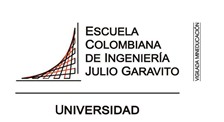
****

**Escuela Colombiana de Ingeniería Julio Garavito**

**Modelos y bases de datos 2024-1**

**Laboratorio VI**

**Cristian Santiago Pedraza Rodríguez**

**Andersson David Sánchez Méndez**

**10 de mayo de 2024**

# MODELOS Y BASES DE DATOS XML y SQL

**2024-01**

# Laboratorio 6/6 [: ) ]

## OBJETIVOS

Evaluar el logro de las competencias adquiridas para:

1. Definir un archivo XML que cumpla con las especificaciones de ejemplares específicos.
2. Validar que un archivo XML está bien formado.
3. Visualizar la estructura de un archivo XML en forma de árbol.
4. Proponer e implementar consultas en XPath
5. Definir el esquema emergente de sus datos usando el estándar DTD
6. Verificar que los datos de un archivo XML cumplen lo definido en un esquema.
7. Integrar información en XML a una base de datos relacional.

## ENTREGA

Publiquen los resultados en un archivo .zip , el nombre de este archivo debe ser la concatenación en orden alfabético de los primeros apellidos de cada uno de los miembros.

## PUNTO UNO. Sólo XML. CATÁLOGO DE PAISES (countries.xml)

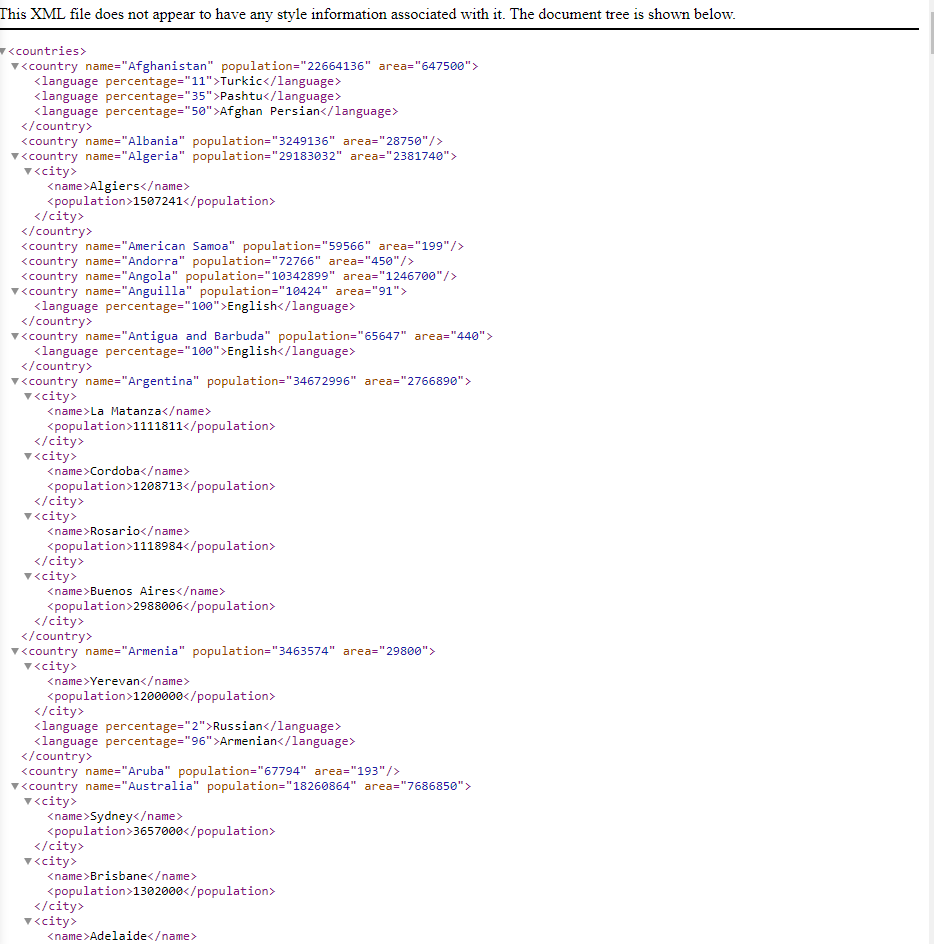
El caso es uno de los propuestos en el curso de bases de datos de **Stanford**.

