



# **UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA**

## **SISTEMAS INFORMATICOS Y COMPUTACION**

**COMPONENTE:** Sistemas Basados en Conocimientos

**ESTUDIANTE:** Luis Alfredo Jaramillo Uday (1719626754)

**DOCENTE:** Janneth Chicaiza Espinosa

**TEMA:** “Tareas previas a la generación de datos RDF”

**FECHA DE ENTREGA:** 16-06-2020

LOJA-ECUADOR

## ACTIVIDAD 1: Definición de URIs

### 1. Definición de URIs

Basándome en lo leído en el documento compartido, y de acuerdo con las definiciones de cada autor, empleado lo siguiente para definir las uris de mi documento.

#### A. 2.3.1 Especificación

##### - Identificación De fuentes

Para esto debemos distinguir bien lo que nos están pidiendo y vamos a recolectar, después con ayuda de las búsquedas y ayudas planteadas debes ver y aprovechar los datos ya publicados por otros, buscar sobre todo datos de libre acceso (dato publicos), que se encontraran en las páginas de las fuentes recomendadas.

Una vez identificado todo esto debemos quedarnos con las fuentes que cumplan a mayor medida con los datos que requerimos, es decir que traigan o contengan los mayores atributos que hemos definido en nuestro modelo, que se trata sobre publicación sobre covid-19, especializado en Keywords.

##### • Ejemplo de fuentes

Id	Fuente
1	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269890/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269890/</a>
2	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269892/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269892/</a>
3	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269897/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269897/</a>
4	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7274583/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7274583/</a>
5	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7274597/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7274597/</a>
6	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269890/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269890/</a>
7	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269892/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269892/</a>
8	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269897/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269897/</a>
9	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7274583/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7274583/</a>
10	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7274597/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7274597/</a>
11	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269890/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269890/</a>
12	<a href="https://www.sciencedirect.com/science/article/pii/S0085253820305329">https://www.sciencedirect.com/science/article/pii/S0085253820305329</a>
13	<a href="https://www.kaggle.com/orgesleka/keywords-similarity-dataset">https://www.kaggle.com/orgesleka/keywords-similarity-dataset</a>
14	<a href="https://www.sciencedirect.com/science/article/pii/S0735109720350087#kwrds0010">https://www.sciencedirect.com/science/article/pii/S0735109720350087#kwrds0010</a>

15	<a href="https://www.sciencedirect.com/science/article/pii/S1130862120303132">https://www.sciencedirect.com/science/article/pii/S1130862120303132</a>
----	---

#### - **Licencias**

Dentro de la las fuente encontradas, existen fuentes como ScienceDirect y IEEE que son sitios web de repositorios que la gran mayoría de veces el acceso es por suscripción para poder acceder una gran base de datos de investigación de los datos que están publicando. Existen algunas licencias que nos encontramos dentro de estos sitios como son

- Open Data: El concepto datos abiertos es una filosofía y práctica que persigue que determinados tipos de datos estén disponibles de forma libre para todo el mundo, sin restricciones de derechos de autor, de patentes o de otros mecanismos de control.,.
- Creative Commons: es una organización sin fines de lucro dedicada a promover el acceso y el intercambio de cultura. Desarrolla un conjunto de instrumentos jurídicos de carácter gratuito que facilitan usar y compartir tanto la creatividad como el conocimiento..

Id_fuente	Licencia
1	Creative Commons
2	Open
3	Open
4	Open
5	Open

### **B. 2.3.2 Modelado**

Para el modelado se ha escogido la ontología mas adecuada para el modelo creado, para ello se usará la mayor cantidad de vocabularios ya implantados y mas conocidos. Así evitar esfuerzos innecesarios a la hora de hay vocabularios para nuestro modelo.

Usaremos vocabularios como

1. Skos

<https://www.w3.org/2009/08/skos-reference/skos.html>

2. Schema

<http://schema.org/>

3. Foaf

<http://xmlns.com/foaf/0.1/>

4. Debepedia

<http://dbpedia.org/resource/>

5. Dublin core

<http://purl.org/dc/elements/1.1/>

- Si ya no encontramos vocabulario adecuado tocara crear uno propio pero preferiblemente debemos utilizar los ya mencionados.

6. Myoonto o mydata

[www.utpl.edu.ec/myonto](http://www.utpl.edu.ec/myonto)

### C. 2.3.3 Generación

Para la generación de datos y para concordar con rdf y vocabularios de linke data he decidido seguir unos pasos para obtener una generación de datos factible para el modelo:

**a. Descarga:** La descarga se ha hecho de un csv el cual contenía en si más de 3000 datos, el cual contenida varias publicaciones realizadas durante este año.

