

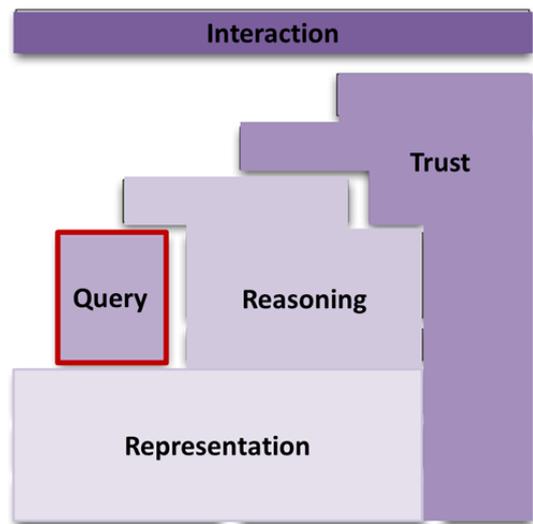
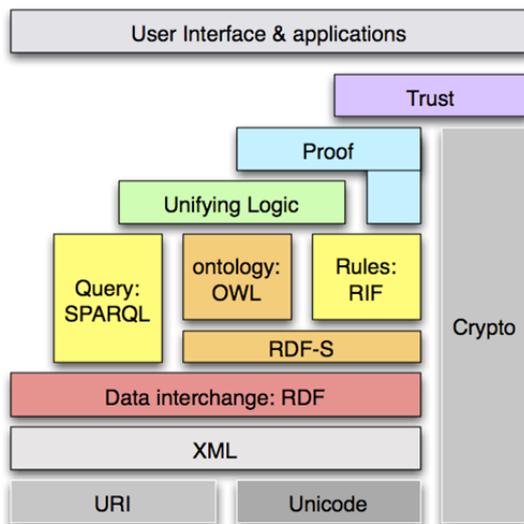
FUNDAMENTOS DE LA WEB SEMANTICA – 2023 UNIDAD 5: SPARQL

Edelweis Rohrer InCo- Fing - Universidad de la República

Qué es SPARQL?

