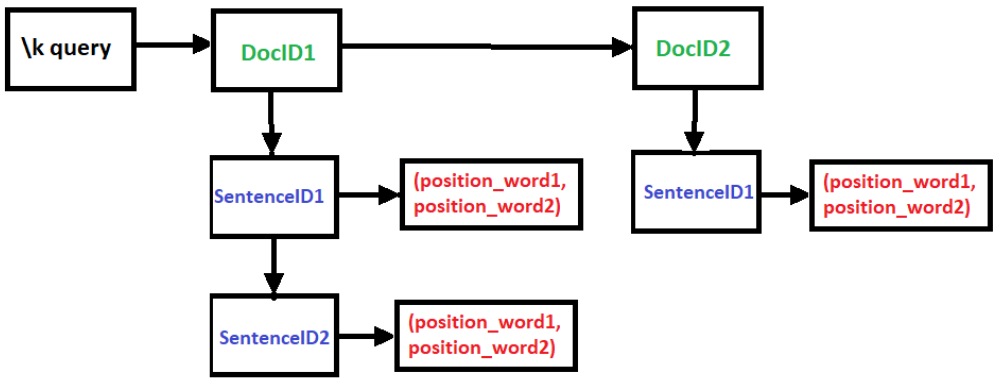
Part 1:

To find the documents in which "I" and "scientist" are at most 2 words apart, a \k query must be implemented. For example, a \k query of the type "word1 \2 word2" will find the occurrences of word1 within k words of word2. Hence, on this case we implement the query "I \2 scientist" to find the documents where the two words are at most two words apart.

Since the terms must be in the same sentence, there must be a pre-parsing step that identifies sentences in documents. Assuming such step exists, the resulting posting list for the \k query considering only same sentence occurrences will indicate the docID in which it found a match, for each docID it will indicate the sentenceID in which the match is and for each sentenceID it will show the tuple (position_word1, position_word2) that indicates the position of each word in that sentence. The following figure shows a general example of what the posting list would look like for a \k query that considers two words:



The output of the query will therefore be in the form:

<1: {2: {6,8}, 50: {1,3}}, 14: {3: {8,10}}>

This result means that the words "I" and "scientist" were found two words apart in document 1, sentence 2 in positions 6 and 9, in document 1, sentence 50 in positions 1 and 4, and in document 14, sentence 3, positions 8 and 11.

Part 2

The index for malware search would contain the fields malware, author and description. Consider the following example of an index:

Type	Author	Description
Ransomware	Author1	Type of malware that encrypts files on the victim's computer and demands payment in exchange for the decryption key.
Virus	Author2	Computer virus that infects executable files and spreads through email attachments and infected USB drives.

The previous index includes two malware objects, each with a type, author and description. The information on the index can be used for search and retrieval of malware objects based on specific criteria. For example, a user can search all the malware objects of the type “ransomware” and use such information for analysis.