

# Notebook

March 8, 2019

```
In [1]: # import modules & set up logging
import gensim, logging
import smart_open, os
logging.basicConfig(format='%(asctime)s : %(levelname)s : %(message)s', level=logging.DEBUG)
import datetime
import pandas as pd
import multiprocessing

# fichier incltu dans le projet
import save_notebook
```

D:\Outil\Anaconda\envs\majeure-ml-env\lib\site-packages\gensim\utils.py:1197: UserWarning: detected Windows; aliasing chunkize to chunkize\_serial)

## 1 D  claration donn  es

```
In [2]: now = str(datetime.datetime.now()).replace(" ", "")
```

```
In [3]: word_embedding = "word2vec"
```

## 2 Prepare data

```
In [4]: filenames = os.listdir("../wikipedia/data")
```

```
In [5]: #Cr       un fichier ou chaque ligne continent tout un fichier
# path="../wikipedia/data"
path="../wikipedia/data/"
with open('../data/wikipedia_informatic.txt', 'w+', encoding="utf8" ) as out_file:

    for fname in filenames:
#         print(fname)
        if "ipynb_checkpoints" in fname:
            continue
        try:
            with open(path + fname, encoding="utf8") as in_file:
                out_file.write(in_file.read().replace("\n", ""))
```

```

except:
    continue

```

```

In [6]: # On lit et on tokenize le fichier
with open('./data/wikipedia_informatic.txt', 'r', encoding="utf8") as f:
    wiki_vocab = f.readlines()
    wiki_vocab = [x.strip() for x in wiki_vocab]

    wiki_vocab_tokenized = []
    # for line in wiki_vocab:
    #     print(gensim.utils.simple_preprocess(line))
    # wiki_vocab_tokenized.append(gensim.utils.simple_preprocess(str(wiki_vocab)))

In [7]: wiki_vocab_tokenized = gensim.utils.simple_preprocess(str(wiki_vocab))

```

### 3 Create model

```

In [8]: # build vocabulary and train model
model = gensim.models.Word2Vec(
    [wiki_vocab_tokenized],
    size=150,
    seed=1234,
    window=10,
    min_count=2,
    workers=multiprocessing.cpu_count())

date_before_learning = datetime.datetime.now()
model.train([wiki_vocab_tokenized], total_examples=len(wiki_vocab_tokenized), epochs=20)
time_training = datetime.datetime.now() - date_before_learning

```

```

2019-02-25 23:20:23,862 : WARNING : consider setting layer size to a multiple of 4 for greater
2019-02-25 23:20:23,863 : INFO : collecting all words and their counts
2019-02-25 23:20:23,864 : INFO : PROGRESS: at sentence #0, processed 0 words, keeping 0 word ty
2019-02-25 23:20:25,662 : INFO : collected 169442 word types from a corpus of 7678696 raw words
2019-02-25 23:20:25,664 : INFO : Loading a fresh vocabulary
2019-02-25 23:20:26,118 : INFO : min_count=2 retains 78229 unique words (46% of original 169442)
2019-02-25 23:20:26,119 : INFO : min_count=2 leaves 7587483 word corpus (98% of original 7678696)
2019-02-25 23:20:26,409 : INFO : deleting the raw counts dictionary of 169442 items
2019-02-25 23:20:26,413 : INFO : sample=0.001 downsamples 28 most-common words
2019-02-25 23:20:26,414 : INFO : downsampling leaves estimated 5844786 word corpus (77.0% of pr
2019-02-25 23:20:26,709 : INFO : estimated required memory for 78229 words and 150 dimensions:
2019-02-25 23:20:26,713 : INFO : resetting layer weights
2019-02-25 23:20:28,072 : INFO : training model with 4 workers on 78229 vocabulary and 150 fea
2019-02-25 23:20:28,082 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:20:28,083 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:20:28,085 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:20:28,132 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:20:28,133 : INFO : EPOCH - 1 : training on 7678696 raw words (10000 effective wo

