

Softwareentwicklung 3
Medizin- und Bioinformatik

Aufwand in Stunden: 26h

Übung 08

Sabine Klammer
Gruppe 1

Beispiel 1: Codierung und Komprimierung nach Huffman

Lösungsidee

Bei dieser Aufgabenstellung geht es darum ein Komprimierungsverfahren nachzubauen bzw. zu simulieren. Hierbei werden Zeichen in Nullen und Einsen „umgewandelt“, somit können ganze Texte auf sehr geringe Datenmengen komprimiert werden. Um dies möglich zu machen, werden Files Zeichen für Zeichen verarbeitet und in eine Tabelle gespeichert, welche die Häufigkeit der Zeichen mitzählt und somit als Erstes eine aufsteigend sortierte Häufigkeitstabelle erstellt. Wurde dies erledigt, geht es zum nächsten Schritt – das Erstellen des Huffmanbaumes. Dieser wird in weiterer Folge benötigt, um die Codes für die einzelnen Zeichen berechnen zu können.

Um den Huffmantree erstellen zu können werden die zwei Zeichen mit der geringsten Häufigkeit entnommen und daraus ein Baum erstellt. Die beiden Zeichen bilden dabei das linke bzw. rechte Kind und die addierten Häufigkeiten der beiden bilden die Wurzel. Dieser neu erstellte Baum wird an der richtigen Position wieder in die sortierte Tabelle eingefügt. Dieser Schritt wiederholt sich so lange bis nur mehr ein einziger Baum in der Tabelle enthalten ist.

Ist nur mehr ein Baum übrig, kann man in den Codierungsschritt übergehen. Hierbei wird bei der Wurzel begonnen und immer dem linken Kind eine Null und dem rechten Kind eine Eins zugeordnet. Will man nun die Codes für die einzelnen Zeichen haben, durchsucht man den Baum bis man das Element gefunden hat und speichert jede Null und jede Eins mit die man auf dem korrekten Weg von der Wurzel bis zu dem gesuchten Element passiert. Daraus ergibt sich eine neue Tabelle mit den jeweiligen Zeichen und den dazugehörigen Codes.

Nach diesem Schritt, kann der eigentlich eingelesene Text wieder aufgenommen werden und jedes Zeichen gegen den entsprechenden Code ersetzt werden, wodurch sich die Komprimierung ergibt.

Das Ergebnis wird auf der Konsole ausgegeben, ebenfalls eine kurze Statistik welche die Komprimierungsrate angibt.

Testfälle

Testfall	Eingabe	Erwartete Ausgabe	Tatsächliche Ausgabe
Ausgabe Häufigkeitstabelle	AACCCNDDDDA ACCCCRRAACC CCNNNADDDDD CCCAAACCNAA AAACDDDDRR AAAANCCRRRA AAAARRRCCCA ARRAAAAAAC C CRRRRNCCC	Korrekte Ausgabe	<pre>{Code: N, Frequency: 8} {Code: D, Frequency: 12} {Code: R, Frequency: 15} {Code: C, Frequency: 32} {Code: A, Frequency: 33}</pre>
Ausgabe Huffmantree	AACCCNDDDDA ACCCCRRAACC CCNNNADDDDD CCCAAACCNAA AAACDDDDRR AAAANCCRRRA AAAARRRCCCA ARRAAAAAAC C CRRRRNCCC	Korrekte Ausgabe	<pre>{HuffmanListNode} --{HuffmanTreeNode Character:A, Value: 33} --{HuffmanTreeNode Character:-, Value: 65} \--{HuffmanTreeNode Character:C, Value: 32} *--{HuffmanTreeNode Character:-, Value: 100} --{HuffmanTreeNode Character:D, Value: 12} --{HuffmanTreeNode Character:-, Value: 20} \--{HuffmanTreeNode Character:N, Value: 8} \--{HuffmanTreeNode Character:-, Value: 35} \--{HuffmanTreeNode Character:R, Value: 15}</pre>
Ausgabe codierter Buchstaben	AACCCNDDDDA ACCCCRRAACC CCNNNADDDDD CCCAAACCNAA AAACDDDDRR	Korrekte Ausgabe	<pre>{CodingTableEntry Char: A, Code: 11} {CodingTableEntry Char: C, Code: 10} {CodingTableEntry Char: D, Code: 011} {CodingTableEntry Char: N, Code: 010} {CodingTableEntry Char: R, Code: 00}</pre>

