

Руководство программиста

1. Назначение и условие применения программы

1.1 Назначение программы

Данная программа была создана для обработки текстовых файлов, а именно для удаления одинаковых слов в текстовом файле(ах). Принцип работы программы основан на свойстве многопоточности. Таким образом, разработанная программа работает для указанного числа потоков, которые параллельно производят обработку текстовых файлов.

1.2 Функции, выполняемые программой

К основным функциям разработанной программы относятся:

- Запуск программы из командной строки (при тестировании использовались средства ввода аргументов командной строки IntelliJ IDEA 13.1.4)
- Создание n-го числа потоков и распределение между ними текстовых файлов, подающихся на обработку. Следует заметить, что если число потоков больше числа файлов (f), тогда число потоков устанавливается в значение f.
- Удаление одинаковых слов из файлов.
- Отображение в консоли хода выполнения программы.
- Вывод результатов файловой обработки в консоль.

2. Описание структуры программы

I. Class FileRuler

Данный класс предназначен для загрузки файлов в память из указанной директории и распределении файлов между потоками для их последующей обработки.

II. Class Distributor

Класс, создающий так называемую «карту распределения» файлов между потоками. Распределение основано на принципе деления без остатка числа файлов на число потоков. Остаток файлов забирает последний поток.

III. Class MyThread

Класс, переопределяющий поведение стандартного потока. Для этого переопределяется метод `run()`, в котором описана функциональность, позволяющая удалять одинаковые слова из коллекции текстовых файлов (потому что поток может обрабатывать не только один файл).

3. Входные и выходные данные

3.1 Входные данные

Входными данными для данной программы являются:

- 1) число потоков;
- 2) директория, содержащая текстовые файлы.

Файл, содержащий слова, отформатирован следующим образом: каждая строка файла содержит ровно одно слово.

3.2 Выходные данные

В качестве выходных данных выводится статистика о работе программы:

- 1) общее число файлов;
- 2) общее число строк во всех файлах;
- 3) число удаленных строк.

Директория, которая была указана в качестве входных данных после работы программы, содержит измененные файлы.