





---

---

View of Product: Functional ----- Symbolic

---

Word	Oxford American Writers Thesaurus	Merriam-Webster	Cambridge Thesaurus	Collins Thesaurus	thesaurus.com
Function	purpose, task, use, role; duty, concern, province, aim; result, outcome, ramification; party	affair, event; capacity, connection, job; position, purpose, role; act, serve	goal, purpose, use, objective, utility; aliasing, bitmap, clickable; arithmetic, percent	purpose, business, job, concern; work, run, operate, perform	activity, duty, objective, purpose, task; affair, gathering, meeting, party; react, serve
Ritual	ceremony, rite, observance; sacrament, liturgy; traditional, routine, usual, habitual	custom, fashion, habit, pattern, practise, manner, ceremony, form, etiquette	cavalcade, initiation, ritual, salute	ceremony, rite, ceremonial, sacrament; custom, tradition, routine, convention	custom, formality, habit, liturgy, practice, procedure, routine, stereotype
Myth	folk tale, story, saga, allegory, parable, tradition; fallacy, fiction; falsehood, lie	fable, legend, mythos, fantasy, fiction, saga, story, tale, tradition, delusion, error	anecdote, saga, story, legend, tale; artifice, lie, perjury, fiction, hypocrisy	legend, story, tradition, fiction, saga; fantasy, delusion, invention, falsehood	fable, fantasy, fiction, illusion, legend, lore, parable, tale, tradition
Sources	Websites for dictionaries listed above, accessed March 2020				

---

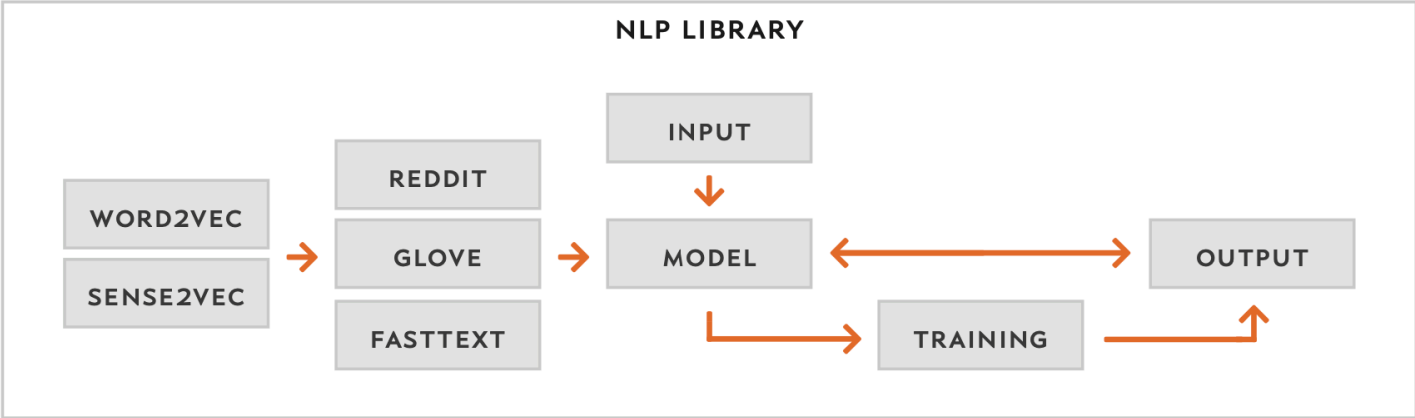
Feature	BNC	Wikipedia	Reddit 2015	WordNet	Brown	Common Crawl
Approx. word count	100 M	3.6 B	1.2 M	200 k	100 M	600 B +
Published	1994	2020 (yearly)	2015	1985	1961	2020
Curated by committee	✓	✓		✓	✓	
Formality						
- Written						
— News	✓		✓		✓	✓
— Dict., Thesauri, Encyclop.		✓		✓		✓
— Books	✓				✓	✓
— Online Comments			✓			✓
— Scientific Papers/Journals	✓				✓	✓
- Spoken						
— Parliament	✓					
— News	✓		✓			
Pre-training						
- vectors		✓	✓			
- Word-sense pairs				✓		
Sources	Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation, Common Crawl, 2020, Weischedel, Ralph, et al. OntoNotes Release 5.0 LDC2013T19. Web Download. Philadelphia: Linguistic Data Consortium, 2013, Wikipedia 2020, Reddit 2015,					



