

Übungsblatt #03

Abgabetermin: Donnerstag, 05.11.2009

Was will uns dieses Übungsblatt sagen?

- 1. Mit Hilfe von `grep` und regulären Ausdrücken lassen sich Textmuster in sehr großen Dateien leicht auffinden. Allerdings sucht das `grep`-Kommando nur.
- 2. Oftmals möchte man aber bestimmte Textmuster durch anderen Text suchen und ersetzen. Dazu dient dann der „Stream Line Editor“, kurz „`sed`“ genannt.
- 3. Der `sed` ist ein vollwertiger Editor, der jedoch nicht interaktiv (d.h. „mit Euch“ an der Tastatur) operiert, sondern als Unix-Filter implementiert ist.
- 4. Seine Stärke spielt der `sed` aus, wenn man ihn in einem Skript verwendet und auf Dateien ansetzt, die einen gleichen und bekannten Aufbau haben (hier im Beispiel: LaTeX-Dateien).
- 5. Dann kann in einer Schleife das `sed`-Kommando auf alle gewünschten Dateien angewendet werden.

Aufgabe 1: Ausländische Verkehrssünder

Auf unserer Homepage findet Ihr die Datei „`kennzeichen.txt`“. Sie enthält eine Million fiktive Einträge von Verkehrssündern: am Beginn der Zeile steht ein Kennzeichen, das entweder ein niederländisches, ein italienisches oder ein belgisches ist. Nach jedem Kennzeichen folgt ein Doppelpunkt und eine Zahl, die angibt, um wieviel km/h zu schnell das Fahrzeug mit dem entsprechenden Kennzeichen gefahren ist. Diese Datei soll für die folgenden drei Teilaufgaben verwendet werden.

a) Niederländische Verkehrssünder



Was bekommt man in Holland, wenn man dreimal durch die Führerscheinprüfung gefallen ist? Richtig: ein gelbes Nummernschild! (So, damit wäre dieser Kalauer auch mal wieder untergebracht...)

Aber bei holländischen Kennzeichen bleiben wir mal für diese Aufgabe:

Lasst Euch nun mit Hilfe des `sed` eine neue Datei namens

„`holland.txt`“ erstellen, in der nur die holländischen Verkehrssünder stehen. Allerdings soll diese neue Datei jeden Eintrag wie folgt führen:

„Das Fahrzeug mit dem niederländischen Kennzeichen XXX fuhr YYY km/h zu schnell.“

(Das heißt, die XXX bzw. YYY werden jeweils durch die echten Daten ersetzt.)

Ach ja: die holländischen Kennzeichen haben (hier vereinfacht) den folgenden Aufbau:

Es gibt drei Gruppen von jeweils zwei Buchstaben oder zwei Ziffern, und diese Gruppen sind von einem „-“ getrennt.

b) Italienische Verkehrssünder

Okay, das war zum Warmwerden. Nun machen wir ein ähnliches Spiel mit den italienischen Kennzeichen. Diese haben den folgenden Aufbau:

zwei Buchstaben, gefolgt von einem Whitespace, gefolgt von drei Ziffern, gefolgt von zwei Buchstaben (ohne Whitespace!)



Lasst Euch - nur mit Hilfe des `sed`! - den 10. bis 20. italienischen Verkehrssünder (d.h. nach Nennung in der Datei) ausgeben. (Eine Änderung des Ausgabebetextes wie bei a) ist nicht nötig.)

c) Belgische Verkehrssünder

Die dritte Kategorie von Kennzeichen, die in unserer Datei enthalten ist, entspricht den belgischen Kennzeichen. Bis zum 01.08.2008 hatten sie den Aufbau



drei Buchstaben, ein „-“, drei Ziffern,

seit dem 01.08.2008 ist dies umgekehrt:

drei Ziffern, ein „-“, drei Buchstaben.

Lasst Euch mit Hilfe des `sed` eine neue Datei namens „`belgien.txt`“ anlegen, in der nur alle belgischen Kennzeichen enthalten sind. Zusätzlich sollen alle neuen belgischen Kennzeichen dadurch hervorgehoben wurden, dass genau eine (die) Zeile vor ihnen und genau eine (die) Zeile nach ihnen leer sind.

Die Schönheitsfehler in unseren LaTeX-Dateien

1.) In LaTeX werden Zeilen, die mit einem Prozentzeichen beginnen, als Kommentarzeilen gewertet. Die in den Beispieldateien enthaltenen Kommentare benötigen wir nicht mehr. Lasst diese Zeilen also ersatzlos entfernen.

2.) Zwar kennt LaTeX auch deutsche Umlaute, und je nach Konfiguration kann man sie auch direkt eingeben. Aber in der Regel wird ein „ö“ als „o“, ein „ä“ als „a“, ein „ü“ als „u“ und ein „ß“ als „s“ eingegeben (analog für Großbuchstaben). Lasst diese Ersetzungen im gesamten Dokument vornehmen.

Aufgabe 2: LaTeX-Dateien

Bestimmt habt Ihr schon von dem Textsatz-System LaTeX gehört. Für wissenschaftliche Arbeiten eignet es sich hervorragend, so dass wir Euch nur empfehlen können, es im Hinblick auf Seminar- und Abschlussarbeiten zu erlernen. Aber keine Sorge: unser Übungsblatt schafft man auch ohne LaTeX-Kenntnisse. Aber wir benutzen für diese Aufgabe mal drei LaTeX-Vorlagen (mit typographischem Blindtext, also bitte keinen Sinn in den Texten suchen!), um eine weitere sinnvolle Anwendung des `sed` kennenzulernen.

Auf unserer Vorlesungshomepage gibt es für diese Aufgabe die drei Dateien „`latex[123].tex`“ (Da! Da war schon wieder ein regulärer Ausdruck!). Sie beinhalten drei vollständige LaTeX-Dokumente, die aber noch kleine Schönheitsfehler haben. Schreibt ein kleines Shell-Skript, das mit Hilfe des `sed` die drei Dateien so korrigiert und unter neuem Namen abspeichert, dass die im Kasten unten genannten Schönheitsfehler ausgebessert sind.

(Wenn Ihr alles richtig gemacht habt, dann sollte der Aufruf „`latex <dateiname.tex>`“ ohne Fehlermeldung durchlaufen.)

Tip: Die Option `-f` des `sed` kann Euch die Arbeit erleichtern!

3.) In dem Bereich, der zwischen „`\begin{document}`“ und „`\end{document}`“ steht, befindet sich der eigentliche Inhalt. (Der Rest wird als „Präambel“ bezeichnet.) Lasst nur in diesem Inhaltsbereich alle Datumsangaben fett drucken. Den Fettdruck erreicht man in LaTeX wie folgt: möchte man XXX fett schreiben, so notiert man `{\bf XXX}` (`bf` steht hier für „bold font“).

4. Es gibt Zeilen, in denen öffnet sich eine geschweifte Klammer, aber sie schließt sich nicht wieder. Korrigiert das, in dem Ihr in alle Zeilen, an denen eine geschweifte Klammer zwar auf- aber nicht wieder zugeht, als letztes Zeichen die schließende geschweifte Klammer notiert. Achtung: korrekte Zeilen dürfen nicht eine weitere Klammer erhalten!