

TEORIJSKE OSNOVE

Hafmanovo kodovanje se primenjuje kao tehnika za kompresiju podataka. 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 na ulazu pojavljuje svaki podatak (frekvencija pojavljivanja) i na osnovu nje za svaki ulazni podatak bira optimalnu vrednost binarnog stringa. Karakteri koji se češće pojavljuju na ulazu su kodovani sa manje bita, dok su karakteri koji se ređe pojavljuju kodovani sa više bita. Odgovarajući kodovani izlaz se dobija prostim spajanjem binarnih vrednosti ulaznih podataka. Primer je prikazan na slici 1.

Više informacija možete pronaći na https://en.wikipedia.org/wiki/Huffman_coding.

Ulazni niz podataka čine karakteri {a, b, c, d, e, f}. Pretpostavimo da na ulazu imamo sekvencu koja daje sledeće frekvencije pojavljivanja:

Podaci	a	b	c	d	e	f
Frekvencija pojavljivanja	45k	13k	12k	16k	9k	5k

Generisanje kodova fiksnim i Hafmanovim kodovanjem:

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

Primer:

Ulaz: a a b d d

Izlaz: 0 + 0 + 101 + 111 + 111 = 0010111111

Slika 1 - Analiza Fiksnog (binarnog) i Hafmanovog kodovanja podataka

Generisanje Hafmanovog koda

Proces generisanja Hafmanovog koda prikazan je na slici 2. Na osnovu frekvencije pojavljivanja svakog od karaktera u ulaznom nizu generiše se Hafmanovo stablo na sledeći način:

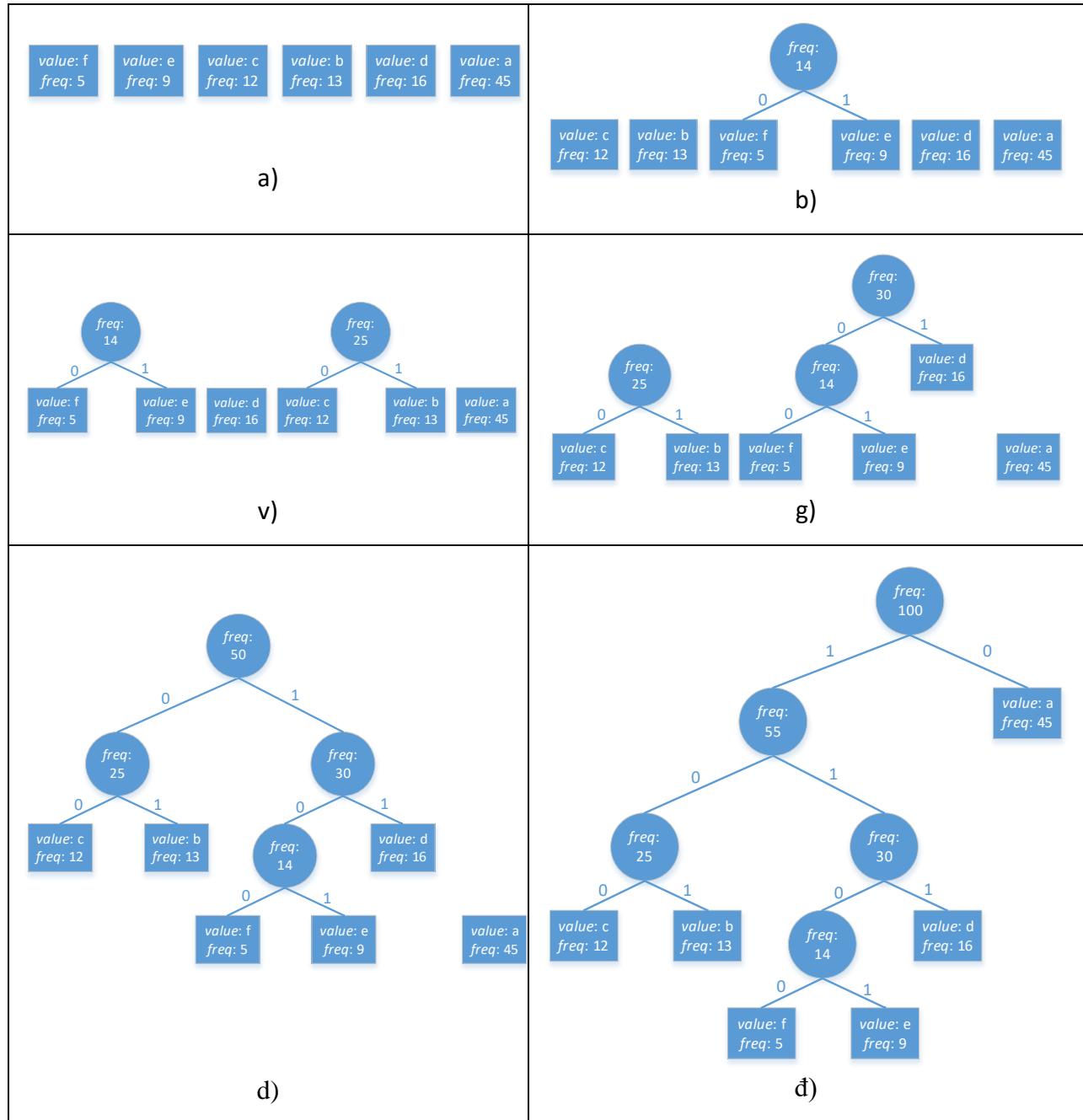
1. svaki od karaktera sa svojim frekvencijama predstavlja početne listove stabla (slika 2, a)
2. dva karaktera sa najmanjim frekvencijama se spajaju i generišu novi čvor čija vrednost je zbir frekvencija čvorova potomaka (na slici 2, b prikazano je spajanje čvorova sa frekvencijama 5 i 9)
3. korak 2 se ponavlja dok nije kreirano celo stablo (odnosno koreni čvor koji ima frekvenciju 100).

