

# UF3

# Persistència en SGBD-XML

## Annex 03

## Introducció al SGBD-XML Sedna

Isidre Guixà i Miranda  
Institut Milà i Fontanals d'Igualada

**Accés a dades**

19 de febrer de 2013  
Isidre Guixà i Miranda  
Institut Milà i Fontanals  
C/. Emili Vallès, 4  
08700 - Igualada

Aquest material està protegit sota llicència de [Creative Commons BY-NC-SA 3.0](#) i és la maqueta dels materials en desenvolupament per l'Institut Obert de Catalunya, que una vegada publicats seran accessibles a l'apartat recursos de la web <http://ioc.xtec.cat>

En cas de suggeriment i/o detecció d'error, podeu informar a [iguixa@xtec.cat](mailto:iguixa@xtec.cat)

#### *Castellano*

Este material está protegido bajo licencia [Creative Commons BY-NC-SA 3.0](#) y es la maqueta de los materiales en desarrollo por el Institut Obert de Catalunya, que una vez publicados seran accesibles en el apartado recursos de la web <http://ioc.xtec.cat>

En caso de sugerencia y/o detección de error, dirigir-se a [iguixa@xtec.cat](mailto:iguixa@xtec.cat)

#### *Euskara*

Material hau [Creative Commons BY-NC-SA 3.0](#) babesa eta eredu materialak garatzen Institut Obert de Catalunya, behin argitaratutako web <http://ioc.xtec.cat> Iradokizuna bada eta / edo errore detekzioa, [iguixa@xtec.cat](mailto:iguixa@xtec.cat) zinen joan

#### *Galego*

Este material está protexido baixo [Creative Commons BY-NC-SA 3.0](#) eo modelo é o desenvolvemento de materiais para o Institut Obert de Catalunya, que unha vez publicado será accesible a sección de recursos do <http://ioc.xtec.cat> web. Se suxestión e / ou detección de erros, indica [iguixa@xtec.cat](mailto:iguixa@xtec.cat)

Els continguts de la unitat formativa “Persistència en BD XML natives” fan referència, en múltiples ocasions, al SGBD XML *Sedna* i, en conseqüència, es fa altament recomanable tenir-lo instal·lat i així poder comprovar la correctesa del codi desenvolupat.

El motiu d'aquest annex és facilitar una introducció a *Sedna*, cobrint les necessitats bàsiques d'instal·lació, configuració i primeres passes en un SO Windows. Per dur a terme la instal·lació en altres SO o per aprofundir en aquest SGBD, cal acudir a la pàgina web oficial del producte (<http://sedna.org/>).

## Instal·lació

Per procedir a la instal·lació en un SO Windows, descarreguem de la pàgina web oficial, la darrera versió disponible per a Windows, que en el moment de redactar aquest annex és la 3.5.101 (2011/11/28) i es facilita amb l'arxiu `sedna-3.5.101-bin-win-x##-tar.gz` on ## correspon a 64 o 84 segons la versió de sistema operatiu. Per si la pàgina web oficial ja no proporciona aquesta versió, se us facilita una còpia [aquí](#).

El procés d'instal·lació consisteix en desempaquetar el contingut de l'arxiu en el directori que interressi, mantenint l'estructura de carpetes inclosa a l'arxiu i tenint en compte que el camí on l'instal·lem no pot contenir caràcters no-ASCII. Una vegada desempaquetat, és convenient (opcional) afegir el camí del directori on hagi quedat instal·lat `sedna/bin` en el camí del sistema.

Amb aquests pocs passos ja estem en situació d'utilitzar el SGBD *Sedna*. Aquest SGBD no incorpora cap interfície gràfica, tot i que n'hi ha alguna de desenvolupada que instal·larem més endavant.

## Primeres passes

La gestió del SGBD que proporciona *Sedna* és via una consola del sistema, executant uns determinats programes instal·lats en el directori `sedna/bin`. Així doncs, obrirem una consola de sistema i, si no hem afegit el camí on és `sedna/bin` dins el camí del sistema, ens caldrà navegar fins aquest directori.

Les ordres bàsiques que hem de conèixer són:

- `se_gov`, per posar en marxa el servidor
- `se_stop`, per aturar el servidor
- `se_cdb nomBD`, per crear una base de dades anomenada `nomBD`
- `se_ddb nomBD`, per eliminar la base de dades anomenada `nomBD`
- `se_sm nomBD`, per engegar la base de dades anomenada `nomBD`
- `se_smsd nomBD`, per aturar la base de dades anomenada `nomBD`.
- `se_rc`, per conèixer l'estat del servidor i si hi ha alguna base de dades oberta.
- `se_term nomBD`, per obrir una terminal amb connexió establerta amb la base de dades anomenada `nomBD`.

