Notebook

February 21, 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.
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: dete
  warnings.warn("detected Windows; aliasing chunkize to chunkize_serial")
```

1 DÃľclaration donnÃľes

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]:
```
#CrÃľÃľ 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=3
time_training = datetime.datetime.now() - date_before_learning
```

```
2019-02-21 17:06:07,830 : WARNING : consider setting layer size to a multiple of 4 for greater
2019-02-21 17:06:07,831 : INFO : collecting all words and their counts
2019-02-21 17:06:07,833 : INFO : PROGRESS: at sentence #0, processed 0 words, keeping 0 word ty
2019-02-21 17:06:09,208 : INFO : collected 133432 word types from a corpus of 5351366 raw words
2019-02-21 17:06:09,209 : INFO : Loading a fresh vocabulary
2019-02-21 17:06:09,603 : INFO : min_count=2 retains 62430 unique words (46% of original 13343
2019-02-21 17:06:09,604 : INFO : min_count=2 leaves 5280364 word corpus (98% of original 53513
2019-02-21 17:06:09,839 : INFO : deleting the raw counts dictionary of 133432 items
2019-02-21 17:06:09,843 : INFO : sample=0.001 downsamples 28 most-common words
2019-02-21 17:06:09,844 : INFO : downsampling leaves estimated 4064182 word corpus (77.0% of pr
2019-02-21 17:06:10,083 : INFO : estimated required memory for 62430 words and 150 dimensions:
2019-02-21 17:06:10,084 : INFO : resetting layer weights
2019-02-21 17:06:11,259 : INFO : training model with 4 workers on 62430 vocabulary and 150 feat
2019-02-21 17:06:11,269 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,271 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,272 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,315 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,318 : INFO : EPOCH - 1 : training on 5351366 raw words (10000 effective wo
```

```
2019-02-21 17:06:11,327 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,327 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,329 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,364 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,364 : INFO : EPOCH - 2 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,373 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,375 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,376 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,406 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,407 : INFO : EPOCH - 3 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,416 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,417 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,417 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,447 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,448 : INFO : EPOCH - 4 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,457 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,458 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,458 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,488 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,489 : INFO : EPOCH - 5 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,490 : INFO : training on a 26756830 raw words (50000 effective words) took
2019-02-21 17:06:11,491 : WARNING : under 10 jobs per worker: consider setting a smaller `batch
2019-02-21 17:06:11,493 : WARNING : Effective 'alpha' higher than previous training cycles
2019-02-21 17:06:11,493 : INFO : training model with 4 workers on 62430 vocabulary and 150 feat
2019-02-21 17:06:11,509 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,510 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,510 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,540 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,541 : INFO : EPOCH - 1 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,541 : WARNING : EPOCH - 1 : supplied example count (1) did not equal expect
2019-02-21 17:06:11,550 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,552 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,553 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,579 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,581 : INFO : EPOCH - 2 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,581 : WARNING : EPOCH - 2 : supplied example count (1) did not equal expect
2019-02-21 17:06:11,590 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,591 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,591 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,620 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,621 : INFO : EPOCH - 3 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,622 : WARNING : EPOCH - 3 : supplied example count (1) did not equal expect
2019-02-21 17:06:11,630 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,632 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,633 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,662 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,663 : INFO : EPOCH - 4 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,663 : WARNING : EPOCH - 4 : supplied example count (1) did not equal expect
```

```
2019-02-21 17:06:11,674 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,675 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,676 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,707 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,708 : INFO : EPOCH - 5 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,709 : WARNING : EPOCH - 5 : supplied example count (1) did not equal expect
2019-02-21 17:06:11,720 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,721 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,722 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,752 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,753 : INFO : EPOCH - 6 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,753 : WARNING : EPOCH - 6 : supplied example count (1) did not equal expect
2019-02-21 17:06:11,763 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,765 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,765 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,796 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,797 : INFO : EPOCH - 7 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,798 : WARNING : EPOCH - 7 : supplied example count (1) did not equal expect
2019-02-21 17:06:11,810 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,811 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,811 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,842 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,843 : INFO : EPOCH - 8 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,844 : WARNING : EPOCH - 8 : supplied example count (1) did not equal expect
2019-02-21 17:06:11,855 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,856 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,857 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,889 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,890 : INFO : EPOCH - 9 : training on 5351366 raw words (10000 effective wor
2019-02-21 17:06:11,890 : WARNING : EPOCH - 9 : supplied example count (1) did not equal expect
2019-02-21 17:06:11,901 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,902 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,903 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,932 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,933 : INFO : EPOCH - 10 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:11,934 : WARNING : EPOCH - 10 : supplied example count (1) did not equal expec
2019-02-21 17:06:11,944 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,945 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,946 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:11,976 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:11,977 : INFO : EPOCH - 11 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:11,978 : WARNING : EPOCH - 11 : supplied example count (1) did not equal expec
2019-02-21 17:06:11,989 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:11,990 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:11,991 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,020 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,021 : INFO : EPOCH - 12 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,022 : WARNING : EPOCH - 12 : supplied example count (1) did not equal expec
```

