



Parallele Programmierung mit C++ (SS 2015)

Abgabe bis zum 22. Mai 2015, 14:00 Uhr

Lernziele:

- Aufteilung und Parallelisierung eines Problems mit Hilfe des Fork-and-Join-Patterns
- Sammlung und Aggregation der Resultate mit Hilfe einer gemeinsamen Datenstruktur, die den gegenseitigen Ausschluss sicherstellt

Aufgabe 6: Schneller Suchen

Die Suche einer Zeichenkette in einer größeren Datei lässt sich gut parallelisieren, indem der Dateiinhalt auf n Threads aufgeteilt wird, die jeweils ihren Bereich durchsuchen.

Zu erstellen ist ein Werkzeug namens *pgrep*, das drei Parameter erhält: Die Zahl der zu verwendenden Threads, die zu suchende Zeichenkette und die Eingabedatei. Auszugeben sind in sortierter Reihenfolge alle Zeilen, in der die Zeichenkette vorkommt, jeweils mit der zugehörigen Zeilennummer.

Hinweise: Um die Suche einfach und effizient zu gestalten, empfiehlt es sich, den Dateiinhalt mit *mmap* in den eigenen Adressraum abzubilden. Wenn Sie möchten, können Sie hierfür die zur Verfügung gestellte Klasse *MappedFile* aus *mapped_file.hpp* verwenden.

Wenn Sie den Dateiinhalt auf n Threads aufteilen, sollten Sie darauf achten, dass Sie den Threads jeweils ganze Zeilen geben. Beachten Sie auch den Sonderfall, dass Sie weniger Zeilen als Threads haben.

Da Sie nicht von vorneherein wissen, wieviele Zeilen die einzelnen Teilbereiche umfassen, sollte jeder Thread zunächst mit relativen Zeilennummern arbeiten. Erst bei der Aggregation können Sie diese in absolute Zeilennummern konvertieren.

Für die Sortierung empfiehlt sich die Verwendung der STL-Container-Klasse *std::set*. Ein Beispiel dazu findet sich im Vorlesungsbeispiel *queens2.cpp*.

Wenn Sie möchten, steht es Ihnen frei, reguläre Ausdrücke zu unterstützen unter Verwendung der entsprechenden C++-Bibliothek, die mit **#include** <regex> zur Verfügung steht.

Für Tests empfiehlt sich die 94 Megabyte große Datei */home/www/htdocs/sai/ws14/cpp/uebungen/12/movies*, in der jede Zeile aus einem Filmtitel aus

den daran beteiligten Schauspielern besteht, jeweils durch Tilden getrennt. Bitte kopieren Sie diese Datei nicht in Ihre eigenen Verzeichnisse bei uns. Verwenden Sie stattdessen symbolische Links.

Packen Sie all Ihre Quellen mitsamt Ihrem *Makefile* in ein kleines *tar*-Archiv ein, das Sie dann einreichen können:

```
tar cvf pgrep.tar *.*pp [mM]akefile
submit pp 6 pgrep.tar
```

Viel Erfolg!