### Explorando

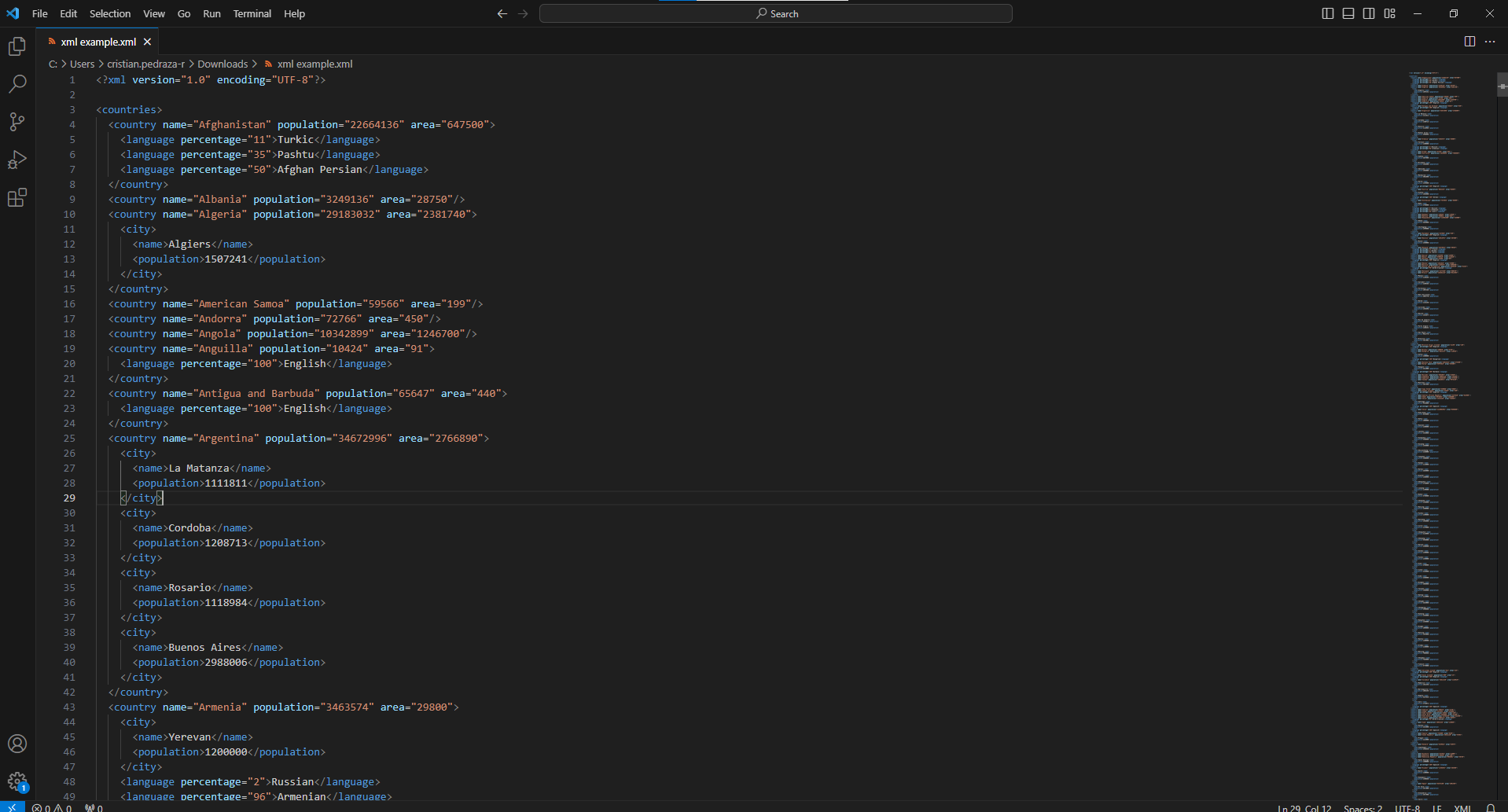
* 1. Explore el archivo usando un navegador y un editor.

¿Cuáles son las ventajas de uno u otro?

En el navegador se ve la estructura del XML, así como labels para desplegar o contraer las etiquetas. La ventaja del navegador es la visualización, pero no se puede editar o modificar el XML.



Ahora, en el apartado del editor, se puede apreciar el código, también se puede expandir o contraer el código, y con este si se puede modificar el XML.



* 1. Describa la información que encuentra en el archivo.

Elemento raíz(<countries>): Contiene todos los datos sobre los países.

Países (<country>): Cada país está representado por un elemento <country>. Cada uno tiene atributos como name (nombre del país), population (población del país) y area (área territorial del país).

Ciudades (<city>): Dentro de cada país, puede haber elementos <city> que representan ciudades importantes. Cada ciudad tiene un elemento <name> que contiene su nombre y un elemento <population> que indica su población.

Lenguajes (<language>): Algunos países tienen elementos <language> que indican los idiomas hablados en ese país. Cada uno de estos elementos tiene un atributo percentage que indica el porcentaje de la población que habla ese idioma. Además, estos elementos contienen el nombre del idioma.

Algunos países muestran:

Afganistán: Muestra la población, área territorial y los idiomas hablados (Turkic, Pashtu, Afghan Persian) con sus respectivos porcentajes de hablantes.

Población: 22,664,136.

Área: 647,500 km².

Idiomas:

Turkic (11%)

Pashtu (35%)

Afghan Persian (50%)

Albania: Muestra la población y el área territorial.

Población: 3,249,136.

Área: 28,750 km².