Les següents línies mostren una possible seqüència d'instruccions a executar:

```
C:\>se_gov
GOVERNOR has been started in the background mode
```

```
C:\>se_cdb mondial
Creating a database (this can take a few minutes)...
The database 'mondial' has been created successfully

C:\>se_sm mondial
Starting database recovery or hot-backup restoration...
Database is in consistent state. Starting...
SM has been started in the background mode

C:\>se_term mondial
Welcome to term, the SEDNA Interactive Terminal. Type \? for help.
mondial> \?
Terminate queries with ampersand + line feed.
\?          - for help
\commit     - to commit transaction
\rollback   - to rollback transaction
\showtime   - to show the time of the latest query execution
\set        - to set the terminal internal variable
              (on help for internal variables type \set?)
\unset      - to unset the terminal internal variable
\quit       - to close session and quit the Sedna Terminal
mondial> \quit
Closing session...Ok

C:\>se_rc
Sedna GOVERNOR is running.

The following databases (SMs) are started:

    mondial, 0 session(s)

C:\>se_smsd mondial
The database 'mondial' has been successfully shut down

C:\>se_rc
Sedna GOVERNOR is running.

C:\>se_stop
SEDNA server has been shut down successfully

C:\>se_rc
SEDNA Message: ERROR SE4400
SEDNA server is not running. Did you forget to run se_gov command?
Details: Can't open GOVERNOR shared memory
```

La seqüència anterior correspon a:

- Engagar el servidor ([se\\_gov](#))
- Crear una base de dades de nom `mondial` ([se\\_cdb](#))

- Posar en marxa la base de dades de nom `mondial` (`se_sm`)
- Obrir una consola de treball (terminal) amb la base de dades `mondial` (`se_term`)
- Obtenir ajuda referent a les ordres que accepta la terminal (`\?`)
- Sortir de la terminal (`\quit`)
- Comprovar l'estat del servidor (`se_rc`)
- Tancar la base de dades (`se_smsd`)
- Comprovar l'estat del servidor (`se_rc`)
- Aturar el servidor (`se_stop`)
- Comprovar l'estat del servidor (`se_rc`)

Com a resultat de la seqüència anterior, ja tenim en el servidor una base de dades de nom `mondial`. No hi tenim, però, cap contingut. En aquests moments ens interessa poder incorporar continguts a la base de dades i, posteriorment executar-hi consultes *XPath* i/o *XQuery*.

Totes les instruccions les haurem d'efectuar des de la terminal `se_term`. Podem fer-ho de tres maneres:

- Posant en marxa una terminal `se_term nomBD` i, anar executant les instruccions dins la terminal, abandonant la terminal amb `\quit` en finalitzar l'execució.
- Executar des del símbol del sistema l'ordre:  

```
se_term -query "instrucció" nomBD
```

on `instrucció` correspon a una expressió *XPath* o *XQuery* i ha d'anar tancada entre dobles cometes.
- Executar des del símbol del sistema l'ordre:  

```
se_term -file nomFitxer nomBD
```

on `nomFitxer` correspon a un document XML que cal carregar a la base de dades o a un document que inclou una instrucció *XQuery* a executar.

Volem, ara, afegir el document `mondial.xml` disponible [aquí](#), a la base de dades `mondial` prèviament creada. Per aconseguir-ho tenim dues possibilitats:

- Obrir una terminal (`se_term`) contra la base de dades `mondial` i executar:  

```
mondial> LOAD 'camíOnResideix\mondial.xml' 'mondial.xml'
```

En aquest cas, estem carregant el document `mondial.xml` dins la base de dades deixant-lo enregistrat, dins la base de dades, amb el nom `mondial.xml`.

- Executar directament l'ordre `LOAD` com l'argument de l'opció `-query` de `se_term`:  

```
se_term -query "LOAD 'camíOnResideix\mondial.xml' 'mondial.xml'"  
mondial
```

En aquest cas, estem utilitzant la sintaxis: `se_term -query "instrucció" nomBD`

En qualsevol cas, si la càrrega s'ha pogut efectuar correctament, ha d'aparèixer el

missatge `Bulk load succeeded`.

Les BD que emmagatzemen documents XML, acostumen a facilitar el concepte de col·lecció per agrupar documents XML en funció de la informació que contenen. Així, les col·leccions juguen dins les BD-XML el paper de les taules dins les BD relacionals. Les col·leccions poden, a la vegada, contenir altres col·leccions. La versió de *Sedna* que acabem d'instal·lar implementa parcialment el concepte de col·lecció, ja que permet definir el que podríem anomenar "col·leccions de 1r. Nivell", en el sentit que una base de dades pot tenir col·leccions però que aquestes ja no poden contenir altres col·leccions en el seu interior.

