For this Etude of finding words under a specific definition of what a word is I wrote 2 programs in 2 different languages. One in python and the second one in java.

I wrote the python program first. I started by taking the file name argument and opening the file and reading each line. For each line I split at the white spaces into an array of tokens. I then ran a for each loop and sent each word to a function to check if it was a word. In java I used a scanner object to get each token and send it to a function to check if it is a word. For this part I preferred the more powerful scanner function in java to get each word. While the same can be achieved in python with a for each loop it reduces the number of arrays to worry and only need to check if there is a next token and if there is get the next token.

In python and java to check if something was a word, I used a regex statement to check if the token met our definition of a word including if it contained allowed punctuation symbols. While the regex was the same this was simpler in python and can be done in 1 line and was therefore just used at the top of an if statement. In java on the other hand this required constructing a pattern object and then a match object and checking if it matches. I therefore made separate function that returned true if it did match and false if not and included it in the if statement.

Similarly, when adding the word to the words set I was using in both languages, I was using the same regex pattern for both. However, in python this involved less setting up of object and therefore I just added straight from the if statement to the words set while in java I made this a separate method.

For both languages a set was used to store the words as a set allows for no duplicates. In python this was a very quick short line: “words = set()”. In java it is required to specify the data type of the set which makes it longer but does make it more clear what the set is being used for.

For this etude that was the main difference between the 2 languages. Python simpler and shorter and faster to write while java takes a little longer to write but code readability is definitely better as each function as a more clear purpose and each variable is more clear on what it does. This means python was easier initial quick write for one specific purpose but if I would like to change how the program works (like add another condition to being a word) java would be easier to go back to as I would only have to change the isWord method.