



M2 Data science

Rapport de projet :

Sujet n2 :

Data linking and keys selection

Réalisé par

Haytham Ait Ettajer

Jean Mohamed Sawadogo

Salma Damak

Encadré par

Fatiha Sais

Année universitaire 2020/2021

Table des matières

| | | |
|----------|--|----------|
| 1 | Analyse et pré-traitement de la base iimb | 5 |
| 1.1 | iimb data | 5 |
| 1.2 | Pré-traitement des données | 5 |
| 2 | Sakey | 7 |
| 3 | Limes | 8 |

Table des figures

| | | |
|---|---|----|
| 1 | Exemple d'article qui n'a pas de sens | 5 |
| 2 | Différence de format | 6 |
| 3 | Résultats de Sakey | 7 |
| 4 | Limes | 8 |
| 5 | Configuration du xml de Limes | 9 |
| 6 | Résultats de Mapping | 10 |
| 7 | Résultats du Limes | 11 |

Introduction

Des volumes élevés d’une grande variété de données précieuses peuvent être facilement collectés et générés à partir d’une large gamme de sources de données, on peut les collecter sous la forme de bases de données relationnelles ou sous la forme de ressources RDF : c’est ce qui nous intéresse dans ce projet. Toutefois, ces données peuvent avoir des anomalies à savoir l’ajout des champs qui n’ont pas de sens et des propriétés qui ne sont pas intéressantes, il demeure donc nécessaire de procéder à un pré-traitement (nettoyer ces ressources) pour ensuite les exploiter dans la mise en place des liens.

Le présent rapport comporte trois parties : La première s’articule autour des pré-traitements pour lesquels on a opté pour notre base large iimb. La deuxième partie, quant à elle, est consacrée à l’analyse des différents résultats obtenus en exploitant sakey sur la base normale (non pré-traitée) et la base pré traitée. Tout au long de la partie finale nous présentons Limes et les résultats obtenus en utilisant ce framework de découverte de liens.

1 Analyse et pré-traitement de la base iimb

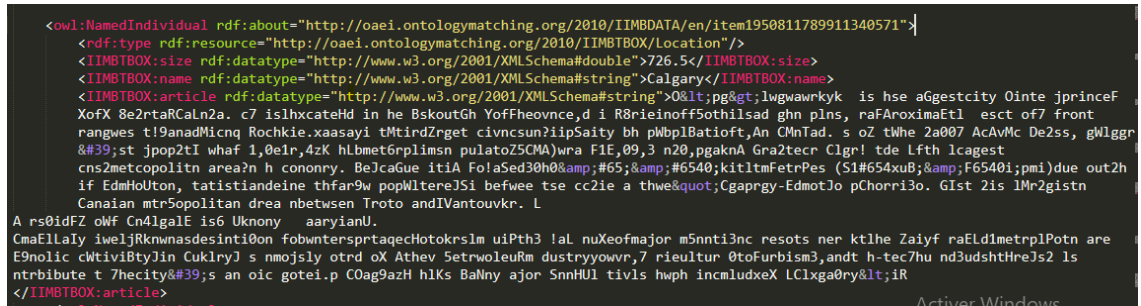
1.1 Iimb data

Cette base de données représente un ensemble de données initiale qui décrit des informations concernant le monde cinématographique (films, acteurs, réalisateurs, etc.) qui ont été extraites du Web.

Différents types de transformations de valeurs, structurelles et logiques, ont été appliquées à cet ensemble de données initial pour générer un ensemble de 80 données de test différentes.

