



displaCy

Dependency Visualizer

Named Entity Visualizer

Similarity

sense2vec: Semantic Analysis of the Reddit Hivemind

Our neural network read every comment posted to Reddit in 2015, and built a semantic map using word2vec and spaCy.

  [Read more >](#)

Sentence Similarity

Semantic Analysis of the Reddit Hivemind

[View on GitHub](#)

Our neural network read every comment posted to Reddit in 2015, and built a semantic map using word2vec and spaCy. Try searching for a phrase that's more than the sum of its parts to see what the model thinks it means. Try your favourite band, slang words, technical things, or something totally random.

natural language processing

Sense

machine learning >

90%

computer vision >

86%

data analysis >

84%

neural nets >

83%

relational databases >

82%

algorithms >

81%

neural networks >

80%

image recognition >

80%

NLP >	80%
big data >	79%
data science >	79%
information retrieval >	79%
programming languages >	78%
database management >	78%
complexity theory >	78%
design patterns >	78%
scientific computing >	77%
programming >	77%
computer programs >	77%
data structures >	76%
bioinformatics >	76%

data analytics >	76%
programming language >	75%
programming background >	75%
computer architecture >	75%
large data sets >	75%
ANNs >	75%
coding >	75%
domain knowledge >	75%
data scientist >	75%
expert systems >	75%
VBA >	75%
web technologies >	75%
finite state machines >	74%

finite state machines >	74%
computational linguistics >	74%
algorithm design >	74%
test automation >	74%
Simulink >	74%
programming skills >	74%
MATLAB >	74%
algorithmic >	74%
software development >	74%
cryptography >	74%
SQL >	74%
writing code >	74%
data scientists >	73%

FPGAs >	73%
SQL. >	73%
embedded software >	73%
compilers >	73%
operations research >	73%
GIS >	73%
web applications >	73%
technical writing >	73%
lambda calculus >	73%
computational >	73%
software engineering >	73%
HTML/CSS >	73%
higher mathematics >	73%

higher mathematics >	15%
Data mining >	72%
specific technology >	72%
computational methods >	72%
artificial intelligence >	72%
econometrics >	72%
unit testing >	72%
NNs >	72%
heuristics >	72%
programmer >	72%
RNNs >	72%
human-computer interaction >	72%
SPSS >	72%

computation >	72%
VBA. >	72%
relational algebra >	72%
computational neuroscience >	72%
large datasets >	71%
computational complexity >	71%
data manipulation >	71%
code structure >	71%
user interfaces >	71%
narrow AI >	71%
pattern matching >	71%
software tools >	71%
debugging >	71%

debugging >	71%
programing >	71%
concurrency >	71%
Microsoft Access >	71%
real-world application >	71%
NLTK >	71%
conceptual ideas >	71%
hard problems >	71%
complex equations >	71%
programming experience >	70%
many programmers >	70%
Prolog >	70%
database stuff >	70%

programming paradigms >

70%

How does this work?

We used spaCy to tag and parse every comment posted to Reddit in 2015, and fed the results to Gensim's word2vec implementation. Using the search above, you can get a lot of interesting insights into the Reddit hivemind. See what spaCy and Gensim think Reddit thinks about almost anything.

[Read the blog post](#)