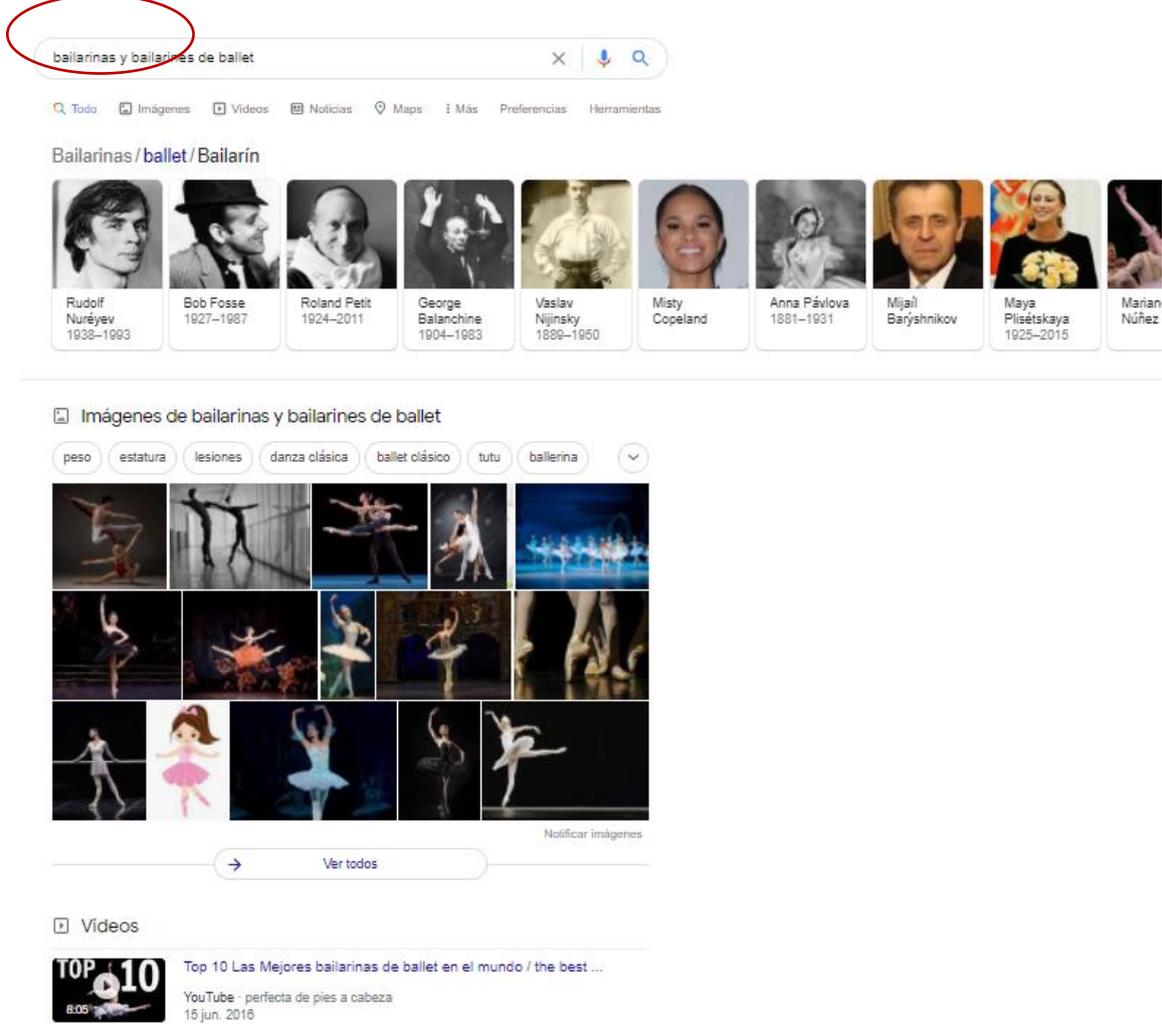
SPARQL en la arquitectura de la Web Semántica

SPARQL es un lenguaje de consulta de grafos RDF. Como se ilustra en la siguiente figura que muestra la arquitectura de la Web Semántica, por encima del bloque de representación en la parte inferior, encontramos el bloque de razonamiento (que comenzamos a abordar en la unidad anterior con el lenguaje RDF) en el que destaca la capacidad de inferir nuevo conocimiento (con RDF Schema y más aún con lenguajes de ontologías que estudiaremos en las unidades siguientes). Al mismo nivel tenemos el bloque de **consulta**, que es lo que tratamos en esta unidad.



Por qué se necesita un lenguaje de consulta de grafos?

Si buscamos en Google “bailarinas y bailarines ballet”, nos devuelve lo que aparece en la figura siguiente:



Los motores de búsqueda devuelven un conjunto de páginas que contienen, con mayor o menor precisión, parte de la información que buscamos. Pero si queremos tener una lista de bailarinas y bailarines de ballet en todo el mundo, hoy en día sabemos que la descripción de la mayoría de ellas va a estar publicada en alguna base de datos RDF, por ejemplo DBPEDIA.

Con los modelos de la web semántica, como RDF y OWL (que veremos más adelante) podemos describir un conjunto de recursos, en este caso correspondientes al conjunto de bailarinas y bailarines de ballet, y podemos inferir nuevo conocimiento, pero esto no es suficiente en la práctica cuando se quiere recuperar un conjunto de recursos que cumplen determinadas condiciones.

En el ejemplo anterior, si un humano o agente automático necesita obtener una lista de las bailarinas de ballet, las bases RDF como la DBPEDIA nos van a ayudar a encontrarlas, ya que además de tener los datos estructurados como grafos RDF, los recursos están descriptos usando propiedades de vocabularios conocidos y consensuados, y se usan constructores RDF Schema que clasifican los recursos (constructores `rdfs:Class`, `rdfs:subClassOf`, `rdfs:range`, `rdfs:domain`, `rdfs:subPropertyOf`).

