

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).
- **Explotar** 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, estructurando el conocimiento bajo el estándar BioLink Model.

Identidad y Origen Genético: La entidad central es la **Proteína Tau**, identificada en UniProt como up:P10636 y tipificada como biolink:Protein., vinculándose a su taxón humano (obo:NCBITaxon_9606) y a su gen codificante **MAPT** (ncbigene:4137) mediante la propiedad biolink:has_gene (The UniProt Consortium et al., 2023).

Contexto subcelular y dinámico: Se han integrado múltiples términos de **Gene Ontology (GO)** para mapear la localización física de la proteína (Citoplasma, Citoesqueleto, Axón, etc.) mediante la propiedad `biolink:active_in`, diferenciándolos de sus capacidades funcionales como la unión a microtúbulos o al RNA, modeladas mediante `biolink:enables` (The Gene Ontology Consortium et al., 2021).

Complejidad Clínica (Tauopatías): Se ha dotado al grafo de una alta relevancia médica integrando diversas patologías de la **Disease Ontology (DOID)**. Mediante la propiedad `biolink:causes_condition`, se conecta la proteína Tau con un espectro de tauopatías que incluye el Alzheimer (`doid:10652`), la demencia frontotemporal (`doid:9255`), la enfermedad de Pick (`doid:11970`) y el parkinsonismo (`doid:678`) (Schriml et al., 2012).

Este modelado evita el uso de literales aislados, tratando cada patología y función como un objeto tipado (`rdf:type biolink:Disease`).

Tabla 1. Definición de Espacios de Nombres (Prefixes)

