

UNIVERSIDAD DE MURCIA

FACULTAD DE BIOLOGÍA

MÁSTER DE BIOINFORMÁTICA



Práctica: Explotación Semántica de Datos – Proteína Tau (MAPT)

Alumno: Alumno05 (puerto 3034)

Fecha: 16 de enero de 2026

Ruta del proyecto: /home/alumno05/entrega_semantica_tau/

Entorno: Servidor Dayhoff – Blazegraph / EasyRDF / FAIRshake

GitHub:

<https://github.com/antonioenriquecoll-design/Knowledge-Graph-Tau-Protein>

AUTOR: Antonio Enrique Coll Meseguer

Curso 2025/2026

ÍNDICE

ÍNDICE.....	2
1. Introducción.....	2
1.1 Fisiopatología de la Proteína Tau (MAPT).....	2
1.2 Tauopatías y manifestaciones clínicas.....	2
1.3 Impacto y necesidad de modelos FAIR.....	3
2. Objetivos.....	3
3. Materiales.....	3
4. Metodología.....	3
4.1 Diseño del Grafo y Selección de URIs.....	3
4.2 Ingeniería de Datos: Conversión y Persistencia.....	4
4.3 Infraestructura: Despliegue y Túnel SSH.....	4
4.4 Publicación FAIR.....	5
5. Resultados y Discusión.....	5
5.1 Evidencia de carga en Blazegraph.....	5
5.2 Validación Técnica (W3C y RDFShape).....	5
5.3 Consultas SPARQL y Automatización en R.....	7
Consulta 1: Verificación de integridad (Total de tripletas).....	7
Consulta 2: Localización celular (Filtro por Gene Ontology).....	9
Consulta 3: Tauopatías asociadas (Disease Ontology).....	10
Consulta 4: Funciones de Unión (Regex Filter).....	11
Consulta 5: Estadísticas del grafo (Count).....	13
5.4 Resultado de publicación FAIRshake.....	14
6. Conclusiones.....	15
7. Referencias.....	15

1. Introducción

1.1 Fisiopatología de la Proteína Tau (MAPT)

La proteína Tau, codificada por el gen *MAPT*, es una proteína asociada a microtúbulos expresada predominantemente en las neuronas del sistema nervioso central. Su función biológica primaria es promover el ensamblaje y mantener la estabilidad de los microtúbulos, estructuras esenciales para el transporte axonal y la morfología neuronal. En condiciones patológicas, Tau sufre modificaciones post-traduccionales, principalmente hiperfosforilación, lo que reduce su afinidad por los microtúbulos y favorece su agregación en filamentos helicoidales emparejados.

1.2 Tauopatías y manifestaciones clínicas

Las enfermedades neurodegenerativas asociadas a la acumulación patológica de agregados de la proteína Tau se conocen como tauopatías. Entre las principales se encuentran la enfermedad de Alzheimer, la enfermedad de Pick, la parálisis supranuclear

progresiva y la demencia frontotemporal. Un rasgo característico de estas patologías es la formación de ovillos neurofibrilares, cuya acumulación muestra una correlación directa con el grado de deterioro cognitivo observado en los pacientes.

1.3 Impacto y necesidad de modelos FAIR

La fragmentación de los datos biomédicos dificulta la investigación. La adopción de tecnologías de la Web Semántica (RDF, SPARQL) y modelos estándar como **BioLink** permite unificar información heterogénea, transformando datos aislados en conocimiento conectado y procesable por máquinas bajo los principios **FAIR** (*Findable, Accessible, Interoperable, Reusable*).

2. Objetivos

- **Diseñar y construir** un dataset RDF que integre la información biológica y patológica de la proteína Tau (P10636).
- **Normalizar** el contenido mediante ontologías de referencia (Gene Ontology, Disease Ontology).
- **Reutilizar** vocabularios estándar (BioLink Model).
- **Implementar** un flujo de trabajo de conversión y validación robusto (Turtle, EasyRDF, W3C).
- **Explorar** semánticamente los datos mediante 5 consultas SPARQL ejecutadas desde R.
- **Publicar** el recurso siguiendo los estándares internacionales en FAIRshake.

