L'algorithme glouton (Greedy Algorithme)

On suppose un treillis de vues associé à un cube de données, avec un coût de traitement associé à chaque vue (le coût de traitement est égal au coût d'espace occupé). Soient c (v) le coût de traitement de la vue v, k le nombre de vues à sélectionner et S l'ensemble qui va contenir les k vues sélectionnées plus la vue sommet qui doit toujours être matérialisée (initialement S = {vue sommet}). Apres avoir sélectionné un ensemble S de vues, le bénéfice de la vue v relatif à S noté B (v,S), est défini comme suit :

1. Pour chaque vue $w \leq v$, on définit la quantité B_w par :

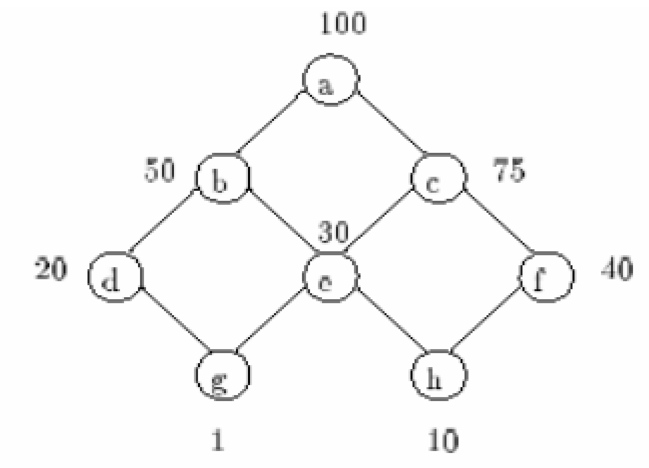
a. Soit u la vue qui a le plus petit coût dans S tel que $w \leq u$ (u est l'ancêtre de w qui a le plus petit coût de traitement dans S).

b. Si $c(v) < c(u)$, alors $B_w = c(u) - c(v)$ Sinon $B_w = 0$.

2. On définit le bénéfice de la vue v relatif à S noté B (v,S) par: $B(v,S) = \sum_{w \leq v} B_w$

On calcule le bénéfice d'une vue v en considérant comment cette vue peut améliorer le coût de traitement des vues. Pour chaque vue w que couvre v, on comparons le coût de w en utilisant v et en utilisant l'évaluation la moins chère à partir de S. Si la présence de v améliore, On retient v pour la matérialisation. L'algorithme glouton pour sélectionner k vues à matérialiser est décrit comme suit :

S = {vue sommet};
Pour i = 1 à k faire
DEBUT
Sélectionner la vue v non dans S qui maximise B(v,S);
S = S U {v};
FIN ;
S est l'ensemble des vues à matérialiser. Nous allons prendre un exemple pour illustrer cet algorithme. Considérons le treillis de la figure . Ce treillis est constitué de huit vues nommées de a à h, ayant des coûts d'espace (coût de traitement) indiqués sur le treillis



:

Remarque Lorsqu'on calcule les bénéfices, on commence avec la supposition que chaque vue est évaluée en utilisant la vue sommet a, car initialement S = {a}. Le tableau suivant représente les bénéfices des vues dans chaque itération de l'algorithme.

	1er choix	2ème choix	3ème choix
b	50* 5= 250		
c	25* 5 =125	25* 2= 50	25* 1 =25
d	80* 2 =160	30* 2= 60	30* 2= 60
e	70* 3 =210	20* 3= 60	20+ 20+ 10 =50
f	60* 2 =120	60 +10= 70	
g	99* 1 =99	49* 1= 49	49* 1= 49
h	90* 1 =90	40* 1= 40	30* 1= 30

Tab. 4.1 - Le bénéfice des vues à chaque itération.

Premier choix : initialement l'ensemble S ne contient que la vue sommet a donc touts les calculs du bénéfice se font à partir de cette vue. Si on prend la vue b pour être matérialisée, on réduit son coût et celui de ses descendants (d , e, g ,h) de 50, donc le bénéfice de B (b, S) = (100-50)*5 = 250. Le même traitement se fait avec toutes les autres vues. Si on prend la vue e pour être matérialisée, on réduit son coût ainsi que le coût de ses descendants (g et h) de 70 donc le bénéfice de e est B(e, S) = (100-30)*3 = 70*3 = 210. Pour le premier choix, on trouve que b a le plus grand bénéfice B(b, S) = 250, b est donc considérée comme la première vue matérialisée après a (la vue sommet). S = {a, b}.

Deuxième choix : Trouver le deuxième choix consiste à refaire la même opération précédente pour toutes les vues, avec S = {a, b}. Par exemple, choisir f réduit son coût de 60 (de 100 qui est le coût de a à 40 qui est le coût de f), la vue f a une seule vue descendante h. h dépend de la vue matérialisée b (b est la vue qui a le plus petit coût dans S). Donc B_h = c(u) – c(v) = c(b) – c(f) = 50 – 40 = 10. Le bénéfice de f est alors B (f, S) = B_f + B_h = 60 + 10 = 70. Comme il est indiqué dans le tableau 4.1 f a le plus grand bénéfice, elle est donc considérée comme la deuxième vue à matérialiser. S = {a, b, f}.

Troisième choix : le troisième choix est indiqué dans la troisième colonne du tableau 4.1. Les mêmes calculs que la première et la deuxième étape se font, mais en considérant l'ensemble S = {a, b, f}. Comme exemple d'un calcul un peut compliqué le calcul du bénéfice de e. Le choix de e réduit son coût de 20. B_e = c(u) – c(v) = c(b) – c(e) = 50 – 30 = 20 (On a pris b car c'est l'ancêtre matérialisé de e qui a le plus petit coût dans S). La vue e a deux vues descendantes g et h, leurs bénéfices sont calculés comme suit : B_h = c(u) – c(v) = c(f) – c(e) = 40 – 30 = 10 (On a pris f car c'est l'ancêtre matérialisé de h qui a le plus petit coût dans S). Bg = c(u)- c(v) = c(b) – c(e) = 50 – 30, (b est l'ancêtre matérialisé de g qui a le plus petit coût dans S) Bg = 20. Donc le bénéfice de e est : B (e, S) = $\sum_{W \leq e}$ Bw = Be + B_h + Bg = 20 + 10 + 20 = 50. Après avoir calculé tous les bénéfices des vues restantes, d est sélectionné pour la matérialisation avec un bénéfice égal à 60. L'ensemble des vues sélectionnées est donc

S = {a, b, f, d}. Coût total = 100 * 8 = 800. (100 est le coût de la vue a, 8 est le nombre de vues) Bénéfice total = B (b, S) + B (f, S) + B (d, S) = 380. Cet ensemble réduit le coût de 800 (le cas où seulement la vue a est matérialisée) à 420. (800 – 380 = 420.)