Algeria: Muestra la población, área territorial y una ciudad con su población (Algiers).

Población: 29,183,032.

Área: 2,381,740 km².

Ciudad:

Argel (población: 1,507,241).

Armenia: Muestra la población, el área territorial, una ciudad con su población (Yerevan) y los idiomas hablados (Russian y Armenian) con sus respectivos porcentajes de hablantes.

Población: 3,463,574.

Área: 29,800 km².

Ciudad:

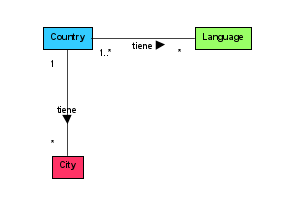
Ereván (población: 1,200,000).

Idiomas:

Ruso (2%)

Armenio (96%)

También se puede modelar en astah los grandes conceptos que se abstrayeron del archivo XML; esto nos da a entender mejor toda la información para validar mejor cuando usemos DTD.



### Consultas iniciales

Implemente las siguientes consultas:

Para evaluar use la herramienta que aparece en moodle.

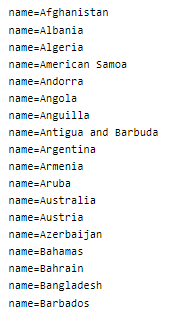
* 1. ¿De cuáles países se tiene información?

[a. Nombre con etiqueta] [b. Nombre sin etiqueta] [c. Nombre con atributo] [d. Nombre sin atributo] [e.Número de países]

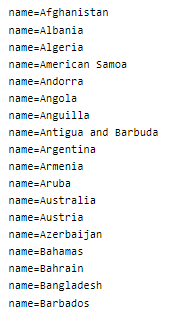
1. countries/country

...

1. /countries/country/@name



1. /countries/country/@name



1. /countries/country/@name/string()



* 1. ¿Cuáles países tienen toda la información? [Nombre]

countries/country[city and language and @name and @population and @area]/@name/string()

Armenia

Australia

Austria

Azerbaijan

Bulgaria

Burma

Chile

Colombia

Cuba

Dominican Republic

France

Georgia

Germany

Hungary

India

Iran

Japan

Kazakstan

Netherlands

Nicaragua

North Korea

Pakistan

Poland

Russia

Saudi Arabia

Serbia and Montenegro

Spain

Uzbekistan

* 1. ¿De qué países se conocen sus ciudades?[Nombre]

countries/country [city]/@name/string()

Algeria

Argentina

Armenia

Australia

Austria

Azerbaijan

Bangladesh

Belarus

Brazil

Bulgaria

Burma

Canada

Chile

China

Colombia

Cuba

Czech Republic

Denmark

Dominican Republic

Ecuador

Egypt

Ethiopia

France

Georgia

Germany

Hungary

India

Indonesia

Iran

Iraq

Italy

Japan

Kazakstan

Kenya

Madagascar

Malaysia

Mexico

Morocco

Netherlands

Nicaragua

Nigeria

North Korea

Pakistan

Peru

Philippines

Poland

Romania

Russia

Saudi Arabia

Senegal

Serbia and Montenegro

Singapore

South Korea

Spain

Sudan

Syria

Taiwan

Tanzania

Thailand

Turkey

Ukraine

United Kingdom

United States

Uruguay

Uzbekistan

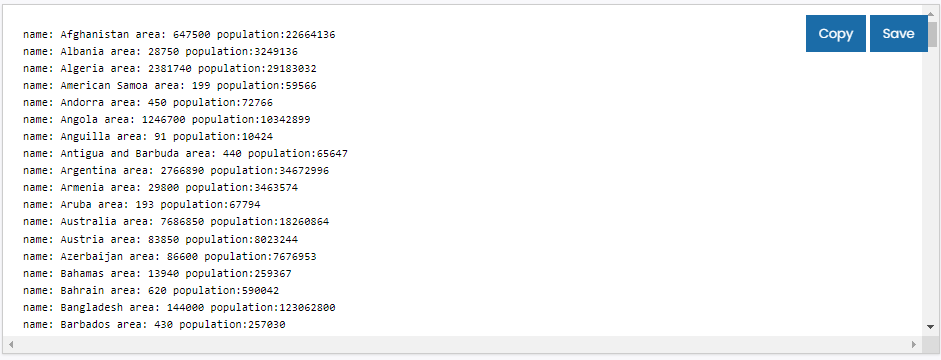
Venezuela

Vietnam

Zaire

* 1. ¿De qué países sólo se conocen sus datos básicos (nombre, área y población)? [Nombre, area y poblacion]

countries/country [@name and @area and @population]/concat("name: ",@name," area: ", @area," population:", @population)



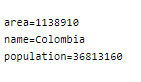
* 1. ¿Cuál es la información de Colombia?

[a. Toda la información]

...

