Purpose

Our query is set for finding a movie that we remember some of its contents while we forget its name. As we were not sure whether there is the Japanese animation in the folder of movie synopsis, and we are supposed to find the movie "Totoro".

Query

```
"animated cat bus umbrella forest girl"
```
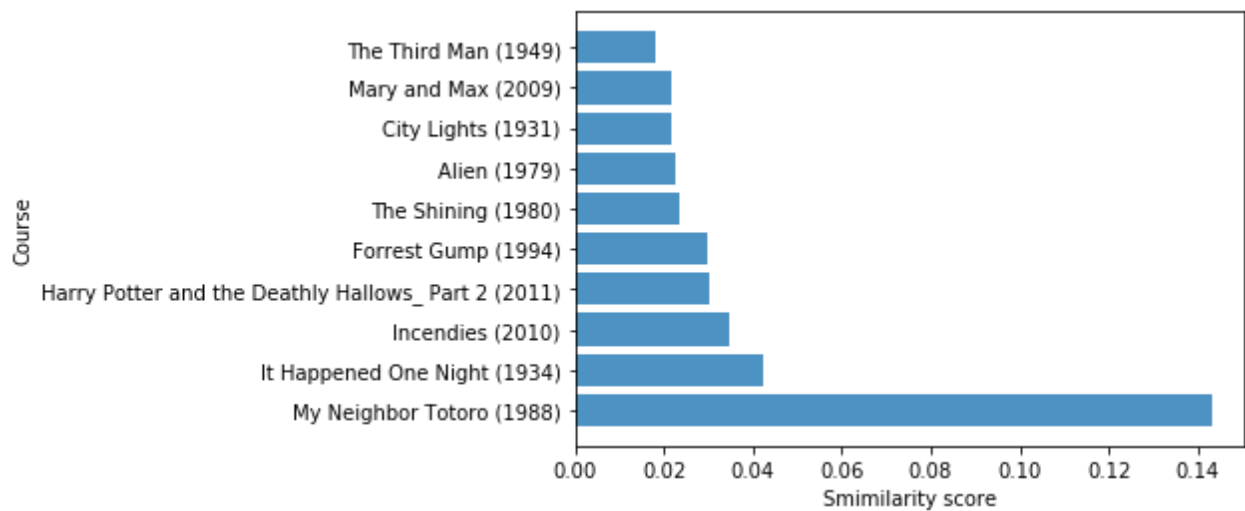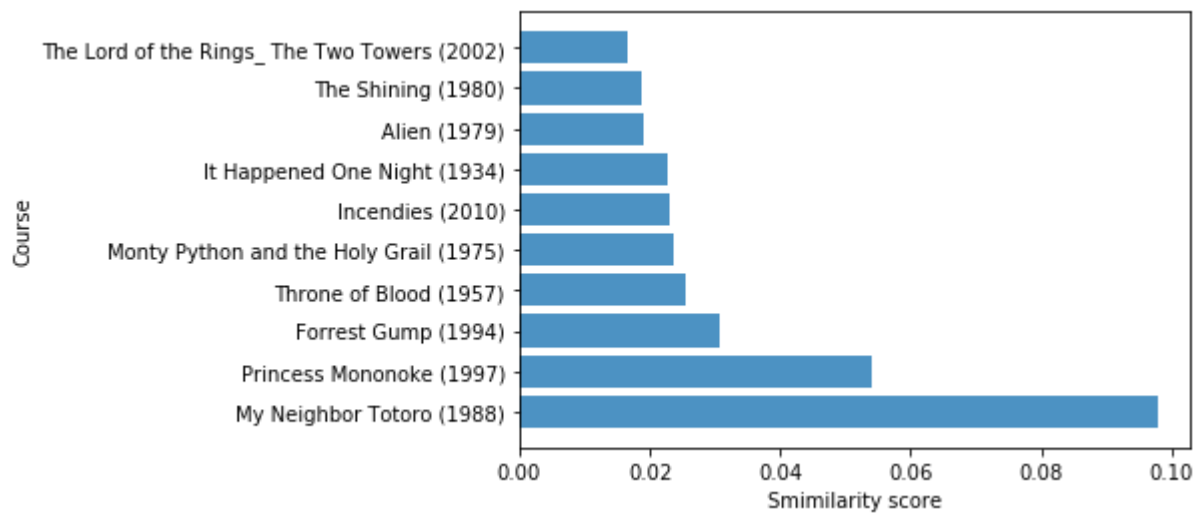
Speed

- Without feature selection: **229** s
- With feature selection: **90** s
- Reduce : **60.7**%

Number of feature

- Without feature selection: **27004** features
- With feature selection: **16539** features
- Reduce : **38.8**%

Query Result

Conclusion

The result is same as what we expect. In our outcome, we can find that "Totoro" get the highest score in the similarity score, which finally meet the goal we set.

However, at the beginning the result of second and third place doesn't fit the features of Japanese animation very well, but after the feature selection, our similarity of "Totoro" is improved significantly.

After the feature selection, our tfidf processing time reduces about 60% and the number of features both decrease.