

Software-Praktikum
„Datenkompression“
SS 05
Aufgabenblatt 2

Bearbeitung: bis 25.5.2005

In der Arbeit „A Locally Adaptive Data Compression Scheme“ von Bentley, Sleator, Tarjan und Wei ist ein Kompressionsschema beschrieben, bei dem ein Wörterbuch angelegt wird, dessen Einträge den einzelnen, im zu komprimierenden Text vorkommenden Wörtern entsprechen und wie eine Liste verwaltet werden. Hierbei wird mit einem leeren Wörterbuch gestartet und jedes Wort bei seinem ersten Auftreten in die neue Datei übertragen.

Wir wollen das Verfahren leicht modifizieren, um auch Nichttextdateien sinnvoll komprimieren zu können: Dazu zerlegen wir die zu komprimierende Datei nicht in Wörter, sondern in Zeichen der Größe von je einem Byte. Ferner gehen wir von einem Initialwörterbuch aus, in dem die 256 möglichen Byte in der kanonischen Anordnung stehen, d.h. an Stelle 1 steht das Zeichen mit der Dualdarstellung 00000000, an Stelle 2 das Zeichen mit der Dualdarstellung 00000001 usw. Hierdurch sind alle möglichen „Wörter“ bereits im Wörterbuch vorhanden, so dass das Übertragen neuer Wörter entfällt. Wie in o.g. Arbeit beschrieben, wandert das angefragte Zeichen von seiner aktuellen Position an den Anfang der Liste, die übersprungenen Einträge rücken dabei um je eine Position nach hinten. In die Kompressionsdatei wird dabei der bisherige Index des aktuellen Zeichens gemäß der in o.g. Arbeit beschriebenen Darstellung geschrieben.

Erweitern Sie Ihr Programm derart, dass bei Auswahl des Eintrags **Komprimieren** des Menüs **BSTW** ein Benutzerdialog erscheint, in dem die zu komprimierende Datei ausgewählt werden kann. Daraufhin ist diese Datei auf oben beschriebene Weise zu komprimieren, wobei sich der Name der neuen Datei aus dem ursprünglichen Namen ergibt, indem die Dateiendung „.BS“ angehängt wird. Nach Auswahl der zu komprimierenden Datei sind folgende Felder im Hauptfenster zu bestücken:

- Name der zu komprimierenden Datei
- Name der komprimierten Datei
- Aktion
- Größe der zu komprimierenden Datei

Während die Kompression läuft, soll ständig eine Fortschrittsanzeige aktualisiert werden, in der angegeben wird, wie viel Prozent der zu komprimierenden Datei bereits bearbeitet sind. Ist die Kompression beendet, so sind die Größe der komprimierten Datei und die hieraus resultierende Kompressionsrate auszugeben.

Bei Auswahl des Eintrags **Dekomprimieren** des Menüs **BSTW** soll ebenfalls ein Benutzerdialog erscheinen, in dem allerdings nur Dateien mit der Endung „.BS“ zur Auswahl angeboten werden. Analog zur Kompression sind nach Auswahl der komprimierten Datei folgende Felder im Hauptfenster zu bestücken:

- Name der ursprünglichen Datei
- Name der komprimierten Datei
- Aktion
- Größe der komprimierten Datei

Während die Dekompression läuft, soll ständig eine Fortschrittsanzeige aktualisiert werden, in der angegeben wird, wie viel Prozent der komprimierten Datei bereits bearbeitet sind. Ist die Dekompression beendet, so sind die Größe der wieder hergestellten Datei und die hieraus resultierende Kompressionsrate auszugeben.

Modellieren Sie in einem UML-Klassendiagramm Ihre bereits implementierte Menüsteuerung sowie die oben beschriebene **BSTW**-Kompression und -Dekompression.

Hinweis: Auf unserer Praktikumsseite gibt es eine **tar**-Datei mit Beispielinstanzen. Diese umfasst folgende Dateien:

- `A12.txt`
- `A16.txt`
- `3letters.txt`
- `blowup.txt`
- `alphabet.txt`
- `alphabet.txt.BS`
- `text1.txt.BS`
- `text2.txt.BS`

Die ersten vier Dateien zeigen, welche unterschiedlichen Kompressionsraten auftreten können. Mit den beiden Alphabetdateien (komprimiert und dekomprimiert) können Sie überprüfen, ob die von Ihnen implementierte Kompression den Vorgaben entspricht. Die Dekompression der beiden letzten Dateien liefert kurze Texte in deutscher Sprache.