Les instruccions que *Sedna* facilita per gestionar les col·leccions són:

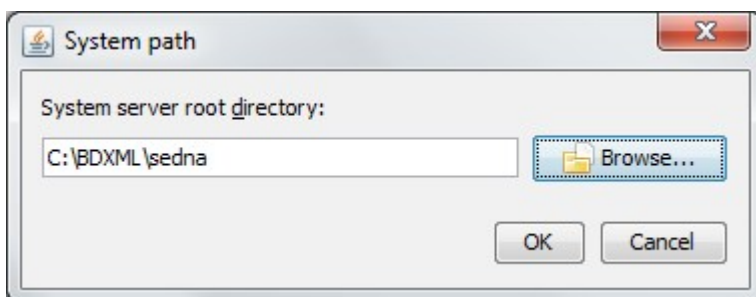
- `create collection "nomCOL"`, per crear una col·lecció anomenada `nomCOL`.
- `load "nomDocumentAmbCamí" "nomDocument" "nomCOL"`, per carregar el document `nomDocumentAmbCamí` a la col·lecció `nomCOL` enregistrant el document com a `nomDocument`.
- `drop document "nomDocument" in collection "nomCOL"`, per eliminar el document `nomDocument` de la col·lecció `nomCOL`.
- `drop collection "nomCOL"`, per eliminar la col·lecció `nomCOL`.

### Client gràfic *SednaAdmin*

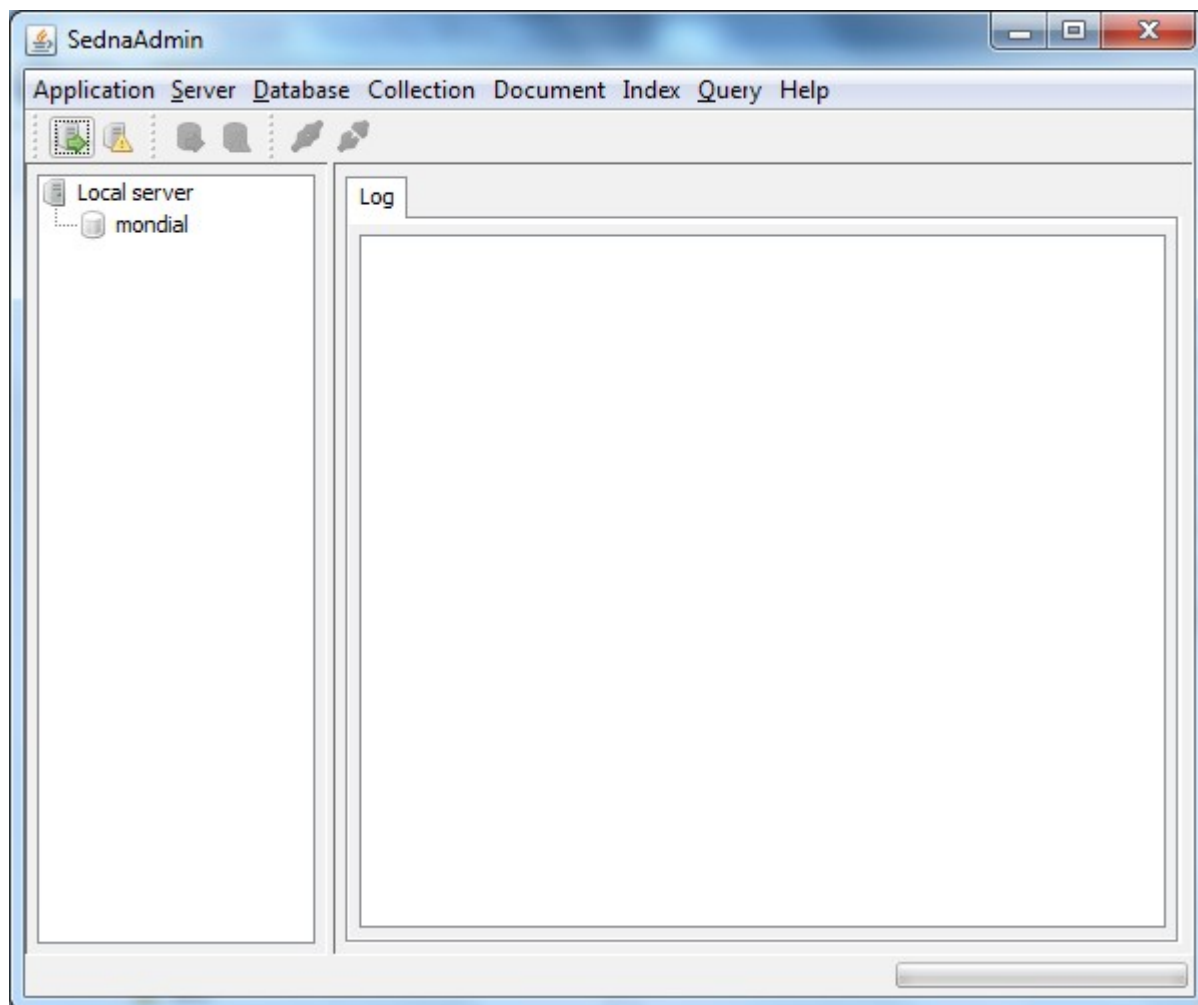
El SGBD *Sedna* no incorpora cap interfície gràfica d'usuari per facilitar la gestió gràfica dels continguts de les bases de dades. Hi ha, però, l'aplicació *SednaAdmin* desenvolupada per Flávio R. C. Sousa, reconeguda pel projecte *Sedna* doncs és accessible des de la zona de descàrrega de *Sedna*.

