

Bibliometric Data

Integrantes:

- **Diego Pinto**
- **Bryant Portilla**

Introducción

Lo que se busca con este proyecto es obtener información de investigaciones científicas relacionadas a covid-19 y mediante visualizaciones se pueda obtener muchos datos para observar qué áreas o qué intereses tiene la ciencia para ayudar a esta situación que está pasando en el mundo.

Fuente de datos

Los datos son obtenidos desde la base de datos científica SCOPUS y para la esto se usó el modelo ETL (extraer, transformar y cargar).

Cadenas de búsqueda:

```
TITLE-ABS-KEY ( ( covid-19 ) AND ( sars-cov-2 ) ) AND ( LIMIT-TO ( PUBSTAGE , "final" ) )
```

```
TITLE-ABS-KEY ( ( covid-19 ) AND ( sars-cov-2 ) ) AND ( LIMIT-TO ( PUBSTAGE , "aip" ) )
```

El resultado de la cadena de búsqueda que se estableció nos devolvió 2000 investigaciones relacionadas al covid-19 y con múltiples características relacionadas al artículo obtenido.

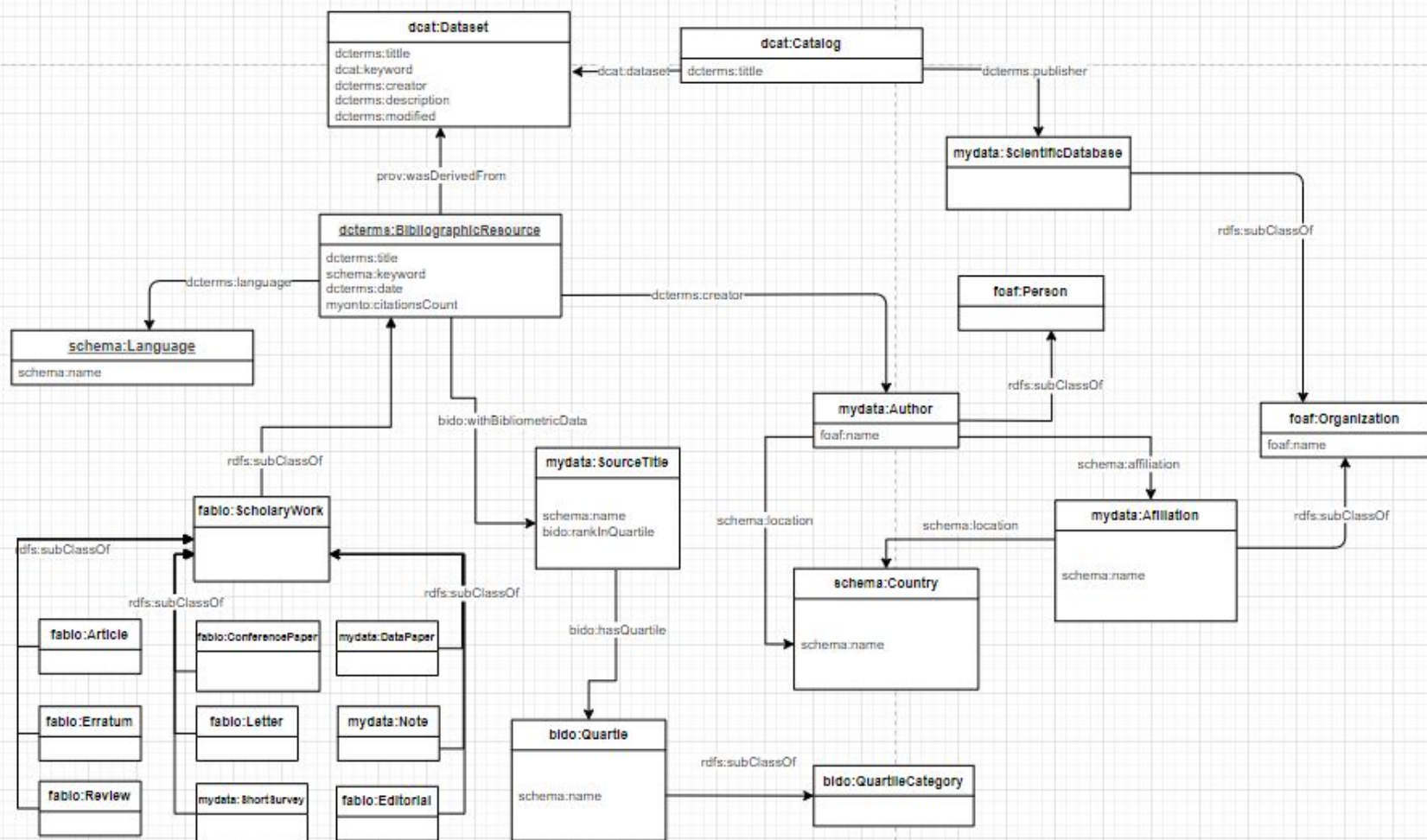
Limpieza de datos

Basándonos en el modelo ETL, dentro de la fase de transformación se realizó una limpieza para mejorar el contenido de la información que se obtuvo.

Los datos de obtuvieron en un csv directo desde la base de datos científica Scopus en donde se ha procedido a realizar:

- Estructurar como un solo atributo todos los autores (nombres, ids)
- Eliminar caracteres especiales (según la lógica de lectura).

Modelo



Generacion RDF (Jena)

Estructura de los datos en excel

Dataset	Authors	Authors_ID	Title	EID	Year	Source title
BiblioData	Byass P.	7006527895	Eco-epidemi	2-s2.0-85084	2020	Global Healt
BiblioData	Liu P.; Cai J.;	57200337904	Dynamic sur	2-s2.0-85086	2020	Emerging mi