3. Materiales

Para la elaboración del grafo de conocimiento se han empleado los siguientes recursos:

- **Fuentes de Datos:** UniProtKB (Acceso P10636), Disease Ontology (DOID), Gene Ontology (GO).
- **Vocabularios:** BioLink Model (esquema de relaciones), RDF, RDFS.
- **Software:** RStudio (paquete SPARQL), Bloc de notas (edición de sintaxis), EasyRDF (motor de conversión).
- **Infraestructura:** Servidor Blazegraph (Clúster Dayhoff), W3C Validator, RDFShape.

4. Metodología

4.1 Diseño del Grafo y Selección de URIs

Para garantizar la interoperabilidad semántica, se han seleccionado identificadores persistentes (URIs) de repositorios internacionales. La entidad central es la **Proteína Tau**, identificada en UniProt como up:P10636 y tipificada como biolink:Protein. Esta entidad se vincula a su gen codificador **MAPT** (ncbi_gene:4137) mediante la propiedad biolink:has_gene (The UniProt Consortium et al., 2023).

En cuanto a las patologías, se han integrado términos de la **Disease Ontology** para representar las tauopatías clave: Enfermedad de Alzheimer (doid:10652) y Enfermedad de Pick (doid:11970), conectadas mediante la relación causal biolink:causes_condition. Las funciones moleculares (*RNA binding*) y localizaciones celulares (Axón) se han normalizado

utilizando **Gene Ontology (GO)** (Schriml et al., 2012; The Gene Ontology Consortium et al., 2021).

Tabla 1. Definición de Espacios de Nombres (Prefixes)

Prefijo	Espacio de Nombres (URI)	Propósito y Relevancia
up:	http://purl.uniprot.org/uniprot/	Identificador de Proteína. Conecta el nodo central con la base de datos UniProtKB, el estándar global de información proteica.
biolink:	https://w3id.org/biolink/vocab/	Modelo de Datos. Proporciona el vocabulario de relaciones (predicados) validado por el consorcio NCATS para grafos de conocimiento biomédicos.
obo:	http://purl.obolibrary.org/obo/	Ontologías Biológicas. Permite integrar términos de Gene Ontology (GO) y Disease Ontology (DOID) bajo el paraguas de la iniciativa OBO Foundry.
rdf / rdfs:	http://www.w3.org/1999/02/22-rdf-syntax-ns#	Sintaxis Básica. Define la estructura de las triplets y las etiquetas de texto (label) legibles por humanos.

4.2 Ingeniería de Datos: Conversión y Persistencia

El grafo se redactó inicialmente en sintaxis Turtle (.ttl) por su legibilidad. Posteriormente, se utilizó la herramienta EasyRDF para transformar el código a formato RDF/XML, asegurando la compatibilidad con sistemas legacy.

Para la persistencia, se realizó una gestión manual del archivo resultante mediante el Bloc de notas, forzando la codificación UTF-8 y la extensión .rdf (eliminando residuos .txt). Esto generó un dataset binario válido de 39 triplets.

4.3 Infraestructura: Despliegue y Túnel SSH

El proyecto se aloja en el clúster dayhoff. La estructura de archivos en el directorio /home/alumno05/entrega_semantica_tau/ incluye los códigos fuente y el script de análisis en R.

```
[alumno05@dayhoff ~]$ cd entrega_semantica_tau/  
[alumno05@dayhoff entrega_semantica_tau]$ ls  
blazegraph.jnl consultas.R practica.rdf practica.ttl rules.log  
[alumno05@dayhoff entrega_semantica_tau]$ █
```

Figura 1. Estructura del directorio de trabajo en el servidor remoto mostrando los archivos .ttl, .rdf y .R.

Dado que Blazegraph se ejecuta en el puerto local del servidor, se estableció un **túnel SSH** para permitir el acceso y la carga de datos desde la máquina local:

```
ssh -L 3034:localhost:3034 alumno05@dayhoff.inf.um.es
```

Este comando redirige el tráfico, permitiendo interactuar con la base de datos a través de <http://localhost:3034/blazegraph/>.

Ejecución de Blazegraph se despliega en el servidor mediante el binario Java:

```
java -server -Xmx4g -jar blazegraph.jar
```

4.4 Publicación FAIR

Siguiendo los principios de Ciencia Abierta, el recurso se registró en la plataforma **FAIRshake**. Se completó la evaluación de métricas para garantizar que el grafo es Encontrable, Accesible, Interoperable y Reutilizable, vinculando los metadatos con el *endpoint* del servidor.

