

Übungsaufgabe: Parallelisierung der Indizierung einer Desktop Suche

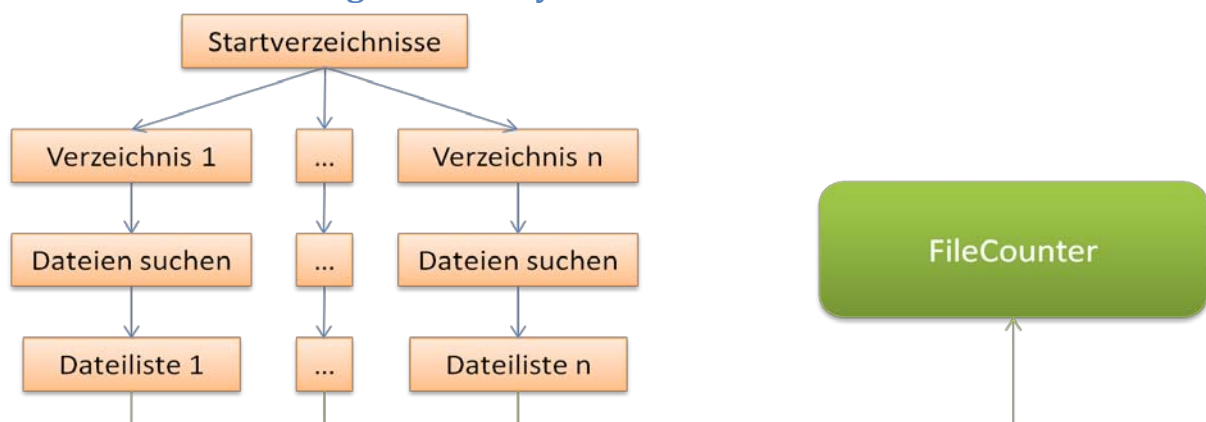
Es soll eine Desktopsuche gebaut werden, welche die Fähigkeiten aktueller Multi-Core-Systeme ausnutzen kann und damit schneller arbeiten kann. Die aktuelle Implementierung ist sequentiell und soll nun in einer neuen Version parallelisiert umgesetzt werden. Dabei sollen Aspekte aus dem aktuellen Workshop Vortrag zur Umsetzung verwendet werden.

Eine Eclipse Installation findet sich unter:

```
C:/Temp/eclipse-iws-multithreading/eclipse/eclipse.exe
```

Diese Umsetzung soll in drei Teilen durchgeführt werden welche nun genauer erläutert werden.

Teil 1: Parallelisierung des Dateisystem-Crawlers



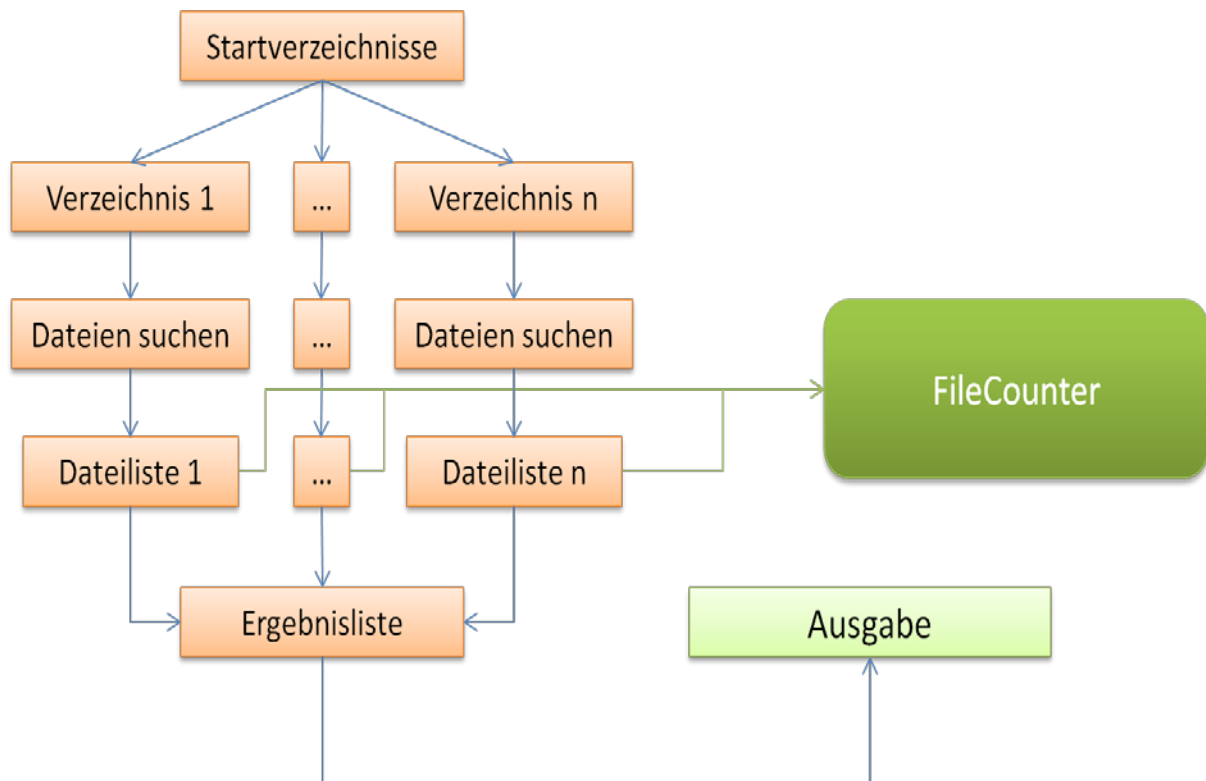
Im ersten Schritt soll das Durchlaufen der verschiedenen für die Desktop Suche verwendeten Verzeichnisse parallelisiert werden. Das Programm erhält für die Indizierung eine Liste der (Toplevel-) Verzeichnisse in denen Dateien gesucht und indiziert werden sollen, welche einem bestimmten Kriterium entsprechen, in unserem Fall, dass es Textdateien sind.

