



UNIVERSIDAD DE MÁLAGA



Graduado en Ingeniería de la Salud

Diseño web de gestor de información de CITES

Realizado por
Rafael Trapero Vílchez

Alejandro Domínguez
Recio

Tutorizado por
Laura Panizo Jaime

Departamento
Lenguajes y Ciencias de la Computación

MÁLAGA, septiembre de 2022



UNIVERSIDAD
DE MÁLAGA



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA
GRADUADO EN INGENIERÍA DEL SOFTWARE

**Diseño web de gestor de información de
CITES**

Realizado por
Rafael Trapero Vílchez
Alejandro Domínguez Recio

Tutorizado por
Laura Panizo Jaime

Departamento
Lenguajes y Ciencias de la Computación

UNIVERSIDAD DE MÁLAGA
MÁLAGA, SEPTIEMBRE DE
2022

Abstract

La información es una herramienta muy poderosa. Pero no basta con que esa información exista; debe ser accesible y entendible para un gran número de personas. Por ello, tanto la integración como la estandarización de dichos datos se han convertido en un requisito fundamental para el análisis de cualquier tipo de datos.

Como hemos dicho anteriormente, la información debe ser accesible para el gran público. Intentado satisfacer esta necesidad, hemos creado este proyecto con el fin de presentar, mediante recursos web, una serie de datos (en este caso sobre el tráfico de especies animales a nivel mundial) a los que el usuario podrá acceder de una forma sencilla por medio de un sistema de consultas gráficas.

Índice

1. Introducción	5
1.1. Motivación	5
1.2. Objetivos	5
1.3. Dataset	5
1.4. Tecnologías usadas	6
2. Diseño web	7
3. XSLT - XQuery	11
3.1. XSLT	11
3.1.1. XML-XML	11
3.1.2. XML-HTML	12
3.2. XQuery	13
4. RDF Schema	15
4.1. Archivo owl	15
4.2. RDF Schema	16
5. Consultas SPARQL	19
5.1. Consultas SPARQL	19
5.2. Consultas SPARQL Extra	20

1

Introducción

1.1. Motivación

Proporcionar a cualquier tipo de usuario un entorno gráfico donde poder trabajar con información (procedente de la Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestres 'CITES') sobre el tráfico de animales, de una forma clara y sencilla.

1.2. Objetivos

Creación de un gestor de información web, el cual permita integrar, explotar, transformar y exportar información de una forma intuitiva.

1.3. Dataset

El dataset con el que se trabaja consiste en una serie de registros de importación/exportación de diversas especies animales a nivel internacional durante los años 2016-2017.

El dataset ha sido descargado de Kaggle (<https://www.kaggle.com/datasets/cites/cites-wildlife-trade-database>) aunque originalmente procede de la web oficial de CITES 'Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestres' (<https://cites.org/esp/disc/what.php>).

El dataset consta de más de 60,000 filas y unas 15 columnas. Cada columna representa una variable a tener cuenta. Encontramos:

- **App:** el apéndice al que la especie pertenece. Siendo Apéndice I aquellas especies más protegidas y Apéndice 2 las que menos.
- **Taxon:** el taxón de la especie.

- **Class:** la clase taxonómica de la especie.
- **Order:** el orden taxonómico de la especie.
- **Family:** la familia taxonómica de la especie.
- **Genus:** nombre científico de la especie.
- **Importer:** país que importa la especie.
- **Exporter:** país que exporta la especie.
- **Origin:** país originario de la especie.
- **Importer reported quantity:** cantidad declarada por el importador.
- **Exporter reported quantity:** cantidad declarada por el exportador.
- **Term:** bien exportado o importado.
- **Source:** describe como el animal o planta fue traída al mercado.
- **Purpose:** propósito por el que esa especie fue importada o exportada.

1.4. Tecnologías usadas

Para la realización del proyecto hemos usado varias herramientas. Visual Studio Code para implementar las transformaciones XSLT, así como el script en python para pasar de CSV a XML.

Adicionalmente hemos usado Protegé para crear el esquema RDF de nuestra base de datos. Finalmente, con la herramienta GraphDB, hemos impotado nuestro esquema y hemos realizado 4 consultas SPARQL.