```

```

2019-02-25 23:20:28,141 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:20:28,146 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:20:28,147 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:20:28,177 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:20:28,178 : INFO : EPOCH - 2 : training on 7678696 raw words (10000 effective words)
2019-02-25 23:20:28,187 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:20:28,188 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:20:28,189 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:20:28,221 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:20:28,222 : INFO : EPOCH - 3 : training on 7678696 raw words (10000 effective words)
2019-02-25 23:20:28,231 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:20:28,232 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:20:28,233 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:20:28,263 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:20:28,264 : INFO : EPOCH - 4 : training on 7678696 raw words (10000 effective words)
2019-02-25 23:20:28,273 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:20:28,275 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:20:28,277 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:20:28,305 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:20:28,309 : INFO : EPOCH - 5 : training on 7678696 raw words (10000 effective words)
2019-02-25 23:20:28,309 : INFO : training on a 38393480 raw words (50000 effective words) took 1.000000 seconds
2019-02-25 23:20:28,309 : WARNING : under 10 jobs per worker: consider setting a smaller `batch_size`
2019-02-25 23:20:28,310 : WARNING : Effective 'alpha' higher than previous training cycles
2019-02-25 23:20:28,311 : INFO : training model with 4 workers on 78229 vocabulary and 150 features
2019-02-25 23:20:28,323 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:20:28,325 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:20:28,326 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:20:28,354 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:20:28,355 : INFO : EPOCH - 1 : training on 7678696 raw words (10000 effective words)
2019-02-25 23:20:28,356 : WARNING : EPOCH - 1 : supplied example count (1) did not equal expected count (7678696)
2019-02-25 23:20:28,365 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:20:28,366 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:20:28,367 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:20:28,396 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:20:28,397 : INFO : EPOCH - 2 : training on 7678696 raw words (10000 effective words)
2019-02-25 23:20:28,398 : WARNING : EPOCH - 2 : supplied example count (1) did not equal expected count (7678696)
2019-02-25 23:20:28,408 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:20:28,409 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:20:28,409 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:20:28,438 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:20:28,439 : INFO : EPOCH - 3 : training on 7678696 raw words (10000 effective words)
2019-02-25 23:20:28,440 : WARNING : EPOCH - 3 : supplied example count (1) did not equal expected count (7678696)
2019-02-25 23:20:28,450 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:20:28,460 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:20:28,460 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:20:28,481 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:20:28,482 : INFO : EPOCH - 4 : training on 7678696 raw words (10000 effective words)
2019-02-25 23:20:28,483 : WARNING : EPOCH - 4 : supplied example count (1) did not equal expected count (7678696)

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]





[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

2019-02-25 23:21:05,776 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:05,777 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:05,778 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:05,806 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:05,807 : INFO : EPOCH - 909 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:05,808 : WARNING : EPOCH - 909 : supplied example count (1) did not equal exp  
2019-02-25 23:21:05,816 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:05,817 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:05,818 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:05,845 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:05,846 : INFO : EPOCH - 910 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:05,846 : WARNING : EPOCH - 910 : supplied example count (1) did not equal exp  
2019-02-25 23:21:05,856 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:05,857 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:05,857 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:05,884 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:05,885 : INFO : EPOCH - 911 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:05,885 : WARNING : EPOCH - 911 : supplied example count (1) did not equal exp  
2019-02-25 23:21:05,894 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:05,895 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:05,895 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:05,921 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:05,922 : INFO : EPOCH - 912 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:05,923 : WARNING : EPOCH - 912 : supplied example count (1) did not equal exp  
2019-02-25 23:21:05,931 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:05,931 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:05,933 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:05,960 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:05,961 : INFO : EPOCH - 913 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:05,963 : WARNING : EPOCH - 913 : supplied example count (1) did not equal exp  
2019-02-25 23:21:05,974 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:05,975 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:05,976 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:06,002 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:06,003 : INFO : EPOCH - 914 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:06,004 : WARNING : EPOCH - 914 : supplied example count (1) did not equal exp  
2019-02-25 23:21:06,012 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:06,014 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:06,015 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:06,043 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:06,049 : INFO : EPOCH - 915 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:06,050 : WARNING : EPOCH - 915 : supplied example count (1) did not equal exp  
2019-02-25 23:21:06,059 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:06,067 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:06,068 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:06,086 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:06,104 : INFO : EPOCH - 916 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:06,106 : WARNING : EPOCH - 916 : supplied example count (1) did not equal exp

[illegible]

[illegible]





[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

2019-02-25 23:21:09,352 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:09,353 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:09,354 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:09,382 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:09,383 : INFO : EPOCH - 997 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:09,384 : WARNING : EPOCH - 997 : supplied example count (1) did not equal exp  
2019-02-25 23:21:09,394 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:09,395 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:09,396 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:09,421 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:09,422 : INFO : EPOCH - 998 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:09,423 : WARNING : EPOCH - 998 : supplied example count (1) did not equal exp  
2019-02-25 23:21:09,440 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:09,441 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:09,442 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:09,476 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:09,479 : INFO : EPOCH - 999 : training on 7678696 raw words (10000 effective v  
2019-02-25 23:21:09,481 : WARNING : EPOCH - 999 : supplied example count (1) did not equal exp  
2019-02-25 23:21:09,494 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:09,495 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:09,496 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:09,526 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:09,527 : INFO : EPOCH - 1000 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:09,528 : WARNING : EPOCH - 1000 : supplied example count (1) did not equal exp  
2019-02-25 23:21:09,539 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:09,540 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:09,542 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:09,576 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:09,580 : INFO : EPOCH - 1001 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:09,584 : WARNING : EPOCH - 1001 : supplied example count (1) did not equal exp  
2019-02-25 23:21:09,611 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:09,613 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:09,614 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:09,645 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:09,649 : INFO : EPOCH - 1002 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:09,654 : WARNING : EPOCH - 1002 : supplied example count (1) did not equal exp  
2019-02-25 23:21:09,684 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:09,687 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:09,688 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:09,716 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:09,717 : INFO : EPOCH - 1003 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:09,719 : WARNING : EPOCH - 1003 : supplied example count (1) did not equal exp  
2019-02-25 23:21:09,732 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:09,745 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:09,747 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:09,764 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:09,767 : INFO : EPOCH - 1004 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:09,768 : WARNING : EPOCH - 1004 : supplied example count (1) did not equal exp



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]





[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

2019-02-25 23:21:29,846 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:29,847 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:29,849 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:29,873 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:29,874 : INFO : EPOCH - 1501 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:29,875 : WARNING : EPOCH - 1501 : supplied example count (1) did not equal exp  
2019-02-25 23:21:29,884 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:29,885 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:29,886 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:29,913 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:29,914 : INFO : EPOCH - 1502 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:29,915 : WARNING : EPOCH - 1502 : supplied example count (1) did not equal exp  
2019-02-25 23:21:29,923 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:29,924 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:29,925 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:29,949 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:29,950 : INFO : EPOCH - 1503 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:29,951 : WARNING : EPOCH - 1503 : supplied example count (1) did not equal exp  
2019-02-25 23:21:29,959 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:29,961 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:29,962 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:29,988 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:29,989 : INFO : EPOCH - 1504 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:29,990 : WARNING : EPOCH - 1504 : supplied example count (1) did not equal exp  
2019-02-25 23:21:29,998 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:29,999 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:30,000 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:30,025 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:30,035 : INFO : EPOCH - 1505 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:30,036 : WARNING : EPOCH - 1505 : supplied example count (1) did not equal exp  
2019-02-25 23:21:30,044 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:30,045 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:30,045 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:30,071 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:30,077 : INFO : EPOCH - 1506 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:30,078 : WARNING : EPOCH - 1506 : supplied example count (1) did not equal exp  
2019-02-25 23:21:30,086 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:30,087 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:30,088 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:30,114 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:30,115 : INFO : EPOCH - 1507 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:30,115 : WARNING : EPOCH - 1507 : supplied example count (1) did not equal exp  
2019-02-25 23:21:30,125 : INFO : worker thread finished; awaiting finish of 3 more threads  
2019-02-25 23:21:30,126 : INFO : worker thread finished; awaiting finish of 2 more threads  
2019-02-25 23:21:30,126 : INFO : worker thread finished; awaiting finish of 1 more threads  
2019-02-25 23:21:30,151 : INFO : worker thread finished; awaiting finish of 0 more threads  
2019-02-25 23:21:30,152 : INFO : EPOCH - 1508 : training on 7678696 raw words (10000 effective  
2019-02-25 23:21:30,153 : WARNING : EPOCH - 1508 : supplied example count (1) did not equal exp

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]





[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

