

Projet Liage de Données

HMIN209 - Web Sémantique et Social

Rendu : En Binôme, Mardi 21 Février, au plus tard minuit*

Dans un unique dossier archivé:

- Fichiers des résultats d'alignement des instances**
- Le fichiervotre configuration Silk-LSL.xml**

**Nommage du dossier : “TP2_2_nomX_nomY”,
où nomX et nomY sont les noms des membres du binôme.**

Le but de ce TP est de se familiariser à l'outil “SILK”.

L'outil permet de découvrir des liens d'équivalences entre des instances au sein de deux jeux de données différents.

Les détails concernant SILK peuvent être consultés sur la page du projet :

<http://silkframework.org/>.

I. Sélection des jeux de données à liées

Les jeux de données à liés peuvent résider dans la même machine ou dans des serveurs distants accessibles à partir d'un SPARQL Endpoint (point d'accès aux RDF triplestores → notion à aborder en détail dans le cours 3).

Dans ce TP, nous travaillerons avec les deux jeux de données suivants :

- INSEE, accessible via SPARQL Endpoint : <http://id.insee.fr/sparql/>
- IGN accessible via SPARQL Endpoint : <http://data.ign.fr/id/sparql/>

Les instances des deux jeux de données sont déjà liés avec des liens owl:sameAs.

Pour extraire les différentes instances liées, il suffit de requêter l'un des SPARQL Endpoint avec la requête suivante :

```
SELECT * WHERE
{
    ?instanceA <http://www.w3.org/2002/07/owl#sameAs> ?instanceB
}
```

Chaque ligne retournée avec cette requête représentera deux instances liées avec le prédicat owl:sameAs

Le but de ce TP est de reconstituer des liens de cette liste en se basant sur l'outil SILK.

II. Installation de Silk Framework et Configuration du fichier LSL

Le jar est téléchargeable depuis le lien: <https://github.com/silk-framework/silk/releases>

Nous travaillons avec la version silk-singlemachine-2.7.0 en mode ligne de commande.

SILK prend en entrée un fichier de configuration « Silk Link Specification Language (Silk-LSL) »

Un exemple expliqué du fichier de configuration Silk-LSL est disponible à l'adresse https://app.assembla.com/wiki/show/silk/Link_Specification_Language

1. Écrivez votre propre fichier de configuration Silk-LSL.xml qui va vous permettre de relier des instances des deux jeux de données INSEE et IGN.
2. Exécutez SILK Single Machine avec la commande suivante :

```
$ java -DconfigFile= Silk-LSL.xml -jar silk.jar
```

3. Vérifiez les résultats obtenus et reconfigurer silk jusqu'à l'obtention du meilleur résultat possible. La vérification peut se faire en consultant les liens générés présents dans le fichier restitué par SILK.