

RDF Store Benchmark

Esteban Sota

RIAM Intelearning Lab – GNOSS

estebansota@gnoss.com



HĒRCULES



Hércules. RDF Store Benchmark

- ☐ RDF Store y SPARQL Endpoint
- ☐ RDF Store Benchmark
- ☐ Casos prácticos



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Una manera de hacer Europa

Hércules ASIO. RDF Store



Hércules ASIO. RDF Store

El RDF Store es el componente de la arquitectura que **almacena el grafo de conocimiento de ASIO en formato RDF**. Dentro de ASIO actúa como un servidor de datos que responde a consultas SPARQL, por lo que también puede ser nombrado como **Servidor RDF o Servidor SPARQL**.

El RDF Store proporciona el servicio de datos y consultas al Linked Data Server y al SPARQL Endpoint.

En el desarrollo de ASIO se ha realizado un *benchmark* para que cada universidad pueda seleccionar la solución que más le convenga. Uno de los requisitos obligatorios definido por el *benchmark* es que debe **cumplir el estándar SPARQL 1.1**, en particular lo referente a actualizaciones de datos.

Hércules ASIO. Reutilización de los datos

ASIO proporciona dos interfaces de reutilización de los datos del grafo de conocimiento modelados por la Red de Ontologías Hércules ROH:

- ❑ **Linked Data Server.** Es el componente que permite la publicación de los datos cargados en el RDF Store como datos abiertos y enlazados (linked open data).
- ❑ **SPARQL Endpoint.** Permite a usuarios y administradores consultar los datos del grafo de conocimiento almacenados en el RDF Store, mediante el lenguaje y protocolo de consultas SPARQL.

Reutilización: Conocer ROH + SPARQL Queries + Linked Data Server

Hércules ASIO. Arquitectura Semántica. SPARQL Endpoint

Además de Linked Data Server, ASIO cuenta con un punto SPARQL (*endpoint*) que permite a usuarios y administradores consultar los datos del grafo de conocimiento almacenados en el RDF Store.

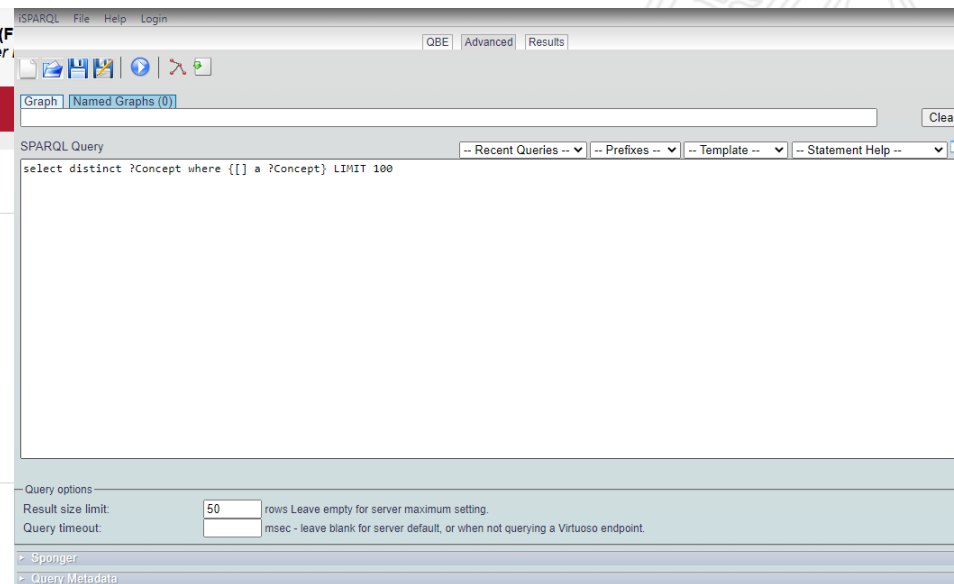
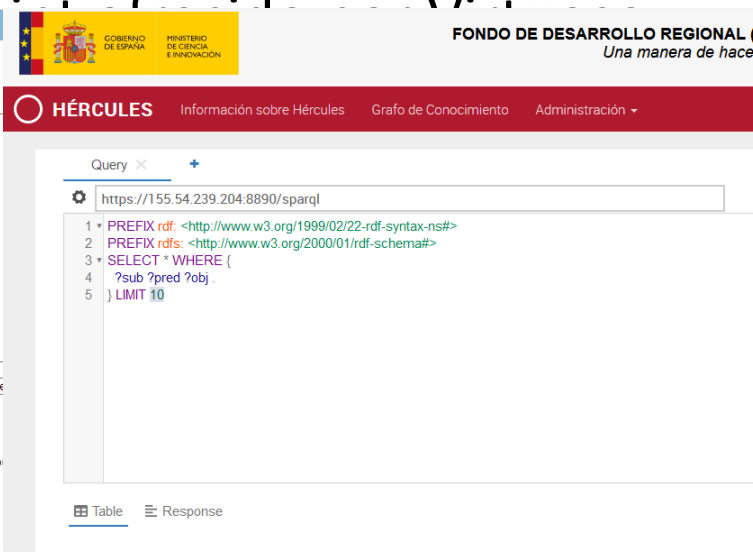
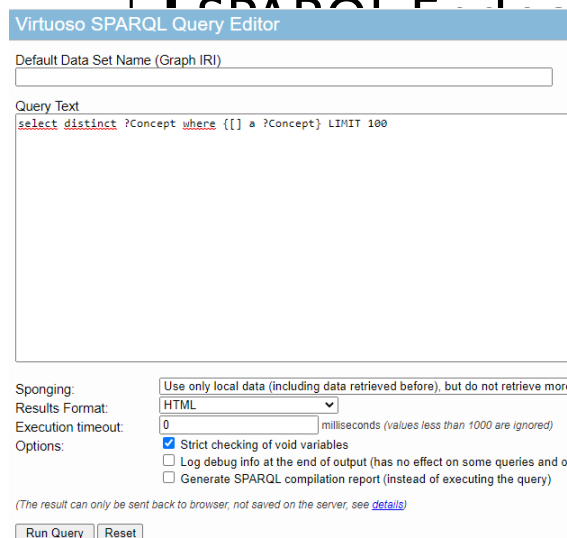
- ❑ **SPARQL**: SPARQL Protocol and RDF Query Language.
- ❑ [SPARQL Versión 1.1 de 2013](#) (la versión 1 es de 2008).
- ❑ **Protocolo de consulta y lenguaje de interrogación** para grafos RDF, normalizado inicialmente por el RDF Data Access Working Group del World Wide Web Consortium (W3C) y por el SPARQL Working Group en su version 1.1
- ❑ *Endpoint* de **sólo lectura para usuarios públicos**, sin acceso a los datos privados.

