SISTEMA DE RECUPERACIÓN SEMÁNTICO

## Recuperación de Información – Francisco Ferraz y Guillermo Cruz

# Resumen

En el desarrollo de las prácticas y el trabajo hemos implementado un sistema de recuperación semántico utilizando RDFS, SKOS, OWL, Jena y SPARQL y lo hemos comparado con el sistema de recuperación tradicional de la anterior entrega.

# Desarrollo

## RDFS

Partiendo del modelo RDFS que se creó para la tutoría de diciembre, se mejoraron algunos detalles:

1. Se asignaron dos subclases a los Documentos, TFG y TFM.
2. Se creó una clase Persona.
3. Se guardaron los subjects como Concepts de SKOS.
4. Se añadieron las propiedades de Organización e Idioma.

## SKOS

Se ha creado un fichero *skos.xml* en el que se desarrollan los distintos temas más relevantes para la creación del grafo RDF según las necesidades de información con las que se opera. Todos ellos han sido ampliados con sinónimos de forma manual de cara a obtener la máxima precisión en los documentos.

## SemanticGenerator

//Rellenar

## SemanticSearcher

Para realizar las búsquedas en SPARQL se ha implementado un módulo que hace uso de las librerías *ModelFactory* de Jena para trabajar con el grafo RDF, y un *HashMap* para almacenar las necesidades de información.

Se itera sobre este mapa para realizar cada una de las consultas sobre el grafo RDF y almacenar los documentos resultantes en un fichero de texto. El formato de almacenamiento es, como se indica en el enunciado, del estilo:

02-2 oai\_zaguan.unizar.es\_5460.xml

02-2 oai\_zaguan.unizar.es\_6453.xml

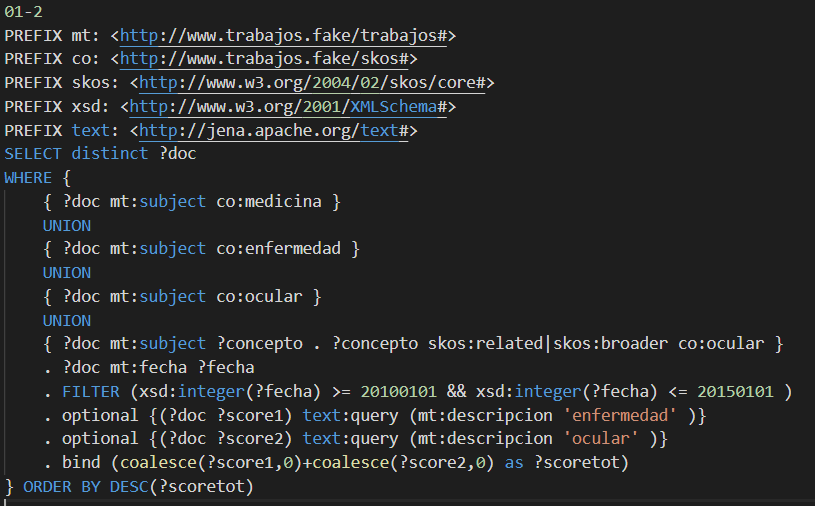
...

Además, los documentos resultantes están ordenados por su ranking en la consulta, utilizando el método que se explica a continuación.

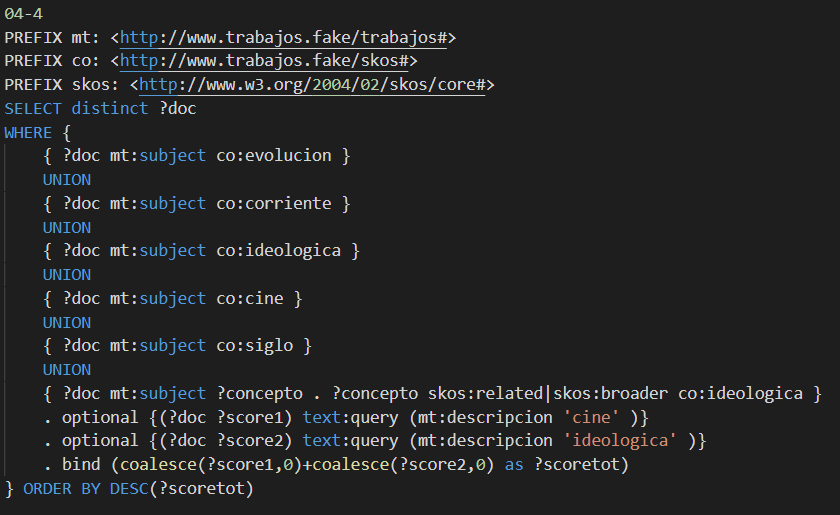
## Consultas SPARQL

Se han creado 5 consultas SPARQL, una por cada necesidad de información, que hacen referencia a los elementos que componen nuestro modelo RDF. En ellas se han intentado recuperar todos los documentos que puedan ser importantes para las consultas, sin dejar de lado las distintas condiciones requeridas en cada necesidad de información. Además, se han añadido índices textuales a todas ellas que apuntan a la descripción de los documentos para clasificarlos en función de su relevancia. Se utiliza la misma técnica que en la práctica 6, una ordenación en base a queries de texto.

