MODELOS Y BASES DE DATOS
XML y SQL
Laboratorio 6/ 6

Mateo Olaya Garzón, Andrés Camilo Oñate Quimbayo

PUNTO UNO. Sólo XML. CATALOGO DE PAISES (countries.xml)

A. Explorando

1. Explore el archivo usando un navegador y un editor. ¿Cuáles son las ventajas de uno u otro?

Es más sencillo visualizar el contenido del archivo XML en el navegador, pero a diferencia de un editor no se pueden realizar modificaciones.

2. Describa la información que encuentra en el archivo.

1. La raiz del archivo es <countries>
2. Con elementos con etiqueta <country> <language> <city> <name> <population>
3. Las etiquetas <country> tienen atributos name, population y area
4. Las etiquetas <language> tiene un atributo percentage y texto asociado
5. Las etiquetas name, population tienen texto

   El archivo representa la descripción de un listado de países con información como: nombre, área y ciudades principales.

B. Consultas iniciales

Implemente las siguientes consultas:
Para evaluar use la herramienta que aparece en moodle.

1. ¿De cúales países se tiene información?

- Nombre con etiqueta

countries/*

```xml
<country name="Afghanistan" population="22664136" area="647500">
<language percentage="11">Turkic</language>
<language percentage="35">Pashtu</language>
<language percentage="50">Afghan Persian</language>
</country>
<country name="Albania" population="3249136" area="28750"/>
<country name="Algeria" population="29183032" area="2381740">
<city>
<name>Algiers</name>
<population>1507241</population>
</city>
</country>
<country name="American Samoa" population="59566" area="199"/>
<country name="Andorra" population="72766" area="450"/>
<country name="Angola" population="10342899" area="1246700"/>
<country name="Anguilla" population="10424" area="91">
<language percentage="100">English</language>
</country>
<country name="Antigua and Barbuda" population="65647" area="440">
<language percentage="100">English</language>
</country>
<country name="Argentina" population="34672996" area="2766890">
<city>
<name>La Matanza</name>
<population>1111811</population>
</city>
```

- Nombre sin etiqueta
  countries/country/@name

```
name="Afghanistan"
name="Albania"
name="Algeria"
name="American Samoa"
name="Andorra"
name="Angola"
name="Anguilla"
name="Antigua and Barbuda"
name="Argentina"
name="Armenia"
name="Aruba"
name="Australia"
name="Austria"
name="Azerbaijan"
name="Bahamas"
name="Bahrain"
name="Bangladesh"
name="Barbados"
name="Belarus"
name="Belgium"
name="Belize"
name="Benin"
name="Bermuda"
name="Bhutan"
name="Bolivia"
name="Bosnia and Herzegovina"
name="Botswana"
name="Brazil"
name="British Virgin Islands"
name="Brunei"
name="Bulgaria"
name="Burkina Faso"
```

- Nombre con atributo
  countries/country/@*

```
name="Afghanistan"
population="22664136"
area="647500"
name="Albania"
population="3249136"
area="28750"
name="Algeria"
population="29183032"
area="2381740"
name="American Samoa"
population="59566"
area="199"
```

- Nombre sin atributo

countries/country/data(@name)

```
Afghanistan
Albania
Algeria
American Samoa
Andorra
Angola
Anguilla
Antigua and Barbuda
Argentina
Armenia
Aruba
Australia
```

- Número de países

  count(countries/country/@name)

```
1  231.0
2
```

2. ¿Cuántos países están detallados? ¿Cuáles son?

