

# TD2 – Requêtes sur une BD non normalisée

1. Décompresser l'archive [dblp.json.zip](#) dans le dossier /tmp/<votreRepo>/data
2. Lancer une nouvelle instance d'un serveur mongoDB avec Docker, exécuter dans un terminal :  
**docker run --name mongoserver -d mongo**
3. Lancer un nouveau conteneur docker qui importe ce fichier json (pensez au -v !) dans mongo  
**docker run -it --link mongoserver:mongo -rm -v /tmp/<votreRepo>/data:/data mongo bash**
4. Puis dans ce container, lancer l'importation des données :  
**mongoimport --db dblp --collection publis --host mongo --port 27017 --file /data/dblp.json --jsonArray**

Cette base DBLP contient des entrées avec des clés pouvant être présentes ou non dans certains cas (p.ex. clés *booktitle*, *cites*,...), voire des données de type différents (p.ex. clé *pages*).

```
{
  "_id" : "books/daglib/0025185",
  "type" : "Book",
  "title" : "MongoDB - The Definitive Guide: Powerful and Scalable Data Storage.",
  "pages" : "I",
  "publisher" : "O'Reilly",
  "year" : 2010,
  "authors" : [ "Kristina Chodorow", "Michael Dirolf" ],
  "isbn" : [ "978-1-449-38156-1" ]
}
{
  "_id" : "conf/icod/BouzeghoubG83",
  "type" : "Article",
  "title" : "The Design of an Expert System for Database Design.",
  "booktitle" : "ICOD-2 Workshop on New Applications of Data Bases",
  "pages" : { "start" : 203, "end" : 223 },
  "year" : 1983,
  "url" : "db/conf/icod/icod83w.html#BouzeghoubG83",
  "authors" : [ "Mokrane Bouzeghoub", "Georges Gardarin" ],
  "cites" : [ "journals/tods/AstrahanBCEGGKLMPTWW76", "books/bc/AtzeniA93",
    "journals/tcs/AtzeniABM82", "journals/jcss/AbiteboulB86" ]
}
```

Cette base contient une extraction des publications de recherche en informatique (sur [DBLP.org](#)) :

- Les articles de conférences/journaux scientifiques sont de *type* « Article », avec leur titre indiqué dans la clé *title*, et l'ouvrage de référence (s'il existe) dans la clé *booktitle* ;
- Les chapitres de livre sont de *type* « Book », avec également les clés *title* et *booktitle* ;
- Les thèses sont de *type* « Phd » ;
- Chaque publication est associée à une liste d'auteurs (*authors*), et éventuellement un éditeur (*publisher*) ;
- Certaines publications peuvent contenir des références internes avec la clé *cites*.

## Questions

Avant de faire les questions ci-dessous, vous devez lancer un client mongo, qui se connectera à l'instance du serveur MongoDB (ici « mongoserver ») afin d'y exécuter des commandes :

```
docker run -it --link mongoserver:mongoserv --rm mongo mongo --host mongoserv test
```

1. Liste de toutes les publications dont le *type* est un livre (« Book »).
2. Liste des publications publiées depuis 2011.
3. Liste des livres depuis 2014.
4. Liste des publications ayant un éditeur (*publisher*).
5. Liste des publications dont l'auteur est « Jeffrey D. Ullman ».
6. Titre des publications dont « Jeffrey D. Ullman » le premier auteur.
7. Nom des éditeurs (*publisher*) et titre des publications dont « Jeffrey D. Ullman » est le seul auteur.
8. Titre des publications dont le titre contient le mot « database ».
9. Liste distincte de tous les éditeurs (*publisher*).
10. Liste de tous les auteurs distincts.
11. Liste des publications contenant le mot « database » et dont le titre de l'ouvrage (*booktitle*) et la page de départ existent. Trier le résultat par titre d'ouvrage et page de départ.
12. Projeter le résultat précédent par titres, titres d'ouvrage et pages.
13. Compter, pour le résultat précédent, le nombre de publications retournées.
14. Pour le résultat précédent, compter par *type* de publication.
15. Compter le nombre de publications par auteur. Trier le résultat par ordre décroissant.
16. Pour chaque éditeur (s'il existe), donner l'année moyenne et le nombre de publications. Trier le résultat par nombre décroissant de publications.
17. Compter par année et éditeur (s'il existe), le nombre de publications. Pour ces derniers, ne retourner que ceux ayant plus 200 publications.
18. Pour chaque éditeur, calculer la moyenne du nombre de publications par année. Trier le résultat par moyenne décroissante.
19. \*\*\*\*\* Modifier toutes les publications ayant des pages pour ajouter le champ *pp* avec pour valeur le motif suivant : « start–end ».