La instal·lació és molt simple. Només cal descarregar l'arxiu que conté els binaris i que en el moment de redactar aquest annex és `sednaadmin-bin-0.1.5.zip`, descomprimir-lo en qualsevol directori del sistema tot mantenint l'estructura de carpetes inclosa a l'arxiu i executar l'arxiu `SednaAdmin.jar` inclòs en el sotsdirectori `sednaadmin-bin`.

La primera vegada que posem *SednaAdmin* en execució, ens mostra una pantalla per introduir el directori on tenim instal·lat el SGBD *Sedna*, com es veu a la imatge.



Una vegada informat el directori on resideix *Sedna*, se'ns mostra la interfície gràfica:

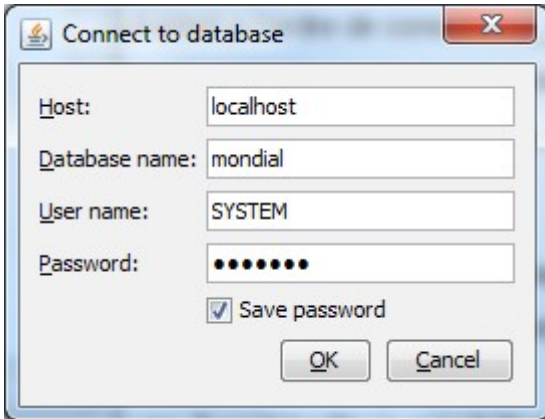


La pantalla ens mostra un panell a l'esquerra, on observem el servidor local i les bases de dades que allotja. En aquests moments, només hi tenim la base de dades `mondial` creada des de la consola.

A la part superior observem una botonera amb sis botons, que corresponen, d'esquerra a dreta, a les accions següents:

