

## OPTIONNEL – GOODREADS

Le fichier `books.db` contient un extrait de la base de données de [goodreads.com](https://www.goodreads.com), un (gros) site sur lequel les utilisateurs peuvent noter et critiquer les livres qu'ils ont lus<sup>1</sup>.

- La table `books` contient divers informations sur 500 livres populaires.
- La table `ratings` contient les notes données par un panel de 1000 utilisateurs à ces 500 livres. Chaque ligne correspond à une note donnée par un utilisateur pour un livre.
- La table `tags` contient une liste d'étiquettes (*tags*) que l'on peut donner aux livres (par exemple "romance" ou "classics" ou "fantasy").
- Chaque ligne de la table `book_tags` indique combien d'utilisateurs ont donné un certain *tag* à un certain livre.
- Chaque ligne de la table `to_read` indique qu'un certain utilisateur a placé un certain ouvrage dans sa liste de « livres à lire ».

### Exercice XLVIII.1 – Structure de la base de données

Pour obtenir plus d'information sur la structure de la base de données, vous pouvez utiliser les requêtes spéciales suivantes :

- `PRAGMA table_info(nom_de_la_table)` donne la liste des colonnes de la table spécifiée avec leur type, et précise l'éventuelle clé primaire ;
  - `PRAGMA foreign_key_list(nom_de_la_table)` donne la liste des clés étrangères présentes dans la table spécifiée avec à chaque fois la colonne à laquelle elles font référence.
1. Exécuter la requête suivante : `PRAGMA table_info(books)`. Quelle est la clé primaire spécifiée dans la base de données ? Essayer de deviner la signification des différentes colonnes (si vous ne trouvez pas tout ce n'est pas très grave. . .). D'autres colonnes pourraient-elles faire office de clé primaire ?
  2. Récupérer le schéma de la table `book_tags`. Indiquer ce que vous pensez être des clés étrangères, et vérifier en exécutant `PRAGMA foreign_key_list(book_tags)`.

### Exercice XLVIII.2 – Requêtes simples

1. Combien de livres<sup>a</sup> ont été publiés depuis l'an 2000 ?
2. La colonne `rating_counts` de la table `books` indique le nombre total de notes reçues par un livre (sans se limiter aux mille utilisateurs de la base). Au total, combien de notes les livres de la base ont-ils reçus ? Combien de ces notes ont été données par les utilisateurs de la base ?  
*Il faut faire deux requêtes différentes pour répondre à ces deux questions.*
3. Donner les 10 livres les mieux notés en moyenne (par l'ensemble des utilisateurs, donc en utilisant la colonne `average_rating`).
4. Combien y a-t-il d'auteurs distincts pour les livres considérés ?
5. Donner la liste des livres dont le titre contient "girl" (on utilisera le mot-clé **LIKE**, se référer au cours).

<sup>a</sup>. On parle systématiquement de livres *de la base*, même quand ce n'est pas précisé

1. ou pas

**Exercice XLVIII.3 – Requêtes intermédiaires**

1. Classer les utilisateurs par la moyenne des notes qu'ils ont données.
2. Classer les auteurs par nombre de notes parfaites obtenues au total (pour chaque livre, ce nombre correspond à la colonne `ratings_5`).
3. Donner la liste des années de publication avec le nombre de livres correspondants.
4. Quelle est la note moyenne donnée par les utilisateurs de la base à 1984 ?
5. Donner les dix livres les mieux notés en moyenne par les utilisateurs de la base.
6. Donner la liste des auteurs ayant au moins 3 livres dans la base.
7. Combien de notes les utilisateurs de la base ont-ils données à des livres publiés en 2013 ?

**Exercice XLVIII.4 – Plus compliqué**

1. Donner la liste des livres ayant reçu au moins 500 fois le tag "`classics`".  
*Optionnel : s'interroger un éventuel déclin de la civilisation occidentale.*
2. Pour chaque livre, donner le tag le plus fréquemment donné.  
*Seuls 61 livres différents apparaissent dans la table `book_tags`, cette requête doit donc renvoyer 61 lignes.*
3. Quels sont les identifiants des livres que personne ne veut lire (c'est-à-dire n'apparaissant pas dans la table `to_read`) ?
4. Quels sont les titres de ces livres ?  
*On pourra utiliser le mot-clé `IN`, après avoir cherché en ligne comment il fonctionnait.*
5. On considère le livre ayant 1 comme `book_id` (il s'agit du premier tome de la trilogie *Hunger Games* et cela nous évitera une jointure supplémentaire). Un certain nombre des utilisateurs de la base ont beaucoup aimé ce livre et lui ont attribué la note de 5 : on peut raisonnablement penser que les goûts de ces utilisateurs diffèrent du goût moyen des utilisateurs du site. Pour vérifier cela, on demande de classer les livres par différence entre la note moyenne que ces utilisateurs leur ont donnée et leur note moyenne globale. Pour éviter le « bruit », on se limitera aux livres qu'au moins 10 de ces utilisateurs ont notés.