Todos los archivos creados en este proyecto se encuentran en nuestro repositorio de GitHub.

https://github.com/GitHubAlejandroDR/open-data-standards_Project.git

2

Diseño web

Este apartado se encargará de mostrar el diseño de la app web.

El primer diseño muestra la página de búsqueda estándar. En la barra de búsqueda introduciremos cualquier palabra y nos mostrará en la tabla de abajo aquellas filas que contengan dicha entrada como valores de alguna de sus columnas.

Adicionalmente, también podemos filtrar por algunas columnas seleccionando los checkboxes deseados. Añadir que, además, podemos ordenar las filas de forma ascendente y descendente por columnas seleccionando las flechas que acompañan al nombre de las mismas. Este diseño corresponde a la primera transformación XSLT (XML-HTML).

CITE

Home About Us Download Search Contact

Appendix

☒ I
☐ II
☐ III

Class

☒ Reptile
☐ Mammalia
☐ Aves
☐ Actinopteri

Term

☒ Live
☐ Leathers
☐ Skins
☐ Bodies
☐ Specimens

Search

Advanced

Appendix	Taxon	Family	Term	Origin
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text

Showing 1-10 of 30 items

1 2 3

Figure 2.1. Página de búsqueda estándar

Seleccionando el botón 'Advanced' pasaríamos al siguiente diseño: búsqueda avanzada. Desplegando el campo 'All fields' podemos seleccionar una de las 15 columnas disponibles y añadir un valor para la misma en la barra de búsqueda.

Pinchando en el botón 'ADD' se añadirá al campo de búsqueda. La finalidad de esta función es que podamos hacer búsquedas más complejas indicando distintos valores para las columnas. Podemos ver en el campo de búsqueda como sería la sintaxis para filtrar aquellas filas cuya clase sea 'Aves' y cuyo país exportador sea 'NT'. Finalmente y una vez hayamos añadido los valores deseados, pinchamos en botón 'SEARCH' para realizar la búsqueda. Este diseño corresponde a la segunda transformación XSLT (XML-HTML).

Figure 2.2. Página de búsqueda avanzada

Podemos apreciar una flecha dentro del botón 'SEARCH'. Pinchando en ella aparece un desplegable que nos da la opción de contar el número de valores distintos (para las columnas) de todas las filas filtradas. Para este caso hemos contado el número de valores distintos para la clase 'Aves'. Este diseño corresponde a la tercera transformación XSLT (XML-HTML).



All fields ▼

Class:Aves

ADD

COUNT(Class:Aves)

SEARCH ▼

COUNT

Class ▼	Count_Total ▼	Count_Purpose ▼	Count_Source ▼	Count_Terms ▼
Aves	35	74	94	8032
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text
Text	Text	Text	Text	Text

Showing 1-10 of 30 items

1

2

3

Figure 2.2. Página de búsqueda avanzada con COUNT

3

XSLT - XQuery

En este apartado se detallarán el conjunto de transformaciones XSLT y consultas XQUERY realizadas para la manipulación del conjunto de datos.

Debido al formato inicial de nuestro conjunto de datos (CSV) se ha requerido de una transformación inicial a formato XML. Para la transformación inicial CSV-XML se ha desarrollado un pequeño script en python 'Convert-CSV-XML.py'.

```
<?xml version='1.0' encoding='utf-8'?>
<data>
  <row>
    <index>0</index>
    <Year>2016</Year>
    <App.>I</App.>
    <Taxon>Aquila heliaca</Taxon>
    <Class>Aves</Class>
    <Order>Falconiformes</Order>
    <Family>Accipitridae</Family>
    <Genus>Aquila</Genus>
    <Importer>TR</Importer>
    <Exporter>NL</Exporter>
    <Origin>CZ</Origin>
    <Importer_reported_quantity/>
    <Exporter_reported_quantity>1.0</Exporter_reported_quantity>
    <Term>bodies</Term>
    <Unit/>
    <Purpose>T</Purpose>
    <Source>C</Source>
  </row>
```

Figure 3.1. Fragmento XML original

3.1. XSLT

Se han realizado un total de tres transformaciones XSLT. Los objetivos de las transformaciones realizadas han sido facilitar la manipulación y representación del conjunto de datos.

