**Synonyms List Establishment**

Jinjin Xiao, Yi Luo

**Introduction**

To achieve the goal of the Search Engine project that to classify materials with 17 sustainable development goals, text classifiers should be build up and applied on the information.

**Data**

Text: the overall text material is 17 sustainable development goals.

**Methodology**

[WordNet](https://wordnet.princeton.edu/) is a lexical database for the English language, which was created by Princeton, and is part of the NLTK corpus. WordNet has been used alongside the NLTK module to find the meanings of synonyms.

**Procedure**

1. The details of 17 goals have been studied, and domain knowledge have been exploited to understand the content of sustainable development in 17 expects.
2. The main idea of each goal has been concluded in to multiple keywords.
3. Keywords have been written into .csv file and formed into a fix format.
4. Input the keywords to python code and run.
5. Each words in the keywords list has a list of synonyms as the output.
6. The total output of synonyms is transformed into a excel file.

**Tool selection**

Python

**Result**

Final excel file is got with all the keywords and synonyms.

**Limitation**

There are some limitations in the keyword list.

1. When the keywords were extracted manually, the accuracy was not certain because of the diversity thinking of the material.
2. The tool was limited because only wordnet is used in python and the wordlist got may not enough.

**Optimization**

The article are supposed to find about 17 goals and we want to traverse the whole article to find the word frequency without stop words, finally add these chosen words to the keyword list.

**Further Research**

1. The synonyms list can also be applied in the searching step in the search engine as the search suggestion.
2. The accuracy may be improved by studying the features of retrieval materials. New rules of classification can be build up to update the synonyms list.