Además, la publicación en GitHub refuerza los principios FAIR, permitiendo el acceso público y permanente a los recursos del grafo.

5. Resultados y Discusión

5.1 Evidencia de carga en Blazegraph

La ingestión de datos en el servidor Blazegraph se realizó mediante la interfaz *Update*, utilizando el formato Turtle validado.

The screenshot shows the Blazegraph Update interface at the URL localhost:3034/blazegraph/#update. The interface has a top navigation bar with links for WELCOME, QUERY, UPDATE, EXPLORE, NAMESPACES, and STATUS. The UPDATE tab is selected. Below the tabs is a code editor containing a SPARQL update query. The query is as follows:

```
17 |   obo:GO_0030425 , obo:GO_0005829 , obo:GO_0043025 ,  
18 |   obo:GO_0045202 .  
19 |  
20 |   obo:GO_0005737 rdfs:label "Cytoplasm" .  
21 |   obo:GO_0005856 rdfs:label "Cytoskeleton" .  
22 |   obo:GO_0030424 rdfs:label "Axon" .  
23 |  
24 | # FUNCIONES Y PROCESOS  
25 | up:P10636 biolink:enables obo:GO_0003723 , obo:GO_0003677 , obo:GO_0008017 ,  
26 |   obo:GO_0001540 , obo:GO_0007489 , obo:GO_0006836 ;  
27 |   obo:GO_0007613 , obo:GO_0050808 , obo:GO_0043621 ,  
28 |   obo:GO_0007018 .  
29 |  
30 |   obo:GO_0003723 rdfs:label "RNA binding" .  
31 |   obo:GO_0008017 rdfs:label "Microtubule binding" .  
32 |  
33 | # ENFERMEDADES  
34 | up:P10636 biolink:causes_condition obo:DOID_10652 , obo:DOID_9255 , obo:DOID_11970 , obo:DOID_678 .  
35 |  
36 |   obo:DOID_10652 rdfs:label "Alzheimer disease" ; rdf:type biolink:Disease .  
37 |   obo:DOID_9255 rdfs:label "Frontotemporal dementia" ; rdf:type biolink:Disease .  
38 |   obo:DOID_11970 rdfs:label "Pick disease" ; rdf:type biolink:Disease .  
39 |   obo:DOID_678 rdfs:label "Parkinsonism" ; rdf:type biolink:Disease .
```

Below the code editor, there are dropdown menus for "Type: RDF Data" and "Format: Turtle". A link "Advanced features" is also present. At the bottom right is a large "Update" button. The status bar at the bottom indicates "Modified: 39" and "Milliseconds: 92".

Figura 2. Interfaz de carga en Blazegraph confirmando la actualización del namespace.

La carga del grafo en Blazegraph finalizó correctamente, registrando **39 tripletas** y un **tiempo de procesamiento de 92 ms**.

La gestión de datos se realizó en el Triple Store **Blazegraph** bajo el namespace predefinido **kb**, actuando como repositorio central de las 39 triplets. Esta configuración en el servidor Dayhoff asegura una indexación eficiente y la consistencia del endpoint de consulta (<http://localhost:3034/blazegraph/namespace/kb/sparql>) para su explotación, tanto mediante interfaz web como de forma programática con el script de **R**.

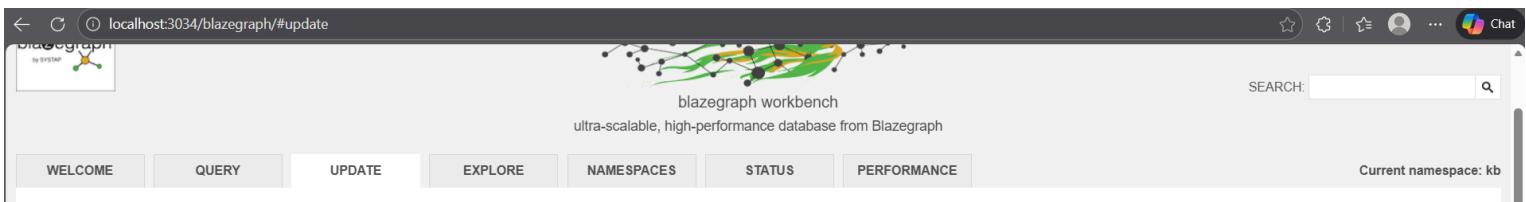


Figura 3. Interfaz de Blazegraph (puerto 3034) con el namespace activo "kb".