---



Technical Characteristic	Conceptnet Numberbatch	word2vec	sense2vec
Architecture			
- Skip-gram		✓	✓
- CBOW		✓	✓
Word disambiguation			
- Polysemy			✓
- word order information			✓
- PoS-tagging	✓		✓
- dependency parsing			✓
- NER	✓	✓	✓



Feature	NLTK	SpaCy	TextBlob	Gensim	Stanford CoreNLP
<b>Word Embedding</b>					
— Tokenisation	✓	✓	✓	✓	✓
— Lemmatisation	✓	✓			
— Stemming	✓	✓			
— POS Tagging	✓	✓	✓	✓	✓
— Parsing	✓	✓	✓		
— Named Entity Recognition	✓	✓			✓
— Text Classification	✓	✓	✓	✓	✓
— Training	✓	✓	✓	✓	✓
<b>Algorithm</b>					
— word2vec	✓	✓		✓	
— sense2vec		✓			
— Conceptnet Numberbatch				✓	
<b>Usability</b>					
— Production-ready		✓	✓		✓
— Good documentation		✓	✓		
— Easy for beginners		✓	✓		✓
<b>Sources</b>	Pinto, Oliveira, Alves, 2014, Active Wizards, 2019, Ferak, 2019, Loria, 2020, Stanford, 2016, Řehůřek, 2014				



---

hammer is capable  
of...

- en strike nail →
- en break glass →
- en drive in nails →
- en break a window →
- en nail a nail →
- en nail nails →
- en break fragile objects →
- en break a wall →
- en be breaking glass →
- en force a nail into a board →
- en hit a nail →
- en nail a board →
- en strike with great force →

## Models

Select one of the available models

- ✓ Finnish 4B wordforms skipgram
- Suomi24 wordforms skipgram
- Suomi24 lemmas skipgram
- English GoogleNews Negative300

## Nearest words

Given a word, this demo shows a list of other words that are similar to it, i.e. nearby in the vector space.

Show nearest

Case sensitive: ☒ Top N:

10

≈

## Word to Vec JS Demo

### Similar Words

Enter a word and see words with similar vectors.

### Word Algebra

Enter all three words, the first two, or the last two and see the words that result.

+

(

-

)

=

## Semantic Analysis of the Reddit Hivemind

We parsed every comment posted to Reddit in 2015 and 2019, and trained different word2vec models for each year. A lot's happened over the last four years, so many words, people or events have different associations. You can also 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.



Term



Sense ?



Year ?





DATA

5 tensors found  
Word2Vec 10K

UMAP

T-SNE

PCA

CUSTOM

Dimension

2D

3D

Perplexity

8

Learning rate

10

Supervise

0

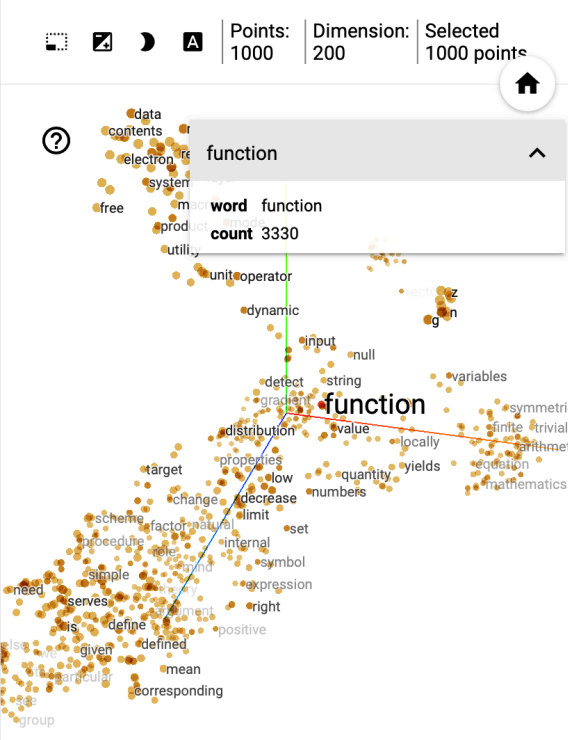
Stop

Pause

Perturb

Iteration: 1506

[How to use t-SNE effectively.](#)



