

## KOLOKVIJUM

29 April 2017

## NAPOMENE

Postavka zadatka se nalazi na intranet portalu <https://moodle-int.rt-rk.com/>. Portalu se pristupa pomoću korisničkog imena **RAxx-yyyy** gde **xx** predstavlja broj indeksa a **yyyy** godinu upisa. Šifra je **TMDlab7**.

Za potrebe kolokvijuma napraviti direktorijum **C:\tmp\student\X\**, gde je X broj indeksa. Rešenje zadatka treba da se nalazi u tom direktorijumu.

Direktorijum **student\X\** zajedno sa rešenjem treba arhivirati i okačiti na lični intranet portal. Direktorijum **C:\tmp\student\X\** ne treba brisati.

## UVOD

Hafmanovo kodovanje se primenjuje kao tehnika za kompresiju podataka i veoma je efikasno, tipične uštede su od 20% do 80% zavisno od karakteristike podataka koji se kompresuju. U našem primeru podaci su predstavljeni kao ulazni niz karaktera. Hafmanovo kodovanje umesto kodnih reči fiksne veličine koristi kodne reči promenljive veličine koje se nazivaju Hafmanovi kodovi. Hafmanovo kodovanje koristi informaciju o tome koliko često se svaki podatak na ulazu pojavljuje (frekvencija pojavljivanja) i na osnovu nje za svaki ulazni podatak bira optimalnu vrednost binarnog stringa. Odgovarajući kodovani izlaz se dobija prostim spajanjem binarnih vrednosti ulaznih podataka. Primer je prikazan na slici 1. Više na [https://en.wikipedia.org/wiki/Huffman\\_coding](https://en.wikipedia.org/wiki/Huffman_coding).

Ulazni niz podataka čine karakteri {a, b, c, d, e, f}						
Podaci	a	b	c	d	e	f
Frekvencija pojavljivanja	45k	13k	12k	16k	9k	5k
Fiksno kodovanje	000	001	010	011	100	101
Hafmanovo kodovanje	0	101	100	111	1101	1100
Ukupno bita potrebno za kodovanja Fiksnim kodovanjem: 300k bita						
Ukupno bita potrebno za kodovanja Hafmanovim kodovanjem: 224k bita						
<u>Ulaz:</u> a a b d d						
<u>Izlaz:</u> 0 + 0 + 101 + 111 + 111 = 00101111111						

Slika 1 – Analiza Fiksnog (binarnog) i Hafmanovog kodovanja podataka.

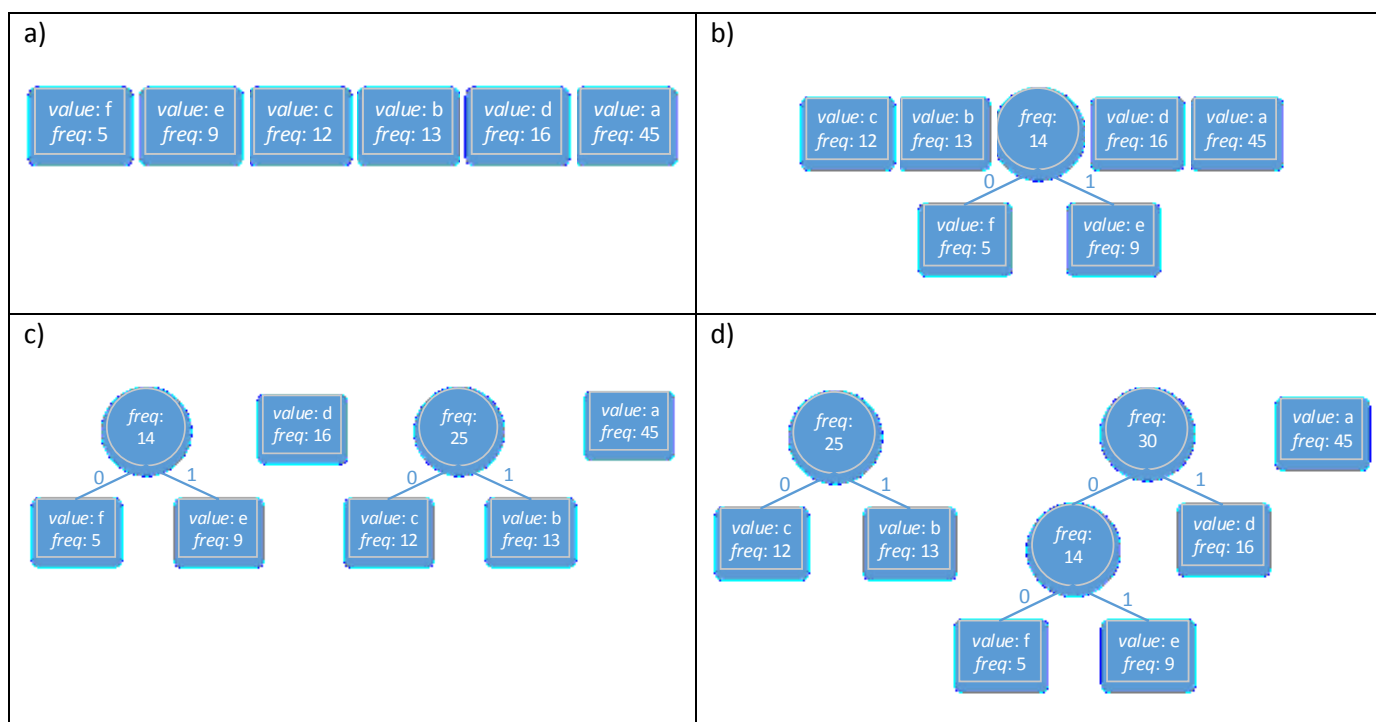
## ZADATAK

1. Napisati funkciju *GetHistogram* koja na osnovu niza ulaznih podataka (karakter) pravi histogram, tj. računa frekvenciju pojavljivanja određenog karaktera. Podatak modelovati kao objekat koji ima najmanje dva atributa, *value* i *freq* koji predstavljaju sam karakter i frekvenciju pojavljivanja konkretnog karaktera, respektivno. Primeri ulaznih podataka dostupan je u *snippets.txt*.
2. Na osnovu histograma formirati Hafmanovo stablo prema sledećoj proceduri:
  - Svi elementi (karakter) su listovi stabla.
  - Dva elementa sa najmanjom verovatnoćom pojavljivanja formiraju novi čvor čiji su potomci ta konkretna dva elementa. Novi čvor ulazi u histogram umesto ta 2 elementa, sa verovatnoćom jednakom zbiru njihovih verovatnoća.
  - Postupak se ponavlja dok se ne isprazni histogram, odnosno dok se ne formira kompletno stablo sa samo jednim elementom koji predstavlja koren.

*Napomena:* Obezbediti funkcije *GetMinFreqElem*, *MakeNewElem*, *PutElem* i *RemoveElem*.

3. Koristeći formirano Hafman stablo kodovati primere dostupne u *snippets.txt*. Hafmanov kod određenog karaktera formira se prostim spajanjem binarnih vrednosti (0 i 1) koje se nalaze na putanji do njegove pozicije u Hafman stablu. Za svaki primer napraviti posebnu funkciju koja ispisuje ulazni niz podataka i dobijeni kodovani izlaz (ceo ulazni niz podataka i ceo kodovani izlaz).

*Napomena:* Obezbediti funkciju *GetEncVal* koja i ispisuje traženi karaktera i njegovu kodovanu vrednost.



Slika 2 – Ilustracija formiranja Hafman stabla.