Volume	Issue	Art. No.	Page start	Page end	Cited by
13	1	1760490			
9	1		1254	1258	

DOI	Link	Document Type	Publication Status	Access Type	Source	Language	keywords
10.1080/1654	https://www	Article	Final	Open Access	Scopus	English	China; Coron
10.1080/2222	https://www	Letter	Final	Open Access	Scopus	English	virus RNA; B

Publisher	ISSN	Affiliation	Country	Quartile	Rank
Taylor and Fr	16549880	Department	Sweden	Q1	47
NLM (Medlin	22221751	Department	China	Q1	24

Generacion RDF (Jena)

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_DAT
Documento (Articulo, Letter, ConferencePaper)	http://utpl.edu.ec/COVIDBiblio/ontology/BibliographicResource/ + EID DOCUMENTO

Generacion RDF (Jena)

<https://github.com/diepinto30/GenerateBibliometricDataCovid>

Incluir los prefijos que se utilizaran en nuestro modelo

```
// 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 ));
```


Generacion RDF (Jena)

Creacion de SCIENTIFIC DATABASE -CATALOG - DATASET

```
// CREACION DE SCIENTIFIC DATABASE
Resource fuenteDocumento = model.createResource(URI_FUENTE)
    .addProperty(FOAF.name, fuente)
    .addProperty(RDFS.subClassOf, FOAF.Organization);

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

Generacion RDF (Jena)

Creación de Recurso bibliográfico (Article-Letter-ConferencePaper...)

```
// 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, dboModel.getResource(dbr+language))
    .addProperty(RDFS.subClassOf, fabioModel.getResource(fabio+ "ScholarlyWork/"))
    .addProperty(prismModel.getProperty(prism + "doi/"), doi)
    .addProperty(prismModel.getProperty(prism + "volume/"), vol)
    .addProperty(provModel.getProperty(prov + "wasDerivedFrom"), datasetInfo));