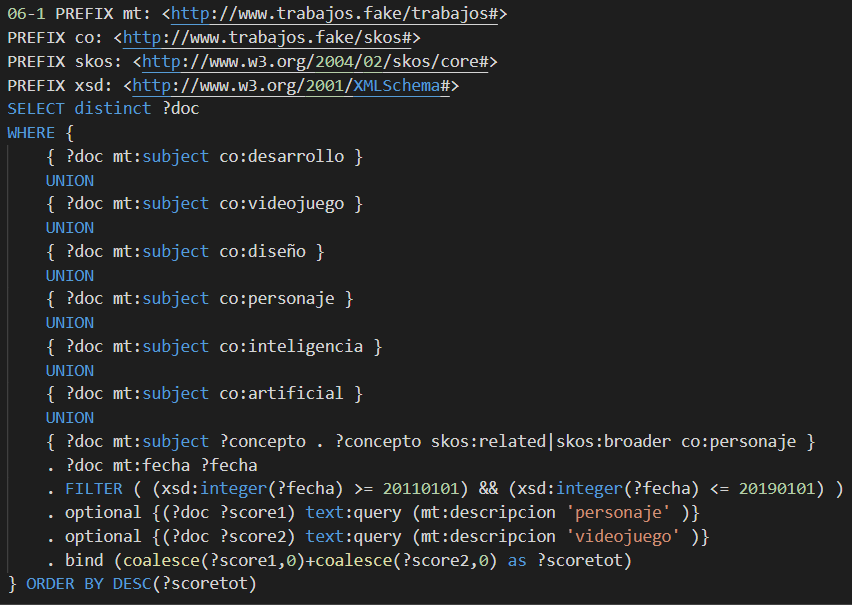
1. **Me gustaría saber qué estudios y avances se depositaron, preferentemente del 2010 al 2015, en la campo de la medicina referentes a las enfermedades oculares**



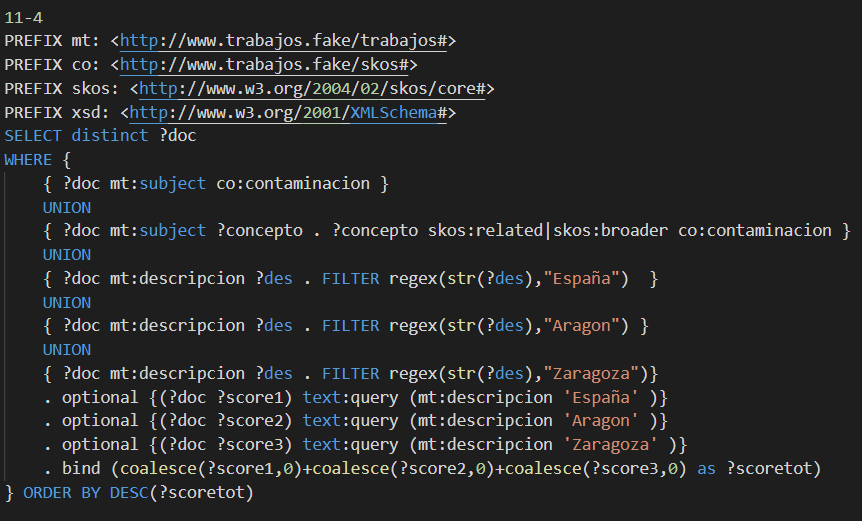
1. **Información acerca de la evolución de las corrientes ideológicas reflejadas en el cine en la segunda mitad del siglo XX**



1. **Busco documentos sobre desarrollo de videojuegos o diseño de personajes que incluyan técnicas de inteligencia artificial en los últimos 8 años**



1. **¿Qué tesis existen que hablen de contaminación en España, Aragón o Zaragoza?**



1. **Busco Trabajos de Fin de Grado dirigidos por un profesor conocido. Sé que se llama Javier, que los trabajos son sobre informática y que se han publicado a partir de 2012**

