

موضوع: الگوریتم‌های حریصانه - کددهی هافمن

مساله: الفبای C شامل n حرف داده شده. به ازای هر $c \in C$

$f_r(c)$ میزان تکرار حرف c را در متن‌های مربوط به الفبای C نشان می‌دهد

هدف ارائه یک روش کد کردن حروف C است که در آن

۱- هر حرف باید محاط با بیتی جایگزین می‌شود \leftarrow طول کد بیتی حرف C

۲- هر متن کد شده به طور یکتا قابل بازبینی باشد

* ۳- متوسط طول متن‌ها تا حد امکان کم شود $\leftarrow \sum_{c \in C} f_r(c) \times d(c)$

مثال ۱) $C = \{s, a, l, m\}$

۵	{	s : ۰۱	۱۰۰	روشن کرده‌ی :
۳		a : ۰۰۱	۰	
۶		l : ۱۰۰	۱۱۱	
۱۰		m : ۱۱	۱۰۱	

Salam o ۰۱۰۰ / ۱۰۰ ۰۰۱۱۱

سوال: آیا هر کدی به طور یکتا قابل بازبینی است؟ بله - چون هیچ کدی پیشوند کدی دیگر نیست

s ← ۰۱
a ← ۰۰۱

سوال: اگر فکانش a صلی زیاد باشد و فکانش بقیه حروف کم باشد، آیا این روش کددهی معتدل است؟ نه

مثال ۲: $C = \{a, b, c, d, e\}$

کددهی: \leftarrow { a = ۱۰, b = ۱۱, c = ۱۱۱, d = ۱۱۱۱, e = ۱۱۱۱۱ }

$$\left\{ \begin{array}{l} a = 11 \\ b = 1 \\ d = 01 \end{array} \right. \leftarrow \left\{ \begin{array}{l} CC \leftarrow 11 \\ b \leftarrow 11 \end{array} \right.$$

روش ۱: کدهی با طول ثابت. هر کاراکتر، یک کد با طول ثابت داشته باشد.

مثال: $C = \{a, b, c, d, e\}$

$$\left\{ \begin{array}{l} a = 000 \\ b = 001 \\ c = 010 \\ d = 011 \\ e = 100 \end{array} \right.$$

سوال: آیا این روش کدهی بیادید می شود: بله

مسئله: زمانی که فواصل حاصل متفاوت باشد: $2 = \text{فر}(\alpha)$
 رقیه: $1 = \text{فر}(\beta)$

روش ۲: کدهی Prefix free

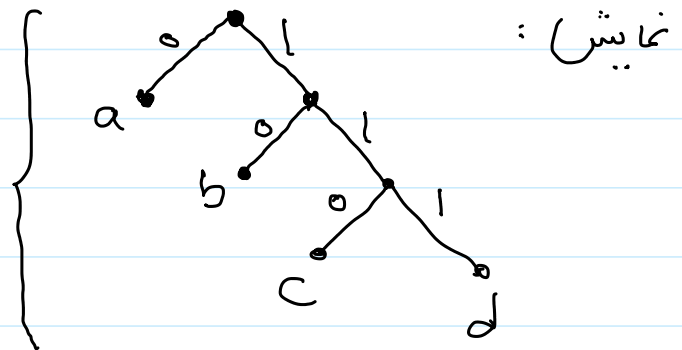
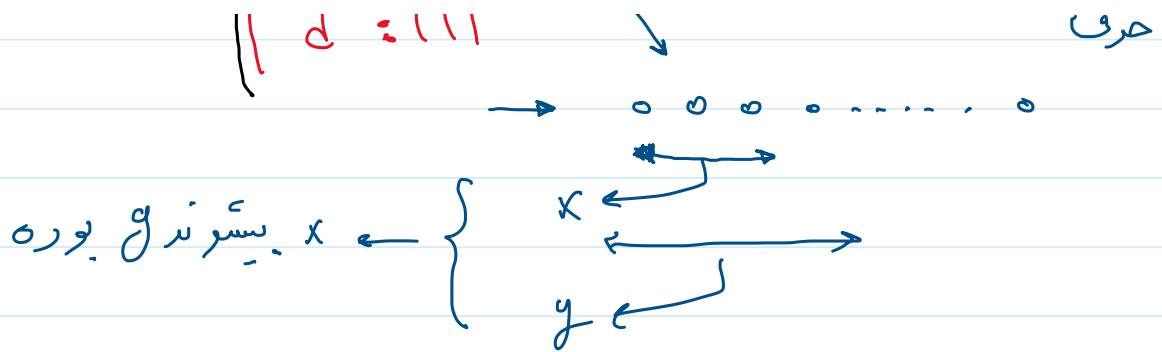
یک روش کدهی پیشوند آزاد است، اگر معادل باینری هیچ حرفی
 پیشوند معادل باینری حرف دیگر نباشد.

$$\underline{C = \{a, b\}} \leftarrow \left\{ \begin{array}{l} a: 0011 \\ b: 00101 \end{array} \right. \rightarrow \text{پیشوند ط است}$$

