Notebook

February 21, 2019

In [16]:
```
# 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
```

1 DÃl'claration donnÃl'es

In [3]:
```
now = str(datetime.datetime.now()).replace(" ","")
```

In [4]:
```
word_embedding = "word2vec"
```

2 Prepare data

In [5]:
```
filenames =  os.listdir("../wikipedia/data")
```

In [6]:
```
#CrÃlÃl 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 [7]:
```
# 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 [8]:
```
wiki_vocab_tokenized = gensim.utils.simple_preprocess(str(wiki_vocab))
```

3 Create model

In [9]:
```
# 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 16:59:12,154 : WARNING : consider setting layer size to a multiple of 4 for greater
2019-02-21 16:59:12,155 : INFO : collecting all words and their counts
2019-02-21 16:59:12,156 : INFO : PROGRESS: at sentence #0, processed 0 words, keeping 0 word ty
2019-02-21 16:59:13,533 : INFO : collected 133432 word types from a corpus of 5351366 raw words
2019-02-21 16:59:13,534 : INFO : Loading a fresh vocabulary
2019-02-21 16:59:13,821 : INFO : min_count=2 retains 62430 unique words (46% of original 13343
2019-02-21 16:59:13,823 : INFO : min_count=2 leaves 5280364 word corpus (98% of original 53513
2019-02-21 16:59:14,055 : INFO : deleting the raw counts dictionary of 133432 items
2019-02-21 16:59:14,059 : INFO : sample=0.001 downsamples 28 most-common words
2019-02-21 16:59:14,060 : INFO : downsampling leaves estimated 4064182 word corpus (77.0% of pr
2019-02-21 16:59:14,301 : INFO : estimated required memory for 62430 words and 150 dimensions:
2019-02-21 16:59:14,302 : INFO : resetting layer weights
2019-02-21 16:59:15,466 : INFO : training model with 4 workers on 62430 vocabulary and 150 feat
2019-02-21 16:59:15,476 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,478 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,479 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:15,525 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,526 : INFO : EPOCH - 1 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,535 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,537 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,539 : INFO : worker thread finished; awaiting finish of 1 more threads
```

```
2019-02-21 16:59:15,572 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,573 : INFO : EPOCH - 2 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,582 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,583 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,584 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:15,615 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,616 : INFO : EPOCH - 3 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,628 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,630 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,631 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:15,663 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,663 : INFO : EPOCH - 4 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,672 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,674 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,675 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:15,705 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,706 : INFO : EPOCH - 5 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,706 : INFO : training on a 26756830 raw words (50000 effective words) took
2019-02-21 16:59:15,708 : WARNING : under 10 jobs per worker: consider setting a smaller `batch
2019-02-21 16:59:15,711 : WARNING : Effective 'alpha' higher than previous training cycles
2019-02-21 16:59:15,712 : INFO : training model with 4 workers on 62430 vocabulary and 150 feat
2019-02-21 16:59:15,722 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,723 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,724 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:15,754 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,755 : INFO : EPOCH - 1 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,756 : WARNING : EPOCH - 1 : supplied example count (1) did not equal expect
2019-02-21 16:59:15,766 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,767 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,767 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:15,798 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,799 : INFO : EPOCH - 2 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,799 : WARNING : EPOCH - 2 : supplied example count (1) did not equal expect
2019-02-21 16:59:15,810 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,811 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,811 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:15,842 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,843 : INFO : EPOCH - 3 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,844 : WARNING : EPOCH - 3 : supplied example count (1) did not equal expect
2019-02-21 16:59:15,856 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,857 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,858 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:15,889 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,890 : INFO : EPOCH - 4 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,891 : WARNING : EPOCH - 4 : supplied example count (1) did not equal expect
2019-02-21 16:59:15,906 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,907 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,908 : INFO : worker thread finished; awaiting finish of 1 more threads
```

```
2019-02-21 16:59:15,936 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,938 : INFO : EPOCH - 5 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,939 : WARNING : EPOCH - 5 : supplied example count (1) did not equal expect
2019-02-21 16:59:15,948 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,949 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,950 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:15,978 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:15,979 : INFO : EPOCH - 6 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:15,980 : WARNING : EPOCH - 6 : supplied example count (1) did not equal expect
2019-02-21 16:59:15,992 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:15,993 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:15,993 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,024 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,025 : INFO : EPOCH - 7 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:16,025 : WARNING : EPOCH - 7 : supplied example count (1) did not equal expect
2019-02-21 16:59:16,035 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,036 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,038 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,069 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,071 : INFO : EPOCH - 8 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:16,072 : WARNING : EPOCH - 8 : supplied example count (1) did not equal expect
2019-02-21 16:59:16,084 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,087 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,102 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,121 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,122 : INFO : EPOCH - 9 : training on 5351366 raw words (10000 effective wor
2019-02-21 16:59:16,123 : WARNING : EPOCH - 9 : supplied example count (1) did not equal expect
2019-02-21 16:59:16,134 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,135 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,137 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,169 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,170 : INFO : EPOCH - 10 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,171 : WARNING : EPOCH - 10 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,190 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,194 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,197 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,226 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,226 : INFO : EPOCH - 11 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,227 : WARNING : EPOCH - 11 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,238 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,242 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,243 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,269 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,271 : INFO : EPOCH - 12 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,272 : WARNING : EPOCH - 12 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,282 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,283 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,284 : INFO : worker thread finished; awaiting finish of 1 more threads
```

```
2019-02-21 16:59:16,314 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,314 : INFO : EPOCH - 13 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,315 : WARNING : EPOCH - 13 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,333 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,334 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,335 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,366 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,372 : INFO : EPOCH - 14 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,373 : WARNING : EPOCH - 14 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,384 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,385 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,386 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,415 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,416 : INFO : EPOCH - 15 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,418 : WARNING : EPOCH - 15 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,429 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,429 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,430 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,458 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,459 : INFO : EPOCH - 16 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,460 : WARNING : EPOCH - 16 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,469 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,470 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,472 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,501 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,502 : INFO : EPOCH - 17 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,503 : WARNING : EPOCH - 17 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,515 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,517 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,517 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,548 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,548 : INFO : EPOCH - 18 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,549 : WARNING : EPOCH - 18 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,563 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,563 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,564 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,595 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,595 : INFO : EPOCH - 19 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,596 : WARNING : EPOCH - 19 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,614 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,616 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,617 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,649 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,650 : INFO : EPOCH - 20 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,650 : WARNING : EPOCH - 20 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,666 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,667 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,668 : INFO : worker thread finished; awaiting finish of 1 more threads
```

```
2019-02-21 16:59:16,699 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,700 : INFO : EPOCH - 21 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,701 : WARNING : EPOCH - 21 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,715 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,716 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,717 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,745 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,746 : INFO : EPOCH - 22 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,747 : WARNING : EPOCH - 22 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,759 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,762 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,763 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,789 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,790 : INFO : EPOCH - 23 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,790 : WARNING : EPOCH - 23 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,801 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,802 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,802 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,832 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,833 : INFO : EPOCH - 24 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,833 : WARNING : EPOCH - 24 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,843 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,845 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,845 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,875 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,876 : INFO : EPOCH - 25 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,877 : WARNING : EPOCH - 25 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,892 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,893 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,894 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,926 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,928 : INFO : EPOCH - 26 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,930 : WARNING : EPOCH - 26 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,941 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,942 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,943 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:16,974 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:16,975 : INFO : EPOCH - 27 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:16,975 : WARNING : EPOCH - 27 : supplied example count (1) did not equal expec
2019-02-21 16:59:16,989 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:16,992 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:16,994 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:17,020 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:17,021 : INFO : EPOCH - 28 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:17,022 : WARNING : EPOCH - 28 : supplied example count (1) did not equal expec
2019-02-21 16:59:17,034 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:17,035 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:17,036 : INFO : worker thread finished; awaiting finish of 1 more threads
```

```
2019-02-21 16:59:17,065 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:17,066 : INFO : EPOCH - 29 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:17,067 : WARNING : EPOCH - 29 : supplied example count (1) did not equal expec
2019-02-21 16:59:17,080 : INFO : worker thread finished; awaiting finish of 3 more threads
2019-02-21 16:59:17,081 : INFO : worker thread finished; awaiting finish of 2 more threads
2019-02-21 16:59:17,081 : INFO : worker thread finished; awaiting finish of 1 more threads
2019-02-21 16:59:17,111 : INFO : worker thread finished; awaiting finish of 0 more threads
2019-02-21 16:59:17,112 : INFO : EPOCH - 30 : training on 5351366 raw words (10000 effective wo
2019-02-21 16:59:17,113 : WARNING : EPOCH - 30 : supplied example count (1) did not equal expec
2019-02-21 16:59:17,114 : INFO : training on a 160540980 raw words (300000 effective words) too
```

```
In [10]: model.save(str("model/" + now.replace(".","-").replace(":","-") + ".model"))
```

```
2019-02-21 16:59:17,125 : INFO : saving Word2Vec object under model/2019-02-2116-58-58-611425.m
2019-02-21 16:59:17,128 : INFO : not storing attribute vectors_norm
2019-02-21 16:59:17,131 : INFO : not storing attribute cum_table
2019-02-21 16:59:19,070 : INFO : saved model/2019-02-2116-58-58-611425.model
```