1.2 Pré-traitement des données

La première chose que nous avons remarqué comme le montre la figure 3 en analysant les différents cas de test c'est la présence des balises qui contiennent des textes qui n'ont pas de sens, des mots générés aléatoirement donc il faut obligatoirement supprimer ces propriétés.



```
<owl:NamedIndividual rdf:about="http://oaei.ontologymatching.org/2010/IIMBDATA/en/item1950811789911340571">
  <rdf:type rdf:resource="http://oaei.ontologymatching.org/2010/IIMBTBOX/Location"/>
  <IIMBTBOX:size rdf:datatype="http://www.w3.org/2001/XMLSchema#double">726.5</IIMBTBOX:size>
  <IIMBTBOX:name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Calgary</IIMBTBOX:name>
  <IIMBTBOX:article rdf:datatype="http://www.w3.org/2001/XMLSchema#string">0<pg&gt;lwgwawrkyk is hse aGgestcity Ointe jprinceF
XofX 8e2rtaRcaLn2a. c7 islhxcateHd in he BskoutGh YofFheovnce,d i R8rieinoff5othilsad ghn plns, raFAroximaEtl esct of7 front
rangwes t19anadMicng Rochkie.xaasayi tMtirdZrget civncsun?iipSaity bh pWbplBatioft,An CMnTad. s oZ tWhe 2a007 AcAvMc De2ss, gWlgg
&#39;st jpop2tI whaf 1,0eIr,4zK hLbmet6rplimsn pulatoZ5CMA)wra F1E,09,3 n20,pgaknA Gra2tecr Clgr! tde Lfth lcagest
cns2metcopolitin area?n h cononry. BeJcaGue itIA FolaSed30h0&amp;#65;&amp;#6540;kitltmFetrPes (S1#654xuB;&amp;#6540i;pmi)due out2h
if EdmHoUton, tatistianeine thfar9w popWltreJSi befwee tse cc2ie a thwe&quot;Cgaprgy-EdmotJo pChorri3o. GIst 2is lMr2gistn
Canaian mtrSopolitan drea nbetwsen Troto andIVantouvk. L
A rs0idFZ oWf Cn4lgaIE is6 Uknony aaryianU.
CmaEllaIy iwe1jRknwnasdesinti0on fobwntersprtagecHotokrs1m uiPth3 !aL nuXeofmajor m5nnti3nc resots ner ktlhe Zaiyf raEld1metrplPotn are
E9nolic cWtiviBtyJin CuklryJ s nmojsly otrd oX Athey SetrwoleuRm dustryowvr,7 rieultur 0toFurbism3,andt h-tec7hu nd3udshtHreJs2 ls
ntrbribute t 7hecity&#39;s an oic gotei.p COag9azH hLks Ballny ajor SnnHUL tivls hwph includxeX LC1xga0ry&lt;IR
</IIMBTBOX:article>
```

FIGURE 1 – Exemple d'article qui n'a pas de sens

En restant toujours sur l'analyse des attributs articles pour les textes qui ne présentent pas des anomalies, on a fixé une valeur seuil à partir de laquelle on décide si c'est un long texte qu'on doit supprimer ou le garder.

- si le nombre de mots total dans un article $>$ seuil : On supprime le texte.
 - si le nombre de mots total dans un article $<$ seuil : on garde le texte.
- Cette valeur seuil n'est pas fixée d'une manière aléatoire, Elle est égale à la moyenne des longueurs des textes présents dans une base étudiée.

Ces bases de tests contiennent aussi des modifications au niveau des formats champs de quelques attributs par exemple la date de naissance elle est parfois écrite sous la forme jour/mois/année ou sous la forme jour-mois-année. Aussi pour le champs gender il est parfois écrit "M" ou Male (comme le montre la figure2), "F" ou Female.

Tout cela s'inscrit dans la problématique vue en cours, qui est de pouvoir définir que 2 éléments sont similaires et de prouver cette logique permettant de dire que cette balise X est la même que la balise Y (Same AS)



```

<owl:NamedIndividual rdf:about="http://oaei.ontologymatching.org/2010/IIMBDATA/en/ken_hutch_hutchinson">
  <rdf:type rdf:resource="http://oaei.ontologymatching.org/2010/IIMBTBOX/Character"/>
  <IIMBTBOX:name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Ken &quot;Hutch&quot;&quot; Hu
  <IIMBTBOX:gender rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Male</IIMBTBOX:gender>
  <IIMBTBOX:acted_by rdf:resource="http://oaei.ontologymatching.org/2010/IIMBDATA/en/owen_wilson"/>
</owl:NamedIndividual>

<owl:NamedIndividual rdf:about="http://oaei.ontologymatching.org/2010/IIMBDATA/en/item20476684090">
  <rdf:type rdf:resource="http://oaei.ontologymatching.org/2010/IIMBTBOX/Actor"/>
  <IIMBTBOX:date_of_birth rdf:datatype="http://www.w3.org/2001/XMLSchema#string">8/23/70</IIMBT
  <IIMBTBOX:name rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Jay Mohr</IIMBTBOX:name
  <IIMBTBOX:gender rdf:datatype="http://www.w3.org/2001/XMLSchema#string">M</IIMBTBOX:gender>
  <IIMBTBOX:article rdf:datatype="http://www.w3.org/2001/XMLSchema#string">L&gt;Jy Mohr7 7(bAuZ

```

FIGURE 2 – Différence de format

Nous avons fait face à un autre souci qui est l'existence de champs ajoutés qui n'existent pas dans la base initiale. Pour résoudre ces anomalies on a utilisé l'ontologie large d'iimb comme référence et on a comparé la base de test en entrée (ontBig) par rapport à cette base (une ontology prise aléatoirement en entrée) pour pouvoir enlever les attributs qui ont été ajoutés, modifiés et qui ne respectent pas le format de base. On a aussi injecté les clés à la place des identifiant dans les balises parce qu'on a remarqué que pour une même entité elle n'a pas la même clé dans deux bases de test (subset) différentes.