5.2 Validación Técnica (W3C y RDFShape)

El dataset RDF/XML fue sometido al servicio de validación de la W3C, confirmando la sintaxis correcta y el conteo de **39 tripleas**.

The screenshot shows the W3C RDF Validator results page. At the top, it says 'Jump To:' with links to Source, Triples, Messages, Graph, Feedback, and Back to Validator Input. Below that, under 'Validation Results', it says 'Your RDF document validated successfully.' A table titled 'Triples of the Data Model' follows, listing 39 triples with columns for Number, Subject, Predicate, and Object. The last row of the table is partially cut off.

Number	Subject	Predicate	Object
1	http://purl.uniprot.org/uniprot/P10636	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	https://w3id.org/biolink/vocab/Protein
2	http://purl.uniprot.org/uniprot/P10636	http://www.w3.org/2000/01/rdf-schema#label	"Tau Protein"
3	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/has_gene	http://identifiers.org/ncbigene:4137
4	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/category	https://w3id.org/biolink/vocab/Protein
5	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/in_taxon	http://purl.obolibrary.org/obo/NCBITaxon_9606
6	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0005737
7	http://purl.obolibrary.org/obo/GO_0005737	http://www.w3.org/2000/01/rdf-schema#label	"Cytoplasm"
8	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0005856
9	http://purl.obolibrary.org/obo/GO_0005856	http://www.w3.org/2000/01/rdf-schema#label	"Cytoskeleton"
10	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0030424
11	http://purl.obolibrary.org/obo/GO_0030424	http://www.w3.org/2000/01/rdf-schema#label	"Axon"
12	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0030425
13	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0005829
14	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0043025
15	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0045202
16	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/enables	http://purl.obolibrary.org/obo/GO_0003723
17	http://purl.obolibrary.org/obo/GO_0003723	http://www.w3.org/2000/01/rdf-schema#label	"RNA binding"
18	http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/enables	http://purl.obolibrary.org/obo/GO_0003677

Figura 4. Validación exitosa en la W3C mostrando la tabla de 39 sentencias semánticas.

Además, se utilizó **RDFShape** para inspeccionar la topología del grafo. La visualización confirma una estructura radial donde el nodo central (Tau) conecta correctamente con nodos de enfermedad y función.

- → ⌂ (✉ [https://rdfshape.weso.es/dataInfo?data=<%3Fxml%20version%3D\"1.0\"%20encoding%3D\"utf-8\"%20%3F>%0A<rdf%3ARDF%20xmlNs%3Ardf%3D\"http%3A%](https://rdfshape.weso.es/dataInfo?data=<%3Fxml%20version%3D\)

Data analysis

Data (RDF)

[Text](#) [URL](#) [File](#)

```

1 <?xml version="1.0" encoding="utf-8" ?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
4   xmlns:ns0="https://w3id.org/biolink/vocab/">
5
6   <rdf:Description
7     rdf:about="http://purl.uniprot.org/uniprot/P10636">
8     <rdf:type rdf:resource="https://w3id.org/biolink/vocab/Protein"/>
9     <rdfs:label>Tau Protein</rdfs:label>
10    <ns0:has_gene
11      rdf:resource="http://identifiers.org/ncbigene:4137"/>
12      <ns0:category
13        rdf:resource="https://w3id.org/biolink/vocab/Protein"/>
14      <ns0:in_taxon
15

```

Format

RDF/XML

Inference

NONE

[Overview](#)

[Prefix Map](#)

[Visualizations](#)

- Well formed RDF
- Number of statements: 39
- Data format: RDF/XML

▶ Full response

[Permalink](#) 

[Overview](#)

[Prefix Map](#)

[Visualizations](#)

Name↑	IRI
ns0	< https://w3id.org/biolink/vocab/ >
rdf	< http://www.w3.org/1999/02/22-rdf-syntax-ns# >
rdfs	< http://www.w3.org/2000/01/rdf-schema# >

