C# Grep

Ing. Andreas Schöngruber, 11725595

April 2020

1 Introduction

This project is a grep command-line implementation with C# using the Task Parallel Library (TPL). It can be used to find regular expressions in files. The program uses several tasks to improve the search performance and prints filename and all lines containing the regular expression.

I chose this for my project because I always wanted to play around with the TPL. As my version of grep does not require to have a sorted output it is a really good use case for the TPL.

2 Program arguments

The program needs 3 arguments to work:

- Grep option: -S for synchronous computation or -P for parallel computation.
- Regular expression we are looking for e.g. "java".
- Search directory: The directory where the program should start with the search. The search includes all sub-directories.

A valid example would be:

```
Grep.exe -p "java" "C:\Program Files\Git"
```

3 Output

The program prints all occurrences of the regular expression in the files to the output. The output has following format:

```
[Thread id]: Absolute file path: line content
```

An example output can be seen in figure 1. The filename is highlighted in purple and the regular expression is highlighted in green.

When the program finishes the search it prints a short summary about the search to the output. This summary is shown in figure 2. It states the total occurrences of the regular expression and the duration of the search. There is also a list of the creation times of threads. This was particularly interesting for me because I wanted to see when the operating system decides to provide another thread to the program. We can see the program started with 4 threads and then created new threads from time to time.

Output formatting and thread information has an impact on performance! Nevertheless as it was intended to gain some experience and information about the TPL in action I left it in the source-code.

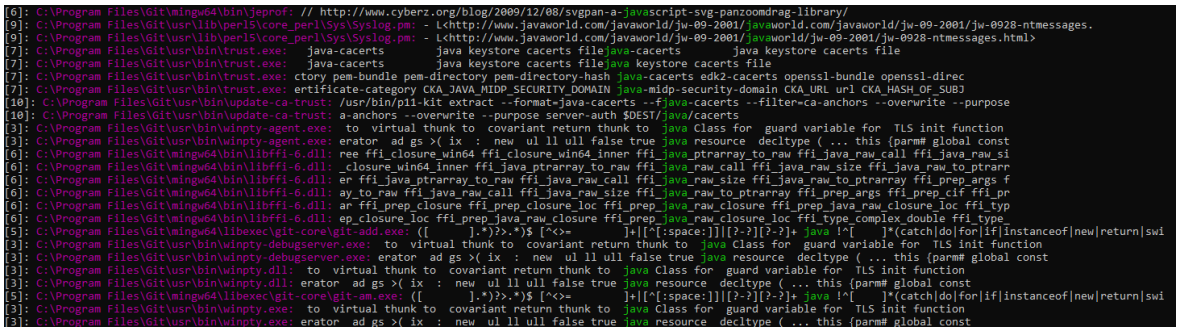

Figure 1: Program output: Search results

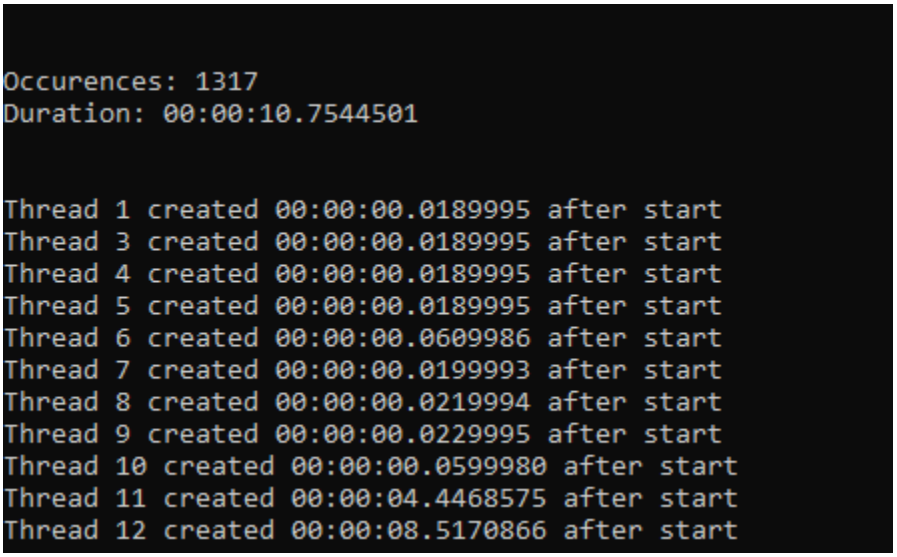

Figure 2: Program output: Search summary

4 Program settings

The program has three settings which can only be set in the source code:

- SHOW_EXCEPTIONS: Determines if exceptions should be printed to the output. Otherwise exceptions e.g. IO Exceptions will be ignored completely. Default = false.
- SHOW_THREAD: Prints the thread id to the output. Default = true.
- LINE_SIZE: How many characters of a line should be printed to the output. Default = 50.