



SISTEMAS BASADOS EN CONOCIMIENTOS

Integrantes:

- Maria Paula Jaramillo
- Karen Torres
- Erika Vásquez
- Stalin Carrión

1. Descripción

Debido a la situación por la que se está pasando en la actualidad referente al virus Covid-19, se ha establecido realizar un análisis, extracción y procesamiento de la información otorgada por el servicio nacional de gestión de riesgos y emergencias del Ecuador. Se han desarrollado informes de situación e infografías desde el 29 de Febrero del 2020, en el cual se puede visualizar diferentes datos a nivel de regiones, provincias y cantones.

Se ha tomado como datos el número de personas contagiadas por cantón, de los cuales se han tomado en cuenta los días desde el 7 de abril del 2020 hasta el 23 de abril del 2020.

2. SKOS

El Sistema simple de organización del conocimiento (SKOS) es un modelo de datos común para compartir y vincular sistemas de organización del conocimiento a través de la Web semántica.

Se ha utilizado este vocabulario utilizando el elemento fundamental, que es el Concept. Permitiendo establecer relaciones entre las etiquetas asociadas a los conceptos. En este caso, se ha hecho uso de las etiquetas prefLabel, altLabel. Además la etiqueta broader, que permite la representación de la relación entre las partes de un todo. Por ejemplo: la región Sierra es parte de Ecuador.

En la Fig.1 se puede observar cómo se ha realizado el uso de este vocabulario, definiendo a las Regiones, Provincias y Cantones como un Concept.

```
eg:Ecuador rdf:type skos:Concept .
eg:Ecuador skos:prefLabel "Ecuador" .
eg:Ecuador skos:altLabel "Ecua"@es .
```

```
eg:Sierra rdf:type skos:Concept .
eg:Sierra skos:prefLabel "Sierra" .
eg:Sierra skos:altLabel "Sierra"@es .
eg:Sierra skos:broader eg:Ecuador .
```

```
eg:Esmeraldas rdf:type skos:Concept .
eg:Esmeraldas skos:prefLabel "Esmeraldas" .
eg:Esmeraldas skos:altLabel "Esmeraldas"@es .
eg:Esmeraldas skos:broader eg:Costa .
```

Fig 1. Uso del vocabulario SKOS

3. Datacube

Este vocabulario permite crear una representación semántica de datos multidimensionales con un enfoque fuerte en sus metadatos. Sin embargo, en el caso de los recursos encontrados en los datasets del Covid-19 en Ecuador, las tablas se presentan en formatos como CSV, XLS o XLSX y presentan datos tabulares con una estructura menos compleja. Por esa razón, la siguiente solución se ha propuesto para abstraer tablas tabulares y ajustarlas de acuerdo con el vocabulario Data cube. Al hacer esto podemos obtener todos los beneficios de la descripción de metadatos y las relaciones que este vocabulario provee.

Clase de Data Cube	Descripción
<i>qb:Dataset</i>	Contiene el conjunto completo de datos junto con su estructura y su metadata que lo describe.
<i>qb:SliceKey</i>	Está asociado a la definición de la estructura de un <i>qb:Dataset</i> representa la dimensión que va a estar fija para poder seleccionar diferentes argumentos de las demás dimensiones y de esa manera extraer una porción de los datos. En este caso la dimensión Provincias ³⁰ siempre estará fija. Luego se selecciona alguna <i>qb:Slice</i> en la dimensión columnas, por ejemplo, el cantón, y podemos obtener la información de todos los cantones de esa columna.
<i>qb:Slice</i>	Permiten agrupar subconjuntos de observaciones de una o más dimensiones que se encuentran fijas, excluyendo a la <i>qb:SliceKey</i> que permanece fija durante toda la representación. En este caso, al definir una dimensión de columnas podemos extraer porciones que representen cada una de las columnas en cualquier recurso y obtener sus valores correspondientes.
<i>qb:Observation</i>	Son los datos medidos o contenidos en cada una de las celdas de la tabla. Cada observación puede ser relacionada con diferentes vocabularios que definan su tipo de dato, que se está midiendo, etc.
<i>qb:DimensionProperty</i>	Representan las dimensiones de la tabla, así como los metadatos asociados a cada una de ellas.

Fig 2. Descripción de clases en Data Cube

Para utilizar este prefijo se propone una tabla donde se presentan las dimensiones identificadas. La tabla 1 presenta los componentes identificados.

	Dim. Periodo		DimensionProperty
	07 abril	08 de abril	
Dim. Area	Dim. Casos	Dim. Casos	
Arenillas	2	3	
Huaquillas	3	5	
Machala	4	4	
Santa Rosa	10	13	
Loja	15	19	
Quilanga	3	5	
Zapotillo	4	4	
Slice Key	Slice		Observation

Tabla 1. Definición de Dimensiones usando el vocabulario DataCube.

