

Word2Vec Paper - <https://arxiv.org/pdf/1301.3781v3.pdf>

GitHub Link: [Word2Vec](#)

- Proposes two model architectures for computing continuous vector representations of words from *very large data sets*
- Quality of representations is measured in a word similarity task - this technique showed vast improvements upon previously best-performing techniques, at a lower computational cost
- Previous techniques treated words as atomic units, that is to say that there is no notion of similarity between words
- For example, one of these traditional techniques is the popular N-gram model which is used for statistical language modelling
- In summary, Word2Vec models can be used for learning high-quality word vectors from huge datasets - i.e. corpora containing billions of words, with millions of words in the vocabulary
- No previously proposed architecture had been successfully trained on more than a few hundred millions of words