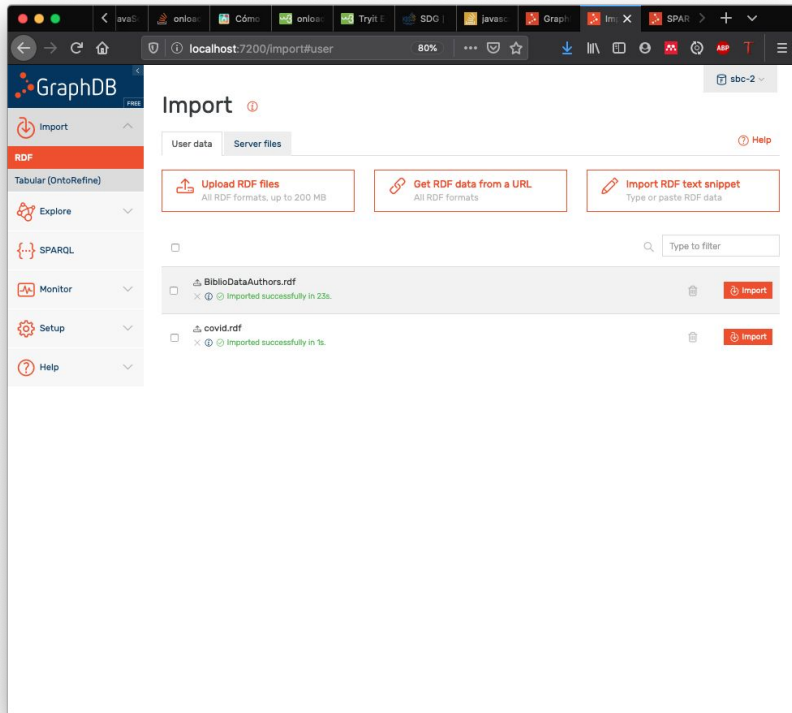
// Se crea el tipo de documento
// Se compara el tipo con los de fabio y si es igual toma la uri de fabio
for (String nombre : tipos_Fabio) {
    if(document_type.replace(" ", "").equals(nombre)){
        documento.addProperty(RDF.type, fabioModel.getResource(fabio + document_type.replace(" ", "")))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

Generacion RDF (Jena)

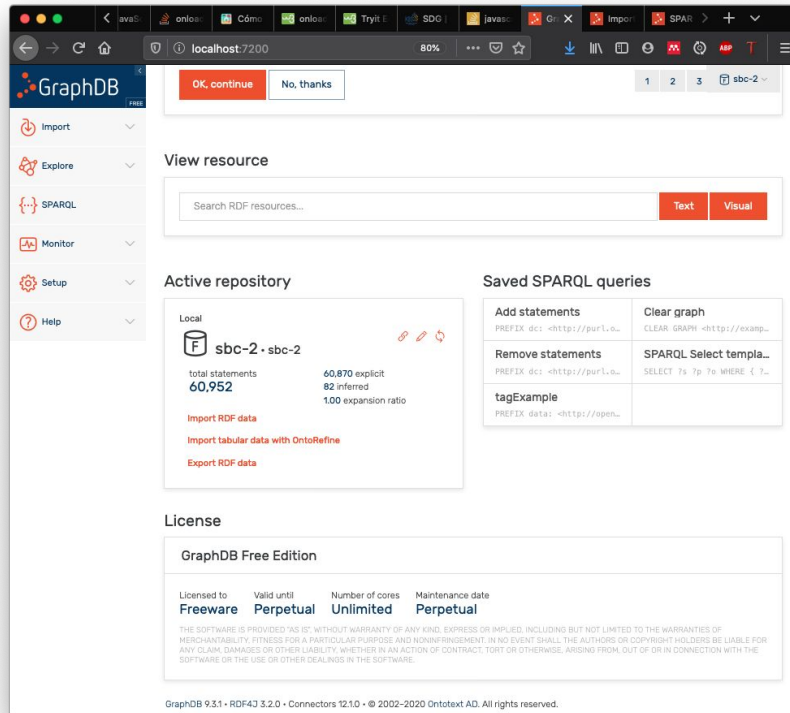
Creación de Autores y vinculación al Recurso Bibliográfico

```
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);  
}
```

Almacenador de datos semánticos - GraphDB



The screenshot shows the GraphDB web interface at localhost:7200/import#user. The left sidebar contains navigation links: Import, RDF (selected), Explore, SPARQL, Monitor, Setup, and Help. The main area is titled 'Import' and has two tabs: 'User data' and 'Server files'. Under 'User data', there are three buttons: 'Upload RDF files' (All RDF formats, up to 200 MB), 'Get RDF data from a URL' (All RDF formats), and 'Import RDF text snippet' (Type or paste RDF data). Below these, there is a list of imported files: 'BiblioDataAuthors.rdf' (Imported successfully in 23s) and 'covid.rdf' (Imported successfully in 1s). Each file has an 'Import' button.