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

```
name="Armenia"
name="Australia"
name="Austria"
name="Azerbaijan"
name="Bulgaria"
name="Burma"
name="Chile"
name="Colombia"
name="Cuba"
name="Dominican Republic"
name="France"
name="Georgia"
name="Germany"
name="Hungary"
name="India"
name="Iran"
name="Japan"
name="Kazakstan"
name="Netherlands"
name="Nicaragua"
name="North Korea"
name="Pakistan"
name="Poland"
name="Russia"
name="Saudi Arabia"
name="Serbia and Montenegro"
name="Spain"
name="Uzbekistan"
```

3. ¿De qué países se conocen sus ciudades y sus lenguas?

countries/country[language and city]

```
<country name="Armenia" population="3463574" area="29800">
<city>
<name>Yerevan</name>
<population>1200000</population>
</city>
<language percentage="2">Russian</language>
<language percentage="96">Armenian</language>
</country>
<country name="Australia" population="18260864" area="7686850">
<city>
<name>Sydney</name>
<population>3657000</population>
</city>
<city>
<name>Brisbane</name>
<population>1302000</population>
</city>
<city>
<name>Adelaide</name>
<population>1050000</population>
</city>
<city>
<name>Melbourne</name>
<population>3081000</population>
</city>
<city>
<name>Perth</name>
<population>1193000</population>
</city>
<language percentage="100">English</language>
</country>
<country name="Austria" population="8023244" area="83850">
<city>
<name>Vienna</name>
<population>1583000</population>
</city>
```

4. ¿De qué países sólo se conocen sus datos básicos (nombre, area y población)?

countries/country[not(language) and not(city)]/@name

```
name="Albania"
name="American Samoa"
name="Andorra"
name="Angola"
name="Aruba"
name="Bahamas"
name="Bahrain"
name="Belize"
name="Benin"
name="Bhutan"
name="Bolivia"
name="Botswana"
name="Brunei"
name="Burkina Faso"
name="Burundi"
name="Cambodia"
name="Cameroon"
name="Cape Verde"
name="Central African Republic"
name="Chad"
name="Comoros"
name="Congo"
name="Cook Islands"
name="Costa Rica"
name="Cote dIvoire"
```

5. ¿Cuál es la información de Colombia?

a. Toda la información
countries/country[@name = "Colombia"]

```
<country name="Colombia" population="36813160" area="1138910">
<city>
<name>Medellin</name>
<population>1621356</population>
</city>
<city>
<name>Barranquilla</name>
<population>1064255</population>
</city>
<city>
<name>Bogota</name>
<population>5237635</population>
</city>
<city>
<name>Cali</name>
<population>1718871</population>
</city>
<language percentage="100">Spanish</language>
</country>
```

b. Todos los atributos

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

c. Todas las etiquetas
countries/country[@name = "Colombia"]/*

```
<city>
<name>Medellin</name>
<population>1621356</population>
</city>
<city>
<name>Barranquilla</name>
<population>1064255</population>
</city>
<city>
<name>Bogota</name>
<population>5237635</population>
</city>
<city>
<name>Cali</name>
<population>1718871</population>
</city>
<language percentage="100">Spanish</language>
```

6. ¿Cuántos habitantes tiene Colombia?

countries/country[@name = "Colombia"]/@population

```
population="36813160"
```

7. ¿Cuáles son las principales cuidades de Colombia?
   countries/country[@name = "Colombia"]/city/name

```
<name>Medellin</name>
<name>Barranquilla</name>
<name>Bogota</name>
<name>Cali</name>
```

8. ¿Cuáles países tienen una única ciudad importante?
   countries/country[count(city) = 1]/@name

```
name="Algeria"
name="Armenia"
name="Austria"
name="Azerbaijan"
name="Belarus"
name="Bulgaria"
name="Burma"
name="Canada"
name="Chile"
name="Cuba"
name="Czech Republic"
name="Denmark"
name="Dominican Republic"
name="Ethiopia"
name="France"
name="Georgia"
name="Hungary"
name="Iraq"
name="Kazakstan"
name="Kenya"
name="Madagascar"
name="Malaysia"
name="Nicaragua"
name="North Korea"
name="Peru"
name="Poland"
name="Romania"
name="Saudi Arabia"
name="Senegal"
name="Serbia and Montenegro"
name="Singapore"
name="Sudan"
name="Syria"
name="Tanzania"
name="Thailand"
name="Uruguay"
name="Uzbekistan"
name="Zaire"
```

