

Föreläsning 15

Träd

Huffman koder

Definition Ett *träd* $T = (V, E)$ är en enkel graf med egenskapen att om v och w är hörn i T så finns det en **unik** enkel väg mellan v och w .

Definition Ett *rotat träd* $T = (V, E)$ är ett träd där ett speciellt hörn v_0 blivit utsedd till rot.

I ett rotat träd med rot v_0 sägs roten vara på nivå 0. Om det finns en enkel väg av längd n från ett hörn $x \in V$ till v_0 sägs x vara på nivå n i trädet.

Ett träds höjd är den högsta nivån bland något av hörnen i T .

Om två hörn v_n på nivå n och v_{n+1} på nivå $n+1$ i ett rotat trä är sammanbundna så kallas

v_n *förälder* till v_{n+1} och

v_{n+1} *för barn* till v_n .

Ett hörn i ett rotat träd som inte har några barn kallas för ett *löv*.

Man kan karaktärisera träd med andra egenskaper än att det finns en unik enkel väg mellan varje par av hörn.

Sats Låt T vara en graf med n hörn. Följande påståenden är ekvivalenta.

- (a) T är ett träd.
- (b) T är sammanhängande och det finns ingen cykel i T
- (c) T är sammanhängande och har $n - 1$ kanter.
- (d) Det finns ingen cykel i T , och T har $n - 1$ kanter.

För att visa en sats som den ovanstående visar man

(a) $\Rightarrow$ (b)

(b) $\Rightarrow$ (c)

(c) $\Rightarrow$ (d)

och

(d) $\Rightarrow$ (a)

Vi visar några av dessa.

'(a) $\Rightarrow$ (b)'

Låt T vara ett träd.

T måste vara sammanhängande eftersom det finns en väg mellan varje par av hörn.

Antag att T innehåller en cykel $C = v_0, e_1, v_1, \dots, v_{k-1}, e_k, v_0$. Man kan anta att det är en enkel cykel (sats 8.2.24 i kursboken).

Det finns i så fall två enkla vägar mellan v_0 och v_1 . v_0, e_1, v_1 och $v_1, e_n, \dots, v_{k-1}, e_k, v_0$. Motsägelse till att T är ett träd. Alltså finns ingen cykel i T . $\square$

'(b) $\Rightarrow$ (c)'

Antag att T är en sammanhängande graf med n hörn och att T inte har någon cykel.

Måste visa att T har $n - 1$ kanter.

Visar med hjälp av induktion på antalet hörn.

$P(n)$: '(b) $\Rightarrow$ (c)' för en graf med n hörn.

Grundsteg: $n = 1$. Om T är en graf med ett hörn utan cykel så kan inte T ha några kanter alls (I så fall är de loopar, alltså cykler).

Induktionssteg. Antar att '(b) $\Rightarrow$ (c)' är sant om T är en graf med n hörn. Visar att det också är sant om T är en graf med $n + 1$ hörn.

Låt T vara en sammanhängande graf med $n + 1$ hörn som inte har någon cykel. Visar att T har n kanter. T har alltså $n - 1 = 0$ kanter.

Välj en väg p i T med maximal längd givet att ingen kant ska förekomma mer än en gång i P .

p innehåller inga cykler eftersom T inte gör det.

I så fall måste p innehålla ett hörn v så att $\delta(v) = 1$

Bilda en ny graf T^* genom att ta bort v och den enda kant som har v som ändpunkt från T .

T^* är sammanhängande och har ingen cykel.

Eftersom T^* har n hörn så har T^* $n - 1$ kanter enligt induktionsantagandet.

T måste alltså ha n kanter. $\square$

Exempel på tillämpning av träd:

Huffman koder

Vanligtvis representeras tecken i ett datorsystem genom att låta varje tecken motsvara en sträng av fix längd. Exvis ASCII

$$a = 1000001$$

$$b = 1000010$$

I en Huffmankod så låter man, för att spara utrymme, vanligt förekommande tecken motsvaras av korta strängar och ovanliga tecken av långa strängar.

Vi ska ge en algoritm för att, mha träd, konstruera en optimal Huffmankod givet frekvensen för förekomsten av olika tecken.

En pseudokod finns given för denna algoritm i kursboken. Enklare att förklara algoritmen genom exempel

Exempel:

Antag att följande tecken används med motsvarande frekvenser

Tecken Frekvens

A 22

B 4

C 3

D 6

E 16

Frekvenserna ska tolkas som att $\frac{22}{22 + 4 + 3 + 6 + 16} \approx 43,1\%$ av tecknen är A .

$\frac{4}{22 + 4 + 3 + 6 + 16} \approx 7,8\%$ av tecknen är B .

Ordna frekvenserna i växande ordning, byt ut de två lägsta mot summan av dessa:

$3, 4, 6, 16, 22 \rightarrow 3+4, 6, 16, 22$

Ordna de nya frekvenserna och byt återigen ut de två lägsta mot summan av dessa

$6, 7, 16, 22 \rightarrow 6+7, 16, 22$

Upprepa tills bara två 'frekvenser' återstår.

13,16,22→13+16,22

22,29

Konstruerar nu ett träd på följande sätt:

Se kursboken för figurer.

Börjar med ett rotat träd med tre hörn, en rot och två löv. Hörnen som inte är roten är märkta 22 resp 29.

Byt sedan ut hörnet märkt med 29 mot ett likadant rotat träd som det man började med.

Nu skall de två nya löven vara märkta med 13 resp. 16. eftersom $29=13+16$ i stegen ovan. etc

OBS Huffmankoden man får är inte unik, det kan finnas flera koder som är optimala

Märk sedan varje kant 'åt vänster' med 1 och varje kant 'åt höger' med 0. Huffman koden för tecknen kan då läsas av från roten till lövet som är märkt med motsvarande frekvens.

I vårt fall

$A = 1$

$B = 0100$

$C = 0101$

$D = 011$

$E = 00$

010010111

BADA

01011001

CAB

000110100

ED

01110111

DADA