



Carrera:

Sistemas Informáticos

Componente:

Sistemas basados en Conocimiento

Integrantes:

Diego Fernando Pinto Orellana

Bryant David Portilla Guamán

Fecha:

28-07-2020

Contenido

Introducción	3
Objetivos	4
Objetivo General	4
Objetivos específicos	4
Contenidos	5
Especificación de la ontología	5
Construcción de glosario de términos	6
Construir taxonomía de conceptos	8
Construir diagrama de relaciones binarias	8
Describir las relaciones binarias ad hoc	8
Describir atributos de instancia	9
Transformación de datos RDF	10
Extracción de datos	10
Limpieza de datos	11
Selección de URIS	11
Generación Jena	12
Estadísticas datos generados	15
Aplicaciones relacionadas	16
Aplicación propuesta	19
Tecnologías	19
Consultas SPARQL	20
Visualización (Capturas de la aplicación)	22
Conclusiones	24
Bibliografía	25

1. Introducción

A lo largo de estos años ha ido variando la web desde su primera versión la 1.0 la misma que se basaba en que las personas puedan intercambiar información en la web, luego da un paso a la web 2.0 se describe como la segunda generación de World Wide Web (WWW) donde las páginas estáticas son reemplazadas por páginas interactivas para una experiencia dinámica para los usuarios, la Web 2.0 es creada con el objetivo de permitir a las personas puedan compartir su información en la red a través de las plataformas de redes sociales, blogs, chats y poder contribuir en la web

La Web 3.0 es el tercer nivel de desarrollo de la web a lo que se denomina la Web Semántica (WS) la que demuestra la naturaleza semántica de los contenidos. Tim Berners-Lee fue el mentor del término de web semántica lo que nos dice que los datos que se encuentran en la Web es información que se la puede tratar y buscar relaciones entre ellas con situaciones reales.

Tim Berners-Lee el padre de la web dice “La Web Semántica es una extensión de la web actual en la que la información tiene un significado bien definido, permitiendo que las computadoras y las personas trabajen mejor en cooperación”, por consiguiente, menciono que la Web Semántica es una red de datos que tiene una relación en búsqueda de tener un trabajo más útil para resolver problemas de manera rápida y oportuna.

El presente trabajo describe todo el proceso necesario para el desarrollo de una aplicación web que permita la visualización de datos bibliométricos sobre publicaciones COVID-19 mediante consultas SparQL hacia un repositorio semántico “GraphDB”. Para lo cual fue necesario en primer lugar realizar la creación de un modelo ontológico, extracción y limpieza de datos para poder realizar la generación de datos RDF. Los repositorios utilizados en el presente proyecto fueron:

- Generación de datos RDF:
<https://github.com/diepinto30/GenerateBibliometricDataCovid>
- Aplicación web: <https://github.com/bdportilla1/BiblioCovid>

2. Objetivos

2.1. Objetivo General

- Desarrollar una aplicación que permita visualizar e interactuar con información de datos RDF generados a partir de un conjunto de dataset recolectada en relación a las investigaciones sobre COVID-19 o SARS-COV-2.

2.2. Objetivos específicos

- Crear un modelo ontológico relacionada a publicaciones COVID-19.
- Extraer metadatos relacionados a publicaciones COVID-19.
- Generar datos RDF con base a un modelo ontológico sobre publicaciones COVID-19.
- Realizar consultas SparQL hacia un repositorio semántico con la finalidad de visualizar los datos a través de gráficas y tablas estadísticas.

3. Contenidos

3.1. Especificación de la ontología

En esta sección se describe la especificación de la ontología que se usará para los datos RDF de publicaciones COVID-19.

Datos Bibliométricos de publicaciones Covid-19	
1. Purpose	El propósito de la construcción de la ontología referente a las publicaciones de covid-19, es para proveer un modelo de conocimiento para aquellas personas, organizaciones e instituciones que requieran esta información semántica.
2. Scope	La ontología tiene que centrarse solo en el dominio de las publicaciones científicas sobre covid-19. El nivel de granularidad está directamente relacionado con las preguntas de competencia y términos identificados.
3. Implementation Language	La implementación de la ontología será en el lenguaje OWL, haciendo uso de la herramienta Protege
4. Intended End-Users	<ul style="list-style-type: none">• Usuario 1: empresas públicas y privadas que requieran analizar datos sobre las publicaciones COVID-19• Usuario 2: universidades que deseen información para potenciar la información sobre publicaciones COVID-19• Usuario 3: grupos de investigación• Usuario 4: aplicaciones agentes para el procesamiento
5. Intended Uses	<ul style="list-style-type: none">• Uso 1: Publicación de datos abiertos• Uso 2: Gráficas estadísticas. (estadística descriptiva e inferencial)• Uso 3: Grafo de conocimiento
6. Ontology Requirements	
a. Non-Functional Requirements	<p>NFR1. La ontología debe admitir un escenario multilingüe en los siguientes idiomas: inglés y español.</p> <p>NFR2. La ontología debe basarse en los estándares internacionales, europeos o de facto existentes o en desarrollo.</p>
b. Functional Requirements: Groups of Competency Questions	<ol style="list-style-type: none">1. ¿Cuáles son los nombres de las publicaciones y el número de citas?2. ¿Cuáles son los idiomas en los que están escritas las publicaciones?3. ¿Cuáles son los nombres de las publicaciones y el doi, issn?4. ¿Cuál es el tipo de documento que pertenecen las publicaciones?5. ¿Cuál es el rank del source al que pertenece cada publicación?6. ¿Cuáles son los nombres de las categorías de cada source?

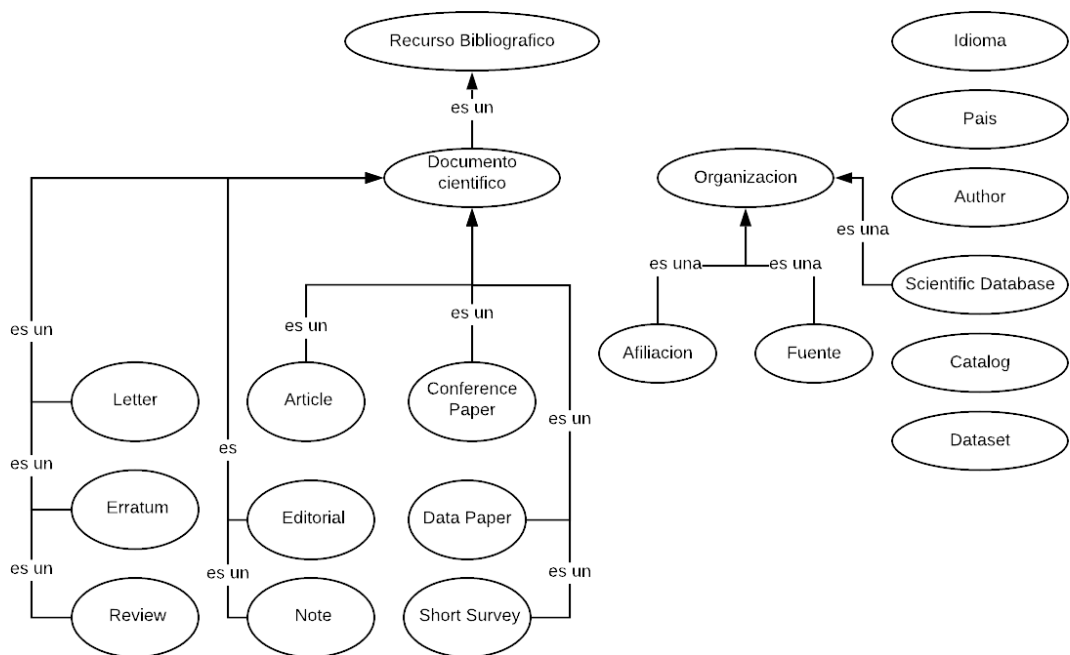
7. Cuales son los cuartiles de cada categoría del source
8. ¿Cuales son las palabras más mencionadas en los títulos de las publicaciones?
9. ¿Cual es el número de publicaciones y el idioma al que pertenece?
10. ¿Cual es el nombre del país y el nombre del Source de las publicaciones?
11. ¿Cuales son los nombres de las categorías de las publicaciones?
12. ¿Cuáles son los nombres de las fuentes y la cantidad de publicaciones que posee acerca de covid-19?
13. ¿Cuáles son los nombres de los autores y el nombre del país al que pertenece?
14. ¿Cuáles son los nombres de los países con mayores publicaciones de COVID19?
15. ¿Cuáles son los nombres de las publicaciones que pertenecen al Source X?
16. ¿Cuáles son los nombres de las publicaciones que son de "x" tipo documento?
17. ¿Cuáles son los nombres de los autores que poseen "x" # de publicaciones COVID?