Die Parallelisierung soll hierbei über die Anzahl der Verzeichnisse gehen und mittels des Executor-Frameworks umgesetzt werden.

Folgende Schritte sollen dabei durchgeführt werden:

1. Implementieren Sie die Methoden der *FileCounter* Klasse so, dass diese threadsicher verwendet werden können.
2. Erzeugen Sie in der *main*-Methode der ***DesktopSearch***-Klasse eine gültige *FileCounter*-Instanz sowie einen sinnvollen ***ExecutorService***.
3. Passen Sie die Klasse ***CrawlJob*** so an, dass diese als Task an den ***ExecutorService*** übergeben werden kann.
4. Lassen Sie in der Klasse ***CrawlJob*** den Counter erhöhen.
5. Fügen Sie der *main*-Methode der *DesktopSearch*-Klasse die Erzeugung der ***CrawlJob***-Instanzen hinzu.

Teil 2: Erweitern des Dateisystem-Crawlers



Nachdem im ersten Teil nur die Anzahl der gefundenen Dateien gezählt wurde, muss dieser Mechanismus nun erweitert werden um eine spätere Indizierung zu ermöglichen.

Deshalb soll im zweiten Schritt nun das Durchlaufen der Verzeichnisse so realisiert werden, dass die Ergebnisse (die gefundenen Dateien) als Liste vorliegen.

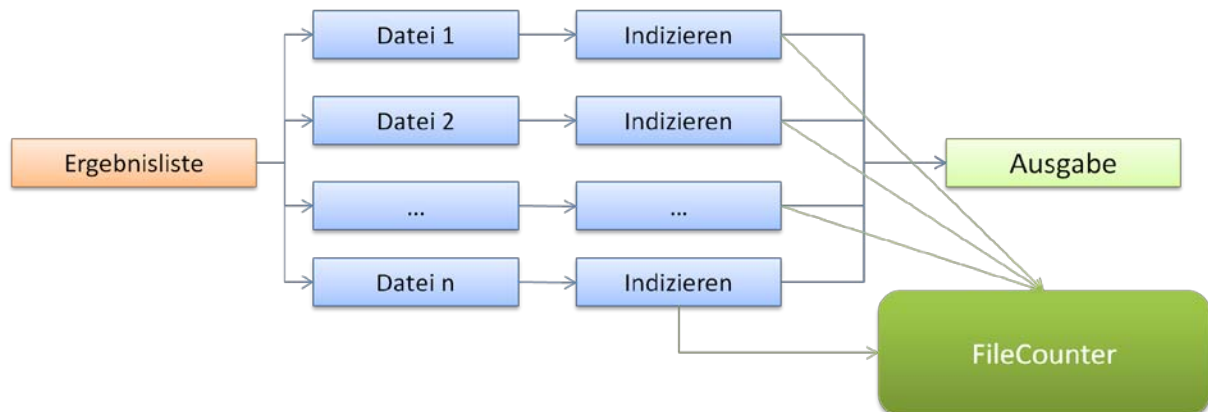
Schließlich sollen die Namen der Dateien sowie die Anzahl der gefundenen Dateien ausgegeben werden.

Folgende Schritte sollen dabei durchgeführt werden:

1. Passen Sie die Klasse **CrawlJob** so an, dass diese eine Liste der gefundenen Dateien zurückliefern kann.
2. Passen sie die Verarbeitung der Tasks (*CrawlJob*) so an, dass die Rückgabewerte übernommen werden.
3. Geben sie alle Rückgabewerte auf der Konsole aus (als Dateinamen)

Tipp: Schritt 2 und 3 erfordern einen Blick in die Zukunft!

Teil 3: Anbinden der Indizierung an den Crawler



Nachdem nun zwei Versionen des Crawlers erstellt wurden, wurde in Teil 3 die Desktopsuche nun um das Indizieren der Dateien erweitert.

Dies geschieht durch die Verwendung eines Threads der das Indizieren übernimmt. Die Dateien werden vom Crawler per Queue an den Thread weitergegeben.

Folgende Aufgaben sind zu untersuchen:

1. Vollziehen sie die Änderungen nach, die zum vorigen Schritt gemacht wurden.
2. Vergewissern sie sich, wie die Übergabe der Werte vom Crawler zum Indizierer funktioniert.
3. Überprüfen sie, wie dafür gesorgt wird, dass das Programm sauber beendet wird.