	AAAANCCRRRA AAAARRRCCCCA ARRAAAAAAAC C CRRRRNCCC		
Leeres File	Leeres File	Korrekte Ausgabe	<pre> ##### Test for file EmptyFile.txt##### File-Content: Coded-Content: Encrypted-Contente: ##### STATISTICS ##### Old Length: 0, New Length: 0 Old Size: 0(bits) , New Size: 0(bits) </pre>
Beispiel aus Aufgabenstellung	AACCCNDDDDA ACCCCRRRAACC CCNNNADDDDD CCCAAACCNAA AAACDDDDRR AAAANCCRRRA AAAARRRCCCCA ARRAAAAAAAC C CRRRRNCCC	Korrekte Ausgabe	<pre> ##### Test for file Example.txt##### File-Content: AACCCNDDDAACCCRRRAACCCNNNADDDDDCCCAAACNAAAAACDDDDRRAAAAANCCRR AAAAARRR CCCCAARRAAAAAACCCRRRRNCCC {Code: N, Frequency: 8} {Code: D, Frequency: 12} {Code: R, Frequency: 15} {Code: C, Frequency: 32} {Code: A, Frequency: 33} {HuffmanListNode} /--{HuffmanTreeNode Character:A, Value: 33} /--{HuffmanTreeNode Character:-, Value: 65} \--{HuffmanTreeNode Character:C, Value: 32} *--{HuffmanTreeNode Character:-, Value: 100} /--{HuffmanTreeNode Character:D, Value: 12} /--{HuffmanTreeNode Character:-, Value: 20} \--{HuffmanTreeNode Character:N, Value: 8} \--{HuffmanTreeNode Character:-, Value: 35} \--{HuffmanTreeNode Character:R, Value: 15} {CodingTableEntry Char: A, Code: 11} {CodingTableEntry Char: C, Code: 10} {CodingTableEntry Char: D, Code: 011} {CodingTableEntry Char: N, Code: 010} {CodingTableEntry Char: R, Code: 00} Coded-Content: 1111101010010010011011111110101010000011111010100100100101101101101101 1101010111111010010111111111101001101101100001111111010101000001111111 10000001010101111000011111111111110100000000010101010 Encrypted-Contente: AACCCNDDDAACCCRRRAACCCNNNADDDDDCCCAAACNAAAAACDDDDRRAAAAANCCRRAAAAARRRCCCCA RRAAAAAACCCRRRRNCCC ##### STATISTICS ##### Old Length: 101, New Length: 220 Old Size: 808(bits) , New Size: 220(bits) Compression-Rate: 73% ##### </pre>
Statistiken ausgeben	AACCCNDDDDA ACCCCRRRAACC CCNNNADDDDD CCCAAACCNAA AAACDDDDRR AAAANCCRRRA AAAARRRCCCCA ARRAAAAAAAC C CRRRRNCCC	Korrekte Ausgabe	<pre> ##### STATISTICS ##### Old Length: 101, New Length: 220 Old Size: 808(bits) , New Size: 220(bits) Compression-Rate: 73% ##### </pre>

