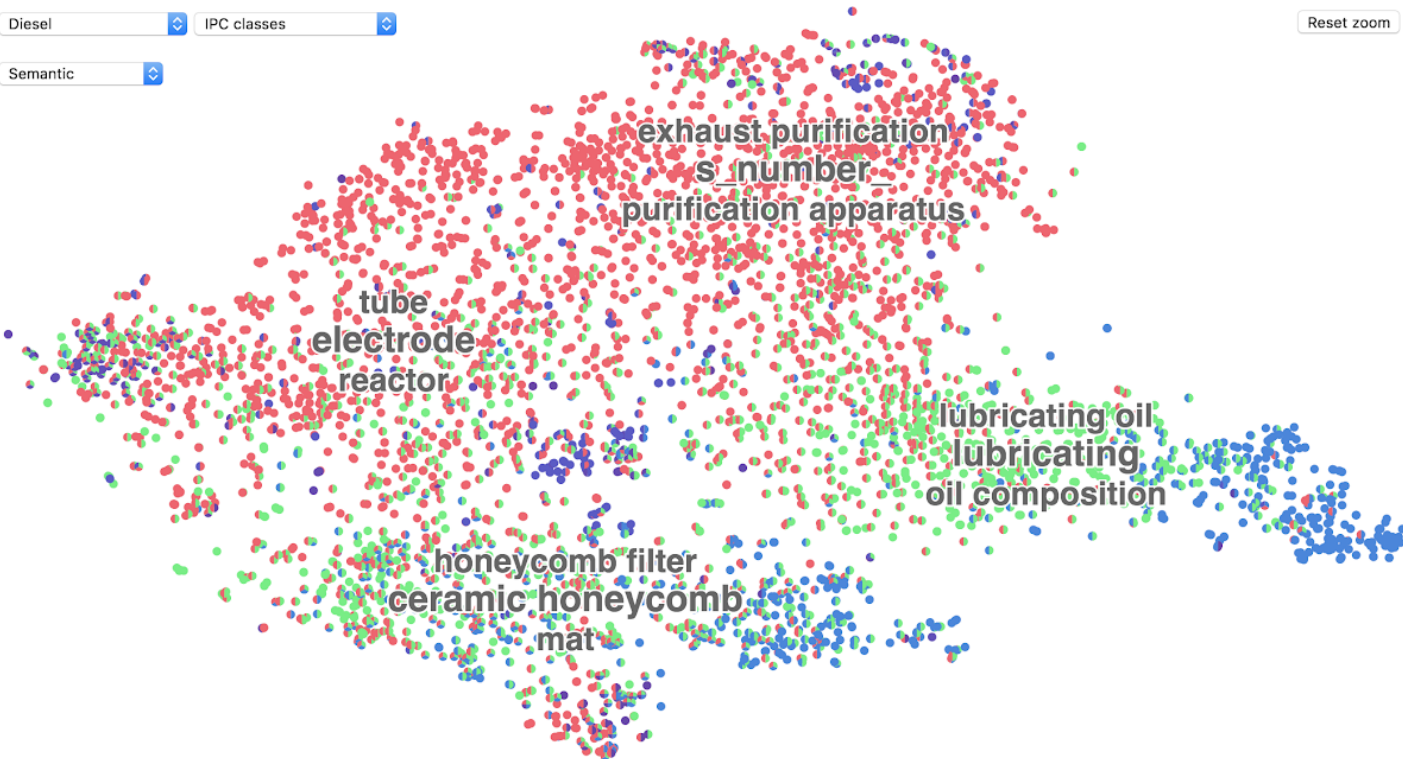
	TITLE	ARTIST	ALBUM
+	Hurt Somebody	Noah Kahan, Julia Michaels	Hurt Somebody
+	islands_	Blaudzun	islands_
+	Run to You	Tom Gregory	Run to You
+	Lonely Together - Acoustic	Avicii, Rita Ora	Lonely Together (Acoustic)
+	Too Much To Ask	EXPLICIT Niall Horan	Too Much To Ask
+	Work On Me	The Tech Thieves	Work On Me
+	Be Your Love	Bishop Briggs	Be Your Love
+	Waiting Here	Jake Isaac	Waiting Here (Remixes)
+	Leave a Light On	Tom Walker	Leave a Light On
+	Nevermind	Dennis Lloyd	Nevermind
+	World Gone Mad	EXPLICIT Bastille	World Gone Mad (From Brig
+	Colors	Isaac Danile	Colors