[b. Todos los atributos]

countries/country [@name = "Colombia"]/@\*

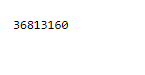


[c. Todas las etiquetas]



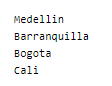
* 1. ¿Cuántos habitantes tiene Colombia? [Habitantes]

countries/country[@name = "Colombia"]/@population/string()



* 1. ¿Cuáles son las principales ciudades de Colombia? [Nombre]

countries/country [@name = "Colombia"]/city/name/text()



* 1. ¿Cuáles países tienen una única ciudad importante? [Nombre]

countries/country [count(city) =1]/@name/string()



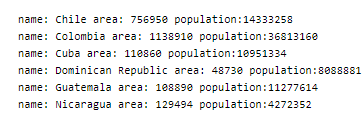
* 1. ¿Cuáles ciudades tienen menos de 2 millones de habitantes? [Nombre]

countries/country /city [count(@population) < 2000000]/name/text()

....

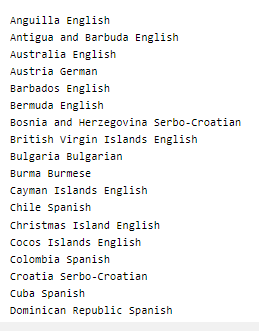
* 1. ¿En qué países se habla únicamente español? [Nombre][Toda la información]

countries/country [language = "Spanish"]/concat("name: ",@name," area: ", @area," population:", @population)



* 1. ¿En qué países se habla un único idioma? [Nombre] [Nombre]

countries/country [count(language) =1]/concat((@name/string())," ",(language/text()))



* 1. ¿En qué países no existe un idioma mayoritario? (Lo conocen más del 50%)[Nombre]

countries/country [language/@percentage > 50]/@name/string()



* 1. ¿Cuál es el país más grande? [Nombre]

/countries/country[@population = max(/countries/country/@population)]/@name

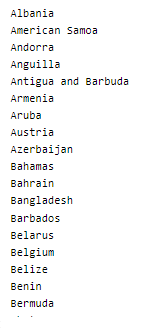


/countries/country[@area = max(/countries/country/@area)]/@name/string()



* 1. ¿Cuáles son los países de área inferior al promedio? [Nombre]

/countries/country[@area < avg(/countries/country/@area)]/@name/string()

...

### Consultas quiz Stanford

* 1. Return to the area of Mongolia.

countries/country [@name = "Mongolia"]/@area/string()



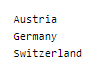
* 1. Return the names of all cities that have the same name as the country in which they are located.

countries/country [(@name/string()) = (city/name/text())]/@name/string()



* 1. Return the names of all countries where over 50% of the population speaks German.

countries/country[language[text() = 'German' and @percentage > 50]]/@name/string()



* 1. Return the name of the country with the highest population.

/countries/country[@population = max(/countries/country/@population)]/@name/string()



* 1. Return the names of all countries that have at least three cities with a population greater than 3 million.

countries/country [(count(city[population > 3000000] ) >= 3) ]/@name/string()



* 1. Return the names of French-speaking and German-speaking countries.

countries/country [language[text() = "German" or text() = "French"]]/@name/string()



* 1. Return the names of all countries containing a city such that some other country has a city of the same name.

//city[name = parent::country/@name]/name/string()



* 1. Return the languages spoken in countries where Russian is spoken.

distinct-values(countries/country [language [text() = "Russian"]]/language/text())



* 1. Return all country name where the country textually contains a language spoken in the country.

//country[language[contains(../@name,.)]]/@name/string()



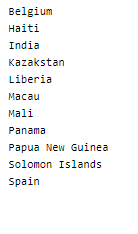
* 1. Return all countries that have at least one city with population greater than 7 million.

countries/country [city[population > 7000000]]/@name/string()



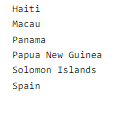
* 1. Return all countries where at least one language is listed, but the total percentage for all listed languages is less than 90%.

countries/country [language and sum(language/@percentage) < 90]/@name/string()



* 1. Return all countries where at least one language is listed, and every listed language is spoken by less than 20% of the population.

countries/country [language and language//@percentage < 20]/@name/string()



* 1. Find all situations where one country's most popular language is another country's least popular, and both countries list more than one language.

/countries/country[language[count(../language) > 1 and not(number(@percentage) = max(../language/number(@percentage)))]]/@name/string()



### Consultas propias

Propongan e implementen tres consultas propias

Return the names of all countries for which the data does not

include any languages or cities, but the country has more than 10

million people.

countries/country[not(language) and not(city)][@population > 10000000]/@name/string()



Return the names of all countries where a city in that country

contains more than one-third of the country's population.

countries/country[city/population > (@population div 3)]/@name/string()



Return the population density of Qatar.