Reutilización: Conocer ROH + SPARQL Queries + Linked Data Server

Hércules. SPARQL Endpoint

ASIO cuenta con un SPARQL Endpoint de acceso público, que:

- ☐ Permite consultas de lectura sobre los datos públicos del SGI.
- ☐ Se pueden configurar limitaciones de uso para evitar saturaciones y ataques.
- ☐ Disponible en varias URLs (pendiente SSL):



Hércules. SPARQL Endpoint

ASIO cuenta con un SPARQL Endpoint de acceso público, que es accesible:

- ☐ Por personas, en un interfaz web (como hemos visto en la diapo anterior)
- ☐ Por máquinas, mediante peticiones HTTP de tipo GET o POST. Por ejemplo, con una petición GET con un comando cURL:

```
curl
```

```
"http://server/sparql?query=SELECT%20*%20WHERE%20%7B%3Fs%20%3Fp%20%3Fo%7D%20LIMIT%208"
```

La consulta enviada es:

```
SELECT * WHERE {?s ?p ?o} LIMIT 8
```


SPARQL. Tutorial de SPARQL

Tutorial de SPARQL recomendado:

<https://docs.data.world/tutorials/sparql/introduction.html>



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Una manera de hacer Europa

Hércules ASIO. RDF Store Benchmark



Hércules. RDF Store. Características comunes

Los RDF Store comparten algunas características generales comunes:

- ☐ Orientados a lectura de datos y no tanto a escritura.
- ☐ Carga de datos masiva muy eficaz.
- ☐ Limitaciones en la transaccionalidad. Con matices en cada producto.
- ☐ (casi) Todo en memoria.
- ☐ Necesario un ajuste fino (tuning) para gestionar volúmenes de datos grandes.
- ☐ Rendimiento geométricamente decreciente en agregaciones de datos.

Hércules. RDF Store Benchmark

[TripleStore Benchmark deliverable report](#) (Mathieu D'Aquin & Alessandro Adamou):

- ☐ Functionalities.
- ☐ Performance and scalability.
- ☐ Management and maintenance.

Hércules. RDF Store Benchmark

[TripleStore Benchmark deliverable report](#) (Mathieu D'Aquin & Alessandro Adamou):

- ☐ Functionalities.
 - ☐ Base handling of RDF and SPARQL
 - ☐ Extensions
 - ☐ Security
- ☐ Performance and scalability.
- ☐ Management and maintenance.

Hércules. RDF Store Benchmark

[TripleStore Benchmark deliverable report](#) (Mathieu D'Aquin & Alessandro Adamou):

- ☐ Functionalities.
- ☐ Performance and scalability.
 - ☐ Query response time.
 - ☐ Loading and update time.
 - ☐ Memory footprint.
 - ☐ Robustness and Scaling.
- ☐ Management and maintenance.

Hércules. RDF Store Benchmark

TripleStore Benchmark deliverable report (Mathieu D'Aquin & Alessandro Adamou):

- ☐ Functionalities.
- ☐ Performance and scalability.
- ☐ Management and maintenance.
 - ☐ Costs and rights.
 - ☐ Ease of deployment.
 - ☐ Administration, updates and maintenance.

