Word Embedding is a term which is used to represent text analysis in the form of real valued vectors.

Word2Vec, as it's name suggests, is a method from 'genism' which allows us to construct an embedding.

We can make use of text file to do so, as well use web scapping. Here in this file, I'm making use of Scrapping. I've used the library such 'beautifulsoup4' to do so.

In [50]:

```
#importing libraries
import bs4 as bs
import urllib.request
import re

#scrapping the data

#picking up any article from web, we can also make use of text douments.

scrapped_data = urllib.request.urlopen('https://en.wikipedia.org/wiki/Main_Page')
article = scrapped_data .read()

parsed_article = bs.BeautifulSoup(article,'lxml')

paragraphs = parsed_article.find_all('p')

article_text = ""

for p in paragraphs:
    article_text += p.text
```

In [51]:

```
#importing nltk and performing data cleaning
import nltk
processed_article = article_text.lower()
processed_article = re.sub('[^a-zA-Z]', ' ', processed_article )
processed_article = re.sub(r'\s+', ' ', processed_article)

#tokenizing
all_sentences = nltk.sent_tokenize(processed_article)

all_words = [nltk.word_tokenize(sent) for sent in all_sentences]
```

In [52]:

```
#removing the stop words
from nltk.corpus import stopwords
for i in range(len(all_words)):
    all_words[i] = [w for w in all_words[i] if w not in stopwords.words('english')]
```

Importing 'Word2Vec from 'gensim'.

In [53]:

```
from gensim.models import Word2Vec

word2vec = Word2Vec(all_words, min_count=2)
```

In [54]:

```
v1 = word2vec.wv['artificial']
```

Finidng out similar words:

In [55]:

```
similar = word2vec.wv.most_similar('intelligence')
```

In [56]:

```
print(similar)
```

```
[('ai', 0.5098832845687866), ('computer', 0.4504174292087555), ('systems', 0.431304395198
822), ('would', 0.4281565248966217), ('human', 0.42633435130119324), ('processes', 0.4180
874824523926), ('could', 0.41014254093170166), ('artificial', 0.4044584035873413), ('data
', 0.39941200613975525), ('ethics', 0.39849933981895447)]
```

In [ ]:

```

```