Por ejemplo, podemos comenzar por tratar de obtener todos los recursos clasificados en la clase *Artist* del vocabulario <http://dbpedia.org/ontology/> que tienen el valor "female" en para la propiedad *gender* del vocabulario <http://xmlns.com/foaf/0.1/>. En la DBPEDIA vamos a encontrar

tanto a Gloria Gaynor (http://dbpedia.org/resource/Gloria_Gaynor) como a María Noel Riccetto (http://dbpedia.org/resource/Mar%C3%ADa_Noel_Riccetto). Se puede entonces agregar condiciones, como por ejemplo filtrar los recursos que en la propiedad *field* del vocabulario <http://dbpedia.org/ontology/> tomen el valor <http://dbpedia.org/resource/Ballet>, o que son de tipo *BalletDancer109834699* o *Ballerina109834592* del vocabulario <http://dbpedia.org/class/yago/>. Imponiendo estas condiciones podemos realizar la siguiente consulta SPARQL en la base DBPEDIA:

```
select distinct ?bailarina ?nombre
where
{
  ?bailarina rdf:type dbo:Artist.
  OPTIONAL {?bailarina dbo:birthName ?nombre. }
  {
    {?bailarina dbo:field dbr:Ballet} UNION
    {?bailarina rdf:type yago:BalletDancer109834699} UNION
    {?bailarina rdf:type yago:Ballerina109834592}
  }
}
```

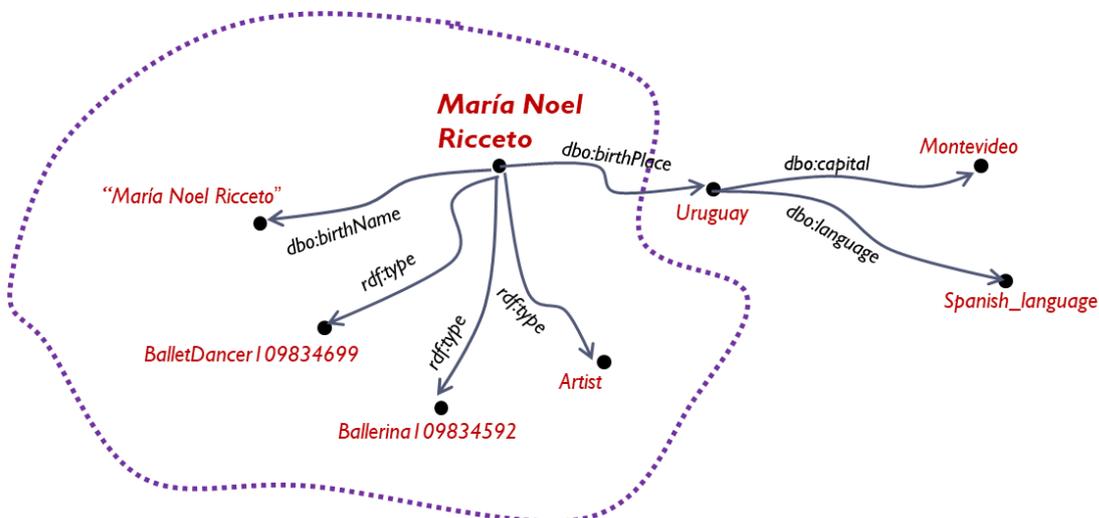
Los prefijos dbo:, dbr: y yago: representan los vocabularios <http://dbpedia.org/ontology/>, <http://dbpedia.org/resource/> y <http://dbpedia.org/class/yago/>.