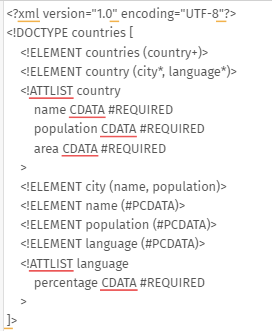
countries/country[@name="Qatar"]/@population div //country[@name="Qatar"]/@area



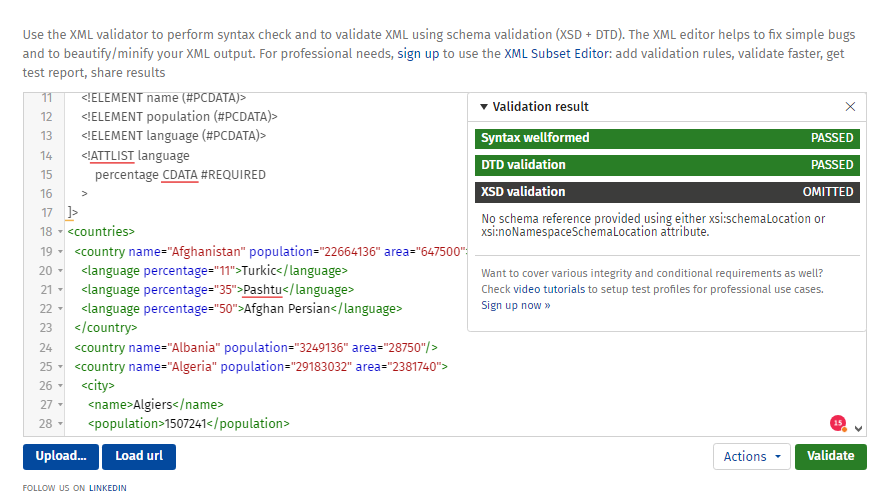
### Esquema

Proponga un DTD para estos datos.

Para verificar use la herramienta que aparece en moodle.

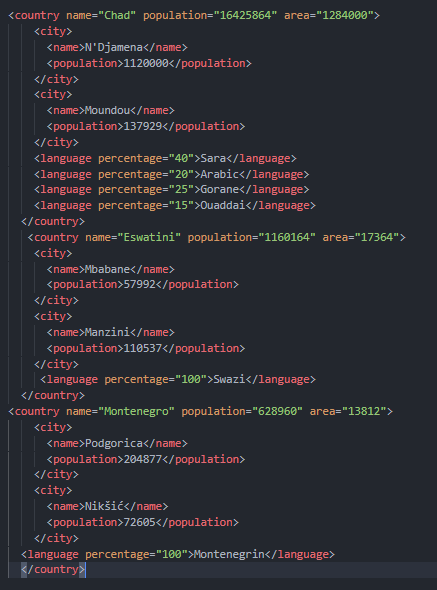


Para la validación se usó: <https://www.truugo.com/xml_validator/>

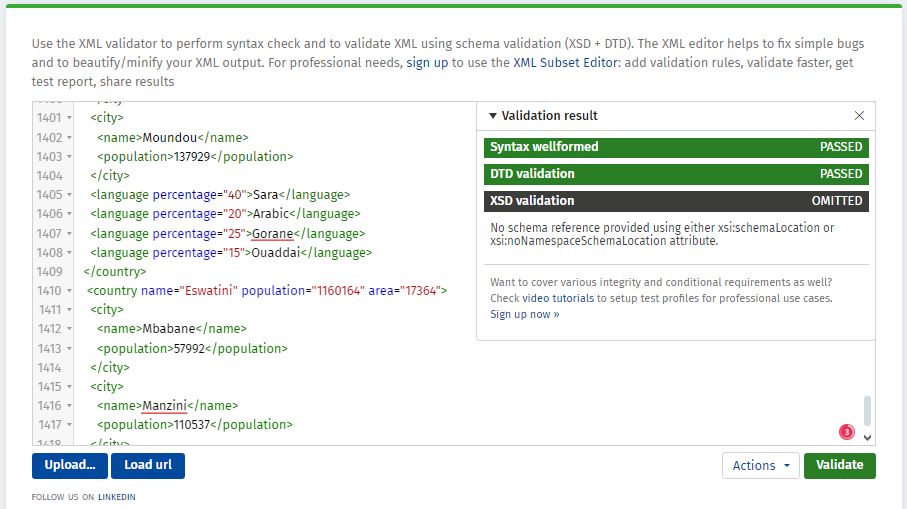


### Nuevos datos

Incluya en el archivo la información de tres países que no existan en el archivo. Verifique que cumple el esquema definido anteriormente.



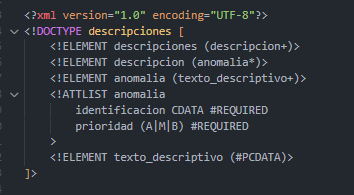
Para la validación se usó: <https://www.truugo.com/xml_validator/>



### PUNTO DOS. (XML – Oracle) trueques Evaluación. TDescripcion

Este punto vamos a perfeccionar uno de los atributos de la base de datos.

