

VALENTINE'S DAY SPECIAL

WHAT COLOUR IS YOUR HEART?

How it works



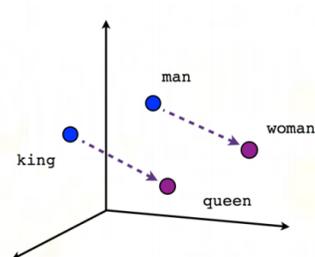
Our Deep Learning model was trained with the whole Wikipedia corpus and it has learnt how to read and understand the human language. It relies on words mapped to vectors of real numbers and leverages the semantic relationship with other encoded elements like colours. The algorithm is able to find which colour is the most similar to a given word because words and colours sharing similar meanings are represented by a similar sequence of digits.

## TECHNICAL FOCUS

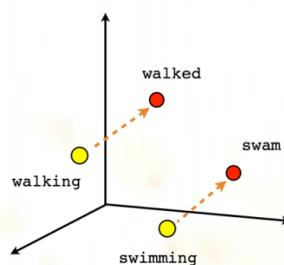
## Word embedding

Word embedding is the collective name for a set of language modeling and feature learning techniques in natural language processing (NLP). Words or phrases from the vocabulary are mapped to vectors of real numbers. Word embedding is a type of word representation that allows words with similar meaning to have a similar representation.

Embedding models allow Neural Networks to be able to map Natural Language elements like sentences, words, n-grams (collections of n letters/words) into an encoding represented by a sequence of numbers (called vectors). One of the most interesting aspects of this encoding is that information about semantics, style and syntax is often preserved. This way, the relationships with other encoded elements are kept, too.

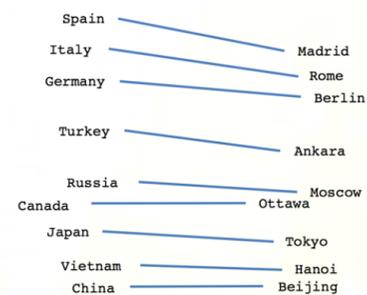


Male-Female



Verb tense

(Image from Analytics Vidhya)



Country-Capital

The image above is taken from the results of Word2vec, presently one of the most famous word encoders.



For instance, not only words defining different countries are clustered close to each other in the vector space, but they also keep similar distances from the words defining their related capitals in the vector space.

Example of similar distances: **Italy - Rome = Germany - Berlin**

If you re-arrange the previous equation in this form **X - Tokyo = Italy - Rome**, then you will find that **X = Japan**

## THE IDEA

## What colour is a word?

What if we compute distances between the embedding of a word and the embeddings of the colours' names?

- First of all, we transform a text input into the corresponding embedding vector

**“Tree”** → [ **0.9, -0.3, 0.4** ]

- Then, we identify twelve main families of colours to be subsequently converted into their corresponding RGB values, and we assign a vector to each colour family

**Blue** → [ **-0.1, 0.6, 0.5** ]

**Green** → [ **0.8, -0.3, 0.3** ]

...

Notice that **“Tree”** and **“Green”** correspond to very similar sequences of numbers.

Why? Because the ML algorithm has seen these pairs of words together in many paragraphs on Wikipedia and this causes a minimum distance between the two terms we are focusing on.

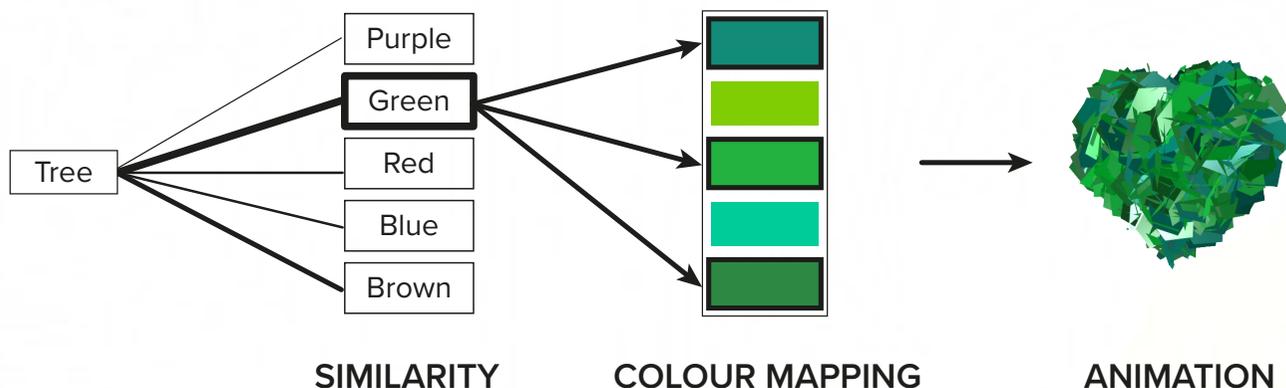


• Finally we pick the colour family according to the minimum distance between the embedded vector ([ **0.9, -0.3, 0.4** ]) and the list of the embedded colour families.

[ **0.9, -0.3, 0.4** ] is very similar to [ **0.8, -0.3, 0.3** ], and this is why the family of **Green** is the best to be related to the word **Tree**.

• We manage twelve main colour families made of five different hues each, totaling sixty different output colours.

The **final colour choice which is related to a single word** relies on the first, second, or third nearest colour computed by the



**Our ML algorithm was trained with the whole Wikipedia corpus and is able to turn simple words into meaningful colours and whilst making them available for any kind of creative purpose.**

## RESOURCES

<https://machinelearningmastery.com/what-are-word-embeddings/>  
[https://www.tensorflow.org/tutorials/text/word\\_embeddings](https://www.tensorflow.org/tutorials/text/word_embeddings)





Reply [MTA, STAR: REY]specialises in the design and implementation of solutions based on new communication channels and digital media. As a network of highly specialised companies, Reply defines and develops business models enabled by the new models of big data, cloud computing, digital media and the internet of things. Reply delivers consulting, system integration and digital services to organisations across the telecom and media; industry and services; banking and insurance; and public sectors.

[www.reply.com](http://www.reply.com)