Prefix codes, dynamic programming 2, matroids

1. Let $a = 0$, $b = 101$, $c = 100$, $d = 111$, $e = 1101$ and $f = 1100$ denote an encoding scheme for the characters $a, b, c, d, e, f$. What is the character sequence encoded by the following string?

   0100100100101111101110001101

You are asked to assign the above code-words to the letters $a, b, c, d, e, f$ in any order you want, to encode a very long sequence of these characters. Which letter should you assign the code-word “0” to? And “1101”?

2. Represent the encoding scheme in the previous exercise with help of a binary tree. What is the depth of this tree? What is the correlation between each code-word and the depth of the corresponding leaf? Compute from this tree what the encoding cost of the sequence $acaccbdbeae$ is. Check your answer by comparing to the above 0-1 sequence.

3. Find the optimal Huffman code of a sequence of characters, where the character frequencies are $f : 5, e : 9, c : 12, b : 13, d : 16, a : 45$. (Hint: create the binary tree representing the code on the way.) How many merges did you execute during the algorithm? How many merges do you need if there are $n$ different characters to encode?

4. Find the longest common subsequence of sequences $X = \langle A, A, B, A, C, D \rangle$ and $Y = \langle C, A, B, C, D, A \rangle$ with help of dynamic programming. Create a table $C$ for this. What does cell $C[i, j]$ in your table stand for? What is the recursive formula you use to compute $C[i, j]$?

5. Write the sets of the graphical matroid of the graph below.

6. Show that all maximal independent sets in a matroid have the same size.

7. (CLRS 16.4-1) Show that $(S, I_k)$ is a matroid, where $S$ is a finite set and $I_k$ is the set of all subsets of $S$ of size at most $k$. 