1. Proponga la estructura XML necesaria para tener la información del atributo descripción. DTD y ejemplos XML OK y XML NoOK. Explique.



La estructura propuesta consta de 3 etiquetas y dos atributos, uno de los cuales fue restringido a una lista enumerada para garantizar que solo las cadenas en dicha lista sean admitidas (prioridad).

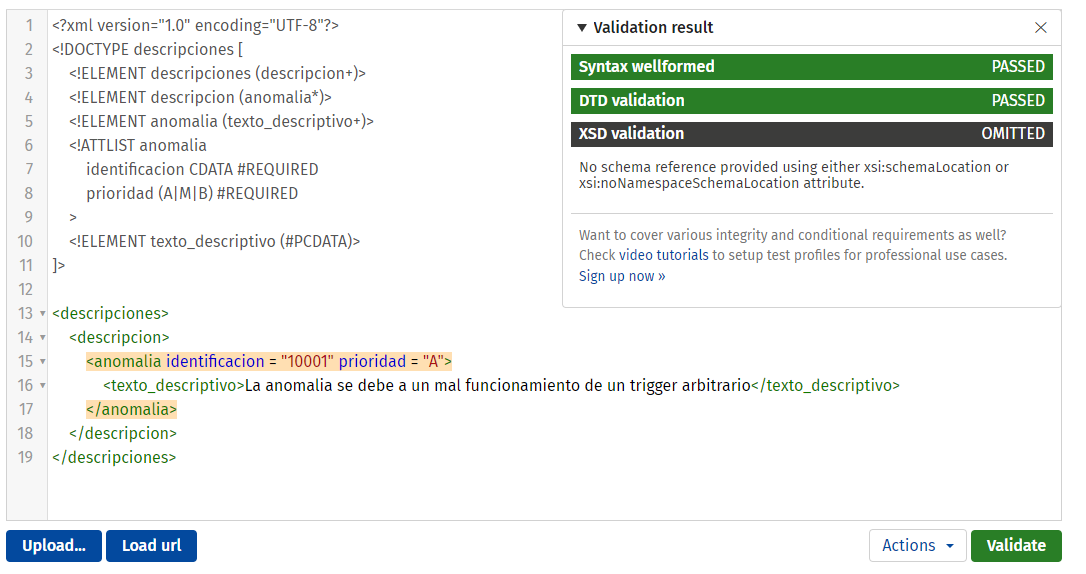
La raíz es "descripciones", la cabeza del árbol. "Descripciones" debe tener al menos un elemento hijo ("descripcion"). Por otro lado, el nodo "descripcion" puede o no tener anomalías.

Pero el nodo "anomalia" debe tener necesariamente un nodo "texto\_descriptivo".

En "anomalia" se establecieron los atributos: "identificacion", que acepta cadenas de caracteres, y "prioridad", que se restringió con la lista de caracteres [A,M,B], y solo se aceptarán este tipo de caracteres.

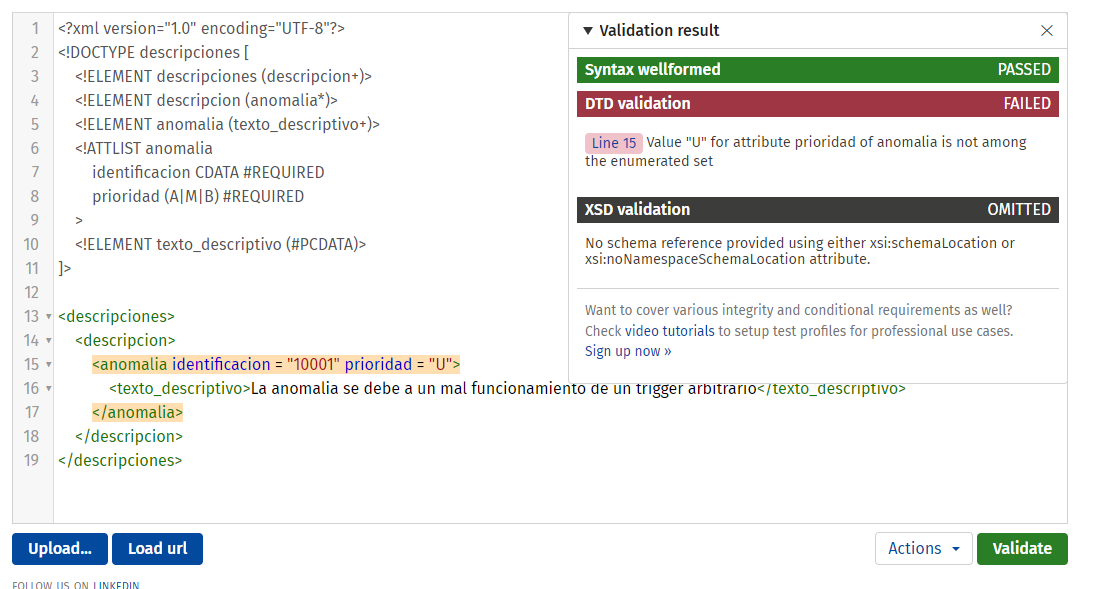
La etiqueta "texto\_descriptivo" es de tipo #PCDATA para aceptar cadenas de caracteres. Se eligió como etiqueta dado que se aceptarán extensas cadenas.