La consulta anterior devuelve un conjunto de bailarines y bailarinas como los que se muestran en la siguiente figura.

bailarina	nombre
http://dbpedia.org/resource/Jeffrey_Golladay	"Jeffrey Golladay"@en
http://dbpedia.org/resource/Arron_Scott	"Arron Scott"@en
http://dbpedia.org/resource/Zaira_Cosico	"Zaira Cosico"@en
http://dbpedia.org/resource/Keith_McMillan	"Keith Lionel McMillan"@en
http://dbpedia.org/resource/Lisa_Macuja-Elizalde	"Lisa Teresita Pacheco Macuja"@en
http://dbpedia.org/resource/Igor_Yebra	"Igor Yebra Iglesias"@en
http://dbpedia.org/resource/Jennifer_Olayvar	"Jennifer Rose Olayvar"@en
http://dbpedia.org/resource/Georgette_Sanchez	"Georgina Johanna Garcia Sanchez"@en
http://dbpedia.org/resource/María_Noel_Riccetto	"Maria Noel Riccetto"@en
http://dbpedia.org/resource/Yvonne_Chouteau	"Myra Yvonne Chouteau"@en
http://dbpedia.org/resource/Luciana_Savignano	
http://dbpedia.org/resource/Michele_Wiles	
http://dbpedia.org/resource/Carolyn_George	
http://dbpedia.org/resource/Christopher_Wheeldon	
http://dbpedia.org/resource/Christopher_d'Amboise	
http://dbpedia.org/resource/Gemma_Bond	
http://dbpedia.org/resource/Sergei_Vikharev	
http://dbpedia.org/resource/Peter_Quanz	
http://dbpedia.org/resource/Esteban_Berlanga	"Esteban Berlanga"@en
http://dbpedia.org/resource/Pyotr_Gusev	"Pyotr Andreyevich Gusev"@en
http://dbpedia.org/resource/Dominic_Antonucci	"Dominic Antonucci"@en
http://dbpedia.org/resource/Jane_Gurnett	"Jane E. Gurnett"@en
http://dbpedia.org/resource/Sabrina_Matthews	"Sabrina Matthews"@en
http://dbpedia.org/resource/Adabel_Guerrero	"Adabel Anahí Guerrero Melachenko"@en
http://dbpedia.org/resource/Viper_(actress)	"Stephanie Green"@en
http://dbpedia.org/resource/Gregori_Lukas	
http://dbpedia.org/resource/Oh_Land	

El lenguaje SPARQL nos permite obtener directamente un subgrafo con los datos que buscamos, aplicando condiciones en el grafo donde se encuentra la información publicada en formato RDF. También nos permite combinar grafos de diferentes bases RDF, para así obtener información más completa y precisa, integrando diferentes fuentes de información.

La siguiente figura muestra el subgrafo que se obtiene al formular la consulta SPARQL presentada anteriormente, considerando solamente el recurso María Noel Riccetto.



SPARQL es un **estándar W3C** desde 2008, dicha especificación comprende:

- Sintaxis y semántica del lenguaje de consulta
- Protocolo de transmisión de consultas y resultados hacia y desde las bases RDF que ofrecen servicios (**endpoints**) que resuelven las consultas SPARQL.

En 2013 se extiende el lenguaje con **nuevos constructores** (subconsultas, agregaciones, entre otros) y se agrega la funcionalidad de **consultas federadas**, que permite formular **consultas distribuidas** en diferentes endpoints, **integrando datos de diferentes bases RDF**.

Existen múltiples servicios de consulta o *endpoints*, uno de ellos es GraphDB que además de permitir publicar y consultar bases de datos RDF, provee un endpoint SPARQL para formular consultas a los grafos publicados, filtrando los datos por diferentes condiciones.

DBPEDIA y WIKIDATA, además de disponibilizar vocabularios e información relevante (por ejemplo, de personalidades conocidas) en formato RDF, proveen endpoints SPARQL, que permiten combinar diferentes grafos y obtener datos que requerirían realizar múltiples consultas a través de un browser. Los resultados de las consultas pueden serializarse en diferentes formatos (HTML, RDF/XML, Turtle entre otros).