

Protokoll: Parallelisierung und Synchronisation in MyFind

Einleitung: Das Programm **MyFind** ist ein Suchwerkzeug, das mit der Programmiersprache C++ für Linux entwickelt wurde. Es hilft dabei, Dateien in einem Verzeichnis oder in dessen Unterordnern zu finden. Damit es funktioniert, muss man mindestens einen Ordnerpfad und einen Dateinamen angeben. Zusätzlich gibt es folgende Optionen:

- **-R:** Sucht auch in Unterverzeichnissen.
- **-i:** Ignoriert Groß- und Kleinschreibung beim Vergleich der Dateinamen.

Wenn eine Datei gefunden wird, gibt das Programm das Ergebnis in diesem Format aus:

<pid>: <filename>: <complete-path-to-found-file>\n

Hierbei steht **<pid>** für die Prozess-ID des Programms, das die Datei gefunden hat.

Wie die Parallelisierung funktioniert: Um mehrere Dateien gleichzeitig suchen zu können, wird für jede Datei ein eigener Kindprozess erstellt. Dazu verwendet das Programm die Funktion **fork()**. Jeder Kindprozess arbeitet dabei eigenständig und durchsucht den angegebenen Ordner (und Unterordner, wenn **-R** aktiviert ist) nach einer bestimmten Datei.

Wichtige Details:

1. **Kindprozesse:**
 - Jeder Prozess sucht nach genau einer Datei.
 - Sobald die Suche abgeschlossen ist, beendet sich der Kindprozess mit **_exit()**.
2. **Hauptprozess:**
 - Der Hauptprozess startet alle Kindprozesse.
 - Mit der Funktion **wait()** wartet er darauf, dass alle Kindprozesse fertig werden. Das verhindert sogenannte Zombie-Prozesse, die Ressourcen verschwenden.

Wie die Ausgabe synchronisiert wird: Die Suchergebnisse werden direkt über **stdout** (Standardausgabe) ausgegeben. Zusätzliche Tools wie Message Queues sind nicht notwendig, da Linux sicherstellt, dass die Ausgaben der Prozesse zeilenweise erfolgen. Dadurch werden Konflikte beim parallelen Schreiben vermieden.

Vorteile dieses Ansatzes:

- **Einfache Struktur:** Es wird keine zusätzliche Infrastruktur benötigt.
- **Fehlertoleranz:** Wenn ein Verzeichnis nicht zugänglich ist, wird der Fehler sauber in **stderr** (Standardfehlerausgabe) gemeldet, ohne dass das Programm abstürzt.

Herausforderungen:

- **Reihenfolge der Ergebnisse:** Da die Prozesse parallel arbeiten, können die Suchergebnisse in unterschiedlicher Reihenfolge erscheinen.
- **Fehler in Kindprozessen:** Der Hauptprozess überwacht die Kindprozesse und meldet, wenn ein Fehler aufgetreten ist.

Zusammenfassung: Mit der Funktion **fork()** kann MyFind mehrere Dateien gleichzeitig suchen, was die Leistung des Programms deutlich verbessert. Durch die Synchronisation der Ausgaben über **stdout** bleibt das Programm einfach und robust. Dieser Ansatz ist besonders gut geeignet für Anwendungen, bei denen viele Dateien effizient durchsucht werden müssen.