

Sistema de Recuperación Semántico

Jaime Ruiz-Borau Vizárraga
546751

Alberto Sabater Bailón
546297

1. El sistema de recuperación semántico

1.1. Cambios realizados sobre el RDFS original

Supresión de la etiqueta 'subclass of': En el modelo que se diseñó inicialmente, en las clases Documento y Persona se especificaba concretamente que eran subclase de **Resource**. Dado que esta especificación resultaba redundante, se optó por suprimirla.

Inclusión de una etiqueta 'Identificador': Con el motivo de devolver el nombre del archivo completo del documento en cuestión, se incluyó una etiqueta en el modelo RDFS para el Documento que incluía este nombre. Dicha etiqueta es **Identificador** (ver fichero rdfs.xml para verla en detalle).

Inclusión de una etiqueta 'Idioma': Esta etiqueta no era estrictamente necesaria para las búsquedas SPARQL semánticas, pero se incluyó para hacer un modelo rdfs más completo.

1.2. Consultas SPARQL

Necesidad de información 13-2: Me interesan los trabajos en relación con el mundo de la musica y els onido cuyo autor o director se llame Javier.

```
PREFIX recinfo: <http://www.recInfo.com/>
PREFIX skos: <http://www.w3.org/TR/skos-primer/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT DISTINCT ?id WHERE {
    ?doc recinfo:Identificador ?id.
    ?doc recinfo:hasConcept ?concept.
    { ?concept recinfo:conceptName "musica". }
    UNION
    { ?concept skos:broader ?sub .
      ?sub recinfo:conceptName "musica" }
    ?doc recinfo:Autor ?autor.
    ?autor recinfo:Nombre ?name
    FILTER regex (?name, ".*Javier.*", "i")
}
```

Listing 1: Consulta SPARQL para la necesidad de información 13-2

Necesidad de información 02-4: Estoy interesado en el periodo histórico de la Guerra de Independencia (1808-1814). Me gustaría saber qué trabajos o artículos hay relacionados con este hecho englobados en la historia de España.

```
PREFIX recinfo: <http://www.recInfo.com/>
PREFIX skos: <http://www.w3.org/TR/skos-primer/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT DISTINCT ?id ?concept where {
    ?doc recinfo:hasConcept ?concept1 .
    ?doc recinfo:hasConcept ?concept2 .
    ?doc recinfo:hasConcept ?concept3 .
    ?doc recinfo:Identificador ?id .
    { ?concept1 recinfo:conceptName "guerra" }
    UNION
    { ?concept1 skos:broader ?sub .
      ?sub recinfo:conceptName "guerra" }
    { ?concept3 recinfo:conceptName "espana" }
    UNION
    { ?concept3 skos:broader ?sub .
      ?sub recinfo:conceptName "espana" }
}
```

Listing 2: Consulta SPARQL para la necesidad de información 02-4

Necesidad de información 09-3: Tesis que traten de energías renovables, en el período de 2010 a 2015.

```
PREFIX recinfo: <http://www.recInfo.com/>
PREFIX skos: <http://www.w3.org/TR/skos-primer/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?id WHERE {
    ?doc recinfo:Identificador ?id.
    ?doc recinfo:hasConcept ?concept.
    { ?concept recinfo:conceptName "energias". }
    UNION
    { ?concept skos:broader ?sub .
      ?sub recinfo:conceptName "energias" }
    ?doc recinfo:Fecha ?year
    FILTER (xsd:integer(?year) > 2010
            && xsd:integer(?year) < 2015)
}
```

Listing 3: Consulta SPARQL para la necesidad de información 09-3

Necesidad de información 07-2: Quiero documentos sobre desarrollo de videojuegos o diseño de personajes en los últimos 5 años.

```
PREFIX recinfo: <http://www.recInfo.com/>
PREFIX skos: <http://www.w3.org/TR/skos-primer/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?id WHERE {
    ?doc recinfo:Identificador ?id.
    ?doc recinfo:hasConcept ?concept.
    { ?concept recinfo:conceptName "videojuegos". }
    UNION
    { ?concept skos:broader ?sub .
      ?sub recinfo:conceptName "videojuegos". }
    ?doc recinfo:Fecha ?year
    FILTER (xsd:integer(?year) > 2010
            && xsd:integer(?year) < 2016)
}
```

Listing 4: Consulta SPARQL para la necesidad de información 07-2

Necesidad de información 05-5: Me gustaría encontrar construcciones arquitectónicas situadas en España con elementos decorativos que pertenezcan tanto a la Edad Media como a la época gótica y cuyo estado de conservación sea óptimo.

```
PREFIX recinfo: <http://www.recInfo.com/>
PREFIX skos: <http://www.w3.org/TR/skos-primer/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?id WHERE {
    ?doc recinfo:Identificador ?id.
    ?doc recinfo:hasConcept ?concept.
    ?doc recinfo:hasConcept ?concept2.
    ?doc recinfo:hasConcept ?concept3.
    { ?concept skos:broader ?sub .
      ?sub recinfo:conceptName "arquitectura" }
    UNION
    { ?concept2 recinfo:conceptName "edad_media". }
    UNION
    { ?concept3 recinfo:conceptName "gotica". }
}
```

Listing 5: Consulta SPARQL para la necesidad de información 05-5

2. Resultados y comparativa con el sistema de recuperación tradicional

2.1. Resultados

Los resultados devueltos por el sistema de evaluación desarrollado en entregas anteriores han sido los siguientes:

```
Information need: 13-2
Precision: 0,182
Recall: 0,267
F1 Score: 0,216
Precision@10: 0,300
Mean Average Precision: 0,283
*****
Information need: 02-4
Precision: 0,094
Recall: 0,750
F1 Score: 0,167
Precision@10: 0,100
Mean Average Precision: 0,115
*****
Information need: 09-3
Precision: 0,280
Recall: 0,292
F1 Score: 0,286
Precision@10: 0,100
Mean Average Precision: 0,262
*****
Information need: 07-2
Precision: 0,240
Recall: 0,444
F1 Score: 0,312
Precision@10: 0,300
Mean Average Precision: 0,328
*****
Information need: 05-5
Precision: 0,000
Recall: 0,000
F1 Score: NaN
Precision@10: 0,000
Mean Average Precision: NaN
*****
TOTAL
Precision: 0,159
Recall: 0,351
F1 Score: 0,219
Precision@10: 0,160
Mean Average Precision: NaN
```

Listing 6: Resultados

2.2. Comparativa de resultados

Los resultados de este recuperador semántico son sustancialmente peores que los del recuperador tradicional implementado anteriormente por diversos motivos.

En primer lugar, la ausencia de un ranking dificulta el acceso a los documentos más relevantes dentro de el conjunto de documentos devuelto. Esto se ve reflejado en la evaluación del sistema, especialmente en la $\text{prec}@10$ y Mean Average Precision.

Por otro lado la construcción de tesauros permite filtrar las búsquedas por familias de palabras (sinonimia, palabras relacionadas, etc.), aunque algunas veces su uso deriva en la devolución de resultados no relevantes.