4 Test

```
In [11]: result = model.wv.most_similar(positive="microsoft",topn=10)
```

```
2019-02-21 16:59:19,081 : INFO : precomputing L2-norms of word weight vectors
```

```
In [12]: print(result)
```

```
[('tel', 0.9967406988143921), ('texte', 0.9967136383056641), ('formats', 0.9966994524002075),
```

5 Save

```
In [21]: import save_notebook
```

```
In [22]: name_notebook_exported = save_notebook.save_notebook("word2vec_with_gensim.ipynb")
```

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

```
OSError                                   Traceback (most recent call last)

<ipython-input-22-4bc3e2481a3a> in <module>
----> 1 name_notebook_exported = save_notebook.save_notebook("word2vec_with_gensim.ipynb")
      2
      3 def write_result(word_embedding, time_training,name_notebook_exported, fname ):
      4     if not os.path.isfile(fname):
      5         fname = "Nombre de mots;Model de word embedding;temps d'apprentissage;Notel


D:\Cours_ingesup\Memoire\word2vec\save_notebook.py in save_notebook(notebook_filename)
     29     file_result_name = "resultats/" + notebook_filename.replace(".ipynb",now).repla
     30
---> 31     with open(file_result_name, "wb") as f:
     32         f.write(pdf_data)
     33         f.close()


OSError: [Errno 22] Invalid argument: 'resultats/word2vec_with_gensim.2019-02-2117:02:1
```

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