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

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

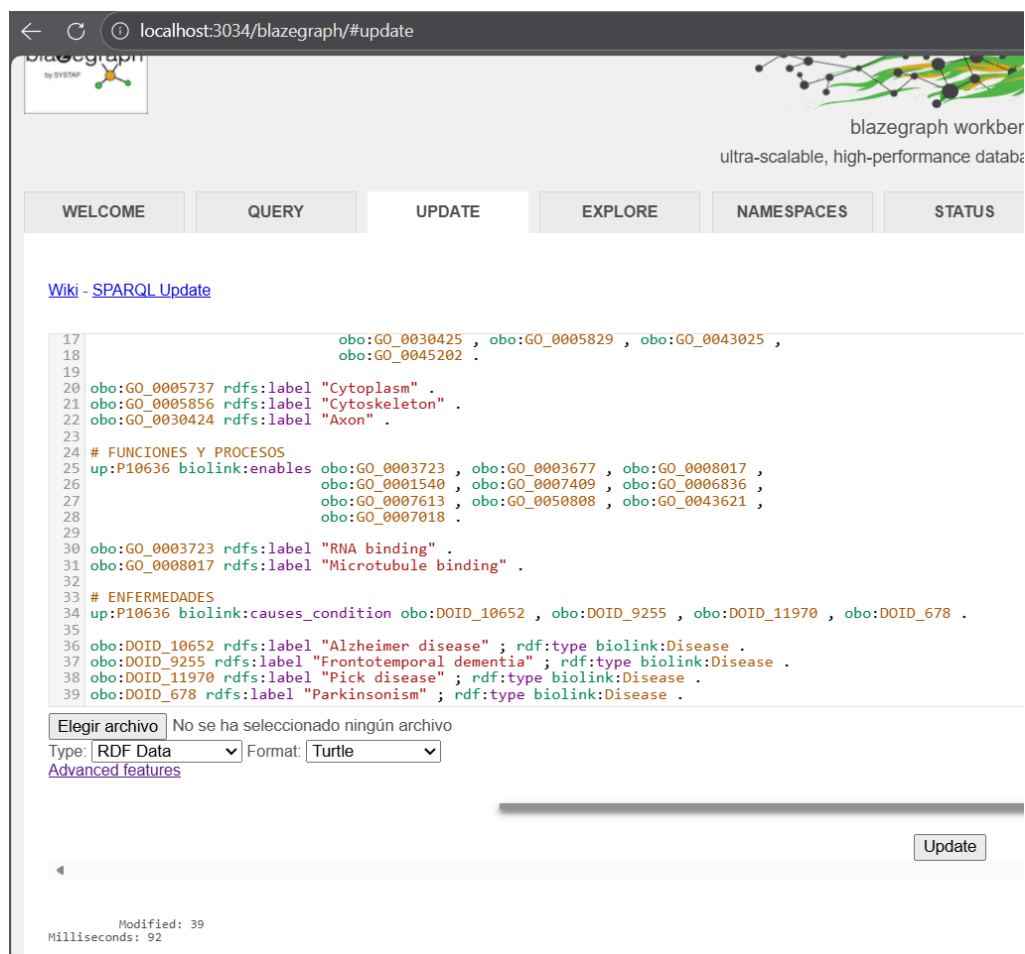


Figura 2. Interfaz de carga en Blazegraph confirmando la actualización del namespace.

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 tripletas. 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**.

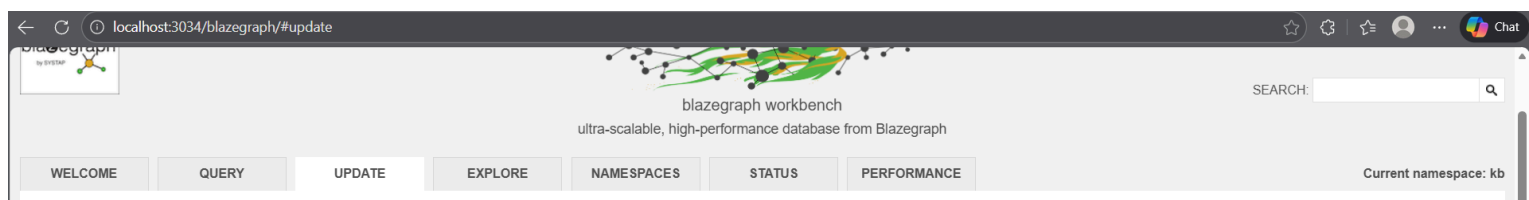


Figura 3. Interfaz de Blazegraph (puerto 3034) con el namespace activo "kb".

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 tripletas**.


```


graph TD
    A("<http://purl.uniprot.org/uniprot/P10636>")
    B("<http://purl.obolibrary.org/obo/GO_0005856>")
    C("<http://purl.obolibrary.org/obo/GO_0005856>")
    D("<http://purl.obolibrary.org/obo/GO_0005856>")
    E("Cytoskeleton")

    A -- ns0:enables --> B
    A -- ns0:category --> B
    A -- a --> D
    A -- ns0:active_in --> D
    A -- ns0:active_in --> E
    D -- rdfs:label --> E
  
```

5.3 Consultas SPARQL y Automatización en R


Consulta 1: Verificación de integridad (Total de tripletas)

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

blazegraph

by blazegraph



blazegraph workbench

ultra-scalable, high-performance database from Blazegraph

WELCOME

QUERY

UPDATE

EXPLORE

NAMESPACES

STATUS

PERFORMANCE

[Wiki - SPARQL Query](#)

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

[Advanced features](#)

Execute

Clear

| s | p | o |
|---|---|---|
| http://purl.obolibrary.org/obo/DOID_10652 | rdfs: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 | rdfs: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 | rdfs:type | https://w3id.org/biolink/vocab/Disease |
| http://purl.obolibrary.org/obo/DOID_678 | rdfs:label | Parkinsonism |
| http://purl.obolibrary.org/obo/DOID_9255 | rdfs: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