9. ¿Cuáles países tienen menos de 500 millones de habitantes?
   countries/country[@population < "50000000"]/@name

```
name="Afghanistan"
name="Albania"
name="Algeria"
name="American Samoa"
name="Andorra"
name="Angola"
name="Anguilla"
name="Antigua and Barbuda"
name="Argentina"
name="Armenia"
name="Aruba"
name="Australia"
name="Austria"
name="Azerbaijan"
name="Bahamas"
name="Bahrain"
name="Bangladesh"
name="Barbados"
name="Belarus"
name="Belgium"
name="Belize"
name="Benin"
name="Bermuda"
name="Bhutan"
name="Bolivia"
name="Bosnia and Herzegovina"
name="Botswana"
name="Brazil"
name="British Virgin Islands"
```

10. ¿En qué países se habla únicamente español?
countries/country[language[@percentage = "100" and (text() = "Spanish")]] /*

```
<city>
<name>Santiago</name>
<population>4318000</population>
</city>
<language percentage="100">Spanish</language>
<city>
<name>Medellin</name>
<population>1621356</population>
</city>
<city>
<name>Barranquilla</name>
<population>1064255</population>
</city>
<city>
<name>Bogota</name>
<population>5237635</population>
</city>
<city>
<name>Cali</name>
<population>1718871</population>
</city>
<language percentage="100">Spanish</language>
<city>
<name>Havana</name>
<population>2241000</population>
</city>
<language percentage="100">Spanish</language>
<city>
<name>Santo Domingo</name>
<population>1400000</population>
</city>
<language percentage="100">Spanish</language>
<city>
<name>Managua</name>
<population>1195000</population>
</city>
<language percentage="100">Spanish</language>
```

11. ¿En qué países se habla un único idioma?

countries/country[language[@percentage = "100"]] /@name

```
name="Anguilla"
name="Antigua and Barbuda"
name="Australia"
name="Austria"
name="Barbados"
name="Bermuda"
name="British Virgin Islands"
name="Bulgaria"
name="Burma"
name="Cayman Islands"
name="Chile"
name="Christmas Island"
name="Cocos Islands"
name="Colombia"
name="Cuba"
name="Dominican Republic"
name="Falkland Islands"
name="France"
name="French Guiana"
name="Germany"
name="Guinea"
name="Iceland"
name="Japan"
name="Montserrat"
name="Mozambique"
name="Nepal"
name="Netherlands"
name="Nicaragua"
name="North Korea"
name="Norway"
name="Poland"
name="Portugal"
name="Russia"
name="Saint Helena"
name="Saint Kitts and Nevis"
name="Saint Pierre and Miquelon"
name="San Marino"
name="Sao Tome and Principe"
name="Saudi Arabia"
name="Sweden"
name="Turks and Caicos Islands"
name="Yemen"
```

12. ¿En qué países no existe un idioma mayoritario?

countries/country[max(*/@percentage) <= 50]/@name

```
name=Afghanistan
name=Haiti
name=India
name=Kazakstan
name=Liberia
name=Macau
name=Pakistan
name=Panama
name=Papua New Guinea
name=Solomon Islands
name=Spain
```

13. ¿Cuál es el país más pequeño?

countries/country[@area = min(/countries/country/@area)]/@name

```
name=Holy See
```

14. ¿Cuáles son los países de área superior al promedio?

    countries/country[@area > avg(/countries/country/@area)]/@name

```
name=Afghanistan
name=Algeria
name=Angola
name=Argentina
name=Australia
name=Bolivia
name=Botswana
name=Brazil
name=Burma
name=Canada
name=Central African Republic
name=Chad
name=Chile
name=China
name=Colombia
name=Egypt
name=Ethiopia
name=Greenland
name=India
```