```
2019-02-21 17:06:12,031 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,032 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,033 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,061 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,062 : INFO : EPOCH - 13 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,063 : WARNING : EPOCH - 13 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,071 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,072 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,073 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,100 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,101 : INFO : EPOCH - 14 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,102 : WARNING : EPOCH - 14 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,113 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,114 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,115 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,145 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,146 : INFO : EPOCH - 15 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,147 : WARNING : EPOCH - 15 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,158 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,159 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,160 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,191 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,192 : INFO : EPOCH - 16 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,193 : WARNING : EPOCH - 16 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,205 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,206 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,207 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,239 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,240 : INFO : EPOCH - 17 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,240 : WARNING : EPOCH - 17 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,250 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,252 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,253 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,283 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,284 : INFO : EPOCH - 18 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,285 : WARNING : EPOCH - 18 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,295 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,296 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,299 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,330 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,331 : INFO : EPOCH - 19 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,331 : WARNING : EPOCH - 19 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,343 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,344 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,344 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,376 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,377 : INFO : EPOCH - 20 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,378 : WARNING : EPOCH - 20 : supplied example count (1) did not equal expec
```

```
2019-02-21 17:06:12,395 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,396 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,397 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,429 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,430 : INFO : EPOCH - 21 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,431 : WARNING : EPOCH - 21 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,446 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,447 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,448 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,480 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,481 : INFO : EPOCH - 22 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,481 : WARNING : EPOCH - 22 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,496 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,497 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,498 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,531 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,532 : INFO : EPOCH - 23 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,534 : WARNING : EPOCH - 23 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,545 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,546 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,547 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,580 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,581 : INFO : EPOCH - 24 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,582 : WARNING : EPOCH - 24 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,593 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,594 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,595 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,629 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,630 : INFO : EPOCH - 25 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,631 : WARNING : EPOCH - 25 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,647 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,648 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,649 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,680 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,681 : INFO : EPOCH - 26 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,682 : WARNING : EPOCH - 26 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,691 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,693 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,694 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,725 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,726 : INFO : EPOCH - 27 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,727 : WARNING : EPOCH - 27 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,745 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,752 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,759 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,785 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,786 : INFO : EPOCH - 28 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,786 : WARNING : EPOCH - 28 : supplied example count (1) did not equal expec
```

```
2019-02-21 17:06:12,795 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,795 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,797 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,826 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,827 : INFO : EPOCH - 29 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,829 : WARNING : EPOCH - 29 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,839 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 17:06:12,839 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 17:06:12,840 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 17:06:12,869 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 17:06:12,870 : INFO : EPOCH - 30 : training on 5351366 raw words (10000 effective wo
2019-02-21 17:06:12,871 : WARNING : EPOCH - 30 : supplied example count (1) did not equal expec
2019-02-21 17:06:12,872 : INFO : training on a 160540980 raw words (300000 effective words) too
```

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

```
2019-02-21 17:06:12,881 : INFO : saving Word2Vec object under model/2019-02-2117-05-54-119339.m
2019-02-21 17:06:12,883 : INFO : not storing attribute vectors_norm
2019-02-21 17:06:12,885 : INFO : not storing attribute cum_table
2019-02-21 17:06:14,815 : INFO : saved model/2019-02-2117-05-54-119339.model
```

4 Test

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

```
2019-02-21 17:06:14,823 : INFO : precomputing L2-norms of word weight vectors
```

```
In [11]: print(result)
```

```
[('tel', 0.9972558617591858), ('pistes', 0.9972215890884399), ('audio', 0.9966777563095093), (
```

5 Save

```
In [12]: import save_notebook
```

```
In [ ]: name_notebook_exported = save_notebook.save_notebook("word2vec_with_gensim.ipynb")
```

```
In [23]: def write_result(word_embedding, time_training,name_notebook_exported, fname ):
             if not os.path.isfile(fname):
                 f=open(fname, "a+")
                 f.write("Nombre de mots;Model de word embedding;temps d'apprentissage;Notebook
                 f.close
             f=open(fname, "a+")
             f.write("\n" + str( str(len(wiki_vocab_tokenized)) + ";" +word_embedding + ";"+ st
             f.close
```

```
In [24]: write_result(word_embedding, time_training,name_notebook_exported, "resultats.csv" )

In [25]: name_notebook_exported
```