3.1.1. Construcción de glosario de términos

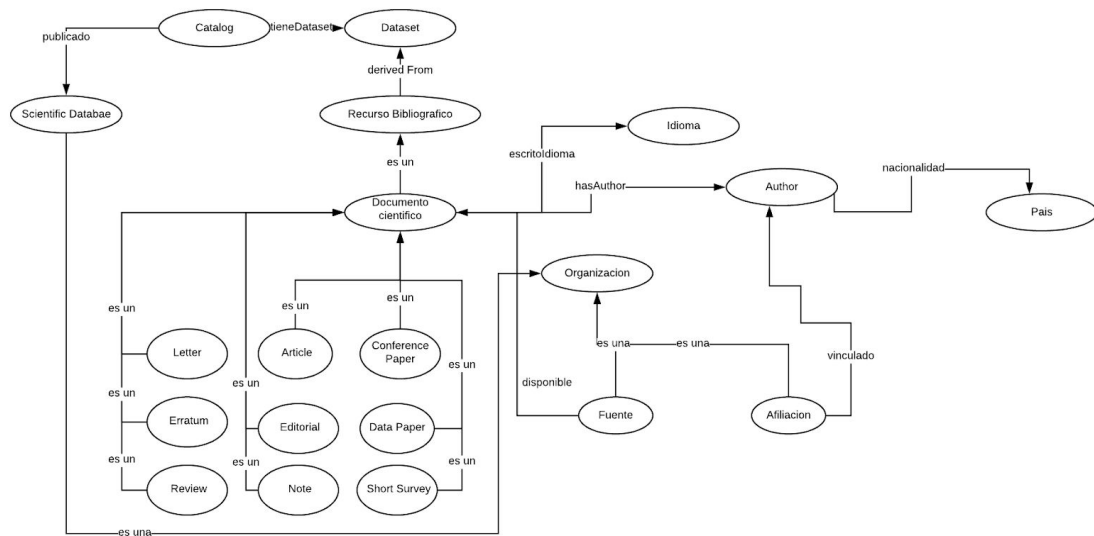
Nombre	Sinónimos	Acrónimos	Descripción	Tipo
Recurso Bibliográfico	Bibliografía científica	BRd	Cualquier documento perteneciente a una base de datos científica	Concepto
title	-	-	El nombre de la publicación	Atributo de instancia
authors	-	-	El nombre de los autores de la publicación	Atributo de instancia
keywords	-	-	Palabras clave de la publicación	Atributo de instancia
year	-	-	Año en el que se realizó la publicación	Atributo de instancia
Source title	-	-	Nombre de la categoría a la que pertenece la publicación	Atributo de instancia
Cite by	-	-	Número de veces que se citó la publicación	Atributo de instancia
publisher	-	-	Nombre de la editorial	Atributo de instancia
link	Dirección	URL	Dirección web de la publicación	Atributo de instancia

publicado			Describe en qué revista se publicó el recurso	Relación
disponible			Describe en donde está disponible el recurso	Relación
hasAuthor			Describe el autor del recurso	Relación
escritoldioma			Describe el idioma del recurso	Relación
Tipo de Documento	Tipo de registro	DTd	Se refiere al tipo de documento científico (Article, Conference Paper y entre otros)	Concepto
name	-	-	Nombre científico del documento	Atributo de instancia
Lenguaje	Idioma	LG	Se refiere al idioma original del documento	Concepto
name	-	-	Nombre del idioma	Atributo de instancia
Fuente			Organización que publica los recursos	Concepto
name	-	-	Nombre de base datos científica	Atributo de instancia
Afiliación	-	-	Revista en la que está disponible el recurso	Concepto
name	-	-	Nombre de la revista	Atributo de instancia
Author	-	-	Persona o personas propietario del recurso	Concepto
names	-	-	Nombre del autor	Atributo de instancia
nacionalidad			Describe la nacionalidad de una persona	Relación
País	-	-	País de origen de la persona	Concepto
name	-	-	Nombre del país	Atributo de instancia

3.1.2. Construir taxonomía de conceptos



3.1.3. Construir diagrama de relaciones binarias



3.1.4. Describir las relaciones binarias ad hoc

Nombre de la relación	Concepto de origen dominio	Cardinalidad máxima	Concepto destino rango
publicada	Documento científico	N	Organizacion
escribididioma	Documento científico	1	Idioma

hasAuthor	Documento científico		Author
disponible	Fuente		Documento científico
vinculado	Afiliacion		Author
nacionalidad	Author		País
Derivado de	Recurso Bibliografico		Dataset
dataset	Catalog		Dataset
publicado	Catalog		Scientific Database

3.1.5. Describir atributos de instancia

Nombre del atributo	Concepto	Tipo valor	Rango de valores	Cardinalidad
title	Recurso Bibliográfico	Cadena de caracteres	-	1,1
authors	Recurso Bibliográfico	Cadena de caracteres	-	1,N
keywords	Recurso Bibliográfico	Cadena de caracteres	-	1,N
year	Recurso Bibliográfico	Cadena de caracteres	-	1,1
issn	Recurso Bibliográfico	Cadena de caracteres	-	1,1
doi	Recurso Bibliográfico	Cadena de caracteres	-	1,1
vol	Recurso Bibliográfico	Cadena de caracteres	-	1,1
Source title	Recurso Bibliográfico	Cadena de caracteres	-	1,1
rank	Recurso Bibliográfico	Cadena de caracteres	-	1,1
quartile	Recurso Bibliográfico	Cadena de caracteres	-	1,1
citeCOUNT	Recurso Bibliográfico	Cadena de caracteres	-	1,N
publisher	Recurso Bibliográfico	Cadena de caracteres	-	1,1
link	Recurso Bibliográfico	Cadena de caracteres	-	1,1
name	Tipo de Documento	Cadena de caracteres	-	1,1

name	Lenguaje	Cadena de caracteres	-	1,1
name	Fuente	Cadena de caracteres	-	1,1
name	Scientific Database	Cadena de caracteres	-	1,1
title	Database	Cadena de caracteres	-	1,1
keywords	Database	Cadena de caracteres	-	1,1
title	Catalog	Cadena de caracteres	-	1,1

3.2. Transformación de datos RDF

En esta sección se describe los pasos necesarios desde la extracción de datos sobre publicaciones COVID-19 hasta la generación de datos RDF con Jena, un framework para la construcción de ontologías.

