word-counter

Приложение предназначено для анализа множества текстовых файлов, и выдачи списка 10 самых часто используемых слов, длина которых входит в переданный диапазон.

Сборка

Сборка проекта выполняется при установленной java 17 командой

```
gradle clean doReadyRun
```

После успешной сборки в корневой папке проекта будет готов к запуску `jar` файл проекта

Запуск

Запуск программы производится из командной строки с параметрами

```
java -jar text-word-counter-1.0-SNAPSHOT.jar --min=5 --max=10 --threads=5 --files
text1.txt text2.txt data
```

Описание параметров:

--min=5 - минимальный размер слова для анализа

--max=10 - максимальный размер слова для анализа

--threads=5 - число потоков для анализа слов

--files text1.txt text2.txt data text* - этот параметр обязательный. Он должен указывается последним в списке параметров, и все последующие значения параметров ассоциируются как его множественные значения. В качестве значений можно указывать как файлы, так и папки, и если ваш командный интерпретатор преобразует указанные маски в конечные имена файлов, то можно указывать и маски файлов при запуске. Обратите внимание, что будет производиться обработка только файлов с расширением `.txt`. Если в качестве парамера указана папка, то будут найдены и обработаны все файлы с расширением `.txt` рекурсивно.

Особенности

Программа обрабатывает текстовые файлы в кодировке uncode. При разборе текста предполагается что любая строка содержит слова целиком, то есть в тексте нет переносов слов с одной строки на другую. Слова выделяются из текста на основании правил: слово

начинается и заканчивается буквой, внутри себя слово между буквами может содержать разделительный символ -. Таким образом, например воспринимаются как целые слова: `по-русски`, `по-пожаски`, `только-только`, `когда-нибудь`. Если же символ - встречается в начале или конце слова, то этот символ на слово не будет влиять, и будет восприниматься как разделительный.