Hércules ASIO. RDF Store Benchmark

RDF Store analizados:

- ☐ Apache Jena.
- ☐ Virtuoso.
- ☐ AllegroGraph.
- ☐ Stardog.
- ☐ Corese.
- ☐ Blazegraph.
- ☐ Amazon Neptune.
- ☐ Neo4J.

Hércules ASIO. RDF Store Benchmark

Benchmark accesible en: <https://herc-as-front-desata.atica.um.es/benchmark>

HERCULES BENCHMARK

Stores Benchmark User Guide About ASIO

Hercules triple store ranking

1	Apache Jena
2	Virtuoso
3	AllegroGraph
4	Stardog
5	Corese
6	Blazegraph
7	Amazon Neptune
8	Neo4J

Not assessed

Assessment of the following triple stores is underway.

1	4store
2	AnzoGraph

HERCULES BENCHMARK

Stores Benchmark User Guide About ASIO

Hercules triple store benchmarking framework

Click on a category name to see its details

Functionalities

Criteria concerning the satisfaction of typical functional requirements

Linked Data Platform (LDP)

Linked Data Platform (LDP) defines a set of rules for HTTP architecture for read-write Linked Data on the web. A score of 5 indicates full compliance with the specification. <https://www.w3.org/TR/ldp/>

CRITERION	DESCRIPTION
Support for the management of LDP RDF resources	The system support the management of LDP RDF resources according to the specification.
Support for the management of LDP Non-RDF resources	The system support the management of LDP Non-RDF resources according to the specification.
Support for HTTP read methods (GET, HEAD and OPTIONS)	The system support the management of LDP Non-RDF resources according to the specification.
Support for HTTP write methods (POST, PUT, DELETE and PATCH)	The system support the management of LDP Non-RDF resources according to the specification.

Base handling of RDF and SPARQL

SPARQL update compliance

Provide SPARQL Update endpoints compliant with the SPARQL Update specification. Other approaches can be used to update the data.

Triple store ranking for this criterion

TRIPLE STORE	SCORE	JUSTIFICATION
Amazon Neptune	5	SPARQL update provided
Apache Jena	5	SPARQL update compliance through SPARQL Update
Blazegraph	5	Compliant with SPARQL Update
Stardog	5	Support SPARQL update query and endpoints
Virtuoso	5	Appears compatible, with example queries
Corese	4	Homepage mentions SPARQL update
AllegroGraph	2	Update queries available, but no endpoints
Neo4J	0	No mention of SPARQL update in documentation

Virtuoso Universal Server

<http://data.datascienceinstitute.ie/software/virtuoso>

Click on a category name to see its details

Functionalities

Criteria concerning the satisfaction of typical functional requirements of triple stores

3.16

Management and maintenance

Criteria concerning the manageability of triple stores over time

3.27

Performance and scalability

Criteria concerning the satisfaction of typical performance requirements of triple stores

3.89

Query response time

CRITERION	SCORE	JUSTIFICATION	PRIMARY SOURCE
Tens of thousands of triples	5.0	Assumed very good results on 10KT	
Hundreds of thousands of triples	5.0	Assumed very good results on 100KT	
Millions of triples	5.0	Very good results on 1MT	
Tens of millions of triples	5.0	Very good results on 10MT	
Hundreds of millions of triples	5.0	Very good results on 100MT	

Hércules. RDF Store Benchmark

Documentación:

- ☐ [TripleStore Benchmark deliverable report](#)
- ☐ [Hercules ASIO Benchmark - User guide](#)
- ☐ [Hercules ASIO Benchmark - Developer documentation](#)
- ☐ [Adding Criteria to the Benchmark Framework](#)



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Una manera de hacer Europa

Hércules ASIO. Casos prácticos



Hércules. RDF Store Benchmark. Caso práctico 1

Identificar los criterios funcionales más importantes para la UM:

- ☐ Funcionalities
 - ☐ Base handling of RDF and SPARQL
 - ☐ Extensions
 - ☐ Security



Hércules. RDF Store Benchmark. Caso práctico 2

Identificar los criterios de rendimiento y escalabilidad más importantes para la UM:

- ☐ Performance and scalability
 - ☐ Query response time.
 - ☐ Loading and update time.
 - ☐ Memory footprint.
 - ☐ Robustness and Scaling.

Hércules. RDF Store Benchmark. Caso práctico 3

Identificar los criterios de administración y mantenimiento más importantes para la UM:

- ☐ Management and maintenance.
 - ☐ Costs and rights.
 - ☐ Ease of deployment.
 - ☐ Administration, updates and maintenance.

Hércules. RDF Store Benchmark. Caso práctico 4

Obtención de la recomendación ajustando los pesos en los criterios:

- ☐ Funcionales
- ☐ Rendimiento y escalabilidad
- ☐ Administración y mantenimiento

Análisis conjunto de los resultados obtenidos.



FONDO EUROPEO DE DESARROLLO REGIONAL (FEDER)

Una manera de hacer Europa

GRACIAS

