# Projet: Découverte de clés pour le liage de Données

## HMIN209 - Web Sémantique et Social

## Rendu : En Binôme, Mardi 28 Février, au plus tard minuit*

## Dans un unique dossier archivé:

**-- Clés découvertes par SAKey sur chaque jeu de données**

**-- Fichiers des résultats de liage de données (deux)**

**--Fichiers de configuration Silk-LSL.xml (deux)**

**Nommage du dossier : "TP3_nomX_nomY", où nomX et nomY sont les noms des membres du binôme.**

Le but de ce TP est de se familiariser à la notion de clés et l'outil SAKey. Cet outil permet la découverte de clés dans des jeux de données RDF en permettant des exceptions. Les clés découvertes par SAKey vont être utilisées pour le liage de données en utilisant l'outil SILK.

### I. Sélection des jeux de données à lier

Les deux jeux de données **Person11.nt** et **Person12.nt** se trouvent sur le lien suivant :

https://drive.google.com/drive/folders/0B95XGTsgngtVWlJqbGtqNl95cXM?usp=sharing

Ces deux jeux de données contiennent des instances de la class Person. Chaque fichier contient 500 instances. Pour chaque instance **a** dans un jeu de données il y a une instance **b** pour la quelle sameAs(a, b).

Le but de ce TP est de découvrir des liens **SameAs** en utilisant des clés découvertes par SAKey pour le liage de données. Les clés découvertes vont être utilisées pour constituer le fichier de configuration.

### II. Exécution de SAKey

Télécharger le fichier exécutable de SAKey sur le lien suivant :

https://www.lri.fr/sakey

Pour lancer SAKey à partir de la ligne de commande :

*java –jar sakey.jar <VotreFichierRDF> <#Exceptions+1>*

Par exemple, pour trouver des 0-almost keys, c.-à-d. des clés sans exceptions, on exécutera la commande *java –jar sakey.jar Person11.nt 1*. Pour trouver des clés avec au plus 2 exceptions on exécutera *java –jar sakey.jar Person11.nt 3.*

1.Lancer SAKey pour chaque de deux jeux de données afin de les 0-almost keys pour chaque fichier.

**III. Silk Framework**

Vous pouvez utiliser la version silk-singlemachine-2.7.0 ou la version graphique qui se trouve sur le lien suivant :

https://github.com/silk-framework/silk/releases/download/release-2.7.1/silk-workbench-2.7.1.tgz

Pour plus d'informations sur la version graphique de Silk, vous pouvez regarder ici : http://personal.sirma.bg/vladimir/misc/silk-book.pdf

2. Écrivez votre propre fichier de configuration Silk-LSL_11.xml en utilisant les clés découvertes sur le jeu de données Person11.nt.

3. Écrivez votre propre fichier de configuration Silk-LSL_12.xml en utilisant les clés découvertes sur le jeu de données Person12.nt.

4. Lancer Silk avec chaque fichier de configuration jusqu'à l'obtention du meilleur résultat possible (C'est à vous de juger si vos réponses sont bonnes ou pas sachant qu' il y a 500 liens corrects (pour chaque instance dans le fichier Person11.nt il y une correspondance dans le fichier Person12.nt).

5. A la fin, il faut fournir deux dossiers contenant chacun a) les clés découvertes par jeu de données, b) le fichier Silk-LSL_1x.xml et c) les fichiers contenant les liens découverts.

**Dépôt du TP sur l'adresse suivante** : danai.symeonidou@inra.fr