2 Sakey

L'exploitation des liens d'identité entre les ressources RDF permet aux applications d'intégrer efficacement les données. Les clés peuvent être très utiles pour découvrir ces liens d'identité. Un ensemble de propriétés est considéré comme une clé lorsque ses valeurs identifient de manière unique des ressources (0-Almost Key). Cependant, ces clés ne sont généralement pas disponibles. Les approches qui tentent de découvrir automatiquement les clés peuvent facilement être dépassées par la taille des données et nécessitent des données propres.

Nous avons utilisé SAKey, un framework java qui détermine les combinaisons de clés dans les données RDF de manière efficace. Pour élaguer l'espace de recherche, SAKey exploite les caractéristiques des données qui sont détectées dynamiquement au cours du processus. De plus, notre approche peut découvrir des clés dans des ensembles de données où des données erronées ou des doublons existent (c'est-à-dire presque des clés). Il faut que nos données respectent le format NT. D'abord nous avons eu recours à `rdf2rdf` dans le but de convertir nos ontologies en fichier NT (composé de triples). `!java -jar rdf2rdf-1.0.1-2.3.1.jar ontBig.owl ontBig.nt` Puis nous utilisons Sakey de la même manière avec une commande bash dans notre code Python. L'approche

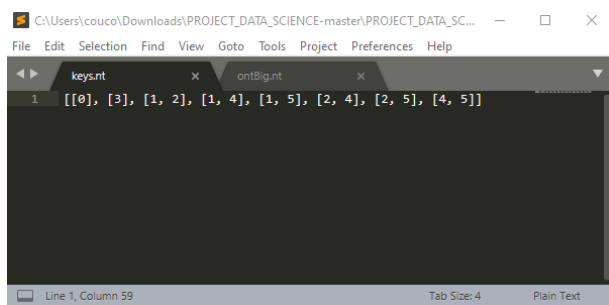


FIGURE 3 – Résultats de Sakey

a été évaluée sur différents ensembles de données synthétiques et réelles. Sa-key nous fournit en sortie des combinaisons de ressources et propriétés qui permettent de déterminer des 0 almost key et d'autres combinaisons approximatives. Les résultats montrent à la fois la pertinence de presque clés et l'efficacité de leur découverte.

3 Limes

LIMES (figure 4) est un framework de découverte de liens pour le Web Semantic ou Web 3.0. Il met en œuvre des approches efficaces en termes de temps pour la découverte de liens à grande échelle en fonction des caractéristiques des espaces métriques.



FIGURE 4 – Limes

Concernant son fonctionnement, nous avons appris en TP à se baser sur un format de fichier en XML qui permet la configuration des approches de découverte de ces liens. Nous avons commencé par sélectionner un encoding adapté à la lecture de nos fichiers, car lors de la récupération des informations de notre ontologie nous avons spécifié ce typage. Puis nous avons mis en place les namespaces de RDF, RDFS, OWL ainsi que celui d'IIMBTBOX.

Dans la balise endpoint nous avons spécifié notre source. Comme axe d'amélioration nous avons pensé à faire en sorte d'uploader nos fichiers NT directement dans un serveur pour que ce soit lu de manière automatique par nos XML.

Nous pouvons voir dans la figure 5 notre configuration du XML de limes et comme pour la conversion ou l'appel de Sakey, nous procédons de la même manière en appelant le fichier à travers une commande Bash en Linux.

