

Blatt 5

(Abgabe am 05.06.2011)

Zum Blatt

Sie dürfen sich wieder wie auf Blatt 4 vier der Aufgaben aussuchen. Beachten Sie, dass Sie für dieses Blatt bis Sonntag um Mitternacht Zeit haben.

Aufgabe 1 (Entropie)

Zeigen Sie: Ist X eine echt unabhängige Zufallsquelle ohne Gedächtnis (die Wahrscheinlichkeit für ein Zeichen ist unabhängig von den vorangegangenen Zeichen), so gilt:

$$2 \cdot H(X) = H(X^2).$$

Dabei ist X^2 , die Zufallsquelle, welche entsteht, wenn man statt einem Zeichen immer zwei aufeinanderfolgende Zeichen betrachtet.

Aufgabe 2 (Verhältnis Signal zu Verzerrung)

Schreiben Sie Routinen, die aus zwei Folgen von double-Werten folgende Werte berechnen: Mittlerer quadratischer Fehler, Signal to Noise und Signal to Noise in Dezibel.

Aufgabe 3 (Huffman)

Analog zu Ihrer Shannon-Fano-Kodierung sollen Sie hier die Huffman Kodierung implementieren.

1. Warum ist Huffman Kodierung (nicht) zur online Komprimierung geeignet? Also z.B. zur Komprimierung von interaktiven Netzwerkverbindungen wie SSH.
2. Implementieren und testen Sie Kodierungs- und Dekodierungsfunktionen zur Huffman Kodierung. Als Ausgabe für den Kodierer dürfen Sie Strings von 0 und 1 statt Bitfolgen verwenden.
3. (*) Implementieren Sie Ihre Huffman Kodierung so, dass Sie Bitfolgen statt 0 und 1 Strings ausgibt.

Aufgabe 4 (BZIP2)

Betrachten Sie das Kompressionsverfahren bei BZIP2: Auf eine Eingabe wird zunächst eine Burrows-Wheeler-Transformation angewandt, danach die Move-to-Front-Transformation und anschliessend die Huffman Kodierung.

1. Warum ist BZIP2 (nicht) zur online Komprimierung geeignet? Also z.B. zur Komprimierung von interaktiven Netzwerkverbindungen wie SSH.
2. Implementieren Sie BZIP2: Basierend auf Ihren Lösungen der letzten Blätter, wenden Sie auf eine Eingabefolge nacheinander die BWT, die MtF und schliesslich die Huffman Kodierung an.
3. Implementieren Sie die Umkehrung von (ihrem) BZIP2.
4. Erweitern Sie Ihren Kodierer um Speicher- und Ladefunktionen. Verwenden Sie dazu wieder die Serialisierungsfunktionen Ihrer Sprache.
5. (*) Stellen Sie Ihr BZIP2 auf die Probe: Wenden Sie Ihr Programm in einer Schleife auf die Beispieldaten vom ersten Blatt an. Wenden Sie ebenfalls das „echte“ BZIP2-Programm auf die Daten an. Vergleichen Sie die Grössen der Eingabe und der Ausgaben gegeneinander. Diskutieren Sie das Ergebnis.

Hinweis: Sollten Sie Aufgabe 3 nicht bearbeitet haben, so nehmen Sie statt dem Huffman Kodierer aus dieser Aufgabe Ihren Shannon-Fano Kodierer.

Aufgabe 5 (BZIP2)

Betrachten Sie das Kompressionsverfahren bei BZIP2 wie bei Aufgabe 4.

Wir erweitern das Verfahren um einen Glättungsschritt zwischen der BWT und der MtF. Das heisst, nach der BWT wird der String durchlaufen und jedes einzeln vorkommende Zeichen wird durch seinen Vorgänger ersetzt. So wird aus „aabaaabbbaaccacaa“ die Zeichenfolge „aaaaaabbbaaccccaa“. Danach läuft der BZIP2 ab wie beschrieben. Durch diese Änderung wird ein höherer Kompressionsgrad bewirkt, allerdings wird beim Dekodieren nicht mehr exakt das Original errechnet sondern nur eine Annäherung.

1. Implementieren Sie diese Variante des BZIP2.
2. Errechnen Sie den Fehler nach Anwendung des modifizierten BZIP2 und seiner Umkehrung. Verwenden Sie die Beispieldaten von Blatt 1 und diskutieren Sie die Ergebnisse.
3. Vergleichen Sie die Kompressionsgrade des normalen BZIP2 und des modifizierten BZIP2. Diskutieren Sie die Ergebnisse.

Aufgabe 6 (Fehlerbetrachtung)

Betrachten Sie den modifizierten BZIP2 aus Aufgabe 5.

Angenommen es wird genau 1 Zeichen im Glättungsschritt verändert. Berechnen Sie, wie viele Zeichen sich nach der Dekodierung vom Original unterscheiden können.

Angenommen, Sie hätten den normalen BZIP2 ohne Glättung, und nach Anwendung des Huffmancodes würde ein Bit verändert. Wie viele Zeichen können sich nun im Original verändert haben?

Vergleichen und diskutieren Sie Ihre Ergebnisse.

Hinweis: Sie müssen hierfür Aufgabe 5 nicht gelöst haben, es ist eine rein theoretische Abgabe gefragt.