50

 per page

☐ Show datatypes ☐ Show languages

| 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 web interface at localhost:3034/blazegraph/#query. The interface has tabs for WELCOME, QUERY, UPDATE, EXPLORE, NAMESPACES, and STATISTICS. The QUERY tab is active, displaying the SPARQL query entered in the text area:

```
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 area, there is an "Advanced features" section with a progress bar and an "Execute" button. The results are displayed in a table with the following content:

| lugar |
|--------------|
| Cytoplasm |
| Cytoskeleton |
| Axon |

At the bottom, it shows "Total results: 3, displaying 1-3 50 per page" and checkboxes for "Show datatypes" and "Show languages".

This screenshot shows the results table from the Blazegraph interface. It includes a table with columns for Time, Query, Results, Execution Time, and Delete. Below the table is a "Clear history" button and a link to the Blazegraph Wiki.

| 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 workbench
ultra-scalable, high-performance distributed graph 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 |

Total results: 4, displaying 1-4 per page
☐ Show datatypes ☐ Show languages

| Time | Query | Results | Execution Time | Delete |
|--------------------------|---|---------|----------------|--------|
| 2026-01-15T13:08:15.808Z | <pre>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 . }</pre> | 4 | 305ms | X |
| 2026-01-15T13:07:10.296Z | <pre>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 . }</pre> | 3 | 315ms | X |
| 2026-01-15T13:05:42.010Z | <pre>SELECT * WHERE { ?s ?p ?o }</pre> | 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 web interface for executing SPARQL queries. At the top, there are tabs: WELCOME, QUERY, UPDATE, EXPLORE, and NAMESPACES. Below the tabs, there is a link "Wiki - SPARQL Query". The main area contains a text editor with the following SPARQL query:

```
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 query editor, there is a section titled "Advanced features". To the right of this section, there is a button labeled "Execute". Below the button, there is a table with the following results:

| funcion |
|---------------------|
| Microtubule binding |
| RNA binding |

At the bottom of the interface, there is a summary of the results: "Total results: 2, displaying 1-2" followed by a dropdown menu set to "50" and the text "per page". Below this, there are two checkboxes: "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, UPDATE, EXPLORE, NAMESPACES, and STATUS. Below the navigation bar, there is a link to "Wiki - SPARQL Query". The main area displays a SPARQL query in a text editor. Below the editor, there is a section for "Advanced features" and a table showing the results of the query. The table has one row with the column "total_relaciones" and the value "26".

blazegraph world
ultra-scalable, high-performance database

WELCOME QUERY UPDATE EXPLORE NAMESPACES STATUS

[Wiki - SPARQL Query](#)

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

[Advanced features](#)

Execute Clear

| total_relaciones |
|------------------|
| 26 |

| Time | Query | Results | Execution Time | Delete |
|--------------------------|---|---------|----------------|--------|
| 2026-01-15T13:10:39.759Z | SELECT (COUNT(?p) AS ?total_relaciones) WHERE { <http://purl.uniprot.org/uniprot/P10636> ?p ?o . } | 1 | 232ms | X |
| 2026-01-15T13:09:11.986Z | PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> SELECT ?funcion WHERE { ?s rdfs:label ?funcion . FILTER regex(?funcion, "binding", ".") } | 2 | 243ms | X |
| 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

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 tripletas ---"
      s
1 <http://purl.obolibrary.org/obo/D0ID_10652>
2 <http://purl.obolibrary.org/obo/D0ID_10652>
3 <http://purl.obolibrary.org/obo/D0ID_11970>
4 <http://purl.obolibrary.org/obo/D0ID_11970>
5 <http://purl.obolibrary.org/obo/D0ID_678>
6 <http://purl.obolibrary.org/obo/D0ID_678>
7 <http://purl.obolibrary.org/obo/D0ID_9255>
8 <http://purl.obolibrary.org/obo/D0ID_9255>
9 <http://purl.obolibrary.org/obo/G0_0003723>
10 <http://purl.obolibrary.org/obo/G0_0005737>
11 <http://purl.obolibrary.org/obo/G0_0005856>
12 <http://purl.obolibrary.org/obo/G0_0008017>
13 <http://purl.obolibrary.org/obo/G0_0030424>
14 <http://purl.uniprot.org/uniprot/P10636>
15 <http://purl.uniprot.org/uniprot/P10636>

35 <http://purl.obolibrary.org/obo/G0_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 (G0) ---"
      lugar      lugar.1 lugar.2
1 Cytoplasm Cytoskeleton Axon
[1] "--- Consulta 3: Patologías relacionadas (D0ID) ---"
      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 tripletas** 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>