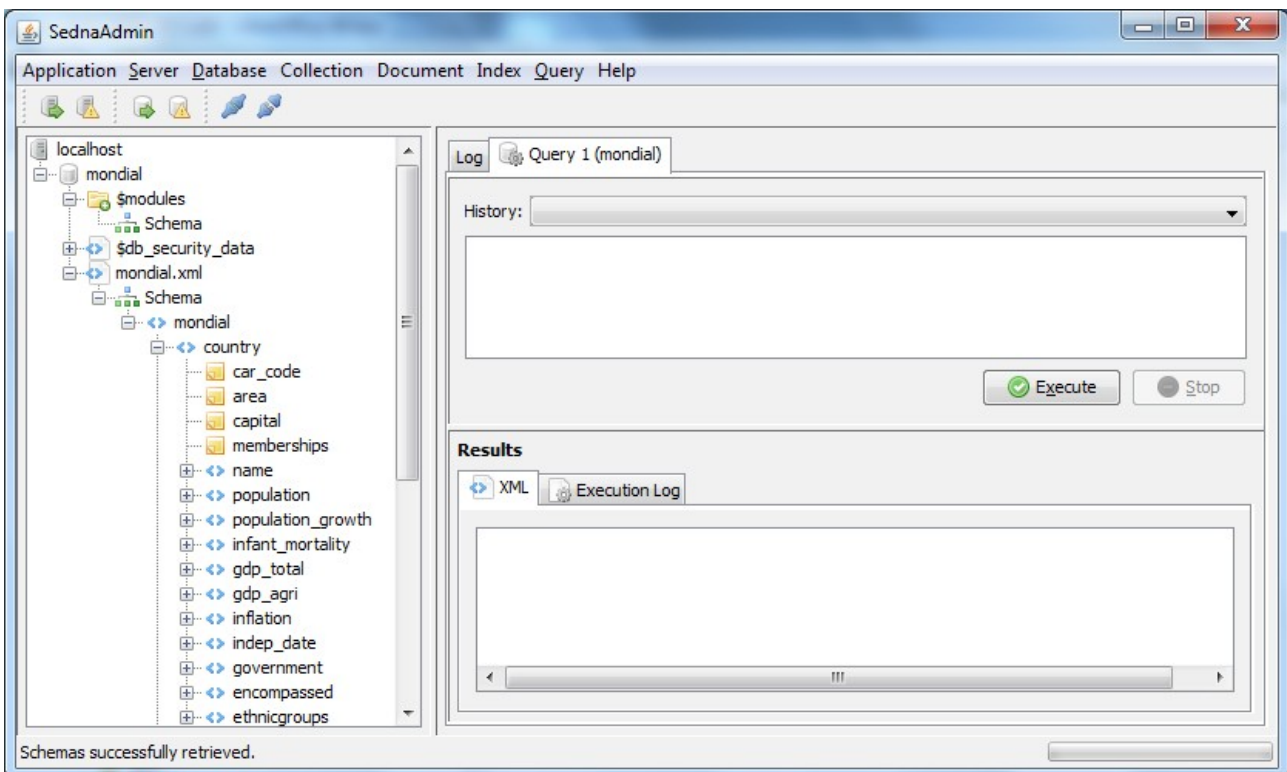
- Engagar servidor, equivalent a l'ordre de consola `se_gov`.
- Aturar servidor, equivalent a l'ordre de consola `se_stop`.
- Engagar base de dades, equivalent a l'ordre de consola `se_sm`.
- Aturar base de dades, equivalent a l'ordre de consola `se_smsd`.
- Connectar amb base de dades, equivalent a l'ordre de consola `se_term`.
- Desconnectar de la base de dades, equivalent l'opció `\quit` de l'ordre de consola `se_term`.

Així doncs, si volem treballar amb la base de dades `mondial`, ens situarem damunt ella en el panell esquerre i procedirem a engagar la base de dades si no és engegada. Posteriorment, ens hi connectem i *SednaAdmin* ens mostra la pantalla de connexió on hem d'introduir l'usuari i la contrasenya.



La pantalla ve emplenada amb l'usuari `SYSTEM` i la contrasenya introduïda (`MANAGER` és la contrasenya amb la que s'instal·la *Sedna*). Cal tenir present que els usuaris i contrasenyes en *Sedna* són *case-sensitive*.

Una vegada connectats a una base de dades, *SednaAdmin* ens mostra en el panell esquerre les col·leccions i documents que conté, ja sigui directament a l'arrel de la base de dades o dins les col·leccions. Cada document va acompanyat de l'esquema que ens mostra la seva estructura. A la següent imatge ho observem amb el document `mondial.xml` introduït a l'arrel de la base de dades `mondial`.



Observem també, a l'anterior imatge, que la base de dades `mondial`, a banda del document `mondial.xml` que hem introduït, incorpora la col·lecció `$modules` i el document `$db_security_data`. Ambdós són per a tasques administratives del SGBD i com que l'administració del SGBD no forma part del nostre objectiu, no els tocarem per a res.

Observem com la banda dreta de la pantalla se'ns ha dividit en dues parts: la superior, per a introduir instruccions *XPath* o *XQuery*; i la inferior, per a veure el resultat de l'execució



de les instruccions.

Si donem una volta pels menús d'aquesta interfície gràfica, observarem que podem crear i eliminar bases de dades i col·leccions i incorporar-hi i eliminar-hi documents.

## Consultes *XPath* i *XQuery*

En el SGBD *Sedna* podem executar les instruccions *XPath* i *XQuery* des de la consola, però com que hem instal·lat l'aplicació *SednaAdmin* ens és més còmode utilitzar aquesta interfície gràfica.

En *SednaAdmin* introduïm l'expressió a avaluar a una de les pestanyes *Query* de la part superior dreta de la pantalla, premem el botó *Execute* de la part inferior dreta de la mateixa pestanya i obtenim el resultat a l'apartat *Results* (part inferior dreta de la pantalla).

Vegem alguns exemples, no pas com a pràctica dels llenguatges *XPath* i *XQuery*, sinó com a pràctica de llur execució dins *Sedna*.

Els exemples que segueixen, pressuposen que tenim oberta la base de dades `mondial` anteriorment instal·lada i que conté un únic document `mondial.xml`. Posteriorment s'indica com actuar en cas que la base de dades contingui més d'un document.