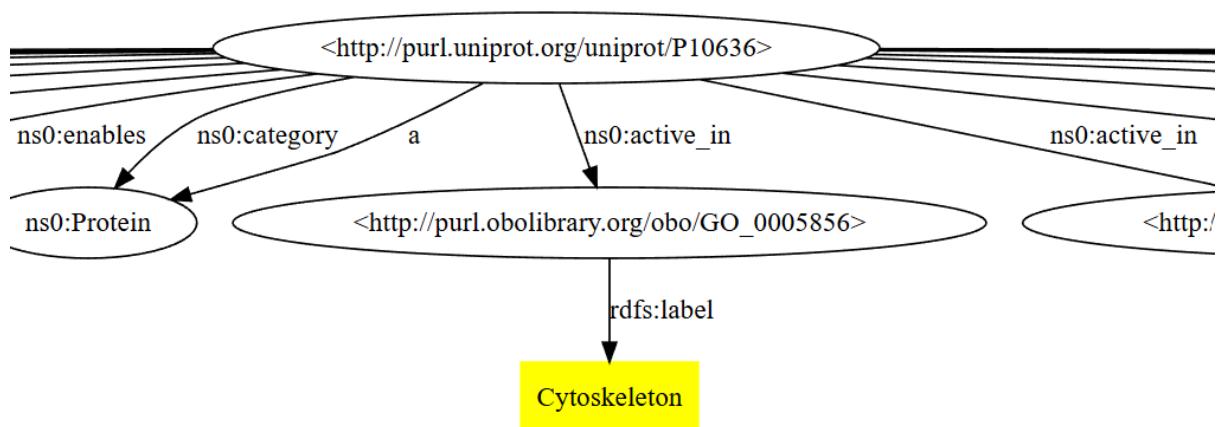


Figura 5. Visualización del Grafo de Conocimiento generado en RDFShape.

5.3 Consultas SPARQL y Automatización en R

Para validar la utilidad biológica del grafo, se diseñaron 5 consultas SPARQL ejecutadas sobre el motor Blazegraph. Estas consultas fueron posteriormente integradas en un script de R (consultas.R) para automatizar la extracción de resultados y facilitar su procesamiento en un entorno de análisis estadístico.

Consulta 1: Verificación de integridad (Total de triplets)

Objetivo: Verificar la integridad de la carga.

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

The screenshot shows the Blazegraph Workbench interface at localhost:3034/blazegraph/#query. The top navigation bar includes a back arrow, a refresh button, and a status indicator. Below the header is a logo for 'blazegraph by SYSTEM' and a large graphic of a network graph. The main area has tabs for WELCOME, QUERY, UPDATE, EXPLORE, NAMESPACES, STATUS, and PERFORMANCE. A sub-header 'Wiki - SPARQL Query' is visible. In the central query editor, the SPARQL query `SELECT * WHERE { ?s ?p ?o }` is entered. At the bottom right of the editor are 'Execute' and 'Clear' buttons. A link 'Advanced features' is located just below the editor. The bottom of the screen shows a horizontal scrollbar.

localhost:3034/blazegraph/#query

s	p	o
http://purl.obolibrary.org/obo/DOID_10652	rdf:type	https://w3id.org/biolink/vocab/Disease
http://purl.obolibrary.org/obo/DOID_10652	rdfs:label	Alzheimer disease
http://purl.obolibrary.org/obo/DOID_11970	rdf:type	https://w3id.org/biolink/vocab/Disease
http://purl.obolibrary.org/obo/DOID_11970	rdfs:label	Pick disease
http://purl.obolibrary.org/obo/DOID_678	rdf:type	https://w3id.org/biolink/vocab/Disease
http://purl.obolibrary.org/obo/DOID_678	rdfs:label	Parkinsonism
http://purl.obolibrary.org/obo/DOID_9255	rdf:type	https://w3id.org/biolink/vocab/Disease
http://purl.obolibrary.org/obo/DOID_9255	rdfs:label	Frontotemporal dementia
http://purl.obolibrary.org/obo/GO_0003723	rdfs:label	RNA binding
http://purl.obolibrary.org/obo/GO_0005737	rdfs:label	Cytoplasm
http://purl.obolibrary.org/obo/GO_0005856	rdfs:label	Cytoskeleton
http://purl.obolibrary.org/obo/GO_0008017	rdfs:label	Microtubule binding
http://purl.obolibrary.org/obo/GO_0030424	rdfs:label	Axon
http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0005737
http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0005829
http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0005856
http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0030424
http://purl.uniprot.org/uniprot/P10636	https://w3id.org/biolink/vocab/active_in	http://purl.obolibrary.org/obo/GO_0030425