Le plus important restera la property qui est l'élément de comparaison par excellence dans notre exemple nous prenons comme base le "name" .

```
<?xml version="1.0" encoding="UTF-8"?>
<!--Sample XML file generated by XMLSpy v2010 rel. 3 sp1 (http://www.altova.com)-->
<!DOCTYPE LIMES SYSTEM "limes.dtd">
<LIMES>
  <PREFIX>
    <NAMESPACE>http://www.w3.org/1999/02/22-rdf-syntax-ns#</NAMESPACE>
    <LABEL>rdf</LABEL>
  </PREFIX>
  <PREFIX>
    <NAMESPACE>http://www.w3.org/2000/01/rdf-schema#</NAMESPACE>
    <LABEL>rdfs</LABEL>
  </PREFIX>
  <PREFIX>
    <NAMESPACE>http://www.w3.org/2002/07/owl/#</NAMESPACE>
    <LABEL>owl</LABEL>
  </PREFIX>
  <PREFIX>
    <NAMESPACE>http://oaei.ontologymatching.org/2010/IIMBTBOX/</NAMESPACE>
    <LABEL>imbt</LABEL>
  </PREFIX>
  <SOURCE>
    <ID>ontBig1</ID>
    <ENDPOINT>C:\Users\couco\Downloads\LIMES.0.6.RC4-2\LIMES.0.6.RC4-2\Release_Examples\ontBig.nt</ENDPOINT>
    <VAR>?x</VAR>
    <PAGESIZE>5000</PAGESIZE>
    <RESTRICTION></RESTRICTION>
    <PROPERTY>imbt:name</PROPERTY>
    <TYPE>N-TRIPLES</TYPE>
  </SOURCE>
  <TARGET>
    <ID>output</ID>
    <ENDPOINT>C:\Users\couco\Downloads\LIMES.0.6.RC4-2\LIMES.0.6.RC4-2\Release_Examples\output_file.nt</ENDPOINT>
    <VAR>?y</VAR>
    <PAGESIZE>5000</PAGESIZE>
    <RESTRICTION></RESTRICTION>
    <PROPERTY>imbt:name</PROPERTY>
    <TYPE>N-TRIPLES</TYPE>
  </TARGET>
  <METRIC>jaro(x.imbt:name,y.imbt:name)</METRIC>
  <ACCEPTANCE>
    <THRESHOLD>0.9</THRESHOLD>
    <FILE>Projet_v1_big.nt</FILE>
    <RELATION>owl:sameAs</RELATION>
  </ACCEPTANCE>
  <REVIEW>
    <THRESHOLD>0.7</THRESHOLD>
    <FILE>Projet_v2_review_big.nt</FILE>
    <RELATION>owl:sameAs</RELATION>
  </REVIEW>
  <EXECUTION>Simple</EXECUTION>
  <OUTPUT>TTL</OUTPUT>
</LIMES>
```

FIGURE 5 – Configuration du xml de Limes

Concernant notre méthode de comparaison nous utilisons Jaro, pour calculer le taux de similarité comme vu sur la figure 6.

```
INFO [main] (PPJoinController.java:146) - Loading target data ...
INFO [main] (HybridCache.java:232) - Checking for file C:\Users\couco\Downloads\LIMES.0.6.RC4-2\LIMES.0.6.RC4-2\cache\1583074885.ser
INFO [main] (HybridCache.java:248) - No cached data found for output
INFO [main] (QueryModuleFactory.java:19) - Generating <n-triples> reader
INFO [main] (FileQueryModule.java:56) - RDF model read from C:\Users\couco\Downloads\LIMES.0.6.RC4-2\LIMES.0.6.RC4-2\Release_Examples\output_file.nt is of size 8678
INFO [main] (ModelRegistry.java:34) - Registry = [C:\Users\couco\Downloads\LIMES.0.6.RC4-2\LIMES.0.6.RC4-2\Release_Examples\ontBig.nt, C:\Users\couco\Downloads\LIMES.0.6.RC4-2\LIMES.0.6.RC4-2\Release_Examples\output_file.nt]
INFO [main] (SparqlQueryModule.java:93) - Properties are [imbt:name]
INFO [main] (SparqlQueryModule.java:139) - Query issued is
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl/#>
PREFIX imbt: <http://oaei.ontologymatching.org/2010/IIMBTBOX/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT DISTINCT ?y ?v0
WHERE {
  ?y imbt:name ?v0 .
}
INFO [main] (SparqlQueryModule.java:141) - Querying the endpoint.
INFO [main] (SparqlQueryModule.java:149) - Getting statements 0 to 5000
INFO [main] (SparqlQueryModule.java:149) - Getting statements 5000 to 10000
INFO [main] (SparqlQueryModule.java:248) - Retrieved 1233 triples and 1233 entities.
INFO [main] (SparqlQueryModule.java:249) - Retrieving statements took 0.014 second.
INFO [main] (HybridCache.java:179) - Serializing 1233 objects to C:\Users\couco\Downloads\LIMES.0.6.RC4-2\LIMES.0.6.RC4-2\cache\1583074885.ser
INFO [main] (PPJoinController.java:159) - Getting links ...
INFO [main] (JaroMapper.java:25) - Running JaroMapper
1649372 = 0.999879362573663
477723 = 0.2896043880499839
INFO [main] (PPJoinController.java:162) - Got links in 487ms.
INFO [main] (SerializerFactory.java:22) - Getting serializer with name TTL
INFO [main] (PPJoinController.java:172) - Using TtlSerializer to serialize
INFO [main] (PPJoinController.java:191) - Returned 1289 links above acceptance threshold.
INFO [main] (PPJoinController.java:192) - Returned 3912 links to review.
INFO [main] (PPJoinController.java:199) - Mapping carried out in 1.988 seconds
INFO [main] (PPJoinController.java:200) - Done.
```