B. Consultas quiz Standford

1. Return the area of Mongolia.
   countries/country[@name = "Mongolia"]/@area

```
1  area="1565000"
2
```

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

   countries/country/city[data(name) = (/countries/country/@name)]/data(name)

```
Mexico
Singapore
```

3. Return the names of all countries where over 50% of the population speaks German.
   countries/country[language[text() = "German" and @percentage > 50]]/data(@name)

```
Austria
Germany
Switzerland
```

4. Return the name of the country with the highest population.

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

China

5. Return the names of all countries that have at least three cities with population greater than 3 million
   countries/country[count(city[population > 3000000]) >= 3]/data(@name)

China
India

6. Return the names of French-speaking and German-speaking countries.
   countries/country[language[text()= "French"] and language[text()= "German"]]/data(@name)

Belgium
Switzerland

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

countries/country[city/name = (preceding-sibling::city/name|following-sibling::city/name)]/@name

8. Return the languages spoken in countries where Russian is spoken.
   countries/country[language = "Russian"]/data(language)

Russian
Armenian
Russian
Armenian
Azeri
Russian
Armenian
Azeri
Georgian
Russian
Russian
Uzbek
Turkmen
Russian
Tajik
Uzbek

9. Return all country name where the country textually contains a language spoken in the country.
   countries/country[language[contains(parent::country/@name, self::language)]]/data(@name)

```
French Guiana
Germany
Kazakstan
Turkmenistan
Uzbekistan
```

10. Return all countries that have at least one city with population greater than 7 million.
    countries/country[city/population > 7000000]/data(@name)

```
Brazil
China
India
Indonesia
Japan
Mexico
Pakistan
Russia
South Korea
Turkey
United States
```

11. 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][sum(language/@percentage) < 90]/@name

```
name="Belgium"
name="Haiti"
name="India"
name="Kazakstan"
name="Liberia"
name="Macau"
name="Mali"
name="Panama"
name="Papua New Guinea"
name="Solomon Islands"
name="Spain"
```

12. 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][max(language/@percentage) < 20]/@name

```
name=Haiti
name=Macau
name=Panama
name=Papua New Guinea
name=Solomon Islands
name=Spain
```

13. 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[count(language) > 1][max(language/@percentage) = (min(following-sibling::language/@percentage)|min(preceding-

sibling::language/@percentage))]

C. Consultas propias

Propongan e implementen cinco consultas propias

a. Países con un área mayor a 500.000 y población mayor a 10´000.000
countries/country[@area > 500000 and @population > 10000000]/@name

```
name="Afghanistan"
name="Algeria"
name="Angola"
name="Argentina"
name="Australia"
name="Brazil"
name="Burma"
name="Canada"
name="Chile"
name="China"
name="Colombia"
name="Egypt"
name="Ethiopia"
name="France"
name="India"
name="Indonesia"
name="Iran"
name="Kazakstan"
name="Kenya"
name="Madagascar"
name="Mexico"
name="Mozambique"
name="Nigeria"
name="Pakistan"
name="Peru"
name="Russia"
name="Saudi Arabia"
name="South Africa"
name="Spain"
name="Sudan"
name="Tanzania"
name="Thailand"
name="Turkey"
name="Ukraine"
name="United States"
name="Venezuela"
name="Yemen"
name="Zaire"
```

b. Países que hablen más de 4 idiomas
countries/country[count(language) > 4]/@name

```
name="Iran"
name="Pakistan"
```

c. Países cuya area es menor al promedio del área de todos los países
countries/country[@area > avg(/countries/country/@area)]/@name

```
name=Afghanistan
name=Algeria
name=Angola
name=Argentina
name=Australia
name=Bolivia
name=Botswana
name=Brazil
name=Burma
name=Canada
name=Central African Republic
name=Chad
name=Chile
name=China
name=Colombia
name=Egypt
name=Ethiopia
name=Greenland
name=India
```

d. El país menos poblado del mundo
countries/country[@population = min(/countries/country/@population)]/@name

```
name=Pitcairn Islands
```

e. País el cual se tienen más ciudades registradas

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

```
name=China
```

D. Esquema

Proponga un DTD para estos datos.
Para verificar use la herramienta que aparece en moodle

