



**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:** “Proyecto E2 - Transformación de datos RDF”

**FECHA DE ENTREGA:** 28-06-2020

**LOJA-ECUADOR**

## 1. Tabla resumen de datos recolectados: Por cada clase del modelo ontológico indicar cuántas instancias generarían (o generaron).

clase	instancia	total
Dce:source	dbo:name	2/150
	Dcat:donwald	150
bibo:Documento	dbo:name	150
	dce::date	150
	dce:identifier	150
	dce:language	3/150
schema:author	foaf:Name	450
	foaf:firstName	450
	foaf:lastName	450
Skos:conceptr	Myonto:keyword	400
	Myonto:tematic	20

## 2. Pre-procesamiento de datos: Indicar qué tareas de limpieza o transformación de datos se realizaron antes de generar RDF.

Primeramente, he procedido a analizar métodos para adaptar los datos al modelo ontológico propuesto previo a la generación de individuos o instancias empleando Jena.

Para ello lo que se ha hecho es definir las fuentes principales de las cuales se va sacar los datos, en función de los atributos que ya fueron definidos con anterioridad.

Se ha clasificado de un dataset los atributos que solamente constan en el modelo, eliminando así los atributos innecesarios o con poca funcionalidad para nuestro modelo, con esto se pudo conseguir una limpieza más optima de datos para poderlos adaptar al modelo propuesto y para observar de una mayor manera dichos datos, se utilizo la herramienta openrefine, la cual presenta los datos con una mayor visibilidad. Todo esto sirvió para que se pueda evitar redundancia en los datos.

1. ESPECIFICACION DE FUENTES DE DATOS		
1.1 Identificación de Fuentes		
Nombre de la Fuente de Datos:		
Proveedor:sciencedirect	Sitio Web:https://www.sciencedirect.com	
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	
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)	75	Incluye la url donde esta el documento y la pagina
Autor(nombre_autor,apellido_autor)	180	incluye los datos del autor responsable del documento
Documento(nombre_documento,date,tipo_documento,idioma_documento)	75	incluye losdatos mas importantes del documento
Palabras clave(palabra_en,palabra_es)	220	incluye las palabras claves de documento, con su traducción al español o viceversa

1. ESPECIFICACION DE FUENTES DE DATOS		
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 ©	
Fomato del archivo:	PDF	
Incluye información de los metadatos?	Multilinguaje	
1.2 Volumetría de datos		
Entidad[atributos]	Cantidad de registros	Descripción
Fuente(nombre_fuente,Url)	70	Incluye la url dodnde esta el documento y la pagina
Autor(nombre_autor,apellido_autor)	200	incluye los datos del autor responsable del documento
Documento(nombre_documento,date,d oi,tipo_documento,idioma_documento )	70	incluye losdatos mas importantes del documento
Palabras clave(palabra_en,palabra_es)	180	incluye las palabras claves del documento, con su traduccion al español o viseversa

### 3. Transformación de datos:

Una vez escogidas las fuentes a trabajar se estará empleando el siguiente modelo de proceso, para la construcción, modelado y limpieza.

Este seria el proceso que se usaría, para finalizar toda la transformacion

