STUDENT NAME: LEE YU XIAN
STUDENT NUMBER: 7233164

Question 1

==

1) Assume both files (crime-stories.txt and patterns.txt) have been uploaded to HDFS.

An objective to you is to implement a MapReduce application in Java that finds the total number of statement in a file that matches a pattern in another file.

Assume crime-stories.txt contains:

text
123
456
789
abc
def
⋮

Assume patterns.txt contains:

Pattern
[1-3]
[a-c]
[4-6]
[d-f]
[7-9]
⋮

Implementation of Mapping phase

Both files crime-stories.txt and pattern.txt are converted into <key,value> pairs, where key = text and value = pattern. In the case of our given objective, a <key,value> pair is created for each line found in crime-stories.txt.

Hence, output of Mapper would be:

```
[123, [1-3]]
[456, [1-3]]
[789, [1-3]]
   ⋮
[123, [a-c]]
[456, [a-c]]
   ⋮
```

STUDENT NAME: LEE YU XIAN
STUDENT NUMBER: 7233164

Question 1

==

Implementation of Reducing phase

Reduce phase operates on one set of <key,value> pair. The <key,value> pair is pass into the function called match(text-line, text-pattern), where the textline = key and text-pattern = value. If the function returns true, a counter will be added. In the end, each pattern with its respective counter will be written to and ouput.

Hence, the result of the reducer is:

[1-3], 1
[a-c], 1
[4-6], 1
⋮