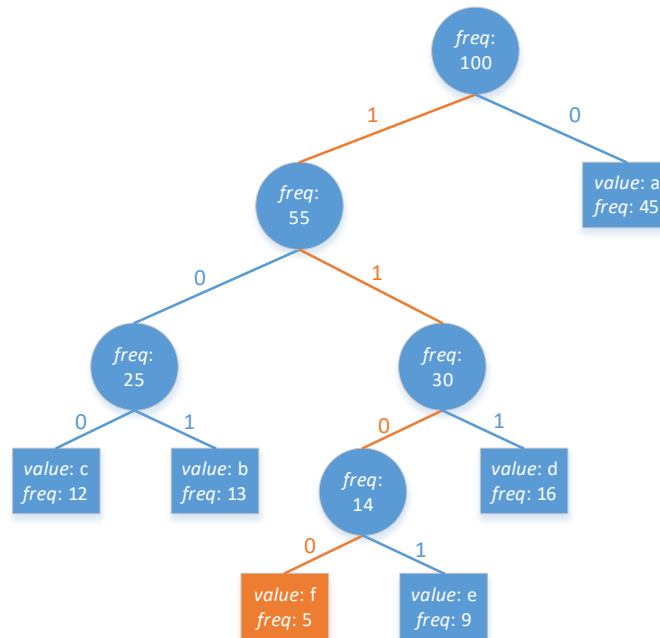
Nakon popunjavanja Hafmanovog stabla, moguće je izračunati kodovane vrednosti za svaki od karaktera na sledeći način:

1. počevši od korena stabla generišemo kod na način da čvoru potomku koji ima veću frekvenciju dodeljujemo kod **1**, a čvoru potomku koji ima manju frekvenciju dodeljujemo kod **0**
2. korak 1 se ponavlja sve dok se ne dođe do lista stabla, odnosno karaktera za koji računamo kod
3. kod karaktera čitamo tako što uzimamo kodovane vrednosti od korena ka listu.

U primeru sa slike 2, primetimo da je kodovana vrednost za karakter f, koji ima frekvenciju pojavljivanja 5 - **1100**. (Ostale karaktere kodujemo vrednostima: a-0, b-101, c-100, d-111, e-1101)

Projektovanje algoritama
- Vežba 8 -





e)

Slika 2 - Ilustracija formiranja Hafmanovog stabla

ZADATAK

1. Napisati funkciju *GetHistogram* koja na osnovu niza ulaznih podataka (karakteri) 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 dostupni su u *snippets.txt*.
2. Na osnovu histograma formirati Hafmanovo **stablo** prema sledećoj proceduri:
 - Svi elementi (karakteri) 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.

Na slici 2 je dat primer kreiranja Hafmanovog stabla.

Napomena: Obezbediti funkcije za:

- za dobavljanje elementa sa minimalnom frekvencijom

- generisanje novog elementa od dva postojeća
 - dodavanje elementa u Hafmanovo stablo
 - brisanje elementa iz stabla.
3. Koristeći formirano Hafmanovo 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 Hafmanovom 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 koja vraća kodovanu vrednost traženog karaktera.