```

2019-02-25 23:21:49,785 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:21:49,787 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:21:49,787 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:21:49,814 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:21:49,815 : INFO : EPOCH - 1997 : training on 7678696 raw words (10000 effective
2019-02-25 23:21:49,816 : WARNING : EPOCH - 1997 : supplied example count (1) did not equal exp
2019-02-25 23:21:49,824 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:21:49,824 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:21:49,825 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:21:49,851 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:21:49,852 : INFO : EPOCH - 1998 : training on 7678696 raw words (10000 effective
2019-02-25 23:21:49,853 : WARNING : EPOCH - 1998 : supplied example count (1) did not equal exp
2019-02-25 23:21:49,862 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:21:49,863 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:21:49,864 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:21:49,889 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:21:49,890 : INFO : EPOCH - 1999 : training on 7678696 raw words (10000 effective
2019-02-25 23:21:49,890 : WARNING : EPOCH - 1999 : supplied example count (1) did not equal exp
2019-02-25 23:21:49,899 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-25 23:21:49,901 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-25 23:21:49,902 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-25 23:21:49,927 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-25 23:21:49,928 : INFO : EPOCH - 2000 : training on 7678696 raw words (10000 effective
2019-02-25 23:21:49,928 : WARNING : EPOCH - 2000 : supplied example count (1) did not equal exp
2019-02-25 23:21:49,929 : INFO : training on a 15357392000 raw words (20000000 effective words)

```

```

In [9]: model.save(str("model/" + now.replace(".", "-").replace(":", "-") + ".model"))

```

```

2019-02-25 23:21:49,937 : INFO : saving Word2Vec object under model/2019-02-2523-19-59-439393.1
2019-02-25 23:21:49,939 : INFO : storing np array 'vectors' to model/2019-02-2523-19-59-439393
2019-02-25 23:21:50,311 : INFO : not storing attribute vectors_norm
2019-02-25 23:21:50,313 : INFO : storing np array 'syn1neg' to model/2019-02-2523-19-59-439393
2019-02-25 23:21:50,793 : INFO : not storing attribute cum_table
2019-02-25 23:21:51,034 : INFO : saved model/2019-02-2523-19-59-439393.model

```

## 4 Test

```

In [28]: result = model.wv.most_similar(positive="redhat",topn=10)

```

```

In [29]: print(result)

```

```

[('decnet', 0.3683986961841583), ('caractÃrizer', 0.3319665789604187), ('tanaka', 0.328825533

```

## 5 Save

```
In [14]: import save_notebook
import utils
name_notebook_exported = save_notebook.save_notebook("word2vec_with_gensim.ipynb")
utils.write_result(word_embedding, time_training, name_notebook_exported, "resultats.")
```