Dr. Andreas Borchert
Christian Ehrhardt

Institut für Angewandte Informationsverarbeitung

30. Januar 2007
Blatt 12

Allgemeine Informatik III (WS 2006/2007)

Abgabetermin 06.02.2006

Archive (10 Punkte)

Ein *Archiv* oder eine *Archivdatei* ist eine Datei, in der mehrere Dateien verpackt sind. Entsprechend gibt es zwei wesentliche Operationen für Archive:

Erzeugen Dabei werden der Name der neu zu erzeugenden Archivdatei sowie die Dateien angegeben, die in das Archiv gepackt werden sollen. Das fertige Archiv enthält dann im Prinzip alle zu archivierenden Dateien hintereinander. Zum genauen Dateiformat siehe weiter unten.

Entpacken Hier wird nur ein bereits existierendes Archiv angegeben. Der Inhalt dieses Archivs wird dann entpackt, d.h. es werden im aktuellen Verzeichnis eine oder mehrere Dateien erzeugt, die den gleichen Namen und den gleichen Inhalt haben, wie die Dateien, die beim Erzeugen in das Archiv gepackt wurden.

Das Dateiformat

Für den Inhalt des Archivs schlage ich folgendes Dateiformat vor:

- Das Archiv besteht aus einer Folge von Dateibeschreibungen.
- Eine Dateibeschreibung beginnt mit zwei Integerzahlen *filesize* und *namesize* (jeweils 4 Byte). Der Wert *namesize*, gibt an, aus wievielen Zeichen der Dateiname der gerade beschriebenen Datei besteht, *filesize* gibt die Größe der Datei selbst an.
- Direkt darauf folgen genau *namesize* Zeichen, die den Dateinamen angeben.
- Anschließend folgt der eigentliche Inhalt der Datei, bestehend aus genau *filesize* Bytes.
- Damit ist die Dateibeschreibung beendet, es folgt entweder nichts mehr (Dateiende) oder eine weitere Dateibeschreibung.

Aufgabe

Schreiben Sie ein Programm, das Archive wie oben beschrieben erzeugen und entpacken kann.

Was zu beachten ist

- Es sollen ausschließlich reguläre Dateien in ein Archiv gepackt werden. Andere auf der Kommandozeile angegebene Dateien werden übersprungen.
- Sollte jemand versuchen, die Archivdatei, die im Moment erzeugt wird, sozusagen in sich selbst einzupacken, dann soll diese Datei übersprungen werden. Ob es sich um die selbe Datei handelt, sieht man an der Inode-Nummer.
- Grundsätzlich soll zur Sicherheit nur in Dateien geschrieben werden, die vorher noch nicht existiert haben.
- Systemaufrufe können grundsätzlich fehlschlagen, daher muß hier immer eine passende Fehlerbehandlung durchgeführt werden.

Hinweise

- Die Größe einer Datei wird am sinnvollsten *vor* dem Lesen des Inhalts ermittelt.
- Interessante Systemaufrufe sind `read`, `write`, `open` und `stat/lstat`.
- Die gepufferte Ein- und Ausgabe, wie sie im letzten Blatt verwendet wurde, eignet sich für die Lösung dieses Blattes **nicht**.
- Da des öfteren eine vorher bekannte Anzahl an Bytes aus einer Datei gelesen oder in eine Datei geschrieben werden muß, empfiehlt es sich, dafür passende Funktionen zu schreiben, die die Fehlerbehandlung gleich mit erledigen.
- Ebenfalls sinnvoll kann eine Funktion sein, die eine vorgegebene Anzahl an Bytes aus einer Datei liest und in eine andere schreibt.
- Archive, die wie oben beschrieben aufgebaut sind, können nicht ohne weiteres zwischen einer Sun und einem Linux-PC ausgetauscht werden!