3.2.1. Extracción de datos

La fuente de datos en donde se extrajeron la información es SCOPUS, una base de datos científica en donde se puede encontrar publicaciones en diferentes categorías. La cadena de búsqueda para el presente trabajo fue **“COVID-19 AND SARS-COV-2”**.

La última fecha de obtención de los datos fue el 20/05/2020.

Para obtener una mejor estructura de datos se optó por descargar diferentes archivos sobre el mismo conjunto de publicaciones pero en cada archivo debemos especificar los campos a obtener. Teniendo un total de seis archivos como se muestra en la siguiente tabla.

# Archivo	Parámetros
archivo_1	id_autores.
archivo_2	nombre_autores.
archivo_3	Document title, Year, EID, Source title.
archivo_4	volume, issue, pages, Citation count, Source & document type
archivo_5	Publication Stage, DOI, Access Type, Serial identifiers (e.g. ISSN), Language.
archivo_6	Author keywords.

Para obtener metadatos necesarios sobre los ranks y cuartiles de las categorías de las publicaciones fue necesario obtener un dataset con esta información desde SCIMAGO.

3.2.2. Limpieza de datos

Una vez extraídos se procede la parte más fuerte que es la limpieza de los datos, en esta parte se realiza múltiples actividades en nuestro caso se hizo la *limpieza* y *validación* de los datos que se extrajo.

Cabe mencionar que todas las actividades de limpieza se realizaron desde el archivo excel. Las actividades generales de limpieza fueron:

- Eliminar caracteres especiales de cada registro.
- Agrupar campos en una sola celda: nombre de autores, id de autores, keywords.
- Estructurar filas de acuerdo a los campos.
- Incluir los datos estructurados de cada archivo en un archivo principal.

En la siguiente figura se observa un fragmento de los datos estructurados en la tabla principal.

idArtículo	Authors	Authors_ID	Title	EID	Year	Source	Volume	Issue	Art. N°	Page	Page end	Cited by	DOI
BiblioData	Byass P.	7006527895;	Eco-epidemiological asses 2-s2.0-8508		2020	Global H	13	1	1760490				10.1080/165
BiblioData	Liu P., Cai J., Jia R., Xia S.	57200337904;572	Dynamic surveillance of S/ 2-s2.0-8508		2020	Emergin	9	1		1254	1258		10.1080/222
BiblioData	Islam H., Rahman A., Ma	57195434373;572	A generalized overview of 2-s2.0-8508		2020	Electroni	17	6	em251				10.29333/ej
BiblioData	Wang C., Li W., Drabek D	57216967065;572	A human monoclonal antil 2-s2.0-8508		2020	Nature C	11	1	2251				12 10.1038/s41
BiblioData	Lau S.-Y., Wang P., Mok	57208238618;545	Attenuated SARS-CoV-2 vi 2-s2.0-8508		2020	Emergin	9	1		837	842	1	10.1080/222
BiblioData	Beloncle F.M., Pavlovsky	57216557219;572	Recruitability and effect o 2-s2.0-8508		2020	Annals o	10	1	55				10.1186/s13
BiblioData	Grimaud M., Starck J., Le	55043167800;572	Acute myocarditis and mu 2-s2.0-8508		2020	Annals o	10	1	69				10.1186/s13
BiblioData	Eslami H., Jalili M.	6506092327;5721	The role of environmental 2-s2.0-8508		2020	AMB Exp	10	1	92				10.1186/s13
BiblioData	Suo T., Liu X., Feng J., Gu	57193390916;572	ddPCR: a more accurate tc 2-s2.0-8508		2020	Emergin	9	1		1259	1268		10.1080/222
BiblioData	Yang X., Dong N., Chan E	57217073220;572	Genetic cluster analysis of 2-s2.0-8508		2020	Emergin	9	1		1287	1299		10.1080/222
BiblioData	Maitra A., Sarkar M.C., R	57217125767;572	Mutations in SARS-CoV-2 2-s2.0-8508		2020	Journal c	45	1	76				10.1007/s12

Por otro lado también fue necesario estructurar la data de SCIMAGO ya que el archivo extraído desde la fuente poseía campos que no son necesarios en nuestro caso. Eliminando las columnas que no eran de nuestro interés, obtuvimos un solo archivo con los siguientes parámetros: rank, idSource, ISSN, título, categorías(quartiles).

3.2.3. Selección de URIS

URIS de los prefijos usados para clases o predicados.

- foaf: <http://xmlns.com/foaf/0.1/>
- fabio: <http://purl.org/spar/fabio/>
- dct: <http://purl.org/dc/terms/>
- rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

- rdfs: <http://www.w3.org/2000/01/rdf-schema#>
- schema: <https://schema.org/>
- mydata: <http://utpl.edu.ec/lod/biblioCOVID/data/>
- xsd: <http://www.w3.org/2001/XMLSchema#>

En cuanto a la generación de URIS para los recursos.

Recurso	URI
Base	http://utpl.edu.ec/COVIDBiblio/ontology/
Author	http://utpl.edu.ec/COVIDBiblio/ontology/Author/ + ID_AUTOR
Scientific Database	http://utpl.edu.ec/COVIDBiblio/ontology/ScientificDatabase/ + NOMBRE FUENTE
Catalog	http://utpl.edu.ec/COVIDBiblio/ontology/Catalog/ + ID_CATALOGO
Dataset	http://utpl.edu.ec/COVIDBiblio/ontology/Dataset/ + ID_DATASET
Documento (Articulo, Letter, ConferencePaper)	http://utpl.edu.ec/COVIDBiblio/ontology/BibliographicResource/ + EID DOCUMENTO
SourceT	http://utpl.edu.ec/COVIDBiblio/ontology/Source/ + ISSN
Quartile	http://utpl.edu.ec/COVIDBiblio/ontology/ + ISSN+"_" + CATEGORIA

3.2.4. Generación Jena

Para la generación de los datos de csv a rdf se utilizó Jena, ya que este tiene múltiples funciones que facilita todo este proceso de creación de datos RDF. En el siguiente link se adjunta el repositorio donde se aprecia cada una de instancia usadas para el desarrollo del proyecto: <https://github.com/diepinto30/GenerateBibliometricDataCovid>

El proceso para la generación de datos RDF consiste en primer lugar hacer lectura de los datos estructurados del archivo principal en donde se obtiene todos los campos y se almacena en un arreglo, así mismo el archivo de scimago. Cada fila hace referencia a una publicación.

