

INF 362, EC1 : Programmation Système

TP : Création d'une bibliothèque pour le traitement de données de recommandation

La mise en place d'un système de recommandation passe par trois étapes importantes qui sont : le prétraitement du jeu de données, la construction d'un modèle pour le système de recommandation et enfin l'évaluation de ce modèle. Ce TP va se concentrer uniquement sur le prétraitement du jeu de données.

Objectifs :

- Mettre en œuvre une bibliothèque logicielle en langage C.
- Traiter des jeux de données au format brut (fichier .txt).
- Manipuler des structures de données et effectuer des analyses simples.
- Comprendre la compilation séparée, les bibliothèques statiques et dynamiques.
- Utiliser un Makefile pour automatiser la compilation.

Les transactions sont stockées dans un fichier texte .txt. Chaque ligne suit le format suivant :

`<user_id> <item_id> <category_id> <rating> <timestamp>`

Exemple :

```
249 1087 21 4.3 969215645
83 1134 34 3.5 1000121189
7 1231 43 2.4 998855800
64 1042 32 3.0 952138789
```

Travail à réaliser

1. Créer la structure du projet

Créez tous les dossiers nécessaires (src, include, lib, test, data, etc.).

2. Fonctions à implémenter

a. Lecture et statistiques sur les données

`void afficher_stats(const char* fichier);`

Elle affiche :

- Nombre d'utilisateurs
- Nombre d'items
- Nombre de catégories
- Plage temporelle (min/max timestamp), convertit en dates compréhensible par l'humain
- Nombre d'items par catégorie

- Catégorie la plus fréquente / la moins fréquente

b. Extraire les transactions dans une période donnée

```
void extraire_transactions_par_periode(const char* fichier_entree, const char*
fichier_sortie, time_t t1, time_t t2);
```

- Lire un fichier de transactions
- Extraire uniquement celles dont les timestamps sont compris entre t1 et t2
- Les enregistrer dans un nouveau fichier (`transactions_t1-t2.txt`)

c. Filtrage par fréquence minimale d'apparition

```
int filtrer_transactions(const char* fichier_entree, const char* fichier_sortie, int minU, int
minI);
```

- Lire les transactions
- Supprimer les transactions des utilisateurs ayant < minU interactions
- Supprimer les transactions des items ayant < minI interactions

d. Nettoyage des données de test

```
int nettoyer_fichier_test(const char* fichier_train, const char* fichier_test, const char*
fichier_clean);
```

- Lire les fichiers d'apprentissage (train) et de test (test)
- Supprimer du test les transactions contenant :
 - Des utilisateurs absents dans le train
 - Des items absents dans le train

3. Fichier de test main.c

Créez un fichier main.c qui :

- Utilise la bibliothèque reco
- Appelle les 3 fonctions principales sur des fichiers d'exemple
- Affiche un message de succès à chaque étape

4. Compilation avec Makefile

Rédigez un Makefile permettant de :

- Compiler la bibliothèque statique .a
- Compiler la bibliothèque dynamique .so
- Compiler le fichier main.c qui teste votre bibliothèque