Rapport de projet MongoDB

Gaspard DANNET

Arthur HOUDANT

**Modélisation**

Compte tenu des besoins exprimés en section 1.3. , nous avons décidé de ne garder que les champs :

Pour les restaurants :

* business\_id
* name
* city
* address
* categories
* attributes {

Ambience

}

Pour les users :

* user\_id
* name
* friends

Pour les reviews :

* review\_id
* user\_id
* stars
* business\_id
* text

Pour simplifier les requêtes exprimées en section 1.3, nous avons ajoutés quelques champs.

Aux restaurants, nous avons ajouté le champ « nbAmbience » qui correspond au nombre d’ambiance dans le restaurant (= nombre d’ambiance de valeur true dans le dictionnaire restaurant[« attributes »][«Ambience »]). Ceci étant fait, le champ « attributes » ne sert plus par la suite, nous l’avons donc retiré.

Aux users, nous avons ajouté le champ « nbFriends » qui correspond au nombre d’ami de l’utilisateur.

L’extraction des données utiles et les modifications apportées ont été effectués dans le script extraire\_donnees.py (à noter que dans ce même fichier .py a été implémenté l’ajout d’un champ « criteres » aux restaurants qui nous sera utile pour la section recommandation et plus particulièrement le calcul du facteur « préférence ». Ce champ représente).

Ensuite, nous avons apportés d’autres modifications à cette base de données afin de faciliter le programme de recommandation.

A la collection de restaurants, nous avons donc rajoutés les champs « nbReview » et « avgStars » correspondant respectivement au nombre de review qui ont été rédigé par un utilisateur sur le restaurant et à la note moyenne du restaurant. Ces deux champs nous serons utiles pour calculer le facteur « appréciation » de la recommandation.

A la collection de users, nous avons ajoutés le champ « criteres » qui correspond aux critères préférés de l’utilisateur. Un critère sera dans « criteres » si l’utilisateur a donné une note supérieure à 2.5 à un restaurant qui possède ce critère dans son champ « criteres ». Ce champ représente .

Au final, la base de donnée ressemble à :

Collection restaurants :

* business\_id
* name
* city
* address
* categories
* nbReview
* avgStars
* criteres

Collection users :

* user\_id
* name
* friends
* nbFriends
* criteres

Collection reviews :

* review\_id
* user\_id
* stars
* business\_id
* text

Pour avoir une information sur le lieu du restaurant, nous avons décidé de garder le champ « address » de la collection restaurants mais nous aurions très bien pu le supprimer.