Kurzer Text	Ich bin ein kurzer Text fuer Testzwecke.	Korrekte Ausgabe	<pre>##### Test for file shortText.txt##### File-Content: Ich bin ein kurzer Text fuer Testzwecke. {Code: ., Frequency: 1} {Code: w, Frequency: 1} {Code: s, Frequency: 1} {Code: r, Frequency: 1} {Code: f, Frequency: 1} {Code: x, Frequency: 1} {Code: b, Frequency: 1} {Code: h, Frequency: 1} {Code: I, Frequency: 1} {Code: k, Frequency: 2} {Code: c, Frequency: 2} {Code: z, Frequency: 2} {Code: t, Frequency: 2} {Code: T, Frequency: 2} {Code: u, Frequency: 2} {Code: n, Frequency: 2} {Code: i, Frequency: 2} {Code: , Frequency: 6} {Code: e, Frequency: 7} {HuffmanListNode} /--{HuffmanTreeNode Character:e, Value: 7} /--{HuffmanTreeNode Character:-, Value: 13} \--{HuffmanTreeNode Character: , Value: 6} /--{HuffmanTreeNode Character:-, Value: 22} /--{HuffmanTreeNode Character:h, Value: 1} /--{HuffmanTreeNode Character:-, Value: 2} \--{HuffmanTreeNode Character:b, Value: 1} /--{HuffmanTreeNode Character:-, Value: 3} \--{HuffmanTreeNode Character:I, Value: 1} /--{HuffmanTreeNode Character:-, Value: 5} \--{HuffmanTreeNode Character:i, Value: 2} \--{HuffmanTreeNode Character:-, Value: 9} /--{HuffmanTreeNode Character:r, Value: 1} /--{HuffmanTreeNode Character:-, Value: 2} \--{HuffmanTreeNode Character:s, Value: 1} \--{HuffmanTreeNode Character:-, Value: 4} /--{HuffmanTreeNode Character:x, Value: 1} \--{HuffmanTreeNode Character:-, Value: 2} \--{HuffmanTreeNode Character:f, Value: 1} *--{HuffmanTreeNode Character:-, Value: 38} /--{HuffmanTreeNode Character:T, Value: 2} /--{HuffmanTreeNode Character:-, Value: 4} \--{HuffmanTreeNode Character:t, Value: 2} /--{HuffmanTreeNode Character:-, Value: 8} /--{HuffmanTreeNode Character:n, Value: 2} \--{HuffmanTreeNode Character:-, Value: 4} \--{HuffmanTreeNode Character:u, Value: 2} \--{HuffmanTreeNode Character:-, Value: 16} /--{HuffmanTreeNode Character:k, Value: 2} /--{HuffmanTreeNode Character:-, Value: 4} /--{HuffmanTreeNode Character:w, Value: 1} \--{HuffmanTreeNode Character:-, Value: 2} \--{HuffmanTreeNode Character:., Value: 1} \--{HuffmanTreeNode Character:-, Value: 8} /--{HuffmanTreeNode Character:z, Value: 2} \--{HuffmanTreeNode Character:-, Value: 4} \--{HuffmanTreeNode Character:c, Value: 2}</pre>
-------------	--	------------------	--

5

bin ein kurzer
Text fuer
Testzwecke.

{ HuffmanListNode

--{HuffmanTreeNode Character:, Value: 98}
--{HuffmanTreeNode Character:-, Value: 181}
 /--{HuffmanTreeNode Character:h, Value: 14}
 /--{HuffmanTreeNode Character:-, Value: 28}
 \--{HuffmanTreeNode Character:b, Value: 14}
 /--{HuffmanTreeNode Character:-, Value: 53}
 /--{HuffmanTreeNode Character:., Value: 14}
 \--{HuffmanTreeNode Character:-, Value: 25}
 \--{HuffmanTreeNode Character:k, Value: 11}
 \--{HuffmanTreeNode Character:-, Value: 83}
 /--{HuffmanTreeNode Character:r, Value: 9}
 /--{HuffmanTreeNode Character:-, Value: 16}
 \--{HuffmanTreeNode Character:c, Value: 7}
 \--{HuffmanTreeNode Character:-, Value: 30}
 \--{HuffmanTreeNode Character:I, Value: 14}
 /--{HuffmanTreeNode Character:-, Value: 293}
 /--{HuffmanTreeNode Character:u, Value: 28}
 /--{HuffmanTreeNode Character:-, Value: 56}
 \--{HuffmanTreeNode Character:T, Value: 28}
 \--{HuffmanTreeNode Character:-, Value: 112}
 /--{HuffmanTreeNode Character:i, Value: 28}
 \--{HuffmanTreeNode Character:-, Value: 56}
 \--{HuffmanTreeNode Character:n, Value: 28}
*-{HuffmanTreeNode Character:-, Value: 503}
 /--{HuffmanTreeNode Character:s, Value: 14}
 /--{HuffmanTreeNode Character:-, Value: 28}
 \--{HuffmanTreeNode Character:w, Value: 14}
 /--{HuffmanTreeNode Character:-, Value: 56}
 /--{HuffmanTreeNode Character:x, Value: 14}
 \--{HuffmanTreeNode Character:-, Value: 28}
 \--{HuffmanTreeNode Character:f, Value: 14}
 /--{HuffmanTreeNode Character:-, Value: 112}
 /--{HuffmanTreeNode Character:t, Value: 28}
 \--{HuffmanTreeNode Character:-, Value: 56}
 \--{HuffmanTreeNode Character:z, Value: 28}
 \--{HuffmanTreeNode Character:-, Value: 210}
 \--{HuffmanTreeNode Character:e, Value: 98}

