

Общие сведения

Приложение должно скачивать HTML-страницу посредством HTTP-запроса, сохранять загруженный файл на жесткий диск компьютера, подсчитывать количество уникальных слов на странице и выводить статистику по словам в консоль.

Технологии

- Java 8+
- Git
- Maven

Задание

Имеется входная строка с адресом web-страницы. Приложение должно скачать web-страницу, сохранить ее на жесткий диск. Затем текстовый контент, который есть на ней (который видит пользователь на странице браузера), разбивать на слова и подсчитывать количество нахождений на странице каждого уникального слова. Разделителями слов считать следующие символы: {' ', ',', '.', '!', '?', '"', ';', ':', '[', ']', '(', ')', '\n', '\r', '\t'}

Пример входной строки:

https://www.simbirsoft.com/

Пример вывода статистики вхождений слов:

- СОЗДАЁМ 1
- ПРОГРАММНОЕ 1
- ОБЕСПЕЧЕНИЕ 2
- ДЛЯ 16
- POCTA 1
- ВАШЕГО 3
- БИЗНЕСА 3

Обязательные требования

- Необходимо учесть, что размер скачанной web-страницы может превышать размер доступной ОЗУ. В случае, если файл больше объема оперативной памяти, подсчет количества слов должен осуществляться корректно;
- В программе необходимо предусмотреть обработку нештатных ситуаций;
- Программа должна обеспечивать корректную обработку исключительных ситуаций (exceptions);



• В решении необходимо использовать стандартные классы – например, если в программе требуется список или словарь, тогда нужно использовать не собственные вручную написанные списки и словари, а имеющиеся в стандартной библиотеке.

Требования к постановке (список файлов для проверки)

- Все файлы поставляются в виде одного архива минимального размера;
- Архив должен содержать:
 - о папка с исходным кодом;
 - папка с готовым к исполнению модулем. Данная папка также включает в себя 2-3 тестовых примера, на которых программу легко можно проверить.
- Readme-файл, содержащий:
 - Руководство по запуску Вашего приложения;
 - Руководство пользователя.

Требования к качеству продукта

- Требования к качеству кода профессиональный уровень. В том числе наличие стиля форматирования обязательное. Комментарии обязательные для классов и для публичных методов-свойств. Разбиение программы на классы, реализация классов в отдельных файлах. И т.д.;
- Исполняемый код должен корректно выполнять вышеописанный функционал.

Дополнительные требования

- Использовать многоуровневую архитектуру;
- Обеспечить расширяемость решения;
- Использовать паттерны проектирования;
- Использовать юнит-тесты;
- Логирование системных ошибок;
- Сохранять результат подсчета частоты нахождения слов на странице в базу данных.