The screenshot shows the GraphDB web interface at localhost:7200. The left sidebar contains navigation links: Import, Explore (selected), SPARQL, Monitor, Setup, and Help. The main area is titled 'View resource' and has a search bar 'Search RDF resources...' with 'Text' and 'Visual' buttons. Below the search bar, there is a section 'Active repository' showing a local repository 'sbc-2 - sbc-2' with statistics: 60,870 explicit statements, 82 inferred statements, and a 1.00 expansion ratio. There are buttons for 'Import RDF data', 'Import tabular data with OntoRefine', and 'Export RDF data'. To the right, there is a section 'Saved SPARQL queries' with a table of queries. Below the queries, there is a 'License' section for 'GraphDB Free Edition'.

License	Valid until	Number of cores	Maintenance date
Freeware	Perpetual	Unlimited	Perpetual

THE SOFTWARE IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Almacenador de datos semánticos - GraphDB

The screenshot shows a web browser window with the address bar displaying 'jsonviewer.stack.hu'. The main content area shows a JSON file named 'graph' with the following structure:

```

[
  {
    "id": "http://localhost:8000/resource/2-a2-0-85079663023",
    "type": "http://purl.org/parifabio/Article",
    "url": "http://utpl.edu.ec/COVIDBiblio/ontology/Article",
    "creator": "http://purl.org/dc/terms/creator",
    "id": "http://localhost:8000/resource/Author/6603787185",
    "date": "http://purl.org/dc/terms/date",
    "value": "2020",
    "language": "http://purl.org/dc/terms/language",
    "title": "http://dbpedia.org/resource/Hungarian",
    "value": "Importance of the imaging techniques in the management of COVID-19infected patients [A kópalko kódi]",
    "value": "http://utpl.edu.ec/COVIDBiblio/ontology/citationsCount",
    "value": "",
    "url": "http://www.w3.org/2000/01/rdf-schema#subClassOf",
    "id": "http://purl.org/parifabio/ScholarlyWork",
    "id": "http://www.ontotext.com/explicit"
  }
]

```

The browser's search bar at the bottom shows 'GO!' and 'Next' buttons, and a 'Previous' button is also visible.