(CodingTableEntry Char: , Code: 111)
(CodingTableEntry Char: h, Code: 110111)
(CodingTableEntry Char: b, Code: 110110)
(CodingTableEntry Char: ., Code: 110101)
(CodingTableEntry Char: k, Code: 110100)
(CodingTableEntry Char: r, Code: 110011)
(CodingTableEntry Char: c, Code: 110010)
(CodingTableEntry Char: I, Code: 11000)
(CodingTableEntry Char: u, Code: 1011)
(CodingTableEntry Char: T, Code: 1010)
(CodingTableEntry Char: i, Code: 1001)
(CodingTableEntry Char: n, Code: 1000)
(CodingTableEntry Char: s, Code: 01111)
(CodingTableEntry Char: w, Code: 01110)
(CodingTableEntry Char: x, Code: 01101)
(CodingTableEntry Char: f, Code: 01100)
(CodingTableEntry Char: t, Code: 0101)
(CodingTableEntry Char: z, Code: 0100)
(CodingTableEntry Char: e, Code: 00)

Coded-Content:

1100011001011011111110101000110010011000111101001011100110100001100111110100001101011101100101100110011111101000011110
101010001110001100101101000011010111110001100110111111010100011001000110001111101001111010000110011111010000110101
01111011001011001100111110100001110101010001110001100101101000011010111110001100110111111010101100011101001
01111001101000011001111101000011010101110110010110011111010000111010101000111000110010110100001101011111000110011011111
110110100110000110010010001111010010111001101000011001111101000011010101110100011001100111101000011101010000111000110010
11010000110101111000110010110111111011010011000111001000111110100101110011010000110011111010011101010111010010110011001
111110100001110101010001110001100101101000011010111100011001011011111010100110001110010011000011101001110100111001101000011001
001100011110100101110110100001100101110100001100101111010010110011111010000111010101000111000110010110100001101011110
001100101101111101010001100011001001111010011100110100001100111110100001101010111010001110100001110101110100001110101
01000111000110010101000011010111110001100101101111101010100011001000111010010111001101000011001111101000011010101
110110010100110011111010000111010101000111000110010101000011010111110001100101011111010100111001100011101010011
110011010000110011110100001101010111010011001111101000011101010100011100110011010101000011010111101001110101111110
11010011000111001001100011111010010111001101000011001111010000110101011101001011001111010000111010100111010000110011010
1000011010111100011001101011111101010011000111001001111010010111001101000011001111101000011010101110100111001100111
11010000111101010100011100011001011010000110101111000110010101111110101001100011101000111001101000011001111
110100001101010111010011001100111101000011101010001110001100101101000110101111000110010101111101010011000111001001
100011110100101110011010000110011111010000110101011101001100111110100001110101000011100011001011000011010111

Encrypted-Content:
Ich bin ein kurzer Text fuer Testzwecke. Ich bin ein kurzer Text fuer Testzwecke. Ich bin ein kurzer Text fuer Testzwecke. Ich bin ein kurz
er Text fuer Testzwecke. Ich bin ein kurzer Text fuer Testzwecke. Ich bin ein kurzer Text fuer Testzwecke. Ich bin ein kurzer Text fue
r Testzwecke. Ich bin ein kurzer Text fuer Testzwecke. Ich bin ein kurzer Text fuer Testzwecke. Ich bin ein kurzer Text fuer Tes
tzwecke. Ich bin ein kurzer Text fuer Testzwecke.
STATISTICS #####
Old Length: 581, New Length: 2352
Old Size: 4648(bits) , New Size: 2352(bits)
Compression-Rate: 40%
#####