Total results: 39, displaying 1-39 <input type="button" value="50"/> per page
<input type="checkbox"/> Show datatypes <input type="checkbox"/> Show languages
Clear history

Time	Query	Results	Execution Time	Delete
2026-01-15T13:05:42.010Z	SELECT * WHERE { ?s ?p ?o }	39	317ms	X
Clear history				

Consulta 2: Localización celular (Filtro por Gene Ontology)

Objetivo: Identificar dónde actúa la proteína según la ontología GO.

PREFIX biolink: <<https://w3id.org/biolink/vocab/>>

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

SELECT ?lugar WHERE {

```

<http://purl.uniprot.org/uniprot/P10636> biolink:active_in ?id .
?id rdfs:label ?lugar .

```

}

The screenshot shows the Blazegraph interface at localhost:3034/blazegraph/#query. The top navigation bar includes WELCOME, QUERY, UPDATE, EXPLORE, NAMESPACES, and STAT. The main area is titled "Wiki - SPARQL Query". A SPARQL query is pasted into the editor:

```

1 PREFIX biolink: <https://w3id.org/biolink/vocab/>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 SELECT ?lugar WHERE {
4   <http://purl.uniprot.org/uniprot/P10636> biolink:active_in ?id .
5   ?id rdfs:label ?lugar .
6 }

```

Below the query, the results are displayed in a table:

lugar
Cytoplasm
Cytoskeleton
Axon

Total results: 3, displaying 1-3 [50] per page
 Show datatypes Show languages

Below the results table, another table shows the history of queries:

Time	Query	Results	Execution Time	Delete
2026-01-15T13:07:10.296Z	PREFIX biolink: <https://w3id.org/biolink/vocab/> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> SELECT ?lugar WHERE { <http://purl.uniprot.org/uniprot/P10636> biolink:active_in ?id . ?id rdfs:label ?lugar . }	3	315ms	X
2026-01-15T13:05:42.010Z	SELECT * WHERE { ?s ?p ?o .}	39	317ms	X

[Clear history](#)

Blazegraph - Wiki

Consulta 3: Tauopatías asociadas (Disease Ontology)

Objetivo: Listar las patologías relacionadas modeladas bajo BioLink.

PREFIX biolink: <<https://w3id.org/biolink/vocab/>>

SELECT ?enfermedad WHERE {

```

<http://purl.uniprot.org/uniprot/P10636>
biolink:causes_condition ?id .

```

```
?id rdfs:label ?enfermedad .
```

```
}
```

blazegraph works
ultra-scalable, high-performance database

WELCOME QUERY UPDATE EXPLORE NAMESPACES STATUS

[Wiki - SPARQL Query](#)

```

1 PREFIX biolink: <https://w3id.org/biolink/vocab/>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 SELECT ?enfermedad WHERE {
4   <http://purl.uniprot.org/uniprot/P10636> biolink:causes_condition ?id .
5   ?id rdfs:label ?enfermedad .
6 }

```

[Advanced features](#)

[Execute](#) [Clear](#)

enfermedad
Alzheimer disease
Pick disease
Parkinsonism
Frontotemporal dementia

localhost:3034/blazegraph/#query

[Advanced features](#)

[Execute](#) [Clear](#)

enfermedad
Alzheimer disease
Pick disease
Parkinsonism
Frontotemporal dementia

Total results: 4, displaying 1-4 [50] per page
 Show datatypes Show languages

Time	Query	Results	Execution Time	Delete
2026-01-15T13:08:15.808Z	PREFIX biolink: <https://w3id.org/biolink/vocab/> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> SELECT ?enfermedad WHERE { <http://purl.uniprot.org/uniprot/P10636> biolink:causes_condition ?id . ?id rdfs:label ?enfermedad . }	4	305ms	X
2026-01-15T13:07:10.296Z	PREFIX biolink: <https://w3id.org/biolink/vocab/> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> SELECT ?lugar WHERE { <http://purl.uniprot.org/uniprot/P10636> biolink:active_in ?id . ?id rdfs:label ?lugar . }	3	315ms	X
2026-01-15T13:05:42.010Z	SELECT * WHERE { ?s ?p ?o }	39	317ms	X

