

Overview

In this practical, we were tasked with creating a program where the user entered a phrase in the command line and the program searched through books and retrieved what phrases were equal to or greater than the jaccard index specified as the user specified.

In my implementation, I successfully implemented the basic requirements of the spec, however I do often see errors in my output.

Design

I decided to make my program with 2 classes, CS1003P4.java and BackEnd.java to separate the back end from the front end.

The main method was in CS1003P4, this class accepted the command line arguments and passed them through to BackEnd.java.

BackEnd.java is made up of 5 methods, Search, ConstructSubsequences, CompareSubsequences, CalculateBiGram and Jaccard Coefficient.

In the search method, this initialises Apache spark, I done it this way since it is the only way I know how to. Afterwards I use the variable phrase word count to hold the value of however many words the output should contain. I thought the most efficient way to read in the files would've been through an array of string object to then using a nested for loop to reads into RDD as well as create and compare the subsequence's.

In regard to creating and comparing subsequence's, I created them by firstly calculating the amount of subsequence's needed, I then used another nested for loop, this is a lot more efficient than making 2 separate loops or using a do while.

In the following method I just checked if they met the requirements, if the Jaccard coefficient was equal to or greater than what was required.

For creating bigrams, it was the exact same as the code we previously used so I recycled that.

I would've preferred to use the Sorensen dice coefficient since I thought it was more accurate however I had to use the Jaccard coefficient since it was a requirement. To do this, I just reused some old code as well as used the formula that can be found on google for it.

Testing

I started my testing by doing the tests provided to us.

What is being tested/name of test method	Expected outcome	Actual outcome	Evidence
			Can be found at end of report

Advice	“She generally gave herself very good advice though she very seldom followed it”		
Not-found			
Sail	“him setting sail to the rising” “sail to the rising wind the” “setting sail to the rising wind”	“setting sail to the rising wind”	
Tree	“Hide the Christmas tree carefully” “The Christmas tree carefully helen”	“Hide the Christmas tree carefully” “The Christmas tree carefully”	
What-50	"adopt this sort of passiveness in" "and that sort of thing is" "guess of what sort it was" "helmer what sort of madness is" "his head what sort of a" "in a sort of hiss what" "is a woe that is madness" "is that it makes one so" "it is something of that sort" "later helmer what sort of madness" "of mine this is what I" "of this man or that man" "of this sort of deliciousness is" "part that sort of thing is" "shipmates when this sort of steady" "sort of hiss what i say"	“Is that the sort of” “So in that broad madness” “what business is that of” “what the wildness of this” “a woe that is madness” “and that sort of thing is” “guess of what sort it was” “his head what sort of” “is a woe that is madness” “is that it makes one so” “it is something of that sort” “of this sort of deliciousness is” “sort of hiss what” “t guess of what sort it”	

	"sort of madness is this nora" "t guess of what sort it" "that manner the shortness of the" "that sort of thing is to" "things of that sort but I this is the sort of weather" "what sort of a dress i" "what sort of girl is miss" "what sort of girl is miss" "what sort of madness is this"	"that manner the shortness of" "This by what sort of" "What sort of a dress I" "what sort of a place this" "what sort of girl is miss" "what sort of girl is miss" "what sort of madness is this" "whether this the sort of"	
What-62	"helmer what sort of madness is" "sort of madness is this nora" "what sort of madness is this"	"What sort of madness is this"	
What-75	"what sort of madness is this"	"What sort of madness is this"	

Through out testing, I suspected that when my program wasn't matching the expected output exactly, it was always when the words spilled over to the next line in the book. I found this true for a few of the books I checked but was unsure why this would affect the output so greatly.

Evaluation

My submission successfully implements all parts of the spec however isn't completely robust or reliable.

Conclusion

In conclusion, can't say I'm happy with how it turned out due to there being so many errors but I done everything I could've done. With more time I would've ran more tests and figured out the cause and solution to these errors.

Evidence

```
ad352@pc7-140-l:~/Documents/CS1003/P4/CS1003-P4 $ Tests/queries/tree/test.sh
hide the christmas tree carefully | hide the christmas tree carefully
the christmas tree carefully      | the christmas tree carefully helen
ad352@pc7-140-l:~/Documents/CS1003/P4/CS1003-P4 $ Tests/queries/sail/test.sh
setting sail to the rising wind   | him setting sail to the rising
                                   > sail to the rising wind the
                                   > setting sail to the rising wind
ad352@pc7-140-l:~/Documents/CS1003/P4/CS1003-P4 $ Tests/not-found/sail/test.sh
bash: Tests/not-found/sail/test.sh: No such file or directory
ad352@pc7-140-l:~/Documents/CS1003/P4/CS1003-P4 $ Tests/queries/not-found/test.sh
ad352@pc7-140-l:~/Documents/CS1003/P4/CS1003-P4 $ Tests/queries/advice/test.sh
                                   > she generally gave herself very good advice though she very s
ad352@pc7-140-l:~/Documents/CS1003/P4/CS1003-P4 $ Tests/queries/what-50/test.sh
is that the sort of               | a sort of hiss what i
so in that broad madness          | adopt this sort of passiveness in
what business is that of          | and that sort of thing is
what the wildness of this         | guess of what sort it was
a woe that is madness             | helmer what sort of madness is
and that sort of thing is         | his head what sort of a
guess of what sort it was         | in a sort of hiss what
his head what sort of             | is a woe that is madness
is a woe that is madness          | is that it makes one so
is that it makes one so           | it is something of that sort
it is something of that sort      | later helmer what sort of madness
of this sort of deliciousness is  | of mine this is what i
sort of hiss what                 | of this man or that man
t guess of what sort it           | of this sort of deliciousness is
that manner the shortness of     | part that sort of thing is
this by what sort of              | shipmates when this sort of steady
what sort of a dress i            | sort of hiss what i say
what sort of a place this         | sort of madness is this nora
what sort of girl is miss         | t guess of what sort it
what sort of girl is miss         | that manner the shortness of the
what sort of madness is this      | that sort of thing is to
whether this is the sort of         | things of that sort but i
                                   > this is the sort of weather
                                   > what sort of a dress i
                                   > what sort of girl is miss
                                   > what sort of girl is miss
                                   > what sort of madness is this
ad352@pc7-140-l:~/Documents/CS1003/P4/CS1003-P4 $ Tests/queries/what-62/test.sh
what sort of madness is this      | helmer what sort of madness is
                                   > sort of madness is this nora
                                   > what sort of madness is this
ad352@pc7-140-l:~/Documents/CS1003/P4/CS1003-P4 $ Tests/queries/what-75/test.sh
what sort of madness is this      | what sort of madness is this
ad352@pc7-140-l:~/Documents/CS1003/P4/CS1003-P4 $ ^C
ad352@pc7-140-l:~/Documents/CS1003/P4/CS1003-P4 $ █
```