Dentro de la codificación es necesario incluir todos los modelos ontológicos externos al framework Jena así mismo nuestra URI base que representa nuestra ontología.

```
String dataPrefix = "http://utpl.edu.ec/COVIDBiblio/ontology/";
model.setNsPrefix("myData", dataPrefix);
Model myOntoModel = ModelFactory.createDefaultModel();

String dbo = "http://dbpedia.org/ontology/";
model.setNsPrefix("dbo", dbo);
Model dboModel = ModelFactory.createDefaultModel();

String dbr = "http://dbpedia.org/resource/";
model.setNsPrefix("dbr", dbr);
Model dbrModel = ModelFactory.createDefaultModel();
```

En la figura se muestra la estructura para la creación de nuevos modelos en donde establecemos la URI de cada uno de ellos. El siguiente proceso es recorrer el arreglo de scimago y crear los “SourceT” que se obtuvieron de la fuente.

```
// CREACION DE SOURCES
for (int l = 0; l < parts_issn.length; l++) {
    Source source = new Source(atributosS[k][1], atributosS[k][2], parts_issn[l], atributosS[k][4],
        atributosS[k][5], categorias, rank, atributosS[k][7], atributosS[k][8]);
    String URI_SourceT = dataPrefix+"Source/" + source.getIssn();
    Resource sourceT = model.createResource(URI_SourceT)
        .addProperty(RDF.type, myOntoModel.getResource(dataPrefix + "SourceT"))
        .addProperty(prismModel.getProperty(prism + "issn"), source.getIssn())
        .addProperty(DCTerms.identifier, source.getSource_id())
        .addProperty(DCTerms.title, source.getTitle())
        .addProperty(dboModel.getProperty(dbo+"country"), dbrModel.getResource(dbr +source.getCountry()).getURI())
        .addProperty(myOntoModel.getProperty(dataPrefix + "rank"), source.getRank());
    for(int contC=0; contC<source.getCategorias().size(); contC++){
        Resource quartile = model.createResource(dataPrefix+ source.getIssn()+"_" +source.getCategorias().get(contC));
        quartile.addProperty(RDF.type, bidoModel.getResource(bido + "Quartile"))
            .addProperty(DCTerms.title, source.getCategorias().get(contC))
            .addProperty(myOntoModel.getProperty(dataPrefix + "quartile"), source.getQuartile().getURI())
            .addProperty(RDFS.subClassOf, bidoModel.getResource(bido + "QuartileCategory"));
        sourceT.addProperty(bidoModel.getProperty(bido + "hasQuartile"), quartile);
    }
}
```

En la figura se muestra la creación de recursos de tipo “SourceT” haciendo uso del “issn” como identificador de cada recurso. Por cada registro también se creó recursos de tipo “Quartile” que hace referencia a las categorías del “SourceT”.

Una vez agregado los datos del archivo de Scimago se procede a crear los recursos, propiedades y relaciones del archivo principal en donde se tiene la información de los metadatos de cada publicación.

Al igual que en Schimago se procede a recorrer el arreglo en donde se guardaron los campos del archivo principal.

En primer lugar se creó los recursos de “Scientific Database”, “Dataset” y “Catalog”.

```
// CREACION DE SCIENTIFIC DATABASE
Resource fuenteDocumento = model.createResource(URI_FUENTE)
    .addProperty(RDF.type, myOntoModel.getResource(dataPrefix + "/ScientificDatabase"))
    .addProperty(FOAF.name, fuente)
    .addProperty(RDFS.subClassOf, FOAF.Organization);

//CREACION DEL DATASET
//String URIDataset = dataPrefix + "Dataset/BiblioDataCovid/";
Resource datasetInfo = model.createResource(URI_DATASET)
    .addProperty(RDF.type, dcatModel.getResource(dcat + "Dataset")
    .addProperty(DCTerms.title, "BiblioDataCovid")
    .addProperty(dcatModel.getProperty(dcat + "keyword"), ("covid19; sars-cov-2"))
    .addProperty(DCTerms.modified, "10-06-2020"));

//CREACION DEL CATALOG
//String URICatalog = dataPrefix+"Catalog/CatalogScopusCOVID/";
Resource catalog = model.createResource(URI_CATALOG)
    .addProperty(DCTerms.title, "CatalogScopusCOVID")
    .addProperty(RDF.type, dcatModel.getResource(dcat + "Catalog")
    .addProperty(DCTerms.publisher, fuenteDocumento)
    .addProperty(dcatModel.getProperty(dcat + "dataset"), datasetInfo ));
```

En la figura se muestra la creación de los recursos mencionados así mismo la relación entre ellos. Después se procede a crear los recursos bibliográficos.

```
// CREACION DE DOCUMENTO BIBLIOGRAFICO
Resource documento = model.createResource(URI_DOCUMENTO)
    .addProperty(DCTerms.title, titulo)
    .addProperty(DCTerms.date, anio)
    .addProperty(myOntoModel.getProperty(dataPrefix + "citationsCount"), num_citas)
    .addProperty(DCTerms.language, dbcModel.getResource(dbr+language))
    .addProperty(RDFS.subClassOf, fabioModel.getResource(fabio+ "ScholarlyWork"))
    .addProperty(prismModel.getProperty(prism + "doi"), doi)
    .addProperty(prismModel.getProperty(prism + "issn"), issn)
    .addProperty(prismModel.getProperty(prism + "volume"), vol)
    .addProperty(provModel.getProperty(prov + "wasDerivedFrom"), datasetInfo)
    .addProperty(bidoModel.getProperty(bido+ "withBibliometricData"), myOntoModel.getResource(dataPref

// Se crea el tipo de documento
// Se compara el tipo con los de fabio y si es igual toma la uri de fabio
tipo = false;
for (String nombre : tipos_Fabio) {
    if(document_type.replace(" ", "").equals(nombre)){
        documento.addProperty(RDF.type, fabioModel.getResource(fabio + document_type.replace(" ", "")));
        tipo = true;
        break;
    }
}
if(!tipo){
    documento.addProperty(RDF.type, myOntoModel.getResource(dataPrefix + document_type.replace(" ", "")));
}
```

En la figura se muestra la creación de recursos bibliográficos con sus respectivas propiedades. Se utilizó el parámetro “EID” para establecer la URI del recurso. También se incluye la lógica que se utilizó para determinar el tipo de documento que es el recurso. Después se realiza la creación de los recursos de tipo Autor.

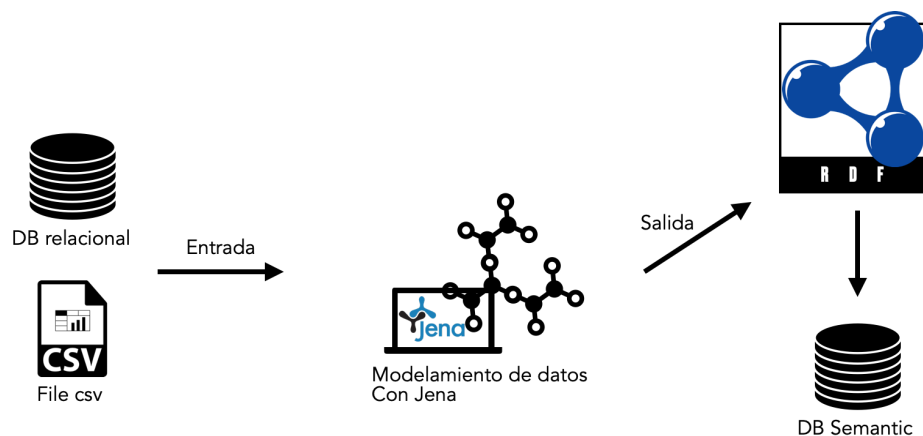

```

for (int j = 0; j < parts_id_autores.length; j++) {
    // CREANDO INSTANCIAS DE LOS AUTORES
    int long_parts_nombres = parts_nombres.length;
    String URI_AUTOR = dataPrefix+ "Author/" + parts_id_autores[j];
    Resource autor = model.createResource(URI_AUTOR)
        .addProperty(RDF.type, myOntoModel.getResource(dataPrefix+"Author"))
        .addProperty(RDFS.subClassOf, FOAF.Person);
    if (j<long_parts_nombres) {
        autor.addProperty(FOAF.name, parts_nombres[j]);
    }
    // Vinculando el autor al documento
    documento.addProperty(DCTerms.creator, autor);
}

```

En la figura se muestra la creación de los Autores estableciendo el campo “**id_Autor**” como identificador para la URI de cada autor. Y finalmente relacionamos cada autor con el documento que se ha creado.