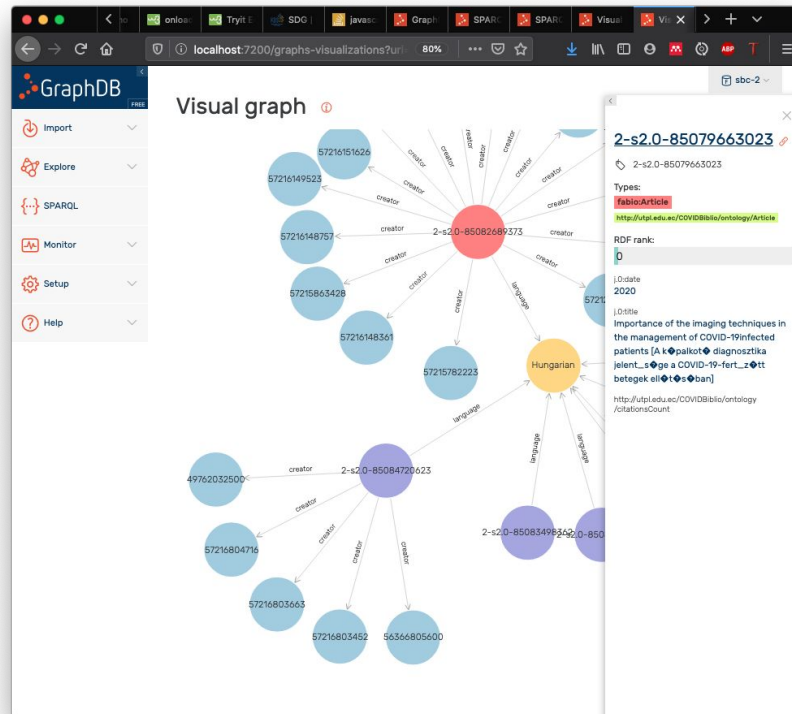
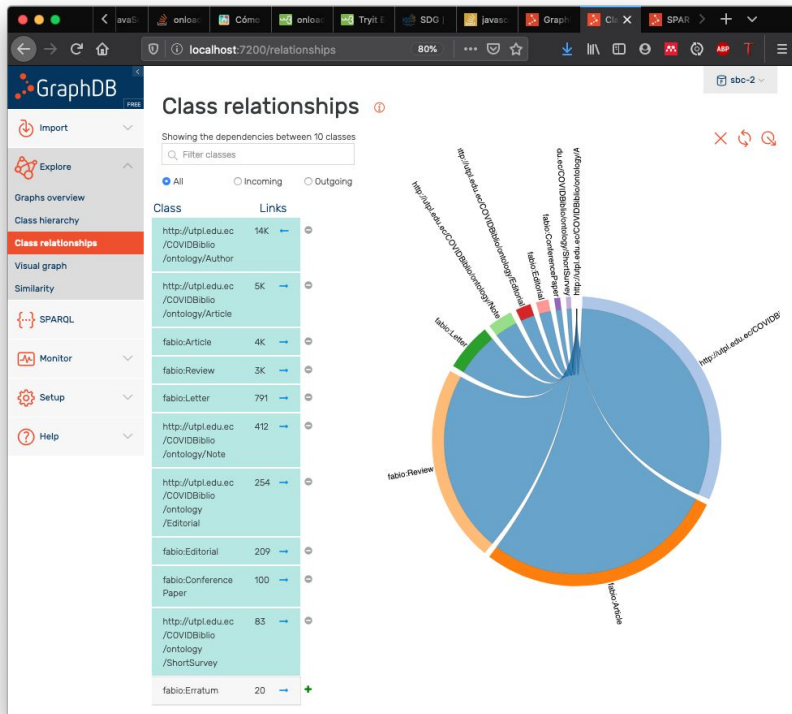
The screenshot shows the GraphDB web application running in a browser. The address bar displays the URL: `localhost:7200/resource?uri=http%3A%2F%2Flocalhost:8000%2Fresource%2F2-s2.0-85079663023`. The page title is "GraphDB". On the left sidebar, there are navigation links: Import, Explore, SPARQL, Monitor, Setup, and Help. The main area displays a SPARQL query result for the URI `http://localhost:8000/resource/2-s2.0-85079663023`.

2-s2.0-85079663023

Source: `http://localhost:8000/resource/2-s2.0-85079663023`

	subject	predicate	object	context
1	<code>http://localhost:8000/resource/2-s2.0-85079663023</code>	<code>j:creator</code>	<code>http://localhost:8000/resource/Author/6603787185</code>	<code>http://www.ontotext.com/explicit</code>
2	<code>http://localhost:8000/resource/2-s2.0-85079663023</code>	<code>j:date</code>	"2020"	<code>http://www.ontotext.com/explicit</code>
3	<code>http://localhost:8000/resource/2-s2.0-85079663023</code>	<code>j:language</code>	<code>dbn:Hungarian</code>	<code>http://www.ontotext.com/explicit</code>
4	<code>http://localhost:8000/resource/2-s2.0-85079663023</code>	<code>j:title</code>	"Importance of the imaging techniques in the management of COVID-19infected patients [A kőpa lkor diagnosztika jelentősége a COVID-19-fertőzött betegek ellátásában]"	<code>http://www.ontotext.com/explicit</code>
5	<code>http://localhost:8000/resource/2-s2.0-85079663023</code>	<code>http://utpl.edu.ec/COVIDBiblio/ontology/citationsCount</code>	"	<code>http://www.ontotext.com/explicit</code>
6	<code>http://localhost:8000/resource/2-s2.0-85079663023</code>	<code>rdf:type</code>	<code>fablo:Article</code>	<code>http://www.ontotext.com/explicit</code>
7	<code>http://localhost:8000/resource/2-s2.0-85079663023</code>	<code>rdf:type</code>	<code>http://utpl.edu.ec/COVIDBiblio/ontology/Article</code>	<code>http://www.ontotext.com/explicit</code>
8	<code>http://localhost:8000/resource/2-s2.0-85079663023</code>	<code>rdfs:subClassOf</code>	<code>fablo:ScholarlyWork/</code>	<code>http://www.ontotext.com/explicit</code>

Almacenador de datos semánticos - GraphDB



Gracias