Diesel IPC classes

Semantic

Reset zoom



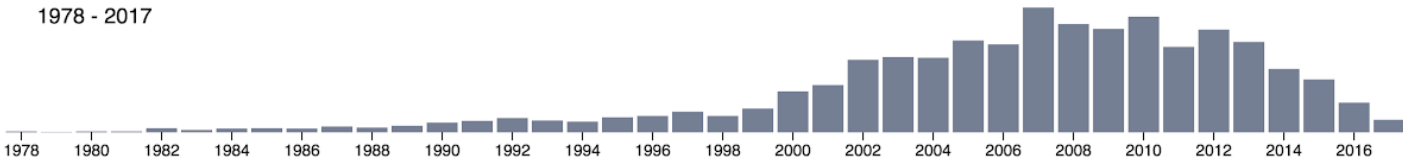
EP-1889812 Method of filtering a liquid used in the exhaust of a diesel engine

2006.08.11 LAB SA

Cites 0 Cited by 0 in this dataset

B01D03/14, B01D03/14, B01D03/96, B01D03/96, C02F01/28, C02F01/28, F01N03/04, F01N03/04

1978 - 2017



350 molecules, filtration, surfactant, pollutants, anionic, carbon black, granular, bed, charge, unit formed

250 The process of filtration of seawater for purifying exhaust fumes of a diesel engine (101) by wet process, comprises adding cationic or anionic surfactant to the seawater, and passing the seawater in a granular filtration bed (103) at a speed of 2-10 cm/s. The seawater contains

200

150

100

50

0

TabNine



code-
generation
algorithm



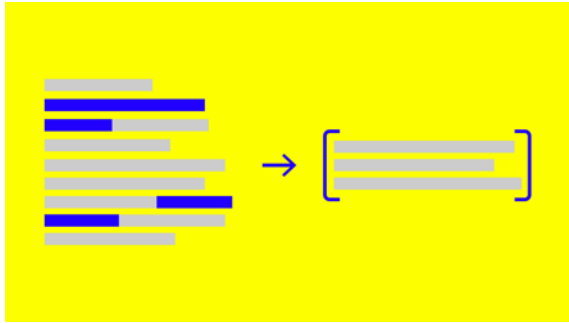
OpenAI GPT-2



Intelligent Code-
Completion

```
1 #include <iostream>
2 #include <vector>
3 #include <unordered_set>
4
5 using namespace std;
6
7 // Check if standard input contains 3 numbers that sum to 0
8 bool three_sum() {
9
10 }
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
```

Other applications we investigate in



Abstractive text summarization using on arbitrary long texts using bidirectional RNNs

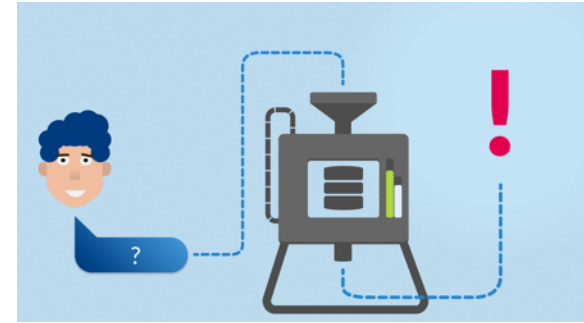


Extract key aspects on long, complicated texts (medical records, legislative texts)



Fine-tune general purpose language models for text classification in special language environments

Limit offensive speech in public communication (bulletin boards, e-mails)



Attention based RNNs for Q&A on database information using reinforcement learning

Access previously not accessible data sources for easy information access (hotlines, chatbots...)

Questions?
Let's talk:

Nico Kreiling



[nicokreiling](https://twitter.com/nicokreiling)



inovex