3.2.5. Estadísticas datos generados







Para almacenar los datos RDF creados es necesario el uso de un repositorio semántico. En el presente trabajo se utilizó GraphDB para el almacenamiento de los datos. En donde podemos apreciar los siguientes datos generados.

Class	Links	
myData:SourceT	103K	⇌ -
bido:Quartile	102K	← -
myData:Author	14K	← -
fabio:Article	10K	→ -
fabio:Review	4K	→ -
fabio:Letter	865	→ -
fabio:Editorial	512	→ -
myData:Note	448	→ -
fabio:ConferencePaper	116	→ -
myData:ShortSurvey	93	→ -
fabio:Erratum	24	→ +

Número de tripletas generadas.

Local


bcovid • bcovid




total statements
837,539

837,443 explicit
96 inferred
1.00 expansion ratio

[Import RDF data](#)
[Import tabular data with OntoRefine](#)
[Export RDF data](#)

3.3. Aplicaciones relacionadas

Web of Science (WOS)

Constituye una de las principales herramientas de búsqueda bibliográfica que existen en el mundo. Más que de una base de datos se trata de un conjunto integrado de ellas con una gran cantidad de prestaciones para el investigador, el docente, los alumnos y las autoridades académicas. Wos permite el acceso a los siguientes recursos:

Colección Principal de Web of Science:

Índices de citas

- SciELO Citation Index --1997-presente
- Science Citation Index Expanded (SCI-EXPANDED) --1900-presente
- Social Sciences Citation Index (SSCI) --1956-presente
- Arts & Humanities Citation Index (A&HCI) --1975-presente
- Conference Proceedings Citation Index- Science (CPCI-S) --1990-presente
- Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH) --1990-presente
- Emerging Sources Citation Index (ESCI) --2015-presente

Scholarly publishing in the Internet age: a citation analysis of computer science literature

Por: Goodrum, AA (Goodrum, AA); McCain, KW (McCain, KW); Lawrence, S (Lawrence, S); Giles, CL (Giles, CL)

INFORMATION PROCESSING & MANAGEMENT
Volumen: 37 Número: 5 Páginas: 661-675

INFORMATION PROCESSING & MANAGEMENT

Impact Factor
1.265 **1.469**
2014 5 años

Categoría de JCR®	Clasificación en la categoría	Cuartil en la categoría
COMPUTER SCIENCE, INFORMATION SYSTEMS	53 de 139	Q2
INFORMATION SCIENCE & LIBRARY SCIENCE	24 de 85	Q2

Datos de la edición 2014 de Journal Citation Reports®

Editorial
PERGAMON-ELSEVIER SCIENCE LTD, THE BOULEVARD, LANGFORD LANE,
KIDLINGTON, OXFORD OX5 1GB, ENGLAND

ISSN: 0306-4573

Dominio de investigación
Computer Science
Information Science & Library Science

Cerrar ventana

Red de citas

91 Veces citado
24 Referencias citadas
[Ver Related Records](#)
[Ver mapa de citas](#)
[Crear alerta de cita](#)
(datos de Colección principal de Web of ScienceTM)

Número de todas las veces citado
91 en Todas las bases de datos
91 en Colección principal de Web of Science
8 en BIOSIS Citation Index
0 en Chinese Science Citation Database
0 en Data Citation Index
0 en Russian Science Citation Index
0 en SciELO Citation Index

Conteo de uso
Últimos 180 días: 4
Desde 2013: 28
[Más información](#)

Cita más reciente
Fairclough, Ruth. National research impact indicators from Mendeley readers. JOURNAL OF INFORMETRICS. OCT 2015.

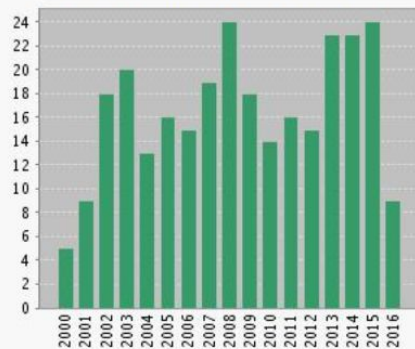
Informe de citas: 281

(de Todas las bases de datos)

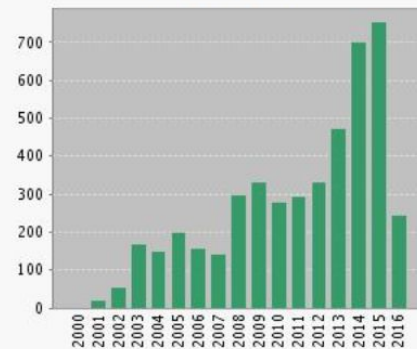
Buscó: **Autor:** (Thelwall, M.) [...Más](#)

Este informe refleja las citas de los elementos origen indexados en todas las bases de datos.