1. Mostrar el contingut del document `mondial.xml` de la base de dades `mondial` (que caldrà tenir oberta):

```
doc("mondial.xml")/mondial
```

Si el document hagués està dins una col·lecció anomenada `xxx`, la instrucció hagués estat:

```
doc("mondial.xml", "xxx")/mondial
```

2. Mostrar els noms de tots els països existents a la base de dades `mondial`:

Tenim diverses possibilitats per obtenir la informació demanada:

- `doc("mondial.xml")/mondial/country/name`
- `doc("mondial.xml")//country/name`

El resultat de l'execució és similar a:

```
<name>Albania</name>
<name>Greece</name>
<name>Macedonia</name>
<name>Serbia</name>
<name>Montenegro</name>
...
```

La pestanya *Execution Log* ens informa que la consulta facilita 238 resultats. En cas que hi hagués països amb dos noms, per a que només ens aparegués el primer, caldria escriure:

- `doc("mondial.xml")/mondial/country/name[1]`
- `doc("mondial.xml")//country/name[1]`

Si el que es desitjava és obtenir únicament el text existent entre les etiquetes

<name> i </name>, calia executar:

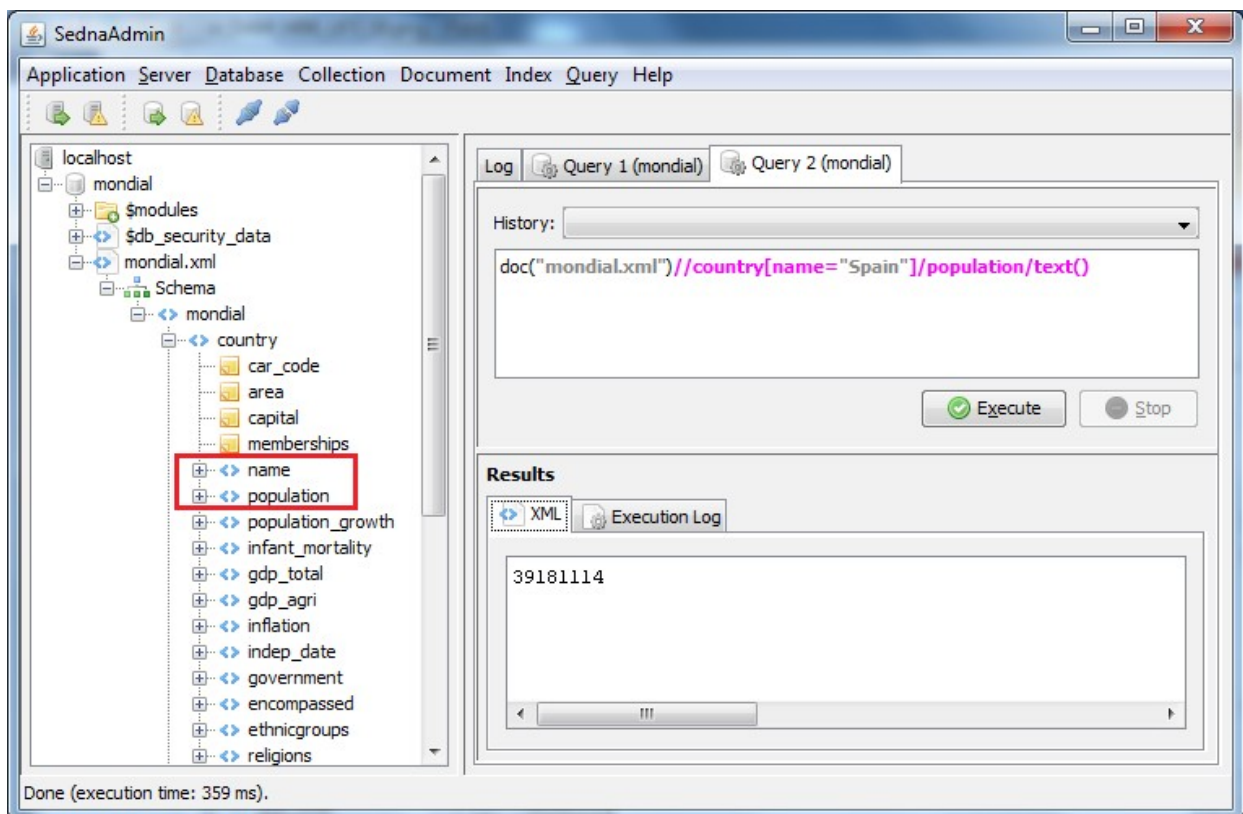
- `doc("mondial.xml")/mondial/country/name[1]/text()`
- `doc("mondial.xml")//country/name[1]/text()`