3.1.1. XML-XML

La transformación inicial CSV-XML nos estructura la información en un nodo principal (data) y dentro de este nodos hijos (row) que identifican cada captura y sus atributos asociados. La estructura de la información dentro de cada nodo hijo carece de un agrupamiento en

base al tipo de información que represente. Con el objetivo de mejorar la estructura inicial del documento XML, se ha agrupado la información de cada captura en base al tipo de esta. Se han desarrollado tres categorías ScientificClassification, Location y CommercialInformation. De esta forma, se pretende facilitar su manipulación permitiendo obtener conjuntos de información asociados a categorías específicas de información.

```
<?xml version="1.0" encoding="UTF-8"?>
<CITIES>
  <Capture id="0">
    <ScientificClassification>
      <App>I</App>
      <Taxon>Aquila hellaca</Taxon>
      <Class>Aves</Class>
      <Order>Falconiformes</Order>
      <Family>Accipitridae</Family>
      <Genus>Aquila</Genus>
    </ScientificClassification>
    <Location>
      <Importer>TR</Importer>
      <Exporter>NL</Exporter>
      <Origin>CZ</Origin>
    </Location>
    <CommercialInformation>
      <Imported_reported_quantity/>
      <Exported_reported_quantity/>
      <Term_bodies</Term>
      <Unit/>
      <Purpose>Commercial</Purpose>
      <Source>C</Source>
    </CommercialInformation>
  </Capture>

```

Figure 3.2. Fragmento XML tras transformación XSLT

3.1.2. XML-HTML

Las siguientes consultas tienen como objetivo la búsqueda de información dentro del conjunto de datos y su representación HTML:

- Búsqueda de capturas coincidente con valor específico en cualquiera de las variables. Salida ordenada descendientemente en base a una variable.

Scientific species information

Id	Taxon	Class	Order	Family	Genus
29225	Turbinaria spp.	Anthozoa	Scleractinia	Dendrophylliidae	Turbinaria
29226	Turbinaria spp.	Anthozoa	Scleractinia	Dendrophylliidae	Turbinaria
29227	Turbinaria spp.	Anthozoa	Scleractinia	Dendrophylliidae	Turbinaria
29228	Turbinaria spp.	Anthozoa	Scleractinia	Dendrophylliidae	Turbinaria
29229	Turbinaria spp.	Anthozoa	Scleractinia	Dendrophylliidae	Turbinaria
29230	Turbinaria spp.	Anthozoa	Scleractinia	Dendrophylliidae	Turbinaria
29231	Turbinaria spp.	Anthozoa	Scleractinia	Dendrophylliidae	Turbinaria
29232	Turbinaria spp.	Anthozoa	Scleractinia	Dendrophylliidae	Turbinaria
29233	Turbinaria spp.	Anthozoa	Scleractinia	Dendrophylliidae	Turbinaria

Figure 3.3. Consulta 1

- Búsqueda de capturas coincidentes en base a valores específicos de variables seleccionadas. Salida ordenada descendientemente en base a una variable.

Scientific species information

Id	Taxon	Class	Order	Family	Genus
36557	Trichoglossus haematodus	Aves	Psittaciformes	Loridae	Trichoglossus
53315	Pyrrhura frontalis	Aves	Psittaciformes	Psittacidae	Pyrrhura
53142	Psittacus erithacus	Aves	Psittaciformes	Psittacidae	Psittacus
53239	Psittacus erithacus	Aves	Psittaciformes	Psittacidae	Psittacus
52462	Poicephalus gulielmi	Aves	Psittaciformes	Psittacidae	Poicephalus

Figure 3.4. Consulta 2

- Recuento de los valores distintos del conjunto de variables asociadas a una captura con una variable especificada.

Scientific species information

Class	CountTotal	CountPurpose	CountSource	CountTerms
Aves	6861	12	9	24

Figure 3.5. Consulta 3

3.2. XQuery

Las siguientes consultas tienen como objetivo la búsqueda de información dentro del conjunto de datos:

- Búsqueda de los términos distintos existentes en una variable fijando un/os valor/es en otra/s.
- Recuento del número de capturas con un valor específico de variable
- Búsqueda de toda la información asociada una categoria donde la captura a la que pertenecen cumple unas variables especificadas.
- Búsqueda de todos los elementos de un categoria específica que empiezan con un patrón específico.