Se han definido tres dimensiones: área, casos y periodo. Además, se ha identificado un slice key por área para luego poder dividir las otras dos dimensiones en periodos.

4. GraphDB

Para la visualización de los datos en forma de grafos, se ha decidido utilizar la herramienta GraphDB, la cual es una base de datos gráficos semánticos. La cual hemos utilizado para almacenar, organizar y gestionar el contenido de los datos desarrollados.

Primero creando un repositorio local, segundo cargando los datos ya sea mediante un archivo, url o directamente en texto, seguidamente se utiliza la visualización de grafos para poder buscar y que la herramienta genera automáticamente los grafos que se ven a continuación.

5. Resultados:

Por una parte se representó las regiones, provincias y cantones del Ecuador utilizando el vocablo skos, en la Figura 3 se puede observar la especialización que se va generando entre las provincias y cantones del Ecuador.

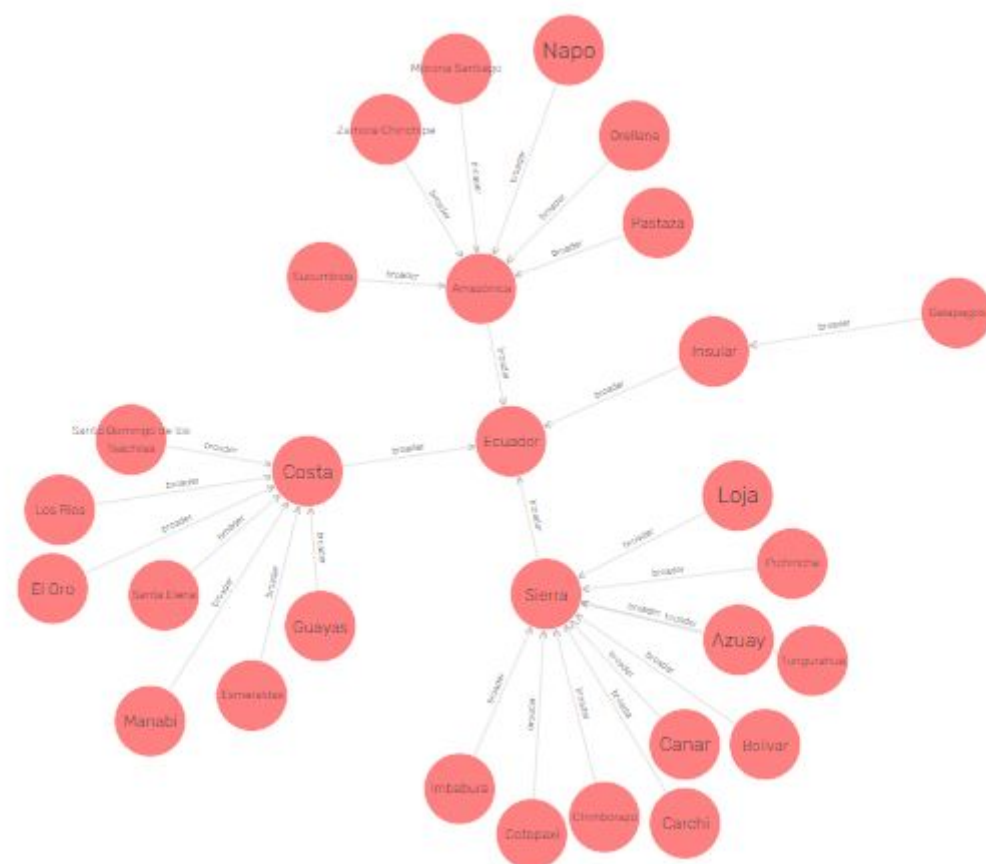


Fig 3. Grafo resultante de las provincias y cantones del Ecuador utilizando el vocabulario skos.

Como ya se mencionó anteriormente, para la representación de los casos confirmados se utilizó Data Cube, desde el cual se crearon varios slice acorde al número de periodos a representar, en este caso se representaron 17 días, en la Figura 4 se puede observar los slice resultantes en forma de grafo.

