

Travail Pratique – SFSD : « Fichier de données non ordonné L~OF avec un Index dense en MC et un fichier index TOF » - Programmation en C

On suppose qu'on souhaite manipuler les documents de la bibliothèque de l'ESI-Alger dans un fichier binaire de données nommé « **DOCUMENTS_LIBRARY.BIN** ». Ce fichier sera organisé selon la méthode « **L~OF ou LnonOF** » : fichier vu comme **Liste**, **non Ordonnée**, avec des enregistrements de taille **Fixe**. Un bloc peut contenir au maximum 1024 enregistrements.

Travail demandé

1. Programmez et utilisez le modèle de la machine abstraite vu en cours. Rappelons-le, le modèle comprend les fonctions suivantes : *ReadBlock*, *WriteBlock*, *getHEADER*, *setHEADER*, *Open*, *Close* et *AllocBlock*.

2. Programmez un module de **création** du fichier **DOCUMENTS_LIBRARY.BIN** où chaque enregistrement est composé de 7 champs : < **Document_ID**, **Title**, **Author**, **Type**, **Domain**, **Pub_Year**, **Available_Qty**>. La création de ce fichier de données s'appuiera sur un module de **chargement initial** qui consiste à remplir le fichier avec N enregistrements (N étant un entier à lire, par exemple : 100 000).

➔ Calculez et affichez le coût de cette opération de création soit C2.

Pour chaque nouvel enregistrement inséré, son identifiant (sa clé) et ses coordonnées (n° de bloc et déplacement) sont insérées dans une table d'index (en MC) avec décalages pour garder l'ordre des clés.

Pour que le chargement initial de **DOCUMENTS_LIBRARY.BIN** soit rapide, vous devrez le réaliser avec des valeurs aléatoires comme suit :

- **Document_ID** (Clé de l'enregistrement): Vous générez un nombre aléatoire entre 110000 et 990000.
- **Title** (Titre du document): Dans un premier temps, vous générez un nombre entre 50 et 70 (ce nombre étant le nombre de caractères du titre). Ensuite, pour chaque caractère du titre, vous générez un nombre entre 1 et 26 puis vous mettez dans ce caractère la lettre qui correspond.
- **Author** (Nom complet du premier auteur): Dans un premier temps, vous générez un nombre entre 4 et 30 (ce nombre étant le nombre de caractères du nom/prénom). Ensuite, pour chaque caractère du nom/prénom vous générez un nombre entre 1 et 26 puis vous mettez dans ce caractère la lettre qui correspond.
- **Type**: Vous générez un nombre entre 1 et 6. Ensuite, vous récupérez le type de document à partir de ce nombre (1 : *Ouvrages et manuels*, 2 : *Revue et périodiques*, 3 : *Actes de conférences*, 4 : *Rapports de recherche*, 5 : *Mémoires et thèses*, 6 : *Polycopiés et support de cours*).
- **Domain**: Vous générez un nombre entre 1 et 12. Ensuite, vous récupérez le domaine abordé par le document à partir de ce nombre (1 : *Algorithmes et structures de données*, 2 : *Intelligence artificielle et apprentissage automatique*, 3 : *Systèmes d'exploitation*, 4 : *Cybersécurité*, 5 : *Bases de données et gestion des données*, 6 : *Réseaux et télécommunications*, 7 : *Génie logiciel, programmation et développement logiciel*, 8 : *Informatique graphique, multimédia, vision par ordinateur et traitement d'images*, 9 : *Science des données et statistiques*, 10 : *Robotique et systèmes embarqués*, 11 : *Blockchain et technologies distribuées*, 12 : *Calcul haute performance et informatique quantique*).
- **Pub_Year**: Vous générez un nombre aléatoire entre 1970 et 2024 pour l'année de publication du document.
- **Available_Qty** (Quantité disponible du document): Vous générez un nombre entre 1 et 10 pour le nombre d'exemplaires du document disponible dans la bibliothèque.