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE countries[
<!ELEMENT countries (country)+>
<!ELEMENT country (city | language)*>
<!ELEMENT city (name, population)>
<!ELEMENT language (#PCDATA)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT population (#PCDATA)>
<!ATTLIST country name CDATA #REQUIRED>
<!ATTLIST country population CDATA #REQUIRED>
<!ATTLIST country area CDATA #REQUIRED>
<!ATTLIST language percentage CDATA #REQUIRED>
]>
```

E. Nuevos datos

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

```
<country name="Democratic Republic of Congo" population="89561000" area="2344858"/>
  <city>
    <name>Kinsasa</name>
    <population>7787832</population>
  </city>
</country>
<country name="Ivory Coast" population="26453542" area="322463"/>
  <city>
    <name>Yamusukro</name>
    <population>200103</population>
  </city>
</country>
<country name=" Republic of South Sudan" population="10975920" area="644329"/>
  <city>
    <name>Juba</name>
    <population>372410</population>
  </city>
</country>
```

PUNTO DOS. (XML – Oracle) youtube. Subscription. Details

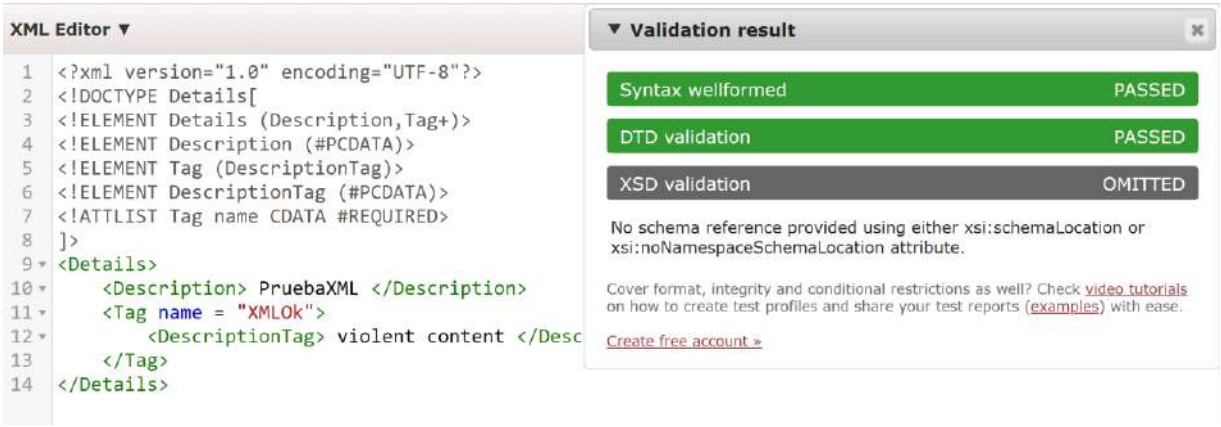
1. Proponga la estructura XML necesaria para tener la información de este atributo. DTD y ejemplos XML OK y XML NoOK. Explique.
   Se propone la siguiente estructura:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Details[
<!ELEMENT Details (Description,Tag+)>
<!ELEMENT Description (#PCDATA)>
<!ELEMENT Tag (DescriptionTag)>
<!ELEMENT DescriptionTag (#PCDATA)>
<!ATTLIST Tag name CDATA #REQUIRED>
]>
```

   La etiqueta raíz es *Details*, que tendrá elementos con etiqueta *Description* y *Tag*.
   Las etiquetas *tag* tienen un *name* como atributo y contienen una etiqueta con la descripción.