[Clear history](#)

Consulta 4: Funciones de Unión (Regex Filter)

Objetivo: Filtrar funciones moleculares que impliquen capacidad de "binding".

```
SELECT ?funcion WHERE {  
    ?s rdfs:label ?funcion .  
    FILTER regex(?funcion, "binding", "i")  
}
```

The screenshot shows a user interface for a SPARQL query. At the top, there are tabs: WELCOME, QUERY (which is selected), UPDATE, EXPLORE, and NAMESPACES. Below the tabs, the title "Wiki - SPARQL Query" is displayed. The main area contains the SPARQL code:

```
1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>  
2 SELECT ?funcion WHERE {  
3     ?s rdfs:label ?funcion .  
4     FILTER regex(?funcion, "binding", "i")  
5 }
```

Below the code, there is a section titled "Advanced features" and a "Execute" button. The results are displayed in a table:

funcion
Microtubule binding
RNA binding

At the bottom, there is a message: "Total results: 2, displaying 1-2 [50] per page" with dropdown menus for page size and checkboxes for "Show datatypes" and "Show languages".

Consulta 5: Estadísticas del grafo (Count)

Objetivo: Cuantificar las relaciones directas de la proteína Tau.

```
SELECT (COUNT(?p) AS ?total_relaciones) WHERE {  
    <http://purl.uniprot.org/uniprot/P10636> ?p ?o .  
}
```

The screenshot shows the blazegraph web interface at localhost:3034/blazegraph/#query. The interface includes a navigation bar with tabs: WELCOME, QUERY (selected), UPDATE, EXPLORE, NAMESPACES, and STATUS. Below the tabs is a search bar with placeholder text "Wiki - SPARQL Query". The main area contains the SPARQL query:

```
1 SELECT (COUNT(?p) AS ?total_relaciones) WHERE {  
2     <http://purl.uniprot.org/uniprot/P10636> ?p ?o .  
3 }
```

Below the query is a "Advanced features" section with a "Execute" button and a "Clear history" link. The results are displayed in a table:

total_relaciones
26

At the bottom, there is a history table showing previous queries:

Time	Query	Results	Execution Time	Delete
2026-01-15T13:10:39.759Z	<code>SELECT (COUNT(?p) AS ?total_relaciones) WHERE { <http://purl.uniprot.org/uniprot/P10636> ?p ?o . }</code>	1	232ms	X
2026-01-15T13:09:11.986Z	<code>PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX biolink: <https://w3id.org/biolink/vocab/> SELECT ?funcion WHERE { ?s rdfs:label ?funcion FILTER regex(?funcion, "binding", "i") }</code>	2	243ms	X
2026-01-15T13:08:15.808Z	<code>PREFIX biolink: <https://w3id.org/biolink/vocab/> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> SELECT ?enfermedad WHERE { <http://purl.uniprot.org/uniprot/P10636> biolink:causes_condition ?id . ?id rdfs:label ?enfermedad . }</code>	4	305ms	X
2026-01-15T13:07:10.296Z	<code>PREFIX biolink: <https://w3id.org/biolink/vocab/> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> SELECT ?lugar WHERE { <http://purl.uniprot.org/uniprot/P10636> biolink:active_in ?id . ?id rdfs:label ?lugar . }</code>	3	315ms	X
2026-01-15T13:05:42.010Z	<code>SELECT * WHERE { ?s ?p ?o }</code>	39	317ms	X