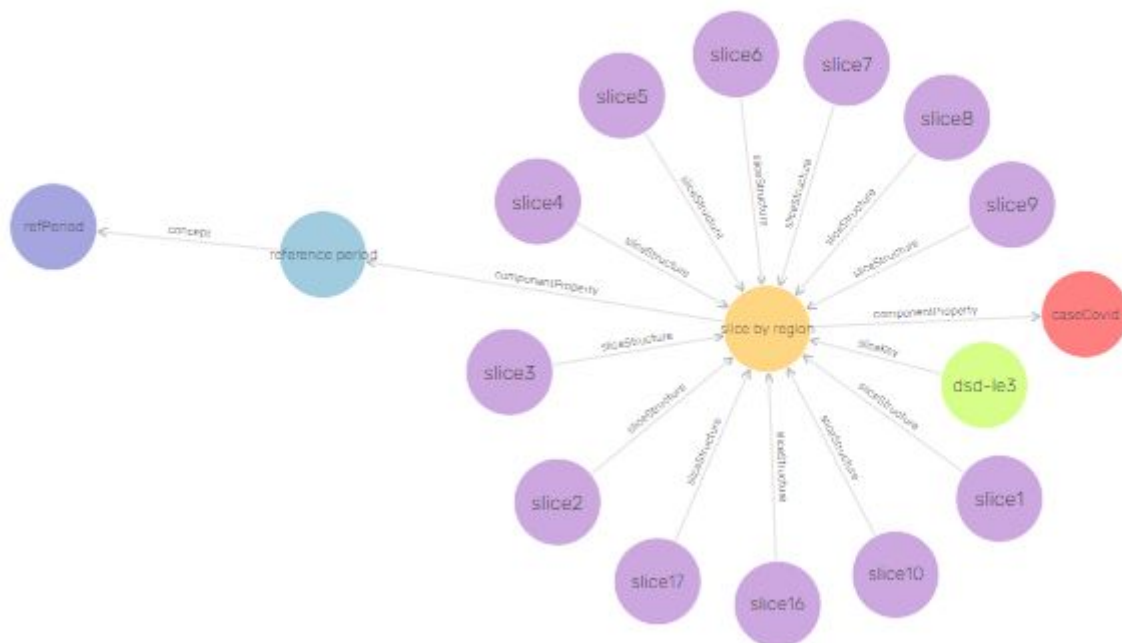


Fig 4. Grafo resultante de los periodos (días) representados por casos de Covid.

Finalmente se extiende las observaciones de cada slice, cada observación representa el número de casos confirmados por cantón, así mismo en la Figura 5 se ilustra el cantón al que corresponde el caso.

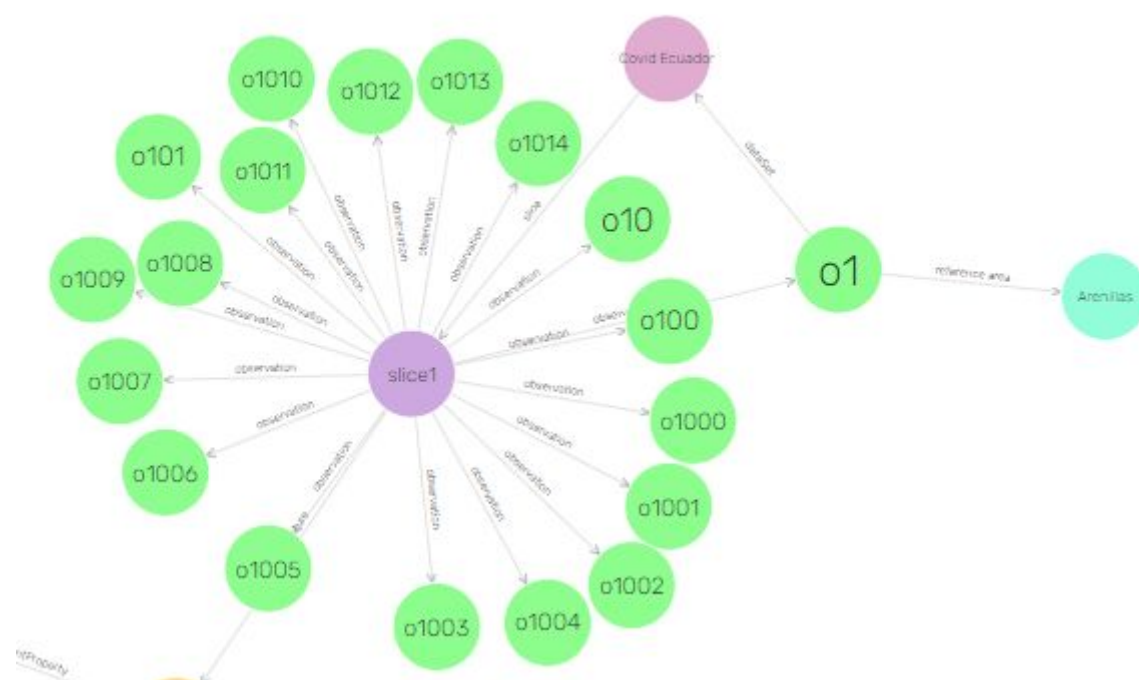


Fig 5. Grafo resultante de las observaciones por casos de Covid.

6. Código

El código Turtle se encuentra en el siguiente repositorio:

<https://github.com/etvasquez/Covid19-DataCube-SKOS>