**Learner:** Nguyen Lam Nhat Anh

**ANSWER**

1. There are 2 primary sub-problems that need to be solved in solving the overall problem:

* Find all the synonym words associated with the keyword in the thesaurus.
* Counting the number of occurrences of the keyword and its synonyms in all the documents in the corpus.

1. - The first pattern: both sub-problems have a similarity solution, which is that we must check each word individually in the thesaurus as well as in the corpus to search for the word we expected.

* The second pattern: In the second sub-problem, we search for the word individually and count its occurrences in a document. Then, we just do the same thing for the remaining documents.

1. - The thesaurus is a set of word synonyms, so there are at least one or more words associated with the keyword, which are input for the second sub-problem: **input-synonym**.

* The corpus is a set of documents, which is a collection of the number of occurrences of an object from all documents: **all-document-object**.
* Each document is a set of the number of occurrences of an object: **document-object**.

1. Algorithm:

* Enter a keyword.
* Add it to a list called ***"input-synonym."***
* Check the first word in the thesaurus if it relates to the key word. If it does, add the word to the input-synonym list. Otherwise, put that word aside and go to the next word.
* Repeat step 3 for each word until reach the end of the thesaurus.
* Start with the first word in the input-synonym and the first document.
* Set the value of ***count*** to 0.
* Check if the first word in the document matches the first word of the input-synonym list. If it does, ***count*** plus 1. Otherwise, put that word aside and go to the next word. Repeat this step until we reach the end of the document and move to the next document.
* Repeat step 7 until we reach the end of the corpus.
* The value of ***count*** now is the number of occurrences of the expected word in the corpus. Output ***count*** and go to the next word in the input-synonym list.
* Repeat steps 6 to 9 until we reach the end of the input-synonyms list.

1. One problem that I may encounter in my day-to-day life with regard to determining the number of occurrences of a word and its synonyms in a corpus is when I do my scientific research. The research paper includes many documents, so it definitely contains a lot of synonyms to avoid repeating words too much. By going through all the text documents for searching for a given word and its synonyms, I can determine the number of occurrences for each word to see whether I'm overusing one word or using its synonymou in inappropriate context or not.