**b. Limpieza de datos:** consiste en dejar los datos limpios, es decir en perfecto estado para que puedan ser leídos por e interpretados en modelos rdfs y hacer que las uris se adapten, también evitar traer datos vacíos que puedan provocar confusión a la hora de ejecutar la conversión o vinculación.

**c. Vinculación:** La vinculación como su nombre indica es vincular elementos para que tengan un sentido y se puedan entender e interpretar con facilidad

## ACTIVIDAD 2: Actualización de datos

### 1. Fuente

Fecha inicial :16/05/2020

1. ESPECIFICACION DE FUENTES DE DATOS		
1.1 Identificación de Fuentes		
Nombre de la Fuente de Datos:		
Proveedor:sciencedirect	Sitio Web: <a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>	
Institución:		
Descripción de la fuente de datos:	una gran base de datos de investigación científica y médica Alberga más de 12 millones de contenidos de 3.500 revistas	
Tamaño del archivo de datos (MB)		
Licencia	Copyright ©	
Formato del archivo:	PDF	
Incluye información de los metadatos:	Multilinguaje	
1.2 Volumetría de datos		
Entidad[atributos]	Cantidad de registros	Descripción
Fuente(nombre_fuente,Url)	35	Incluye la url donde esta el documento y la pagina
autor(nombre_autor,apellido_autor)	35	incluye los datos del autor responsable del documento
documento,date,tipo_documento	35	incluye los datos mas importantes del documento
palabras clave(palabra_en,palabra_es)	155	incluye las palabras claves del documento, con su traduccion al español o viceversa

Fecha actualizacion:06/06/2020

1. ESPECIFICACION DE FUENTES DE DATOS		
1.1 Identificación de Fuentes		
Nombre de la Fuente de Datos:		
Proveedor:sciencedirect	Sitio Web: <a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>	
Institución:		
Descripción de la fuente de datos:	una gran base de datos de investigación científica y médica. Alberga más de 12 millones de contenidos de 3.500 revistas	
Tamaño del archivo de datos (MB)		
Licencia	Copyright ©	
Formato del archivo:	PDF	
Incluye información de los metadatos:	Multilenguaje	
1.2 Volumetría de datos		
Entidad[atributos]	Cantidad de registros	Descripción
Fuente(nombre_fuente,Url)	100	Incluye la url donde esta el documento y la pagina
autor(nombre_autor,apellido_autor)	450	incluye los datos del autor responsable del documento
_documento,date,tipo_documento)	100	incluye losdatos mas importantes del documento
abras clave(palabra_en,palabra_es)	450	incluye las palabras claves del documento, con su traduccion al español o viseversa

Fuente 2

Fecha original:15/05/2020

1.1 Identificación de Fuentes		
Nombre de la Fuente de Datos:		
Proveedor:SciELO	Sitio Web: <a href="https://scielo.org">https://scielo.org</a>	
Institución:		
Descripción de la fuente de datos:	ión para el Apoyo a la Investigación del Estado de São Paulo,	
Tamaño del archivo de datos (MB)		
Licencia	Copyright ©	
Formato del archivo:	PDF	
Incluye información de los metadatos:	Multilenguaje	
1.2 Volumetría de datos		
Entidad[atributos]	Cantidad de registros	Descripción
Fuente(nombre_fuente,Url)	10	Incluye la url donde esta el documento y la pagina
Autor(nombre_autor,apellido_autor)	19	incluye los datos del autor responsable del documento
e_documento,date,tipo_documento,ic)	10	incluye losdatos mas importantes del documento
palabras clave(palabra_en,palabra_es)	51	incluye las palabras claves del documento, con su traduccion al español o viseversa

Fecha actualizacion:04/06/2020

1.1 Identificación de Fuentes		
Nombre de la Fuente de Datos:		
Proveedor:SciELO	Sitio Web:https://scielo.org	
Institución:		
Descripción de la fuente de datos:	ión para el Apoyo a la Investigación del Estado de São Paulo,	
Tamaño del archivo de datos (MB)		
Licencia	Copyright ©	
Fomato del archivo:	PDF	
Incluye información de los metadatos:	Multilenguaje	
1.2 Volumetría de datos		
Entidad[atributos]	Cantidad de registros	Descripción
Fuente(nombre_fuente,Url)	200	Incluye la url dodnde esta el documento y la pagina
Autor(nombre_autor,apellido_autor)	80	incluye los datos del autor responsable del documento
e_documento,date,tipo_documento,id	200	incluye losdatos mas importantes del documento
alabras clave(palabra_en,palabra_es)	510	incluye las palabras claves del documento, con su traduccion al español o viseversa

- Aumentado los datos obtenidos respecto al anterior dado que se ha encontrado mas documentos en las fuentes

#### Bibliografía

<https://creativecommons.org>