Name: Ronaldlee Ejalu Course Name: CSC 583 Assignment number: HW#5

- 1. Part 1:
  - My Development environment was Google Colab and the libraries and packages
     I used were csv, re, os, numpy, pandas and pairwise metric
  - Results of Part 1 were:
  - → Question 1
    - 1. word similar to dog were cat as shown below:

```
similar-words cosine_value
5448 dog-cat 0.879808
```

2. words similar to whale was shark as shown below:

```
similar-words cosine_value
9860 whale-shark 0.784017
```

3. words similar to before was again:

```
similar-words cosine_value
375 before-again 0.858747
```

4. words similar to however was although as shown below:

```
similar-words cosine_value
375 however-although 0.965755
```

5. words similar to fabricate was invent and this is shown below:

```
similar-words cosine_value
24072 fabricate-invent 0.704018
```

→ Question 2

```
a. dog : puppy :: cat : ?

dog : puppy :: cat : animal

b. speak : speaker :: sing : ?

speak : speaker :: sing : sang

c. France : French :: England : ?

France : French :: England : scotland

d. France : wine :: England : ?

france : wine :: england : britain
```

The results on part Question 2(d) surprised me because I expected to see something like beer or a type of drink enjoyed by the British.

## Part 11

- My Development environment was Google Colab and the libraries and packages
  I used were torch, nn, optin, numpy, pandas, os, nltk, RegexpTokenizer,
  word tokenize, stopwords and itertools.
- Results of Task 1 were:

 The top three words that are closest to processes are our, programs and abstract as shown below:

The top three words that are closest to 'processes' by cosine similarity:

	similar-words	cosine_value
46	processes-our	0.624608
36	processes-programs	0.545113
12	processes-abstract	0.525402

Task 2 failed to run as I was trying to run the model:

I became a bit disappointed because I couldn't proceed past this step.

The top three words closest to titanic were brock, comes and could as shown:

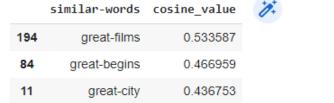


The top three words closest to acting were real, hollywood and footage as shown below:



The top three words closest to great were films, begins and city as shown:

The top three words that are closest to 'great' by cosine similarity:



The top three words closest to poor were director, action and long as shown:



General reflections about this assignment were learning how to prepare the data in the right format, the CBOW model expects it.

The difficulty I encountered was on Part 2 Task 2 with the error I encountered; I tried to debug it but I couldn't succeed.