Show All Data

Isolate 1000 points

Clear selection

Search

function

by

word

neighbors

995

distance

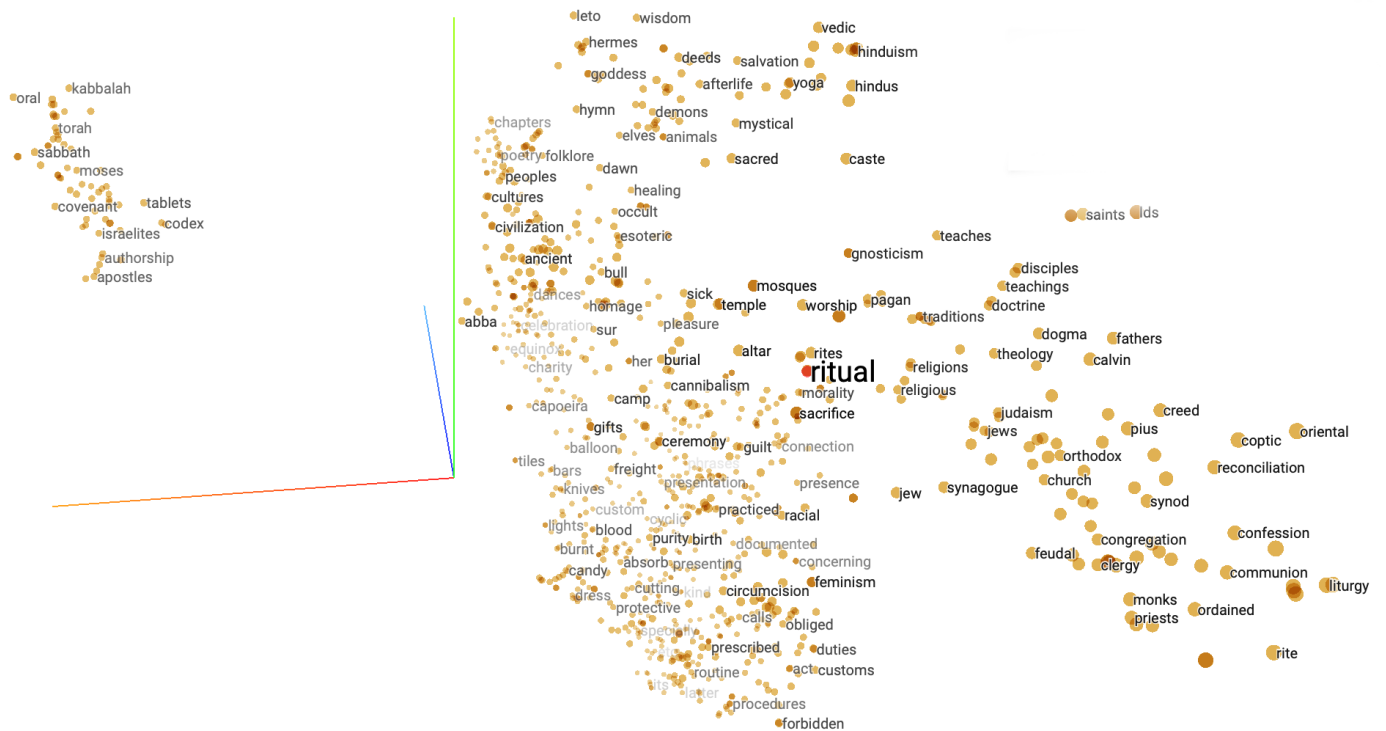
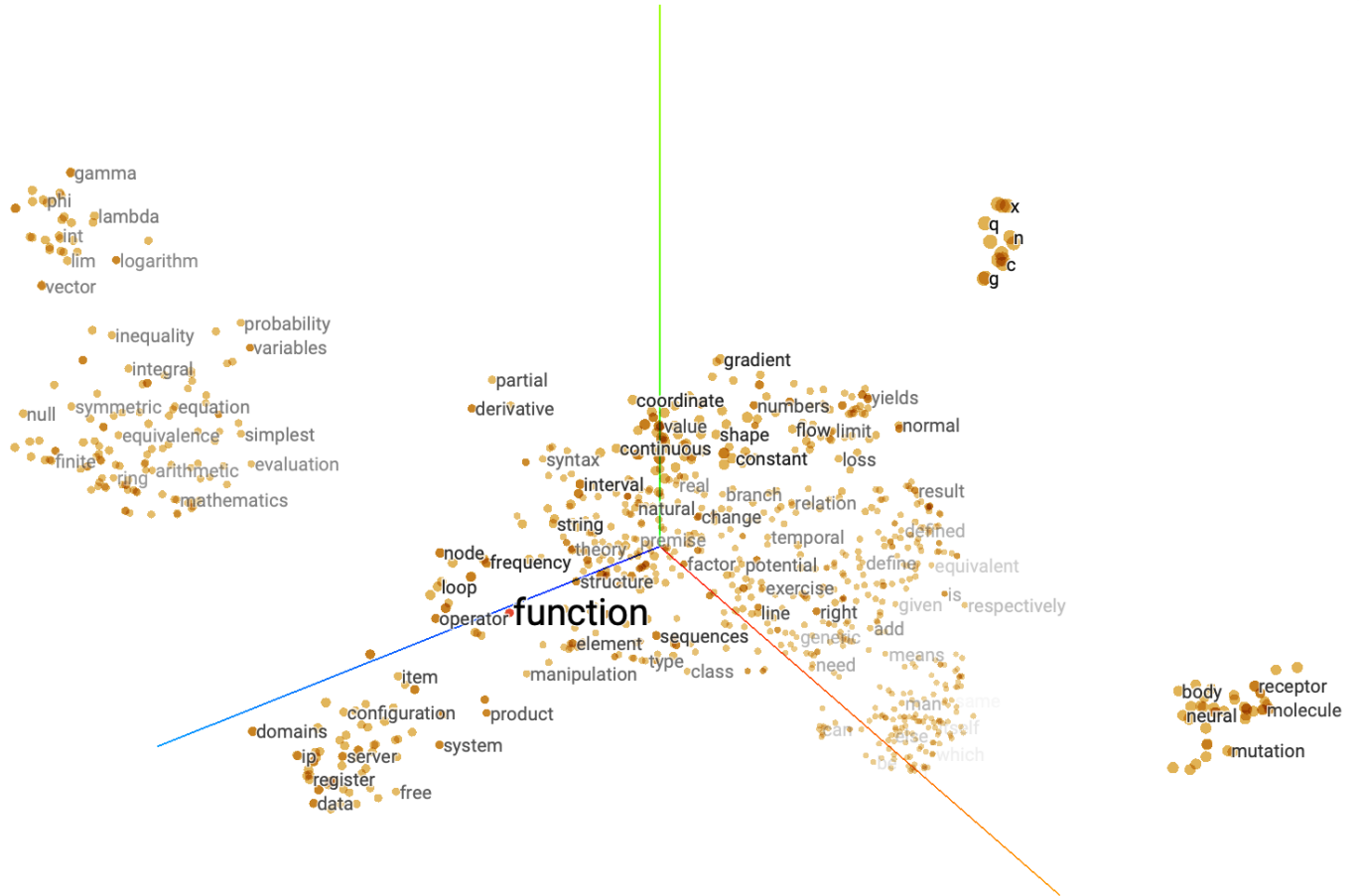
COSINE

EUCLIDEAN

Nearest points in the original space:

