

**UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA**

*Sistemas basados en conocimiento*



**Integrantes:**

* Elvis Ayala
* Fernando León

**Tutor:** Ing. Janneth Chicaiza Espinosa

**Tema:** Definición de URIs y actualización de datos



**1 Especificación de Uri**

En este se definen las **URI’s** a utilizar en el proyecto, para lo cual se van a reutilizar los principales vocabularios a usar en la ontología. Asi como tambien a crear una **URI** base y sus diferentes derivaciones para poder describir los diferentes datos de la ontología.

**1.1 URI’s reutilizadas:**

|  |  |  |
| --- | --- | --- |
| **Recurso** | **Uri** | **Descripción** |
| **DoCO** | http://purl.org/spar/doco | en una ontología que proporciona un vocabulario estructurado escrito de componentes del documento, tanto estructurales, como retóricos. |
| **PRO** | http://purl.org/spar/pro/ | es una ontología para la caracterización de los roles de los agentes: personas, organismos corporativos y agentes computacionales en el proceso de publicación. |
| **Foaf** | http://xmlns.com/foaf/0.1/ | Esta especificación describe el lenguaje FOAF, definido como un diccionario de propiedades y clases con nombre utilizando la tecnología RDF de W3C. |
| **DCMI** | http://purl.org/dc/terms/ | Estos términos están destinados a usarse en combinación con términos de metadatos de otros vocabularios compatibles en el contexto de los perfiles de aplicación. |
| **FRAPO** | http://purl.org/cerif/frapo | es una ontología para describir la información administrativa de proyectos de investigación, por ejemplo, solicitudes de subvenciones, organismos de financiación, socios de proyectos, etc. |
| **FABIO** | http://purl.org/spar/fabio | es una ontología para registrar y publicar en la Web Semántica descripciones de entidades que están publicadas o potencialmente publicables, y que contienen o son referidas por referencias bibliográficas, o entidades utilizadas para definir tales referencias bibliográficas. |

**1.2 Diseño de Uri**

Para el diseño de la uri vamos a tomar una estructura base y a partir de esta crear otras para cada recurso del modelo ontológico.

|  |  |
| --- | --- |
| **Uri base** | **http://utpl.edu.ec/sbc/COVID19publications/ontology/** |
| **Recurso** | **Uri** |
| **catalog** | http://utpl.edu.ec/sbc/COVID19publications/ontology/catalog/ |
| **dataset** | http://utpl.edu.ec/sbc/COVID19publications/ontology/dataset |
| **publication** | http://utpl.edu.ec/sbc/COVID19publicationsontology//publication |
| **person** | http://utpl.edu.ec/sbc/COVID19publications/ontology/person |
| **author** | http://utpl.edu.ec/sbc/COVID19publications/ontology/person/author |
| **organization** | http://utpl.edu.ec/sbc/COVID19publications/ontology/organization |
| **journal** | http://utpl.edu.ec/sbc/COVID19publications/ontology/journal |

**1.3 Definición de la licencia**

**Las licencias Creative Commons:** Luego de analizar cada una de las licencias propuestas se definió usar esta debido a que permite adaptar y compartir la información siempre y cuando se de atribución a los autores de la misma.

**2. Fuentes de Datos Actualizadas**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre de la fuente de datos** | COVID-19 Open Research Dataset Challenge | | |
| **Proveedor** | Kaggle | | <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/data> |
| **Última actualización:** | 09/06/2020 | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre de la fuente de datos** | COVID-19 Open Research Dataset Challenge | | |
| **Proveedor** | bioRxiv | | <https://www.biorxiv.org/search/covid%252B19> |
| **Última actualización:** | 01/06/2020 | | |

**2.1 Preparación de Datos**

**2.1.1 Exploración**

Para la exploración y limpieza de archivos se usó python y algunas librerías necesarias para la exploración como pandas, numpy, json, etc.

Los archivos se encuentran en formato Json, esta estructura resulta complicada para realizar un análisis. Para la exploración de datos se cargan todos los archivos json en una lista de diccionarios anidados, para obtener el total de archivos existentes.

Luego se sacó los datos necesarios que contienen los archivos Json acordes con el modelo conceptual que se planteó, ahora actualizados al modelo ontológico propuesto en “Protege”, se aplicaron algunas funciones necesarias para obtener los datos como autores, organización, locación, abstract, el tipo de recurso bibliográfico.

Una vez obtenidos y limpiados los datos necesarios se procedió a convertir los archivos Json en CSV esto con la finalidad de hacer los datos más legibles al momento de procesarlos por medio de Jena.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CLASE | SUBCLASE | Propiedades Objeto |  | Propiedades Data |
| dc:BibliographicResource | fabio:Article | dc:create | **Dominio:** dc:BibliographicResource | dcterms:identifier  fabio:doi  dcterms:title  dcterms:abstract  dcterms:date  fabio hasURL |
|
|
|
| **Rango:** foaf:Agent |
| prov:wasDerivedFrom | **Dominio:** dc:BibliographicResource |
| **Rango:** dcat: Dataset |
| frbr:part | **Dominio:** dc:BibliographicResource |
| **Rango:** fabio: Journal |
| foaf:Organization | dbo:Institue | dcterms:publisher | **Dominio:** dcat:Catalog | foaf:name |
| **Rango:** foaf:Organization |
|
|
| dcat:Catalog |  | dcat:dataset | **Dominio:** dcat:Catalog | dcat: catalog  dcat:dataset |
| **Rango:** dcat:Dataset |
|
| dcat:Dataset |  |  |  | dcterms:title |
| foaf:Person | foaf:Agent | rbr:part | **Dominio:** dcat:Catalog | foaf:name  foaf:lastname |
| **Rango:** dcat:Dataset |
| fabio:Journal |  |  |  | foaf:name |