Elementos publicados cada año



Citas cada año



Resultados encontrados: 281

Total de veces citado [?] : 4612

Total de veces citado sin citas propias [?] : 3367

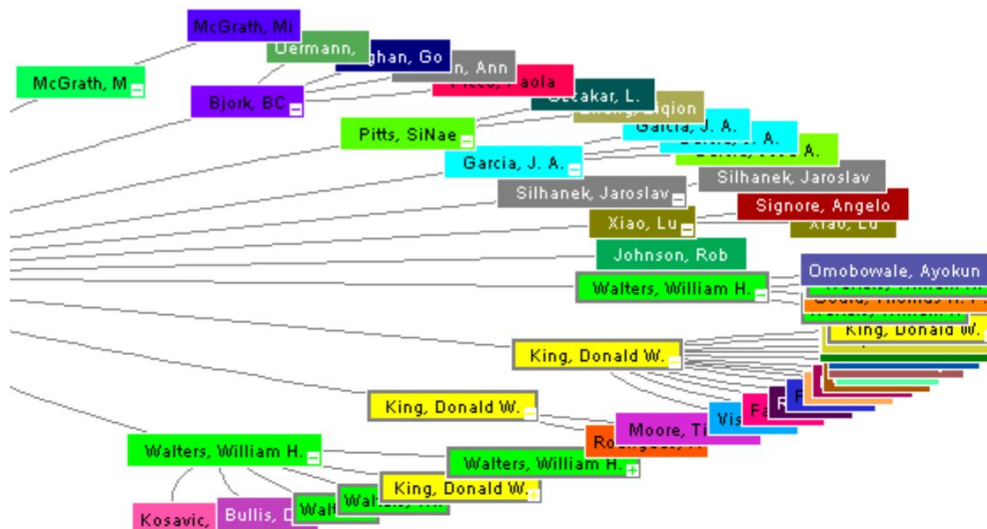
Artículos en que se cita [?] : 2264

Artículos totales en que se cita sin citas propias [?] : 2031

Promedio de citas por elemento [?] : 16.41

h-index [?] : 36

		2012	2013	2014	2015	2016	Total	Promedio de citas por año
Use las casillas de verificación para eliminar elementos individuales de este informe de citas o restrinja la búsqueda a elementos publicados entre <input type="text" value="2000"/> y <input type="text" value="2016"/> <input type="button" value="Ir"/>		334	475	703	753	247	4612	271.29
<input type="checkbox"/>	1. Sentiment in Short Strength Detection Informal Text Por: Thelwall, Mike; Buckley, Kevan; Paltoglou, Georgios; et al. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY Volumen: 61 Número: 12 Páginas: 2544-2558 Fecha de publicación: DEC 2010	16	28	47	53	21	171	24.43
<input type="checkbox"/>	2. Sentiment in Twitter Events Por: Thelwall, Mike; Buckley, Kevan; Paltoglou, Georgios JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY Volumen: 62 Número: 2 Páginas: 406-418 Fecha de publicación: FEB 2011	15	26	35	34	9	121	20.17
<input type="checkbox"/>	3. Extracting macroscopic information from Web links Por: Thelwall, M JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY Volumen: 52 Número: 13 Páginas: 1157-1168 Fecha de publicación: NOV 2001	6	4	7	2	1	120	7.50
<input type="checkbox"/>	4. Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis Por: Kousha, Kayvan; Thelwall, Mike JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY Volumen: 58 Número: 7 Páginas: 1055-1065 Fecha de publicación: MAY 2007	9	11	18	12	2	109	10.90
<input type="checkbox"/>	5. Search engine coverage bias: evidence and possible causes Por: Vaughan, L; Thelwall, M INFORMATION PROCESSING & MANAGEMENT Volumen: 40 Número: 4 Páginas: 693-707 Fecha de publicación: JUL 2004	7	6	12	5	1	100	7.69



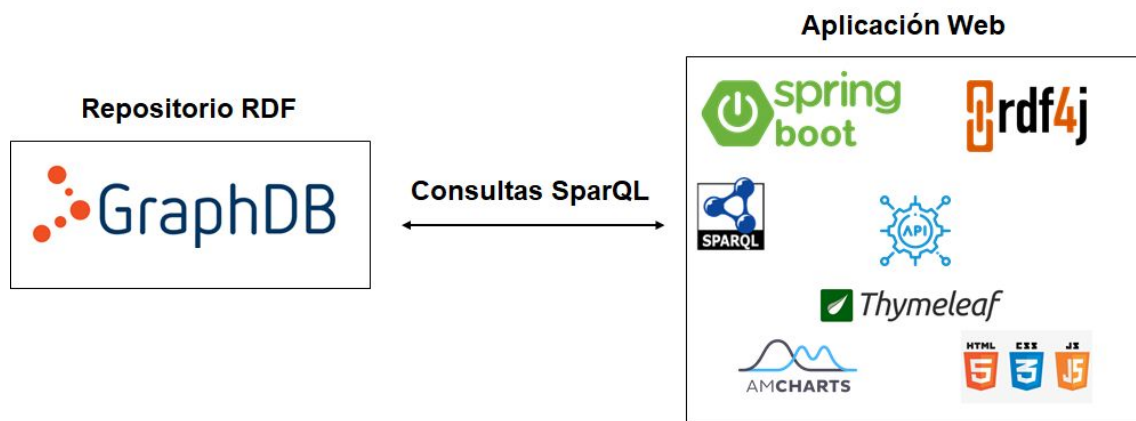
3.4. Aplicación propuesta

En esta sección se describe las tecnologías y funcionamiento general de la aplicación propuesta.

3.4.1. Tecnologías

- **GraphDB:** El repositorio RDF en donde se almacena los datos RDF generados por la aplicación en Java con Jena explicado en la sección 3.3. Se ha elegido este repositorio semántico debido a su fácil integración con diferentes tipos de aplicaciones. También porque es orientada a grafos lo que facilita el entendimiento de los datos RDF generados y todas sus relaciones

- **SpringBoot:** Es un módulo de Spring Framework que nos permite la creación y configuración de aplicaciones web de manera rápida y sencilla. Los lenguajes que se pueden utilizar son Java, Kotlin y Groovy.
- **RDF4J:** Es un marco de trabajo basado en el lenguaje Java que nos permite acceder a repositorios de datos RDF y realizar consultas o manejar los datos en el repositorio.
- **Thymeleaf:** El motor de plantillas que se usa para la creación de todas las vistas o interfaces que se le presentará al usuario. Las plantillas tienen el formato html.
- **JavaScript:** Es el lenguaje utilizado en las vistas de la aplicación para generar mayor interacción con el usuario,
- **Chart4m:** Es una biblioteca de acceso libre que permite la visualización de datos a través de gráficas simples y avanzadas.



3.4.2. Consultas SPARQL

Se ha creado apis en el servidor en donde cada una genera consultas hacia el repositorio semántico. Y con la información que se se obtiene se procede a generar gráficas y tablas estadísticas en las interfaces de usuario.

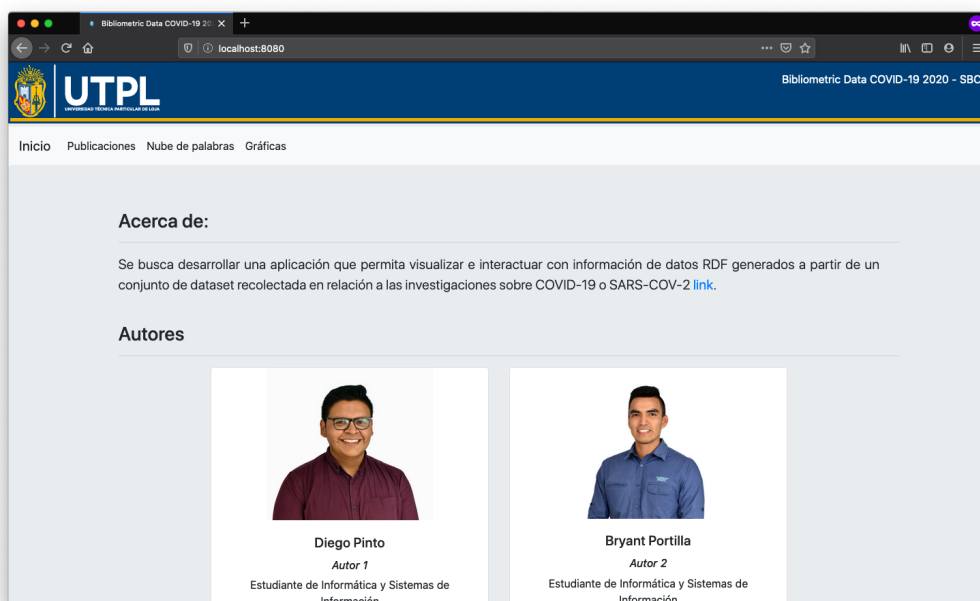
Las consultas que se han generado a partir de los datos generados fueron las siguientes:

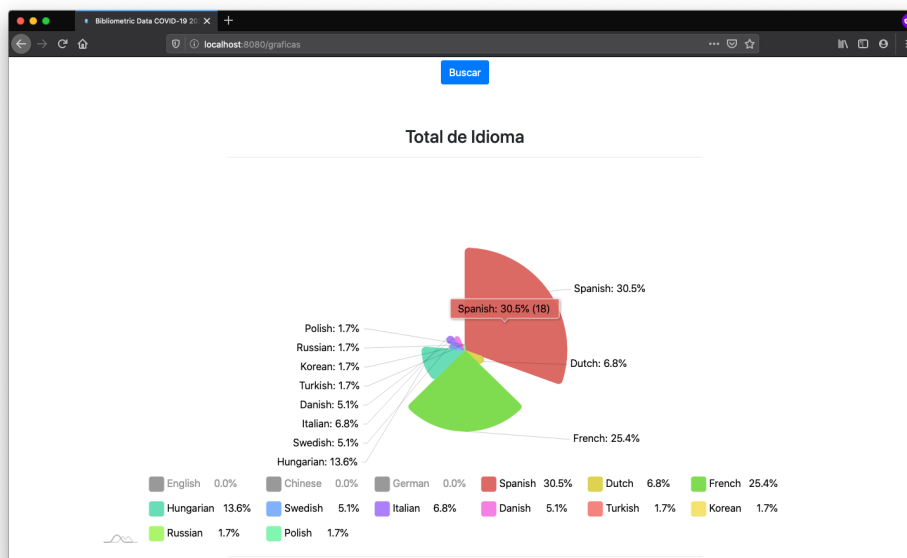
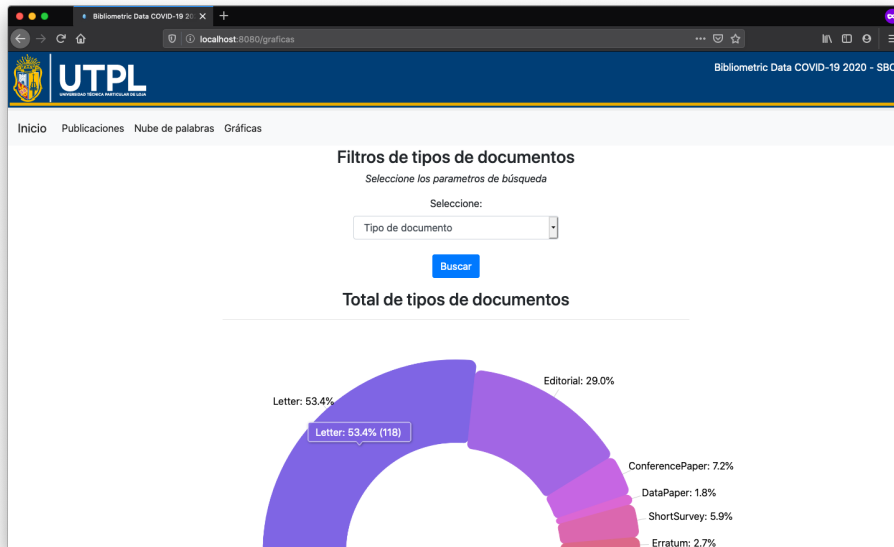
Descripción	Consulta
-------------	----------

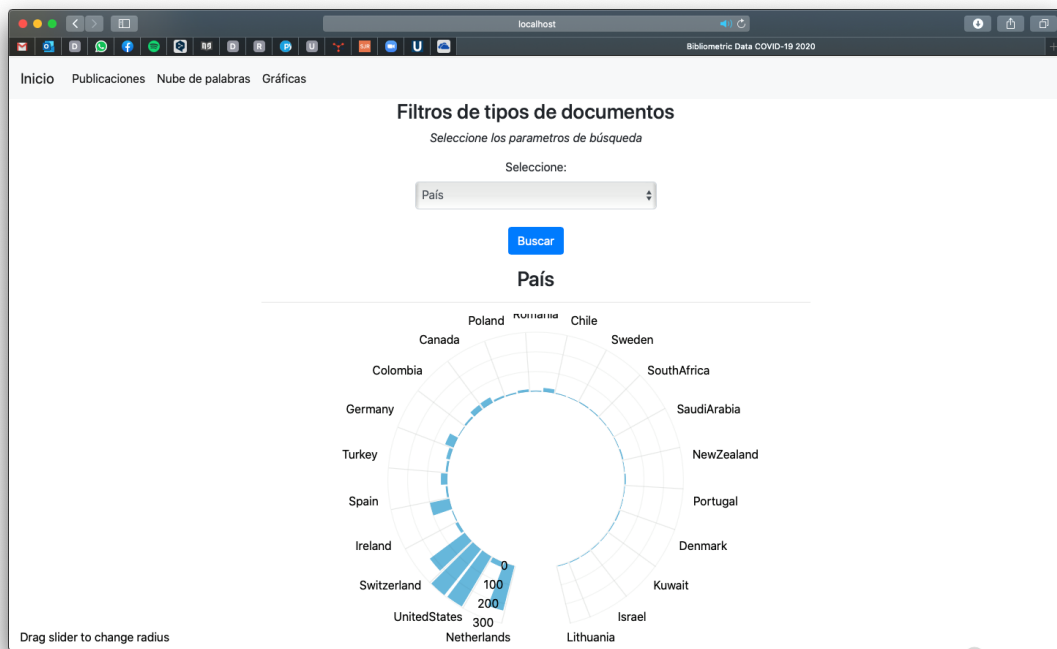
<p>Consultar el total de recursos de tipo Scholarly Work, Autores, Lenguajes y Sources.</p>	<pre> 1 PREFIX fabio: <http://purl.org/spar/fabio/> 2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> 3 PREFIX dct: <http://purl.org/dc/terms/> 4 PREFIX myData: <http://utpl.edu.ec/COVIBiblio/ontology/> 5 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> 6 PREFIX schema: <http://schema.org/> 7 select * {{ select (count(?Recursos) AS ?count_Recursos) 8 { ?Recursos rdfs:subClassOf fabio:ScholarlyWork .}} 9 {select (count(?Autores) AS ?count_Author) 10 { ?Autores a myData:Author .} } 11 {select (count(?language) AS ?count_language) 12 { ?language a schema:Language.} } 13 {select (count(?Citas) AS ?count_Citas) 14 { ?Citas a myData:SourceT.}}}</pre>
<p>Consultar el recurso el titulo de las publicaciones (Scholarly Work) y nombre de los autores relacionados a la publicación</p>	<pre> 1 PREFIX foaf: <http://xmlns.com/foaf/0.1/> 2 PREFIX fabio: <http://purl.org/spar/fabio/> 3 PREFIX dct: <http://purl.org/dc/terms/> 4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> 5 select DISTINCT ?recurso ?titulo ?autor WHERE { 6 ?recurso rdfs:subClassOf fabio:ScholarlyWork ; 7 dct:title ?titulo; 8 dct:creator ?creador. 9 ?creador foaf:name ?autor. 10 }</pre>
<p>Consultar la cantidad de publicaciones agrupadas por el lenguaje en el que están escritos</p>	<pre> 1 PREFIX fabio: <http://purl.org/spar/fabio/> 2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> 3 PREFIX dct: <http://purl.org/dc/terms/> 4 SELECT DISTINCT ?lenguaje (count(*) AS ?cantidad) WHERE { 5 ?Recursos rdfs:subClassOf fabio:ScholarlyWork; 6 dct:language ?lenguaje. 7 } GROUP BY ?lenguaje </pre>
<p>Consultar la cantidad de publicaciones agrupadas por el título del Source al que pertenecen</p>	<pre> 1 PREFIX fabio: <http://purl.org/spar/fabio/> 2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> 3 PREFIX dct: <http://purl.org/dc/terms/> 4 PREFIX myData: <http://utpl.edu.ec/COVIBiblio/ontology/> 5 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> 6 PREFIX bido: <http://purl.org/spar/bido/> 7 select DISTINCT ?source (count(*) AS ?cantidad) WHERE { 8 ?Recurso rdfs:subClassOf fabio:ScholarlyWork; 9 bido:withBibliometricData ?Source. 10 ?Source dct:title ?source 11 } GROUP BY ?source </pre>
<p>Consultar la cantidad de publicaciones agrupadas por el país al que pertenece el Source de cada publicación</p>	<pre> 1 PREFIX fabio: <http://purl.org/spar/fabio/> 2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> 3 PREFIX dct: <http://purl.org/dc/terms/> 4 PREFIX myData: <http://utpl.edu.ec/COVIBiblio/ontology/> 5 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> 6 PREFIX bido: <http://purl.org/spar/bido/> 7 PREFIX dbo: <http://dbpedia.org/ontology/> 8 select DISTINCT ?country (count(*) AS ?cantidad) WHERE { 9 ?Recurso rdfs:subClassOf fabio:ScholarlyWork; 10 bido:withBibliometricData ?Source. 11 ?Source dct:title ?source; 12 dbo:country ?country. 13 } GROUP BY ?country </pre>

