

# 1. Übungsblatt

Ausgabe: 15.11.04

Abgabe: 01.12.04

---

## 1] Aufräumarbeiten

20 Punkte

Deutsche Festplatten sind voll von Dateileichen. Fast jedes Programm legt Sicherungskopien an, ohne sie irgendwann wieder zu entfernen. Wie eine Studie des Beratungsunternehmens *Frei & Erfunden* ergab, sind in Deutschland mehrere tausend Terabyte Festplattenplatz mit Dateien belegt, die nie wieder gebraucht werden. Dadurch entsteht ein volkswirtschaftlicher Schaden in Höhe von bis zu drei Milliarden Euro. Aber zum Glück gibt ja die Studenten der Uni Bremen, die sich dieses Problems in Form dieses Übungsblattes annehmen.

Es ist also ein Programm zu implementieren, welches solche Sicherungskopien entfernt. Dieses Programm wird von der Kommandozeile aus aufgerufen, traversiert die als Argumente angegebenen Dateibäume und löscht alle Sicherungskopien, die einem spezifizierten Muster entsprechen. Der Aufruf erfolgt wie folgt:

```
cleanup [options]
```

Das Programm sollte ein Logfile seiner Aktivität (also insbesondere der gelöschten Dateien) erzeugen, und auf der Standardausgabe ausgeben.

Als Optionen sind folgende vorgesehen:

**-p** *pattern*: Spezifikation für ein Muster zu löschender Dateien.

Eine Spezifikation besteht aus einem Muster, optional gefolgt von einer Zeitspezifikation:

$$\begin{aligned} \textit{patternspec} &::= \textit{pattern} [-\textit{mtime} \textit{timespec}] \\ \textit{timespec} &::= \textit{numberh} | \textit{numberd} | \textit{numberm} \end{aligned}$$

Ein Muster (*pattern*) ist eine gewöhnliche Zeichenkette, mit den folgenden besonderen Zeichen:

- Das Muster `*` passt auf jeden String einschließlich des leeren;
- Das Muster `?` passt auf jedes Zeichen;
- Eine Liste von Zeichen `[c1 . . . cn]` passt auf jedes der Zeichen `c1, . . . , cn`. Wenn das erste Zeichen ein `^` oder `!` ist, passt es auf *keines* der Zeichen `c1, . . . , cn`.

Das Manual der Bourne-Again-Shell (man bash) enthält im Abschnitt *pattern matching* noch eine genauere Spezifikation (aber es müssen nicht alle dort aufgeführten Sonderfälle implementiert werden).

Ferner bedeutet `-mtime`, dass diese Datei seit der angegebenen Zeit (angegebenen in Stunden (`h`), Tagen `d` oder Monaten `m`) nicht mehr verändert wurde.

Es ist möglich, mehrere `-p` Optionen anzugeben.

`-f` *force*: Direkt löschen.

`-i` *interactive*: Vor dem Löschen nachfragen (default).

Der Default hier ist der interaktive Modus, wenn die Standardeingabe ein Terminal ist, ansonsten der direkte Modus.

Das Programm sollte die Muster, die zu löschende Dateien kennzeichnet, auch aus der Datei `cleanup` im Heimatverzeichnis des Benutzers lesen; dort steht pro Zeile ein Muster nach obiger Spezifikation.

Hier ein Benutzungsbeispiel:

```
cleanup -p *~ -p .### -p tmp/* -mtime 1d -f /home/cx1
```

Löscht ausgehend vom Verzeichnis `/home/cx1` ohne Nachfrage (`-f`) alle Sicherungskopien des Emacs (Muster `*~`), und alle Dateien im Verzeichnis `tmp`, die seit einem Tag nicht mehr modifiziert wurden.

Achten Sie auf die Fehlerbehandlung — das Programm sollte nicht abstürzen, weil Verzeichnisse unlesbar sind, Dateien nicht gelöscht werden können, symbolische Verweise ins Leere zeigen etc; dieses würde beim vorgesehenen Test mit der Computerzeitschrift *x'y* einen schlechten Eindruck machen.

Die Markteinführung ist in vier Wochen vorgesehen; die Werbekampagne wird in zwei Wochen anlaufen. Viel Glück!

*Hinweise*: Viele der benötigten Funktionen finden Sie in den Haskell98-Standardbüchereien `Directory`, `System` und `Time`.

Das Heimatverzeichnis eines Benutzers steht in der Umgebungsvariable `HOME`.

Der Glasgow-Haskell-Compiler verfügt über eine *regular expression*-Bücherei (`Text.Regex`), allerdings haben diese regulären Ausdrücke eine andere Syntax gegenüber den Mustern oben; sie lassen sich allerdings in diese übersetzen.

Um zu prüfen, ob die Standardeingabe eine Terminal ist oder nicht, können Sie die Funktion `queryTerminal` aus dem Modul `Posix` verwenden.