Per finalitzar aquest exemple, comentar que la següent expressió seria un error, doncs facilitaria els continguts de totes les etiquetes `name`, és a dir, noms de països, noms de ciutats,... doncs dins la base de dades `mondial` hi ha informació diversa sota etiquetes `name`:

- `doc("mondial.xml")//name`

3. Mostrar la quantitat d'habitants del país anomenat `Spain` a la base de dades `mondial.xml`:

En el panell esquerre de *SednaAdmin* podem analitzar l'estructura del document XML per veure on resideix la informació que se'ns demana, com es mostra a la imatge següent, on trobem que `name` i `population` són elements sota el node `country`.



Com es veu a la imatge, la consulta a executar és:

```
doc("mondial.xml")//country[name="Spain"]/population/text()
```

4. Mostrar, a partir de la base de dades `mondial`, el nom de tots els països acompanyat de la seva població.

En aquest cas es tracta d'una consulta *XQuery*, amb una instrucció `FLWOR`:

```
for $i in doc("mondial.xml")//country
return concat($i/name[1], "-", $i/population)
```

Observem que és convenient indicar `name[1]` per si la base de dades incorpora, per alguns països, dos noms, doncs en tal situació la funció `concat()` detecta un error (XPTY0004).

Una altra possibilitat, per obtenir el resultat amb millor format, és:

```
for $i in doc("mondial.xml")//country
return <país>{concat($i/name[1], "-", $i/population)}</país>
```

- Mostrar, a partir de la base de dades `mondial`, el nom de tots els països acompanyat de la seva població, ordenada per la població en ordre descendent.

```
for $i in doc("mondial.xml")//country
order by number($i/population) descending
return <país>{concat($i/name[1], "-", $i/population)}</país>
```

Fixem-nos que a l'hora d'ordenar cal utilitzar la funció `number()` per a ordenar la població segons ordre numèric, doncs del contrari s'ordenaria segons l'ordre lexicogràfic.

- Mostrar, a partir de la base de dades `mondial`, el nom del país que tingui més població.

```
let $nHab:=max(doc("mondial.xml")//country/number(population))
for $i in doc("mondial.xml")//country
where $i/number(population)=$nHab
return $i/name/text()
```

## Col·leccions i documents

El SGBD-XML *Sedna* obliga a estar connectat a una base de dades per a poder treballar. És a dir, no hi ha consultes que permetin, estant connectat a una base de dades, accedir a informació d'altres bases de dades. Cada base de dades pot tenir col·leccions (un únic nivell) i aquestes poden tenir diversos documents (però no altres col·leccions). En els exemples anteriors hem vist com sempre hem estat accedint al document `mondial.xml` ubicat a l'arrel de la base de dades.

Per obtenir un llistat de les col·leccions i/o documents existents a la base de dades activa, el sistema ens facilita dos documents:

- Document `$collections` amb informació sobre totes les col·leccions existents.

Així, per veure les col·leccions escriuríem:

```
doc("$collections")
```

i obtindríem quelcom similar a:

```
<collections>
  <collection name="$modules"/>
```

```
<collection name="..." />
</collections>
```

La col·lecció `$modules` engloba tots els documents de la base de dades ubicats a l'arrel, és a dir, sense pertànyer a cap col·lecció específica.

- Document `$documents` amb informació de totes les col·leccions amb els documents de cada col·lecció i els documents ubicats a l'arrel de la base de dades

Per veure'n el seu contingut escriurem:

```
doc("$documents")
```

i obtindríem quelcom similar a:

```
<documents>
  <document name="$db_security_data" />
  <collection name="$modules" />
  <collection name="..." />
    <document name="..." />
  <collection name="...">
    <document name="..." />
  </collection>
  <document name="mondial.xml" />
  <document name="..." />
</documents>
```

Així doncs, si volem efectuar una consulta sobre tots els documents de la base de dades, haurem d'escriure quelcom similar a:

```
for $doc in doc("$documents")//document/@name/string()
where $doc != "$db_security_data"
return doc($doc) / <consultaQueCorrespongui>
```

Si volem la consulta sobre els documents d'una col·lecció concreta, escriurem:

```
for $doc in doc("$documents")//document/@name/string()
where $doc != "$db_security_data"
return doc($doc) / <consultaQueCorrespongui>
```