Consultar los Scholarly Works y todas sus propiedades principales (title, date, lenguaje, tipo, doi, issn, volume, numero de citas)	<pre> 1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> 2 PREFIX dct: <http://purl.org/dc/terms/> 3 PREFIX myData: <http://utpl.edu.ec/COVIDBiblio/ontology/> 4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> 5 PREFIX fabio: <http://purl.org/spar/fabio/> 6 PREFIX prism: <http://prismstandard.org/namespaces/1.2/basic/> 7 select DISTINCT ?recurso ?titulo ?date ?lenguaje ?tipo ?doi ?issn ?volume ?numCitas where { 8 ?recurso rdfs:subClassOf fabio:ScholarlyWork; dct:title ?titulo; 9 dct:date ?date; dct:language ?lenguaje; 10 rdf:type ?tipo; prism:doi ?doi; 11 prism:issn ?issn; prism:volume ?volume; 12 myData:citationsCount ?numCitas. 13 }</pre>
Consultar la cantidad de publicaciones agrupadas por el tipo de documento	<pre> 1 PREFIX fabio: <http://purl.org/spar/fabio/> 2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> 3 PREFIX dbo: <http://dbpedia.org/ontology/> 4 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> 5 SELECT DISTINCT ?tipo (count(*) AS ?cantidad) WHERE { 6 ?Recursos rdfs:subClassOf fabio:ScholarlyWork . 7 ?Recursos rdf:type ?tipo . 8 } GROUP BY ?tipo</pre>
Consultar los Scholarly works, título del recurso y el source con su rank y sus respectivas categorías y cuantiles	<pre> 1 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> 2 PREFIX dcat: <http://www.w3.org/ns/dcat#> PREFIX fabio: <http://purl.org/spar/fabio/> 3 PREFIX data: <http://opendata.org/resource/> 4 PREFIX foaf: <http://xmlns.com/foaf/0.1/> PREFIX dbo: <http://dbpedia.org/ontology/> 5 PREFIX dct: <http://purl.org/dc/terms/> PREFIX bido: <http://purl.org/spar/bido/> 6 PREFIX myData: <http://utpl.edu.ec/COVIDBiblio/ontology/> 7 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> 8 select DISTINCT ?Recurso ?tituloRecurso ?Source ?titulo ?country ?titleQuartile ?ValueRank ?ValueQuartile 9 where { 10 ?Recurso rdfs:subClassOf fabio:ScholarlyWork; dct:title ?tituloRecurso; 11 bido:withBibliometricData ?Source. 12 ?Source dct:title ?titulo ; dbo:country ?country ; 13 myData:rank ?ValueRank; bido:hasQuartile ?Quartile . 14 ?Quartile dct:title ?titleQuartile ; myData:quartile ?ValueQuartile. 15 }</pre>

3.4.3. Visualización (Capturas de la aplicación)







localhost

Bibliometric Data COVID-19 2020

Inicio Publicaciones Nube de palabras Gráficas

Filtros de tipos de documentos

Seleccione los parametros de búsqueda

Seleccione:

Buscar

Source title

Copy PDF Print

Buscar:

#	Source	Cantidad
1	International Journal of Infectious Diseases	54
2	Reproductive and Developmental Medicine	1
3	Journal of Clinical Endocrinology and Metabolism	2
4	Trials	13
5	Viruses	15
6	Nephron	1
7	Journal of Thrombosis and Haemostasis	3
8	Journal of the American College of Surgeons	1
9	Medical Journal of Australia	4
10	Molecules	4

4. Conclusiones

Al terminar el presente trabajo se puede concluir que:

- La web semántica y las funciones que ofrece son importantes, debido a que el gran número de datos que se encuentran misma no poseen una estructura abierta, por lo tanto el objetivo de la web semántica es de solventar este problema.
- Los datos RDF tienen una estructura que permite la integración con múltiples aplicaciones con el objetivo principal de generar más conocimiento.
- Jena es un framework que nos facilita la creación de datos RDF de manera sencilla y rápida.
- GraphDB es un motor de base de datos semántica orientada a grafos lo que permite el fácil entendimiento de los datos a almacenar.

5. Bibliografía

Candela, G., Escobar, P., Carrasco, R. C., & Marco-Such, M. (2019). A linked open data framework to enhance the discoverability and impact of culture heritage. *Journal of Information Science*, 45(6), 756–766. <https://doi.org/10.1177/0165551518812658>

Hazaël-Massieux, D., & Berners-Lee, T. (2003). The Semantic Web and its applications at W3C. <https://www.w3.org/2003/Talks/simo-semwebapp/all.htm>

Ontotext. (2020). GraphDB Documentation. <http://graphdb.ontotext.com/documentation/standard/>

Pfenninger, S., DeCarolís, J., Hirth, L., Quoilin, S., & Staffell, I. (2017). The importance of open data and software: Is energy research lagging behind? *Energy Policy*, 101, 211–215. <https://doi.org/10.1016/j.enpol.2016.11.046>

Sharma, V. (2015). What are WEB 2.0 & WEB 3.0? <https://www.quora.com/What-are-WEB-2-0-WEB-3-0>

Lapiente, M. J., & Lamarca, C. (2018). RDF.

Enlaces

<https://dzone.com/articles/getting-started-with-ontotext-graphdb-and-rdf4j>