* **XML OK:**

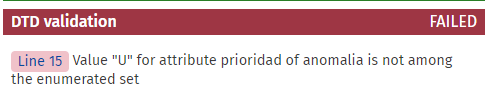
****

La validación fue exitosa dado que se cumple con el DTD definido.

* XML NO OK:



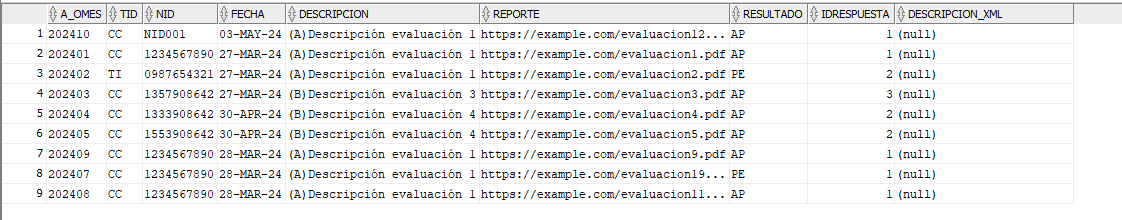
La validación no fue exitosa dado que no se cumple con el DTD definido. En este caso, se está estableciendo como prioridad U, pero en la definición de DTD, solo se aceptan prioridades A, M o B.



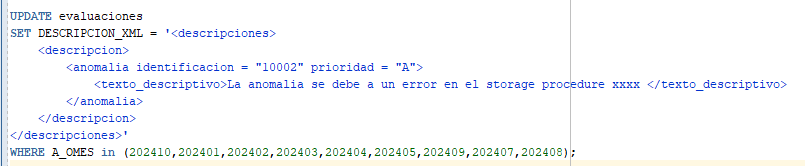
1. Actualice la tabla y los datos adicionados a la base de datos. (PoblandoOK, PoblandoNoOK).

