CSI2531, Printemps 2001

Devoir 1

A remettre Lundi le 29 Janvier à 11h30

(Valeur: 7.5%   Devoir noté sur 100)

1 Exercice de Programmation: Tri Interne (65 Points)

Dans ce devoir vous allez avoir à traiter les données d'un fichier contenant les mots de passe d'une classe d'étudiants. En particulier, on vous demande de trier le fichier par numero d'étudiant. L'objectif de cet exercice est de vous familiariser avec le langage et le compilateur de C++.

Votre programme doit performer un "tri interne", ce qui veut dire que le fichier est trié en mémoire primaire après y avoir été amené d'une traite (ceci est possible dans ce programme car le fichier est assez petit pour pouvoir tenir complètement en mémoire).

L'utilisateur du programme fournira le fichier d'entrée. Ce fichiers sera un fichier texte contenant des enregistrements à longueur fixe et à champs de longueur fixe comme decrit ci-dessous.

Nom du Champ	**Type et Longueur du Champ**	**Note**
Numero d'Etudiant	7 caractères (numeric)	Clé unique
Mot de Passe	6 caractères	
Nom de l'Etudiant	20 caractères	

Après chaque enregistrement, un caractère "newline" est présent de manière a être facilement lisible par un être humain (comme votre TA :-)).
Votre programme doit faire les choses suivantes:

- Lire les enregistrements (a partir d'un fichier d'entrée), enregistrement par enregistrement, et les sauvegarder en mémoire primaire.

- Trier les donneés en mémoire primaire (par numéro d'étudiant croissant) en utilisant un algorithme de tri efficace de votre choix (n'importe quel algorithme d'ordre $O(n\ log\ n)$ ou Quicksort).
- Ecrire les enregistrements en ordre croissant de numéro d'étudiant dans un fichier de sortie, qui devra utiliser le même format que le fichier d'entrée decrit ci-dessus.

Veuillez créer des classes de C++ qui vous permettrons de pérformer toutes ces tâches. **Note:** Veuillez noter que ces structures seront ré-utilisées dans le prochain devoir. Dans Devoir 1, les données n'ont besoin d'être traitées que de manière sequentielles. Dans Devoir 2, on vous demandera de ré-utiliser certaines parties de votre code afin de faire une rechèrche directe qui permettra de changer les mots de passes initiaux des etudiants. Votre code doit donc être assez clair et assez bien documenté pour vous permettre de le ré-utiliser ultérieurement.

Votre travail sera évalué sur sa justesse, son efficacité et son style de programmation. Par example, un problème d'efficacité a considerer est le suivant: supposons que vous avez sauvegardé les données dns un tableau d'objets `Etudiant`. Le triage explicite du tableau entier causera beaucoup de déplacement d'octets lorsque deux objets sont échangés. Pouvez-vous trouver un moyen plus efficace et élégant pour traiter ce problème?
IMPORTANT: Vous devez suivre les standards pour ce devoir qui seront postés sur la page Web un peu plus tard [Mais vous pouvez commencer votre devoir dès maintenant].

2 Problem Ecrit A: Disques (15 Points)

L'unité de disque de l'IBM 3350 utilise l'addressage par bloc. Les deux organisations de sous-blocs suivantes peuvent être utilisées:

Count-data, dans laquelle les éspaces perdus a l'organisation et utilisés par le sous-bloc de compte et les espaces interblocs sont équivalent à 185 octets; et

Count-key-data, dans laquelle les éspaces perdus a l'organisation et utilisés par les sous-blocs de compte et de clé ou les espaces accompagnants sont équivalents à 267 octets, plus la taille de la clé.

Un IBM 3350 a 19,069 octets utilisables par piste, 30 pistes par cylindre, et 555 cylindres par unité de disque. Veuillez supposer que vous avez un fichier de 350,000 enregistrements de taille 80 octets que vous voulez sauvegarder sur l'unité de disque d'un IBM 3350. Veuillez répondre aux questions suivantes en supposant (à moins que d'autres diréctives soient données) que le facteur de bloc est 10 et que l'organisation en sous-bloc count-data est utilisée. Expliquez vos réponses soigneusement.

a. Combien de blocs peuvent être sauvegardés sur une piste? Combien d'enregistrements?

b. Combien de blocs peuvent être sauvegardés sur une piste si l'organisation en sous-blocs count-key-data est utilisée et la taille de la clé est de 13 octets?

c. Combien de cylindres sont nécéssaires pour contenir le fichier (format count-data et facteur de bloc de 10)? Quel montant d'éspace sera inutilisé à cause de la fragmentation interne?

d. Si le fichier est sauvegardé sur des cylindres contigus et s'il n'y a pas d'interference avec d'autres processus qui utilisent l'unité de disques, le temps de recherche (seek time) moyen pour un accès au hazard (random acces) du fichier est d'environ 12 msec. Veuillez utiliser ce taux pour calculer le temps moyen nécéssaire a l'accès d'un enregistrement au hazard. Veuillez supposer que le disque pivote a une vitesse de 3600 r.p.m. (rotations par minutes). **Note:** En moyenne, le délai de rotation est d'une demi révolution.

3 Problem Ecrit B: Bandes Magnétiques (20 Points)

Veuilez considérer un fichier de 1,000,000 d'enregistrements dans lequel chanque enregistrement contient 100 octets. Le fichier doit être sauvegardé sur des bobines de bandes de 2,400 pieds a densité de 6,250-bpi (octet par pouce) et avec des espaces inter-blocs de 0.3 pouces. La vitesse de la bande est de 200 pouces par seconde. Veuillez répondre aux questions suivantes. Expliquez vos réponses soigneusement.

a. Veuillez montrer qu'une seule bande est nécéssaire pour sauvegarder le fichier si un facteur de bloc de 50 est utilisé.

b. Si un facteur de bloc de 50 est utilisé, combien d'enregistrements supplémentaires peuvent être accomodés sur une bande magnétique de 2,400 pieds?

c. Quelle est la densité effective d'enregistrement lorsqu'un facteur de bloc de 50 est utilisé?

d. Quelle doit être la taille optimale du facteur de bloc afin d'atteindre une densité effective d'enregistrement maximum? Quels résultats négatifs peuvent résulter de l'augmentation du facteur de bloc? (Note: une mémoire tampon d'entrée/sortie assez large pour accomoder un bloc doit être allouée)

e. Quel serait le facteur de bloc minimal requis pour accomoder le fichier sur une bande magnétique?

f. Si un facteur de bloc de 50 est utilisé, combien de temps la lecture d'un bloc (l'espace inter-bloc inclus) prendrait-elle? Quelle serait le taux effectif de transmission? De combien de temps aurait-on besoin pour lire le fichier en entier?