4

RDF Schema

Para la realización del esquema RDF hemos usado la herramienta 'Protégé'. El archivo .owl resultante esta disponible en nuestro repositorio de github.

4.1. Archivo owl

Nuestro archivo 'CITES.owl' consta de una clase principal 'Captura', que indica el tipo de todas las instancias creadas. A parte, encontramos las clases 'Especie', 'Medio', 'País', 'Proposito' y 'Term'. Las clases 'País' y 'Especie' constan de subclases que dependen de ellas. Dentro de la primera encontramos 'Exportador', 'Importador' y 'Origen'. Dentro de la segunda encontramos 'Apendice', 'Genus' y 'Taxon'. A su vez, dentro de 'Taxon' encontramos varias clases que atienden a características propias del taxon, como 'Clase', 'Familia', 'Nombre' y 'Orden'. En la siguiente imagen podemos apreciar la jerarquía de clases anteriormente descrita:

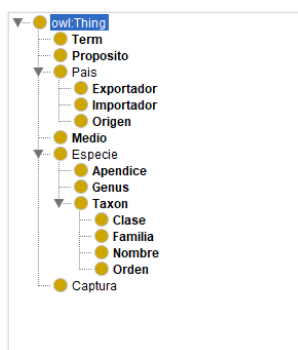


Figure 4.1. Jerarquía clases Protégé

Cada una de estas clases (excepto 'Captura') haría referencia a una columna dentro de nuestro dataset. Para los posibles valores que pueden tener cada una de estas columnas, hemos creado diversas instancias que pertenecen a una o varias (en función de la jerarquía anteriormente descrita) clases. Por ejemplo, para la clase 'Apendice' encontramos tres instancias (1,2 y 3) que hacen referencia al grado de protección de las distintas especies.

Hemos creado un total de 15 relaciones que permiten modelar los datos de nuestra aplicación. La finalidad de estas relaciones es enlazar las clases anteriormente descritas con las distintas instancias. Encontramos, por ejemplo, 'clasetaxon' que indica la clase de taxón de una determinada instancia.

4.2. RDF Schema

En la siguiente imagen podemos ver el esquema principal. En ella no se ven todas las instancias disponibles de cada clase pues el esquema tendría muchas más relaciones y su comprensión sería compleja. Tampoco se ven todas las subclases por la misma razón. Por ejemplo, para la clase 'País' encontramos 3 subclases más.