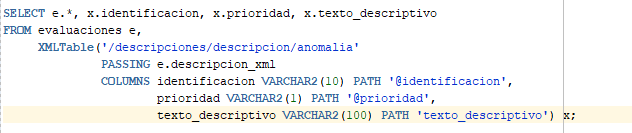


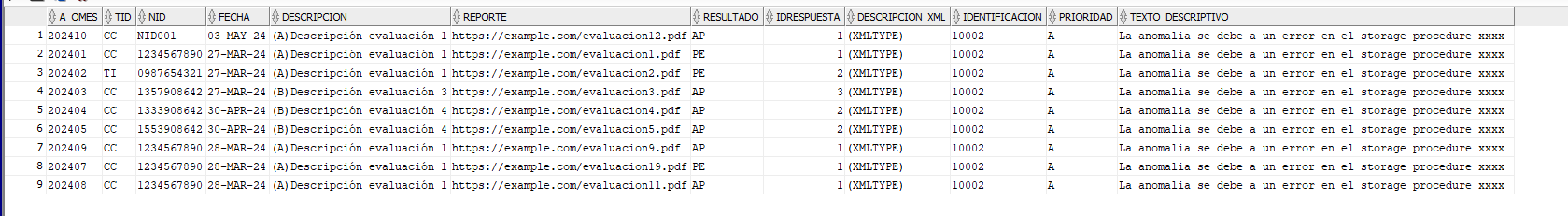




POBLANDO OK

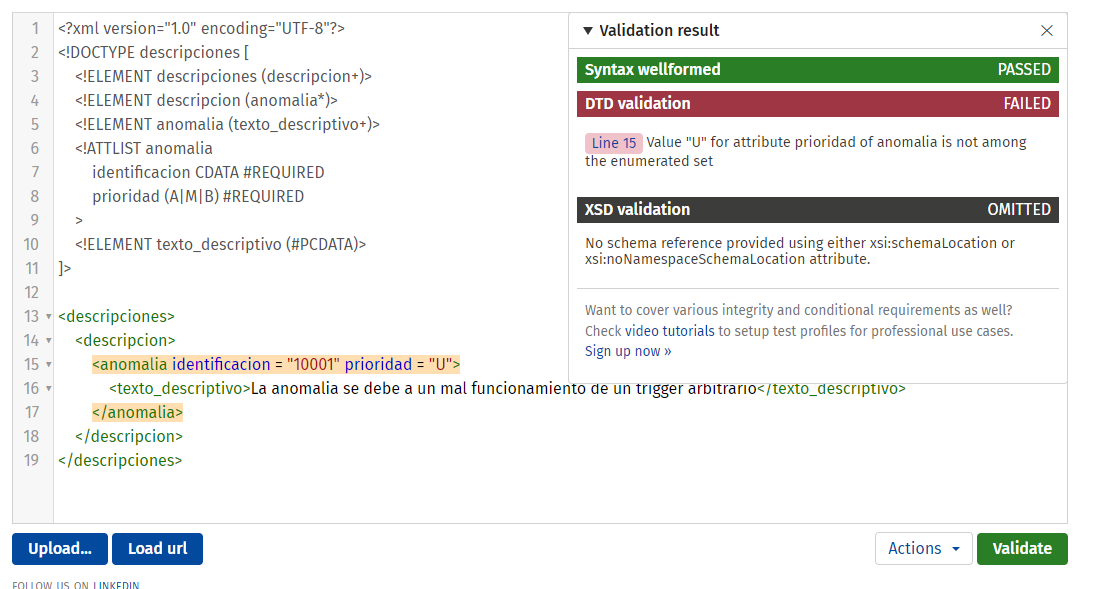


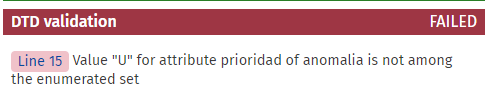




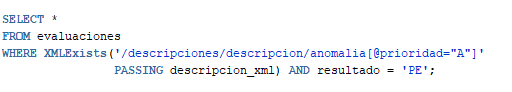
POBLANDO NO OK







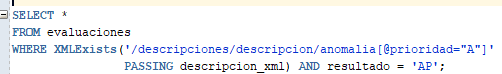
1. Implemente la consulta *Consultar anomalías de prioridad alta en estado pendiente.*





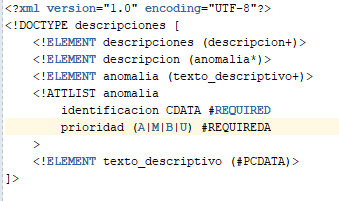
1. Proponga otra nueva consulta que use ese atributo (Diseño e implementación)

Anomalía con estado aprobado y prioridad alta

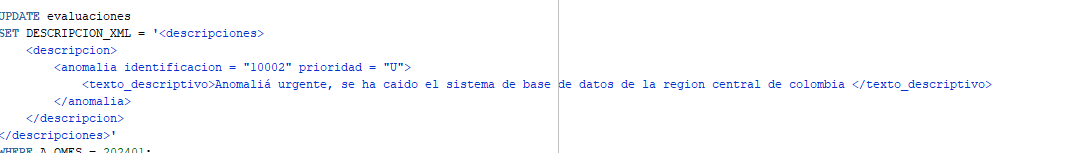




1. Extienda la información de este atributo (DTD) y proponga una nueva consulta que ilustre la pertinencia de la nueva información registrada en XML. (Diseño e implementación)

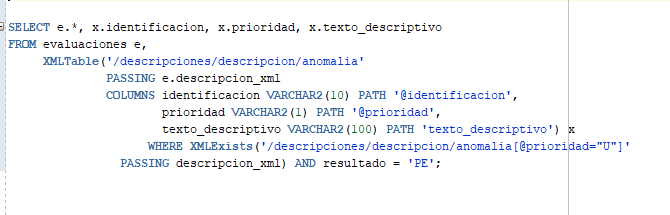


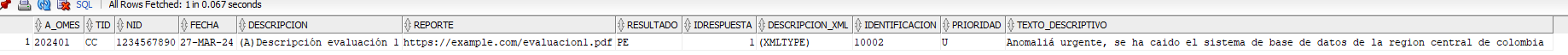
Se añadió al conjunto de caracteres ‘U’ de urgente, en donde se clasifica una anomalía como urgente.





### 





### RETROSPECTIVA

* 1. ¿Cuál fue el tiempo total invertido en el laboratorio por cada uno de ustedes? (Horas/Hombre)

El tiempo total invertido fue de 15 horas.

* 1. ¿Cuál es el estado actual del laboratorio? ¿Por qué?

Completamos todos los requisitos de entrega haciendo que todo funcionara perfectamente.

* 1. ¿Cuál consideran fue el mayor logro? ¿Por qué?

Implementar algunas consultas que costaron más trabajo porque tocó investigar algunas otras maneras que se pueden hacer uso en XML para dar solución a esos problemas.

* 1. ¿Cuál consideran que fue el mayor problema técnico? ¿Qué hicieron para resolverlo?

Verificar que el XML cumpla el DTD porque en sintaxis estaba bien, pero marcaba ya error en el DTD, y eso es necesario para verificar que el XML esté bien. Se solucionó investigando y llegar hasta el punto de resolver el problema.

* 1. ¿Qué hicieron bien como equipo? ¿Qué se comprometen a hacer para mejorar los resultados?

El trabajo en equipo fue vital para el desarrollo de este y todos los laboratorios, así, fue muy importante trabajar de esta manera, y en cada laboratorio aumentar más habilidades y conocimiento.