## Instruccions Update

El SGBD *Sedna* facilita una extensió del llenguatge *XQuery* per a poder efectuar canvis en els documents XML emmagatzemats a la base de dades. L'extensió *Update* facilitada per *Sedna* (versió 3.5.101) no segueix el llenguatge [XQuery Update Facility \(XQUF\)](#) recomanat per W3C. Cal tenir en compte que els diversos SGBD XML van iniciar la implementació d'extensions *Update* abans que hi hagués cap recomanació i això provoca que, en ocasions, es faci difícil canviar el rumb iniciat. Ningú sap, però, si *Sedna* facilitarà, en un futur, una extensió XQUF. La sintaxis de l'extensió *Update* que proporciona *Sedna* es pot consultar a la documentació de *Sedna* i segueix la sintaxis proposada per [Patrick Lehti](#) (a l'igual que el SGBD *eXist-db*)

Els exemples que segueixen, pressuposen que tenim oberta la base de dades `mondial` anteriorment instal·lada i que conté un únic document `mondial.xml`

1. Inserir, entre els continents Àsia i Amèrica, el node següent:

```
<institut nom="IOC">IOC</institut>
```

Per poder efectuar aquesta inserció ens cal mirar el document `mondial.xml` i observar com conté la informació.

```
...
<continent id="europe">
  <name>Europe</name>
  <area>9562488</area>
</continent>
<continent id="asia">
  <name>Asia</name>
  <area>45095292</area>
</continent>
<continent id="australia">
  <name>Australia/Oceania</name>
  <area>8503474</area>
</continent>
<continent id="africa">
  <name>Africa</name>
  <area>30254708</area>
</continent>
<continent id="america">
  <name>America</name>
  <area>39872000</area>
</continent>
...
```

Ara ja podem procedir a inserir el nou node:

```
UPDATE insert <institut nom="IOC">IOC</institut>
following doc("mondial.xml")//continent[name="Asia"]
```

O també:

```
UPDATE insert <institut nom="IOC">IOC</institut>
preceding doc("mondial.xml")//continent[@id="australia"]
```

En qualsevol cas, el resultat és:

```
...
<continent id="asia">
  <name>Asia</name>
  <area>45095292</area>
</continent>
<institut nom="IOC">IOC</institut>
<continent id="australia">
  <name>Australia/Oceania</name>
  <area>8503474</area>
</continent>
```

...

## 2. Substituir el valor del node que acabem d'inserir pel valor "Institut Obert de Catalunya"

```
UPDATE replace $n in doc("mondial.xml")//institut[@nom="IOC"]
with
<institut nom="IOC">Institut Obert de Catalunya</institut>
```

L'extensió *Update* de *Sedna* no facilita la possibilitat de modificar el valor d'un element... Cal substituir tot el node.

El resultat és:

```
...
<continent id="asia">
  <name>Asia</name>
  <area>45095292</area>
</continent>
<institut nom="IOC">Institut Obert de Catalunya</institut>
<continent id="australia">
  <name>Australia/Oceania</name>
  <area>8503474</area>
</continent>
...
```

## 3. Substituir el node `institut` amb atribut `nom="IOC"` pel node:

```
<cicle>
  <codi>DAM</codi>
  <nom>Desenvolupament d'aplicacions multiplataforma</nom>
</cicle>
```

Instrucció:

```
UPDATE replace $n in doc("mondial.xml")//institut[@nom="IOC"]
with
<cicle>
  <codi>DAM</codi>
  <nom>Desenvolupament d'aplicacions multiplataforma</nom>
</cicle>
```

Resultat:

```
...
<continent id="asia">
  <name>Asia</name>
  <area>45095292</area>
</continent>
<cicle>
  <codi>DAM</codi>
  <nom>Desenvolupament d'aplicacions multiplataforma</nom>
</cicle>
```

```
<continent id="australia">
  <name>Australia/Oceania</name>
  <area>8503474</area>
</continent>
...
```

#### 4. Reanomenar els nodes `cicle` per `cf`:

```
UPDATE rename doc("mondial.xml")//cicle on cf
```

#### Resultat:

```
...
<continent id="asia">
  <name>Asia</name>
  <area>45095292</area>
</continent>
<cf>
  <codi>DAM</codi>
  <nom>Desenvolupament d'aplicacions multiplataforma</nom>
</cf>
<continent id="australia">
  <name>Australia/Oceania</name>
  <area>8503474</area>
</continent>
...
```

#### 5. Eliminar els nodes `cf`:

```
UPDATE delete doc("mondial.xml")//cf
```

#### Resultat:

```
...
<continent id="asia">
  <name>Asia</name>
  <area>45095292</area>
</continent>
<continent id="australia">
  <name>Australia/Oceania</name>
  <area>8503474</area>
</continent>
...
```

Comentar que Sedna facilita una opció `delete` que elimina els nodes indicats però manté els fills dins el document. Aquesta opció no és contemplada en XQUF 1.0.