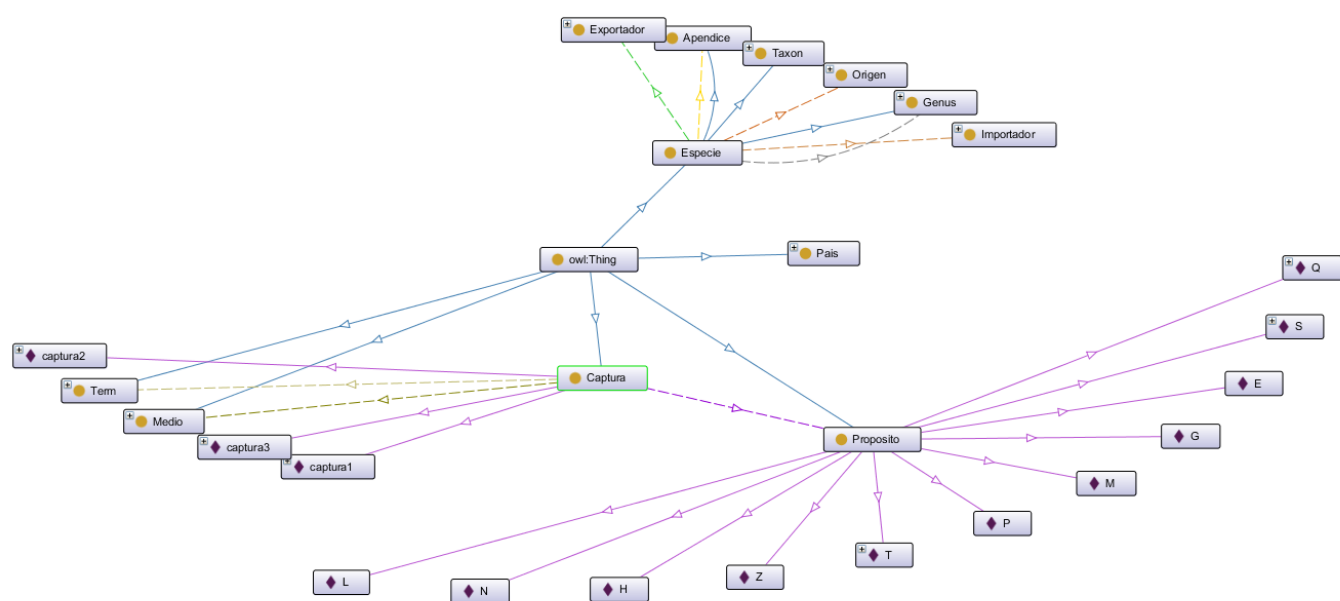


Figure 4.2. Esquema RDF Global

Encontramos también 3 individuos (captura 1, 2 y 3), que hacen referencia a las tres primeras filas de nuestro dataset. Desplegando el primero de ellos podemos ver toda la información asociada y a que clase pertenece dicha información. Para poder ver exactamente las relaciones asociadas al individuo, bastaría con pinchar en las líneas discontinuas. Con el archivo .owl asociado es posible.

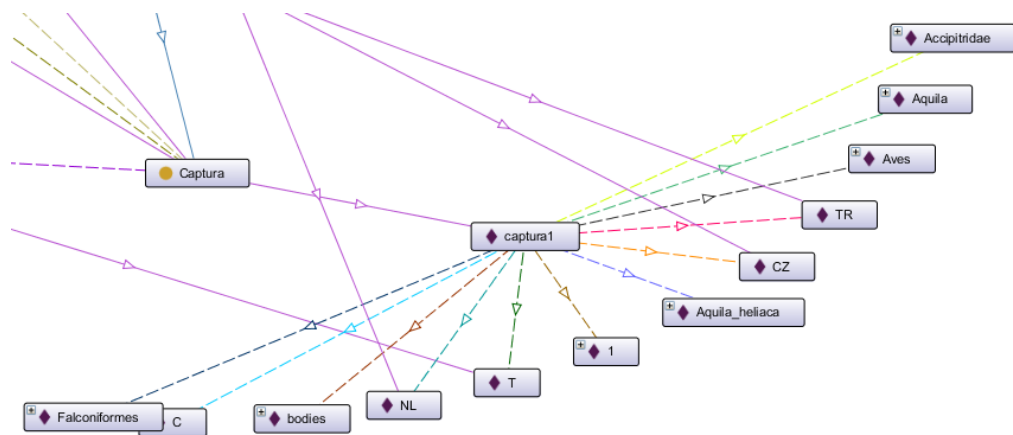


Figure 4.3. Esquema RDF: captura 1

Consultas SPARQL

5.1. Consultas SPARQL

- Búsqueda de capturas coincidente con valor específico en cualquiera de las variables.
Salida ordenada descendientemente en base a una variable.

```
PREFIX cites: <http://www.semanticweb.org/rafalete/ontologies/2022/10/cites#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select ?c ?term ?medio ?pais_exportador ?pais_importador ?pais_origen ?proposito ?genus ?clase_taxon ?familia_taxon
where {
  ?c cites:Captura.
  ?c cites:apendice ?a.
  ?a rdfs:label ?label_a.
  FILTER(?label_a=1)
  ?c cites:term ?term.
  ?c cites:medio ?medio.
  ?c cites:pais_exportador ?pais_exportador.
  ?c cites:pais_importador ?pais_importador.
  ?c cites:genus ?genus.
  ?c cites:clase_taxon ?clase_taxon.
  ?c cites:proposito ?proposito.
  ?c cites:familia_taxon ?familia_taxon.
  ?c cites:pais_importador ?pais_importador.
} ORDER BY DESC(?medio)
```