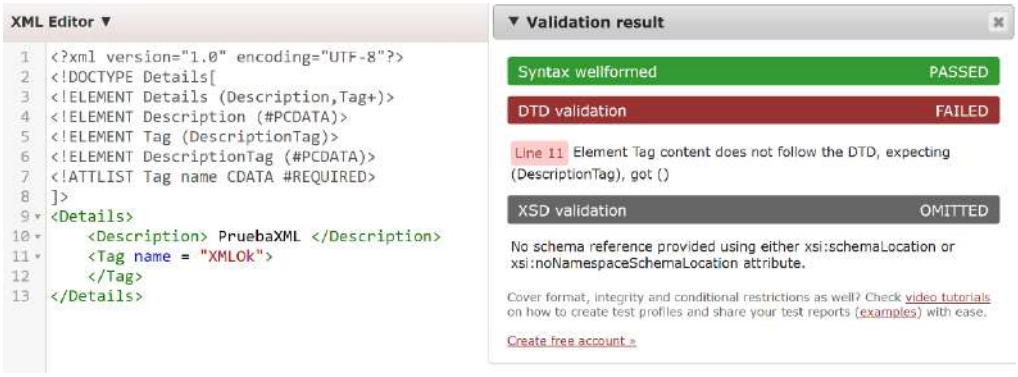
   XML OK

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Details[
<!ELEMENT Details (Description,Tag+)>
<!ELEMENT Description (#PCDATA)>
<!ELEMENT Tag (DescriptionTag)>
<!ELEMENT DescriptionTag (#PCDATA)>
<!ATTLIST Tag name CDATA #REQUIRED>
]>
<Details>
    <Description> PruebaXML </Description>
    <Tag name = "XMLOk">
        <DescriptionTag> violent content </DescriptionTag>
    </Tag>
</Details>
```



XML No Ok

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Details[
<!ELEMENT Details (Description,Tag+)>
<!ELEMENT Description (#PCDATA)>
<!ELEMENT Tag (DescriptionTag)>
<!ELEMENT DescriptionTag (#PCDATA)>
<!ATTLIST Tag name CDATA #REQUIRED>
]>
<Details>
    <Description> PruebaXML </Description>
    <Tag name = "XMLOk">
    </Tag>
</Details>
```



2. Actualice la tabla y los datos adicionados a la base de datos. (PoblandoOK, ProblandoNoOK)

```
---suscripciones
CREATE TABLE suscripciones(
   id NUMBER(5),
   createdAt DATE NOT NULL,
   details XMLTYPE,
   cuentas_id NUMBER(5) NOT NULL,
   suscritoA NUMBER(5) NOT NULL
);
```

Poblando Ok

```
insert into suscripciones (id, createdAt, details, cuentas_id, suscritoA) values (669, '20/06/2013',
'<?xml version="1.0"?>
<Details>
    <Description> Ver reseñas de peliculas </Description>
        <Tag name = "Reseña">
        <DescriptionTag> Peliculas y series </DescriptionTag>
    </Tag>
</Details>', 383, 2258);
```

3. Proponga otra nueva consulta que use Details (Diseño e implementación) Implemente la consulta Subscriptions requiring violent content

```
---------------------- Laboratorio 6-------------------------------------
SELECT cuentas.id, COUNT(suscripciones.id) AS REQUIRINGVIOLENT
FROM cuentas JOIN suscripciones ON(cuentas.id = suscripciones.cuentas_id)
WHERE EXTRACTVALUE(details,'/Details/Tag/DescriptionTag/text()') = ' violent '
GROUP BY cuentas.id;
```