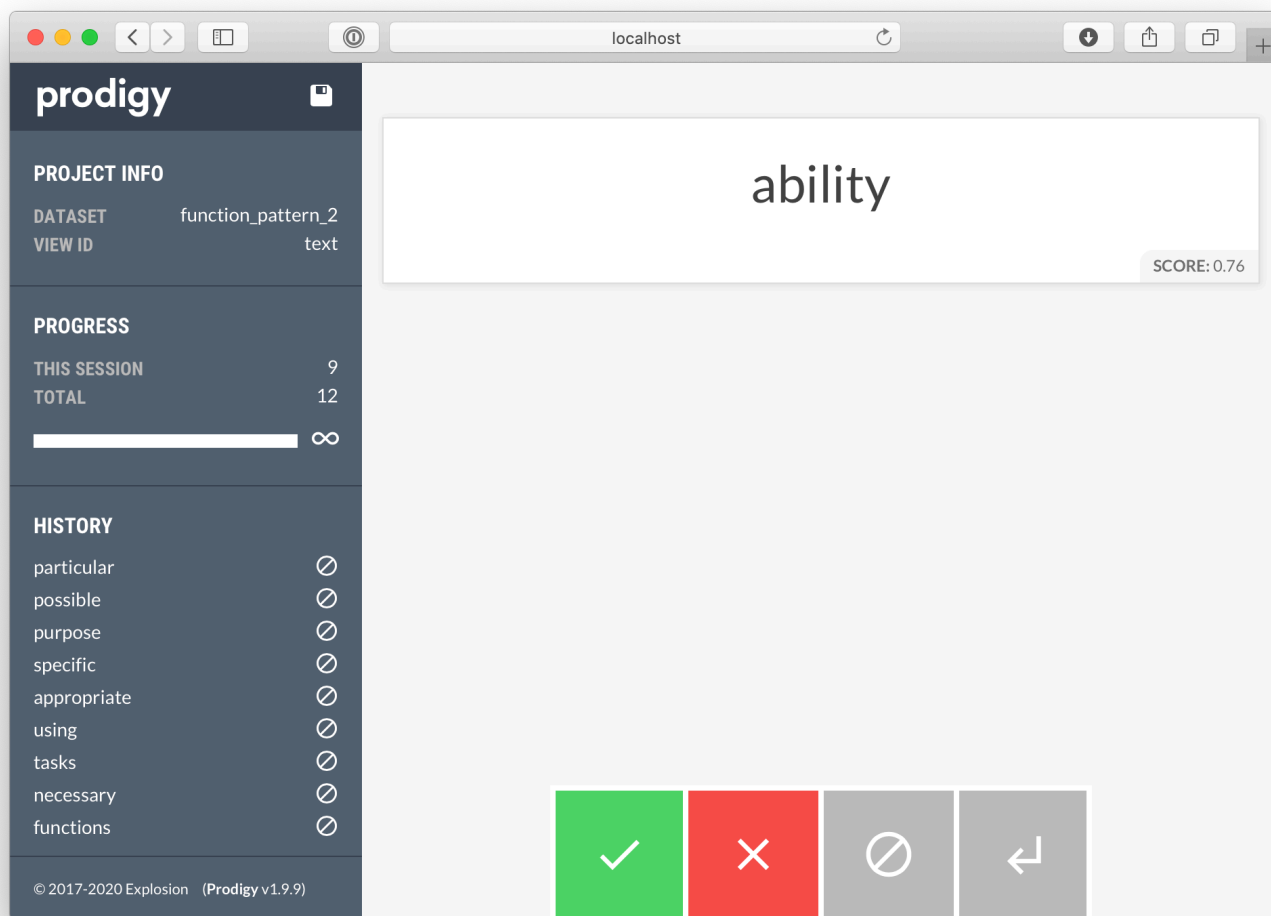
functions	0.486
value	0.686
arithmetic	0.760
constant	0.760
continuous	0.763
positive	0.777
defined	0.787
variables	0.795

BOOKMARKS (0)









[illegible]



