

Z Wikipedie, otevřené encyklopedie

**Word2Vec** je způsob [zpracování přirozeného jazyka](#), kdy jsou slova reprezentována vektory.<sup>[1]</sup> Bezkontextový model publikovali v roce 2013 společně výzkumníci [Tomáš Mikolov](#), Kai Chen, Greg Corrado, [Ilya Sutskever](#) a [Jeff Dean](#), kteří pracovali pro Google. Algoritmus word2vec využívá model [neuronové sítě](#) k učení slovních asociací z velkého [korpusu](#) textu. Po natrénování může takový model odhalit [synonyma](#) nebo navrhnut další slova pro další část věty. Word2vec reprezentuje každé odlišné slovo určitým seznamem čísel, kterému se říká [vektor](#). Vektory jsou pečlivě vybrány tak, aby zachycovaly sémantické a syntaktické vlastnosti slov; jednoduchá matematická funkce ([kosinus](#) úhlu mezi vektory) tak může indikovat úroveň [sémantické podobnosti](#) mezi slovy reprezentovanými těmito vektory.

Bezkontextové modely, jako je word2vec nebo [GloVe](#), vytvářejí pro každé slovo ve slovníku jedinečnou reprezentaci, přičemž [BERT](#) bere v úvahu kontext pro každý výskyt daného slova. Zatímco vektor pro slovo anglické slovo "run" bude mít stejnou reprezentaci z word2vec pro oba jeho výskyty ve větách "*He is running a company*" a "*He is running a marathon*", BERT poskytne kontextové osazení, které se bude lišit v závislosti na větě. Nově se právě pro [vnoření slov](#) používají novější modely jako [BERT](#) či [GPT-3](#).<sup>[2]</sup>

## Odkazy

### Reference

V tomto článku byl použit [překlad](#) textu z článku [Word2vec na anglické Wikipedii](#).

- ↑ A Beginner's Guide to Word2Vec and Neural Word Embeddings. *Pathmind* [online]. [cit. 2022-10-29]. [Dostupné online](#). (anglicky)
- ↑ ALAMMAR, Jay. The Illustrated Word2vec. *jalammar.github.io* [online]. [cit. 2022-10-29]. [Dostupné online](#).

### Externí odkazy

- Obrázky, zvuky či videa k tématu [Word2Vec](#) na Wikimedia Commons

- [Oficiální stránka projektu](#)

Citováno z „<https://cs.wikipedia.org/w/index.php?title=Word2Vec&oldid=24823261>“

Kategorie:

- [Počítačová lingvistika](#)
- [Zpracování přirozeného jazyka](#)
- [Umělé neuronové sítě](#)

Skryté kategorie:

- [Monitoring:Články přeložené z enwiki](#)
- [Monitoring:Autoritní kontrola s 0 identifikátory](#)

Hledání

Speciální:Hledání

Hledat

Word2Vec

18 jazyků

[Přidat téma](#)