Resultado de la Consulta

SQL | Todas las Filas Recuperadas: 5 en 0,055 segundos

	ID	REQUIRINGVIOLENT
1	2383	1
2	133	1
3	1383	1
4	508	1
5	2008	1

4. Proponga otra nueva consulta que use Details (Diseño e implementación)

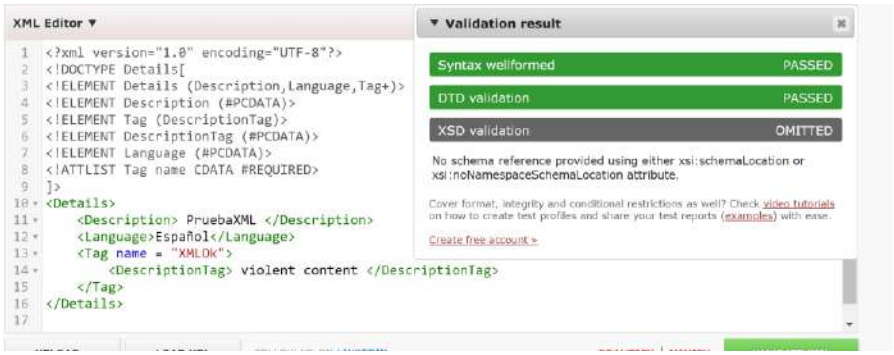
```
1336
1337  ----- Consulta 2.
1338  SELECT suscripciones.id, EXTRACTVALUE(details,'/Details/Tag/@name') AS NameTag
1339  FROM suscripciones
1340  WHERE EXTRACTVALUE(details,'/Details/Tag/DescriptionTag/text()') = ' violent ';
1341
```

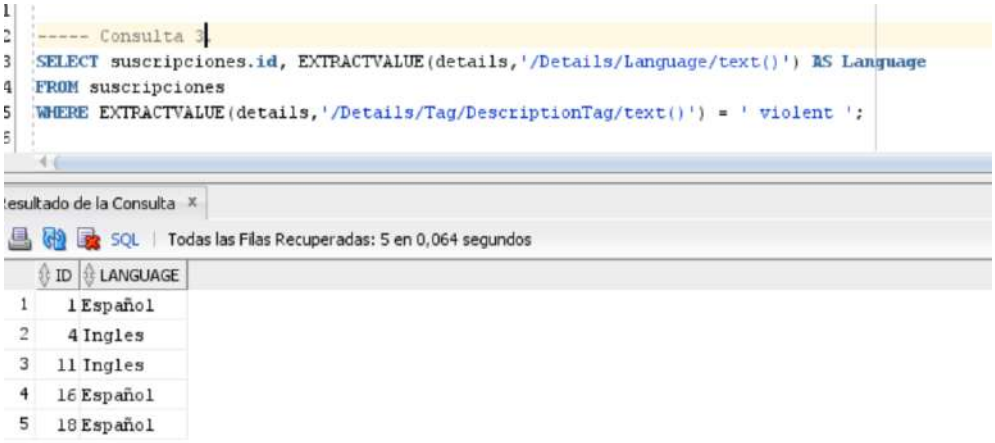
Resultado de la Consulta | Resultado de la Consulta 1

SQL | Todas las Filas Recuperadas: 5 en 0,048 segundos

	ID	NAMETAG
1	1	Asoka
2	4	Asoka
3	11	Chis
4	16	Kun
5	18	Asoka

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





RETROSPECTIVA

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

2. ¿Cuál es el estado actual del laboratorio? ¿Por qué?
   Completo, se dedicó el tiempo necesario para terminar el laboratorio.

3. ¿Cuál consideran fue el mayor logro? ¿Por qué?
   Realizar las consultas, entender la sintaxis XML y como conectarlo con ORACLE.

4. 4. ¿Cuál consideran que fue el mayor problema técnico? ¿Qué hicieron para resolverlo?
   La base de datos de la escuela, al trabajar desde casa es muy común que no ejecute las instrucciones.

5. ¿Qué hicieron bien como equipo? ¿Qué se comprometen a hacer para mejorar los resultados?
   Establecer horarios para la terminación del laboratorio.