

**Dr. Jürg M. Stettbacher**

Margrit Rainer Strasse 12a  
CH-8050 Zürich

Telefon: +41 43 299 57 23  
Fax: +41 43 299 57 25  
E-Mail: [dsp@stettbacher.ch](mailto:dsp@stettbacher.ch)

# Huffman Code

## Praktikum

Version 2.00  
2013-06-12

**Zusammenfassung:** In diesem Praktikum geht es darum, ein Programm zu schreiben, das für gegebene Daten in einer Datei den Huffman Code entwickelt. Zu diesem Zweck muss zuerst die Auftretenswahrscheinlichkeit der einzelnen Symbole in den Daten ermittelt werden.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>2</b>
<b>2</b>	<b>Dateien</b>	<b>2</b>
<b>3</b>	<b>Aufgabe</b>	<b>3</b>
<b>4</b>	<b>Zusatzaufgabe</b>	<b>3</b>

## 1 Einleitung

Der Huffman Encoder ist ein typischer Quellenencoder. Er wandelt Codeworte gleicher Länge, zum Beispiel ASCII-Zeichen, in Codeworte unterschiedlicher Länge um. Dabei erhalten häufig auftretende Zeichen kurze Codeworte und seltene Zeichen lange Codeworte.

In dieser Übung haben wir verschiedene Datenquellen in der Form von Dateien zur Verfügung. Es handelt sich um ASCII-Dateien (\*.txt) und jedes ASCII-Zeichen daraus stellt ein Symbol dar. Gesucht ist der Huffman Code für diese Daten.

## 2 Dateien

Die folgenden Dateien stehen Ihnen für das Bearbeiten des Praktikums zur Verfügung:

- *huffman\_template.php* (Hauptaufgabe)
- *data\_1.txt* bis *data\_6.txt* und *deutsch.txt* (Testdaten)

## 3 Aufgabe

Schreiben Sie ein PHP-Skript *huffman.php*, das eine bestimmte Datei öffnen und zeichenweise lesen kann. Der Name der ASCII-Datei soll auf der Kommandozeile übergeben werden, so dass der Aufruf folgendermassen aussieht.

```
> php huffman.php data.txt
```

Sie erhalten eine Vorlage *huffman\_template.php*, in welcher einige Funktionsblöcke bereits vorhanden sind. Sie müssen nur an den bezeichneten Stellen ergänzen. Für die Ermittlung des Huffman Codes gehen Sie so vor:

- Erstellen Sie ein Histogramm für alle Symbole.
- Ordnen Sie das Histogramm nach Häufigkeiten, resp. nach der Wahrscheinlichkeit.
- Kombinieren Sie wie im Huffman-Baum die Elemente des Histogramms mit der kleinsten Wahrscheinlichkeit und bauen Sie während dem Kombinieren den Code für die betroffenen Symbole auf. Überlegen Sie sich dabei, ob Sie den Code von links nach rechts oder von rechts nach links aufbauen müssen. Beachten Sie auch, dass in der resultierenden Kombination alle bisher zusammenkombinierten Symbole bekannt sein müssen. Bilden Sie also einen String aus diesen Symbolen.
- Wiederholen Sie den Schritt der Kombination bis das Histogramm nur noch aus einem Element besteht.
- Das Skript soll für jedes Symbol seine Auftretenswahrscheinlichkeit, sowie das ermittelte binäre Codewort anzeigen.

Diese erste Aufgabe ist zu erfüllen, und wenn nötig im Selbststudium fertig zu stellen.

## 4 Zusatzaufgabe

Dies ist eine Bonusaufgabe. Da wir bereits einen Huffman Code für eine bestimmte Datendatei entwickelt haben, sollte es möglich sein, die Daten zu codieren, abzuspeichern und anschliessend wieder zu decodieren. Die decodierte Datei muss dann natürlich mit der ursprünglichen überein stimmen.

Schreiben Sie zwei weitere Skripte, nämlich *huffman\_encode.php* und *huffman\_decode.php*, welche dies tun. Beachten Sie beim Decoder, dass in der codierten Datei Codewort an Codewort stösst, und dass auf Grund der unterschiedlichen Längen das Ende des nächsten Codeworts nicht von Anfang an klar ist. Wurde der Huffman Code aber korrekt entwickelt, so ist er präfixfrei. Daher lassen sich Codeworte zweifelsfrei identifizieren.