FIGURE 6 – Résultats de Mapping

À la fin de cette tâche, nous avons un fichier qui est différent du fichier mis en entrée comme peut le montrer la figure 7.

```

1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix owl: <http://www.w3.org/2002/07/owl#> .
3 @prefix imbt: <http://oaei.ontologymatching.org/2010/IIMBTBOX/> .
4 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
5 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/vince_vaughn> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/vince_vaughn> .
6 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/comoros> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/comoros> .
7 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/port_vila> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/port_vila> .
8 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/glendale> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/glendale> .
9 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/winter_park> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/winter_park> .
10 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/stanley_tong> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/stanley_tong> .
11 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/idar-oberstein> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/idar-oberstein> .
12 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/brazzaville> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/brazzaville> .
13 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/james_dooan> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/james_dooan> .
14 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/sondra_pransky> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/sondra_pransky> .
15 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/berkeley_california> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/berkeley_california> .
16 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/sam_huntington> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/sam_huntington> .
17 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/laurence_olivier> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/laurence_olivier> .
18 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/ota_tokyo> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/ota_tokyo> .
19 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/ota_tokyo> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/kita_tokyo> .
20 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/mogadishu> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/mogadishu> .
21 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/judi_dench> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/judi_dench> .
22 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/shia_labeouf> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/shia_labeouf> .
23 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/erin_grant> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/erin_grant> .
24 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/leonard_nimoy> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/leonard_nimoy> .
25 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/phenixville> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/phenixville> .
26 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/landstuhl> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/landstuhl> .
27 <http://oaei.ontologymatching.org/2010/IIMBDATA/m/0bvbjw3> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/m/0bvbjw3> .
28 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/belize> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/belize> .
29 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/hikaru_sulu> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/hikaru_sulu> .
30 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/havana> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/havana> .
31 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/queens> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/queens> .
32 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/ving_rhames> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/ving_rhames> .
33 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/bert_convay> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/bert_convay> .
34 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/zoe_washburne> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/zoe_washburne> .
35 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/jean_simmons> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/jean_simmons> .
36 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/baltimore> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/baltimore> .
37 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/australia> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/australia> .
38 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/australia> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/m/09pkq0> .
39 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/andrew_mclaglen> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/andrew_mclaglen> .
40 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/dorking> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/dorking> .
41 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/spider-man> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/spider-man_3> .
42 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/spider-man> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/spider-man_3> .
43 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/lord_wessex> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/lord_wessex> .
44 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/eugene_levy> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/eugene_levy> .
45 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/barbados> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/barbados> .
46 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/darth_vader> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/darth_vader> .
47 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/memphis_tennessee> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/memphis_tennessee> .
48 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/stanley_donen> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/stanley_donen> .
49 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/windsor_ontario> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/windsor_ontario> .
50 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/shanghai_knights> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/shanghai_knights> .
51 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/kenley> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/kenley> .
52 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/dr_snaute> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/dr_snaute> .
53 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/scotland> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/scotland> .
54 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/charlie_yeung> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/charlie_yeung> .
55 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/islamabad> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/islamabad> .
56 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/dorothy_christy> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/dorothy_christy> .
57 <http://oaei.ontologymatching.org/2010/IIMBDATA/m/0bthgtc> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/m/0bthgtc> .
58 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/steve_forrest> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/steve_forrest> .
59 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/olney> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/olney> .
60 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/marc_platt_1913> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/marc_platt_1913> .
61 <http://oaei.ontologymatching.org/2010/IIMBDATA/en/alan_ladd> owl:sameAs <http://oaei.ontologymatching.org/2010/IIMBDATA/en/alan_ladd> .

```

FIGURE 7 – Résultats du Limes

Conclusion

Ce travail était très enrichissant mais on a aussi fait face à des difficultés pour comprendre le sujet, et savoir ce que l'on devait faire. SAkey est aussi très compliqué à manier

Mais on a appris à manipuler les ontologies ainsi que les bibliothèques java telle que Limes et SAkey, a exploité nos connaissances en web sémantique et en knowledge discovery in graph data pour atteindre ces résultats.

On a pu exploiter toutes nos connaissances sous une seule problématique ce qui permet de voir la force de ces outils et avoir un aperçu du Web 3.0