Figure 5.1 Consulta SPARQL 1

El resultado sería el siguiente:

	c	term	medio	pais_exportador	pais_importador	pais_origen	proposito	genus	clase_taxon	familia_taxon
1	cites:captura3	cites:feathers	cites:W	cites:DE	cites:CH	cites:CH	cites:S	cites:Acipenser	cites:Actinopteri	cites:Acipenseridae
2	cites:captura2	cites:bodies	cites:O	cites:NO	cites:BE	cites:BE	cites:Q	cites:Haliaeetus	cites:Aves	cites:Accipitridae
3	cites:captura1	cites:bodies	cites:C	cites:NL	cites:TR	cites:TR	cites:T	cites:Aquila	cites:Aves	cites:Accipitridae

Figure 5.2 Resultado Consulta SPARQL 1

- Recuento de los valores distintos del conjunto de variables asociadas a una captura con una variable especificada.

```

PREFIX cites: <http://www.semanticweb.org/rafalete/ontologies/2022/10/cites#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select (count(distinct ?term) as ?count_term) (count(distinct ?medio) as ?count_medio) (count(distinct ?pais_exportador) as ?count_pais_exportador) (count(distinct ?pais_importador) as ?count_pais_importador)
where {
  ?c a cites:Captura.
  ?c cites:apendice ?a.
  ?a rdfs:label ?label_a.
  FILTER(?label_a=1).
  ?c cites:term ?term.
  ?c cites:medio ?medio.
  ?c cites:pais_exportador ?pais_exportador.
  ?c cites:pais_importador ?pais_importador.
  ?c cites:genus ?genus.
  ?c cites:clase_taxon ?clase_taxon.
  ?c cites:proposito ?proposito.
  ?c cites:familia_taxon ?familia_taxon.
  ?c cites:pais_importador ?pais_importador.
}

```

Figure 5.3 Consulta SPARQL 1

El resultado sería el siguiente:

	count_term	count_medio	count_pais_exportador	pais_importador	pais_origen	proposito	count_genus	count_clase_taxon	familia_taxon
1	*2**xsd:integer	*3**xsd:integer	*3**xsd:integer	*3**xsd:integer	*3**xsd:integer	*3**xsd:integer	*3**xsd:integer	*2**xsd:integer	*2**xsd:integer

Figure 5.4 Resultado Consulta SPARQL 1

5.2. Consultas SPARQL Extra

- Búsqueda de las capturas, cuya cantidad reportada por el país exportador sea mayor a uno, ordenadas descendientemente en base a la cantidad reportada.

```

Select distinct ?c ?cantidad ?term ?medio ?pais_exportador ?pais_importador
Where {
  ?c a cites:Captura.
  ?c cites:cantidad_exportador ?cantidad.
  filter(?cantidad > 1)
  ?c cites:term ?term.
  ?c cites:medio ?medio.
  ?c cites:pais_exportador ?pais_exportador.
  ?c cites:pais_importador ?pais_importador.
} Order by desc (?cantidad)

```

Figure 5.5 Consulta SPARQL Extra 1

	c	cantidad	term	medio	pais_exportador
1	cites:captura2	*43**xsd:integer	cites:bodies	cites:O	cites:NO
2	cites:captura3	*4**xsd:integer	cites:feathers	cites:W	cites:DE

Figure 5.6 Resultado consulta SPARQL Extra 1

- Búsqueda de las capturas cuya variable orden taxon comience por "F".

```
Select distinct ?c ?orden_taxon ?term ?medio ?pais_exportador ?pais_importador
Where {
  ?c a cites:Captura.
  ?c cites:orden_taxon ?orden_taxon.
  ?orden_taxon rdfs:label ?label
    filter(STRSTARTS(?label, "F")).
  ?c cites:term ?term.
  ?c cites:medio ?medio.
  ?c cites:pais_exportador ?pais_exportador.
  ?c cites:pais_importador ?pais_importador.
}
```

Figure 5.7 Consulta SPARQL Extra 2

	c	orden_taxon	term	medio	pais_exportador
1	cites:captura1	cites:Falconiformes	cites:bodies	cites:C	cites:NL
2	cites:captura2	cites:Falconiformes	cites:bodies	cites:O	cites:NO

Figure 5.8 Resultado consulta SPARQL Extra 2



UNIVERSIDAD
DE MÁLAGA

| **uma.es**

E.T.S. DE INGENIERÍA INFORMÁTICA

E.T.S de Ingeniería Informática
Bulevar Louis Pasteur, 35
Campus de Teatinos
29071 Málaga