3. Programmez les modules énumérés ci-dessous (avec affichage du coût de chaque opération) :

- 3.1. Sauvegarde de l'index dans un fichier index TOF nommé « DocumentID_INDEX.idx » (Coût C31)
- 3.2. Chargement de l'index en MC à partir du fichier DocumentID_INDEX.idx (Coût C32)
- 3.3. Recherche d'un document par son identifiant. Ce module doit retourner l'adresse du bloc ainsi que sa position dans le bloc. (Coût C33)
- 3.4. Insertion d'un nouveau document au fichier DOCUMENTS_LIBRARY.BIN. Son identifiant doit être généré aléatoirement et doit être unique (pas de doublons). (Coût C34)
- 3.5. Suppression d'un document donné par l'identifiant. (Coût C35)
- 3.6. Modification de l'état d'un document donné (Available_Qty) lors par exemple d'un emprunt. (Coût C36)

4. Programmez les modules énumérés ci-dessous (avec affichage du coût de chaque opération). Afin d'accélérer les traitements, vous pouvez utiliser d'autres Index adéquats en MC et des fichiers de sauvegarde de ces index.

- 4.1. Suppression de tous les enregistrements relatifs à un type de document donné. (Coût C41)
- 4.2. Affichage de tous les documents relatifs à un type donné et datant de moins de 5 ans dont les années de publication appartiennent à l'intervalle [Y1, Y2]. (Coût C42)
- 4.3. Affichage de tous les enregistrements appartenant à un domaine donné. (Coût C43)

5. A partir du fichier DOCUMENTS_LIBRARY.DAT, créez un fichier TOF nommé JOURNAL_MAGAZINE.BIN contenant tous les document de type « Revues et périodiques ». (Coût C5).

6. A partir du fichier JOURNAL_MAGAZINE.BIN, programmez les modules suivants (avec affichage du coût de chaque opération) :

- 6.1. Recherche d'une revue ou périodique (Journal or Magazine) donnée par son identifiant et affichage de tous ses champs. (Coût C61).
- 6.2. Création d'une table d'index primaire non dense dans laquelle chaque entrée contient la plus grande valeur d'identifiant par bloc du fichier JOURNAL_MAGAZINE.BIN ainsi que le numéro du bloc et la position dans le bloc où se trouve le document. Cette création engendre la création du fichier index soit INDEX_JOURNAL_MAGAZINE.IDX. (Coût C62).
- 6.3. Recherche d'une revue ou périodique donnée à partir de l'index et affichage de tous ces champs. (Coût C63).

7. Implémentez les modules **facultatifs** suivants :

- 7.1. Affichage de l'en-tête d'un fichier donné (DOCUMENTS_LIBRARY.DAT ou JOURNAL_MAGAZINE.BIN).
- 7.2. Affichage du contenu d'un bloc donné d'un fichier donné.
- 7.3. Affichage du contenu de tout le fichier bloc par bloc pour un fichier donné.



Points importants :

- Tous les modules demandés doivent être exécutés via un menu principal.
- Vous devez me remettre vos codes sources en les téléversant sur mon *Drive* via un formulaire en ligne qui vous sera communiqué prochainement. Le dernier délai de remise du TP est fixé pour le **mardi 7 janvier 2025 avant 23h59**. Aucun retard n'est toléré. Si vous dépassiez ce délai, vous seriez pénalisés.

- Vous devez respecter le format suivant pour l'identificateur de vos fichiers : **NOM1_NOM2_TP2_Gi.zip** où i représente de numéro de groupe (fichier compressé contenant uniquement vos codes sources, vos fichiers de données et fichiers d'index utilisés). Si vous ne respectiez pas ce format, vous seriez sanctionnés.
- Plusieurs aspects seront pris en considération lors de l'évaluation de vos TP (la démonstration, les codes sources (les structures, les commentaires, etc.), l'interface homme-machine (IHM), etc.).

Bon courage ! بالتوفيق