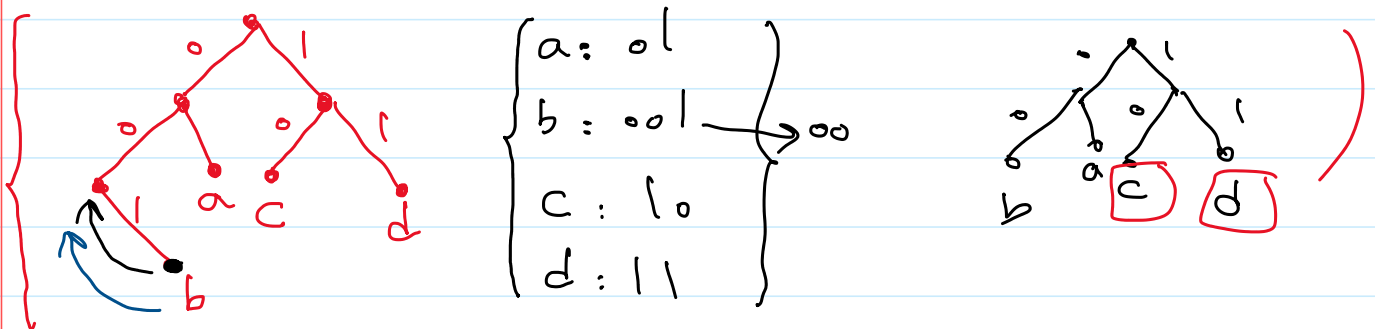
$$C = \{a, b, c, d\}$$

$$\left\{ \begin{array}{l} a: 0 \\ b: 10 \\ c: 110 \\ d: 111 \end{array} \right.$$

حرف



مثال: $C = \{a, b, c, d\}$



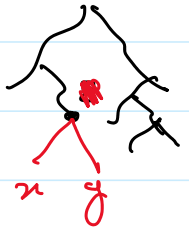
الگوریتم هافمن: ورودی: الفبای C به همراه فرکانس‌ها
 خروجی: یک روش کددهی بی‌شوند آزاد که
 میزان \leftarrow عمق مروف C در درخت بی‌شوند

$$B(T) = \sum_{c \in C} f_r(c) \times d(c)$$

راکمینی کند

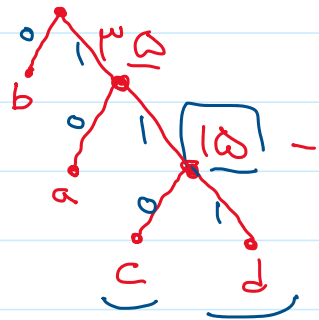
1. دو حرف با کمترین فرکانس را انتخاب کنید (x و y)
2. حرف جدید z با فرکانس $f_r(z) = f_r(x) + f_r(y)$ بسازید

۲. حرف جدید z با فرکانس $f_r(z) = f_r(x), f_r(y)$ بسازید
۳. قرار دهید $C' = C / \{x, y\} \cup \{z\}$ و مساله را به طور بازگشتی برای C' حل کنید
۴. x و y را فرزند چپ و راست z در T قرار دهید



مثال ۱) $C = \{a, b, c, d\}$

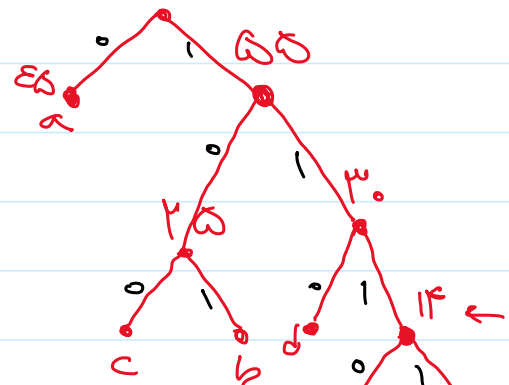
$$\begin{cases} \times f_r(a) = 20 \\ \cdot f_r(b) = 15 \\ \times f_r(c) = 5 \\ \times f_r(d) = 10 \end{cases}$$



$$\begin{cases} b = 0 \\ a = 10 \\ c = 110 \\ d = 111 \end{cases}$$

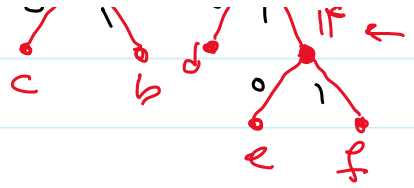
۲. $C = \{a, b, c, d, e, f\}$

$$\begin{cases} \rightarrow f_r(a) = 45 \\ f_r(b) = 13 \\ f_r(c) = 12 \\ f_r(d) = 14 \\ f_r(e) = 9 \\ f_r(f) = 5 \end{cases}$$



$$\begin{cases} a = 0 \\ b = 101 \\ c = 100 \\ d = 110 \\ e = 111 \\ f = 1111 \end{cases}$$

$f = \underline{1111}$



اثبات، الوریتم ها فیس رضی تولیدی کند مقدار $\sum_{c \in C} \text{freq}(c) \times d(c)$ را کمینه می کند

وض کیند ا حرف با معنی پستینه باشد

$a \neq b$ جای a و b تعید بهم، کل عبارت $B(T)$ کم می شود

$$r \quad r \quad r \quad r \quad y, b \quad r \quad b \neq y$$

لم c فرض کنید به ازای اعضای C ، x و y دو کاراکتر با فاکانس یکجمله باشد و قرار دهید

$$\begin{aligned} \hookrightarrow C' &= C / \{x, y\} \cup \{z\} \\ \text{pr}(z) &= \text{pr}(x) + \text{pr}(y) \end{aligned}$$

همچنین فرض کنید T درخت بینه برای C باشد. در این صورت درخت T که با اضافه کردن z فرزند به راس مربوط به z برای کاراکترهای x و y تولید می شود، درخت بینه C است.

*

↔