A continuación, se muestra la ejecución exitosa de este script en el entorno del clúster:

```
[alumno05@dayhoff entrega_semantica_tau]$ Rscript consultas.R
Loading required package: SPARQL
Loading required package: XML
Loading required package: RCurl
[1] "--- Consulta 1: Listado completo de triplets ---"
   s
1 <http://purl.obolibrary.org/obo/DOID_10652>
2 <http://purl.obolibrary.org/obo/DOID_10652>
3 <http://purl.obolibrary.org/obo/DOID_11970>
4 <http://purl.obolibrary.org/obo/DOID_11970>
5   <http://purl.obolibrary.org/obo/DOID_678>
6   <http://purl.obolibrary.org/obo/DOID_678>
7   <http://purl.obolibrary.org/obo/DOID_9255>
8   <http://purl.obolibrary.org/obo/DOID_9255>
9   <http://purl.obolibrary.org/obo/GO_0003723>
10  <http://purl.obolibrary.org/obo/GO_0005737>
11  <http://purl.obolibrary.org/obo/GO_0005856>
12  <http://purl.obolibrary.org/obo/GO_0008017>
13  <http://purl.obolibrary.org/obo/GO_0030424>
14    <http://purl.uniprot.org/uniprot/P10636>
15    <http://purl.uniprot.org/uniprot/P10636>

35   <http://purl.obolibrary.org/obo/GO_0050808>
36     <http://identifiers.org/ncbigene:4137>
37 <http://purl.obolibrary.org/obo/NCBITaxon_9606>
38   <https://w3id.org/biolink/vocab/Protein>
39                               Tau Protein
[1] "--- Consulta 2: Localización celular (GO) ---"
     lugar      lugar.1 lugar.2
1 Cytoplasm Cytoskeleton Axon
[1] "--- Consulta 3: Patologías relacionadas (DOID) ---"
     enfermedad enfermedad.1 enfermedad.2          enfermedad.3
1 Alzheimer disease Pick disease Parkinsonism Frontotemporal dementia
[1] "--- Consulta 4: Funciones de tipo 'binding' ---"
     funcion      funcion.1
1 Microtubule binding RNA binding
[1] "--- Consulta 5: Conteo total de relaciones ---"
     total_relaciones
1           26
[alumno05@dayhoff entrega_semantica_tau]$ █
```

Figura 6. Ejecución del script 'consultas.R' contra el endpoint, recuperando patologías y funciones.

5.4 Resultado de publicación FAIRshake

La evaluación final en la plataforma FAIRshake:

Assessments (1)



Figura 7. Evaluación FAIR (Cuadrado Azul) obtenida tras el registro del proyecto.

El recurso obtuvo la certificación FAIR, con el siguiente identificador:

[Assessment of RDF Dataset: Neurodegenerative Diseases \(Biolink Model\) - Alumno05 with FAIR metrics by fairmetrics.org](#)

Blazegraph endpoint (Dayhoff UMU): <http://dayhoff.inf.um.es:3034/blazegraph/>

Para asegurar la transparencia y reutilización, el código y el dataset se han publicado en: <https://github.com/antonioenriquecoll-design/Knowledge-Graph-Tau-Protein>.

6. Conclusiones

El desarrollo de esta práctica ha permitido implementar un flujo de trabajo para la gestión de datos semánticos utilizando como modelo la proteína **Tau (MAPT)**. El proceso se inició con el modelado en formato **Turtle (.ttl)** y su posterior conversión a un archivo **.rdf (RDF/XML)**, logrando un grafo final de **39 triplets** validadas mediante el servicio del W3C. La persistencia de los datos se realizó en una instancia de **Blazegraph** alojada en el servidor **Dayhoff**, accesible mediante un **túnel SSH**. La explotación del recurso se ha verificado con la ejecución de **5 consultas SPARQL**, incluyendo la automatización de las mismas a través de un script en **R**. Finalmente, el conjunto de datos se ha publicado en **GitHub** y ha sido evaluado bajo los estándares **FAIR** en la plataforma **FAIRshake**, garantizando su interoperabilidad y capacidad de reutilización en el dominio biomédico.

7. Referencias

- Schriml, L. M., Arze, C., Nadendla, S., Chang, Y.-W. W., Mazaitis, M., Felix, V., Feng, G., & Kibbe, W. A. (2012). Disease Ontology: A backbone for disease semantic integration. *Nucleic Acids Research*, 40(D1), D940-D946. <https://doi.org/10.1093/nar/gkr972>
- The Gene Ontology Consortium, Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L.-P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., ... Elser, J. (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research*, 49(D1), D325-D334. <https://doi.org/10.1093/nar/gkaa1113>
- The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., Da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., ... Zhang, J. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1), D523-D531. <https://doi.org/10.1093/nar/gkac1052>