

Travaux Dirigés n°4 : Les bases de données NoSQL - MongoDB et Python

Outils à utiliser :

- Anaconda (Python): <https://www.anaconda.com>
- MongoDB : <https://www.mongodb.org/>
- PyMongo : pip install pymongo

Dans ce TD, nous allons créer une base, une collection, d'y insérer des données et de l'interroger. Les documents fournis correspondent à un extrait d'une base de publications scientifiques, The DBLP Computer Science Bibliography.

Télécharger la base de données **[DBLP.json.zip](https://sites.google.com/site/lessonsfeerrag/home/courses-for-master-1-STIC/dblp.json.zip)**
(<https://sites.google.com/site/lessonsfeerrag/home/courses-for-master-1-STIC/dblp.json.zip>)

Partie 1 : Gestion de collection

Voici le résumé des commandes à effectuer.

- Se connecter à la base DBLP : use DBLP;
- Créer une collection « publis » : db.createCollection('publis');
- Créer le document suivant (astuce pour le client mongo: le mettre sur une ligne; c'est plus facile avec un client graphique type RoboMongo):

```
{  
  "type": "Book",  
  "title": "Modern Database Systems: The Object Model, Interoperability,  
and Beyond.",  
  "year": 1995,  
  "publisher": "ACM Press and Addison-Wesley",  
  "authors": ["Won Kim"],  
  "source": "DBLP"  
}
```

- Insérer le document dans la collection publis avec **db.publis.save(...)**;
- Consulter le contenu de la collection : **db.publis.find()**;
- Importer les données dans MongoDB :
 - Télécharger le fichier contenant les données : **[DBLP.json.zip](https://sites.google.com/site/lessonsfeerrag/home/courses-for-master-1-STIC/dblp.json.zip)**
 - Décompresser le fichier **dblp.json.zip**
 - Dans le même répertoire, lancer l'importation du fichier :

```
mongoimport --host localhost:27017 --db dblp --collection  
publis < dblp.json --jsonArray
```

Note

Le chemin vers l'exécutable `mongoimport` est nécessaire, ou la variable d'environnement `PATH` contenant le chemin vers `mongo/bin`. L'opération peut prendre quelques secondes (118000 items à insérer)

- Dans la console mongo vérifier que les données ont été insérées : `db.publis.count()`;

Partie 2 : Interrogation simple

Exprimez des requêtes simples pour les recherches suivantes utilisant un programme python :

- Question1.** Liste de tous les livres (type « Book ») ;
- Question2.** Liste des publications depuis 2011 ;
- Question3.** Liste des livres depuis 2014 ;
- Question4.** Liste des publications de l'auteur « Toru Ishida » ;
- Question5.** Liste de tous les éditeurs (type « publisher »), distincts ;
- Question6.** Liste de tous les auteurs distincts ;
- Question7.** Trier les publications de « Toru Ishida » par titre de livre et par page de début ;
- Question8.** Projeter le résultat sur le titre de la publication, et les pages ;
- Question9.** Compter le nombre de ses publications ;
- Question10.** Compter le nombre de publications depuis 2011 et par type ;
- Question11.** Compter le nombre de publications par auteur et trier le résultat par ordre croissant ;