

Ejercicio Final Modulo Web de Datos

A partir de la información disponible en la LOD Cloud e información embebida en las páginas HTML se procederá a enriquecer las entidades que tenemos en nuestro dataset local (instancia1, instancia2, instancia3, instancia 4).

- 1) Crear un repositorio en Sesame "SocialNetwork"
- 2) Añadir el fichero "Entidades.n3" (mostrado más abajo) al repositorio "SocialNetwork".
- 3) Para la instancia 2 se procederá a enriquecer la información con información embebida en páginas HTML. Partiendo de la información que se habrá recopilado en el archivo "manuChao.n3", proceder a volcarlo en el repositorio de SocialNetwork y hacer una query que liste los "MusicRecording".
- 4) Partiendo del ejercicio 5_1, y según la lista de datasets sugeridos (ver más abajo), se determinará que propiedades nos interesa incorporar a nuestras entidades.
- 5) Para la instancia1, instancia3 e instancia 4 se procederá a enriquecer la información con información de los datasets previamente recomendados de LOD Cloud. La información se recuperará programáticamente sirviéndose de fichero "enquierer.py" que hay en la carpeta de Ejercicio Final.
 - a. En él se muestra los pasos inconclusos para la recuperación de información de dbpedia para enriquecer la instancia1.
 - b. Para proceder a enriquecer la instancia3, e instancia4 se procederá a elaborar las SPARQL queries de los métodos "getLinkedmdbResource" y "getWebenemasunoResource".
- 6) Para superar la prueba a nivel básico bastará con imprimir los recursos de los repositorios remotos que coinciden con el label de mis instancias locales, esto es:
 - a. Para el caso de dbpedia y tal y como está definido, se obtiene el recurso "http://dbpedia.org/resource/Alicia_Keys" para la instancia 1.
 - b. Para el caso de Linkedmdb, se obtendrá aquella instancia (o instancias) cuyo nombre coincida con el (los) label de la instancia 3.
 - c. Para el caso de WebN+1, se obtendrá aquella instancia (o instancias) cuyo nombre coincida con el (los) label de la instancia 4.
- 7) (Opcional – Nivel Medio). Se procederá a obtener alguna información extra (alguna propiedad) de los recursos obtenidos remotos.
- 8) (Opcional – Nivel Alto). Insertar las tripletas obtenidas con el enquierer en el repositorio SocialNetwork:
 - a. Volcar las tripletas a un fichero y volcarlas al repositorio SocialNetwork via el workbench
 - b. O insertarlas directamente en el repositorio mediante la librería SPARQLWrapper.

Fecha de Entrega

30 de Junio 2016

Outcomes

- Enquirer.py completado
- Fichero txt incluyendo la query SPARQL para la instancia 4
- Opcional: Enquirer_con_volcado.py conteniendo el código del volcado. Y fichero socialNetwork_enriched.n3 conteniendo una exportación (a través del workbench) del repositorio.

Fichero de entidades (Entities.n3)

```
# Default graph
```

```
@prefix sn: <http://ciff.curso2015/ontologies/owl/socialNetwork#> .
```

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
```

```
sn:instancia1 a sn:Person;
```

```
    rdfs:label "Alicia Keys"@en .
```

```
sn:instancia2 a sn:Person;
```

```
    rdfs:label "Manu Chao"@en-gb .
```

```
sn:instancia3 a sn:Film;
```

```
    rdfs:label "Batman" ;
```

```
    rdfs:label "Batman"@es ;
```

```
    rdfs:label "Batman"@en .
```

```
sn:instancia4 a sn:Guide;
```

```
    rdfs:label "Un vino cosmopolita" ;
```

```
    rdfs:label "Un vino cosmopolita"@es ;
```

```
    rdfs:label "Un vino cosmopolita"@en .
```

Dataset Recomendados

- BBPedia_ES (<http://es.dbpedia.org/sparql>)
 - o http://www.dbpedia.org/page/Alicia_Keys
- MusicBrainz (<http://dbtune.org/musicbrainz/sparql>)
 - o <http://dbtune.org/musicbrainz/resource/artist/704acdbb-1415-4782-b0b6-0596b8c55e46>
 - o <http://dbtune.org/musicbrainz/page/artist/a992aada-7108-455d-9747-0b7b6a089e8d>
- Web n+1 el viajero: (<http://webenemasuno.linkeddata.es/sparql>)
 - o http://webenemasuno.linkeddata.es/page/elviajero/resource/Guide/20060513ELPVIALBV_5.TES
- MDB: (<http://data.linkedmdb.org/sparql>)
 - o <http://data.linkedmdb.org/page/film/300>